

DOKTORI ÉRTEKEZÉS

Fordított szövegek számítógépes összevetése

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

SEIDL-PÉCH OLÍVIA

2011

Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar

DOKTORI ÉRTEKEZÉS

SEIDL-PÉCH OLÍVIA

Fordított szövegek számítógépes összevetése

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

Nyelvtudományi doktori iskola, vezetője: BAÑCZEROWSKI JANUSZ, DSc.

Fordítástudományi doktori program vezetője: KLAUDY KINGA, DSc.

A bizottság tagjai:

Elnök: KLAUDY KINGA, DSc.

Opponensek: Károly Krisztina, PhD. és Balaskó Mária, PhD.

Titkár: FÓRIS ÁGOTA, PhD.

További tagok: HELTAI PÁL, CSc.

Kurtán Zsuzsa, CSc. és Székely Gábor, DSc.

Témavezető: PRÓSZÉKY GÁBOR, DSc.

Budapest, 2011

Tartalom

I.	BEVEZETÉS	9
1	EREDETISÉG NYILATKOZAT	9
2	KÖSZÖNETNYILVÁNÍTÁS	10
3	AZ ÉRTEKEZÉS TÉMÁJA ÉS CÉLJA	11
4	AZ ÉRTEKEZÉS FELÉPÍTÉSE	13
5	A KUTATÁS MÓDSZERE	14
5.1	Előzmény	14
5.1.1	Új megközelítés a kohézió-kutatásban: összehasonlító elemzés	14
5.1.2	Szövegválasztás	15
5.1.3	A Károly-féle taxonómia	15
5.1.4	A gépi feldolgozás előkészítése	16
5.1.5	Elemzés és annotálás	17
5.1.6	Lekérdezések	19
5.2	A kiválasztott módszer	20
5.2.1	Miért a szöveg produkció, miért nem a percepció van a kutatás fókuszában?	20
5.3	Újszerűség	21
5.4	Eredmények	22
6	KUTATÁSI KÉRDÉSEK ÉS HIPOTÉZISEK	24
6.1	Kutatási kérdések	24
6.2	Kutatási hipotézisek	24
7	A KUTATÁS HASZNA	25
II.	ELMÉLETI HÁTTÉR	27
1	A FORDÍTOTT SZÖVEG MINT A TUDOMÁNYOS VIZSGÁLAT TÁRGYA	27
1.1	Bevezetés	27
1.2	A fordított szövegek vizsgálata	27
1.3	A leíró fordításkutatás megjelenése	29
1.4	A fordítási szituáció mint interkulturális tevékenység	31
1.5	A fordított szövegek hatása az anyanyelv használatra	32
1.6	A célnyelvi szövegek keletkezését befolyásoló tényezők	34

1.7	A célnyelvi szövegalkotásra jellemző sajátosságok	37
1.7.1	Fordításnyelv – kvázi helyesség	39
1.7.2	A szövegszintű eltolódások okai	41
1.8	A fordítói kompetencia	43
1.9	A norma és az ekvivalencia fogalma a fordításkutatásban	46
1.9.1	A fordítási norma	46
1.9.2	Az ekvivalencia	49
1.10	A fordítási univerzálé	53
2	A KUTATÁS SZÖVEGNYELVÉSZETI BEÁGYAZOTTSÁGA	56
2.1	A szövegkutatás megjelenése	56
2.2	A szövegkutatás feladata	57
2.3	A fordítások szövegnyelvészeti megközelítése	60
2.4	Mit tekintünk szövegnek?	63
2.5	A szövegszerűség kritériumai	65
3	A LEXIKAI KOHÉZIÓ VIZSGÁLATA FORDÍTOTT SZÖVEGEKBEN	69
3.1	Bevezetés	69
3.2	A kohézió	69
3.2.1	A kohézió meghatározása és a kohéziós eszközök	70
3.2.2	A lexikai kohézió	75
3.2.3	A lexikai egység	77
3.2.4	A kohéziókutatás legismertebb képviselői	77
3.3	A lexikai kohézió kutatása és a fordítás	80
3.3.1	Miért érdekes számunkra a lexikai kohézió vizsgálata fordított szövegekben?	80
3.3.2	Lexikai kohézió és fordítás	82
3.3.3	Leíró fordításkutatás	84
3.3.4	Fordításoktatás	85
4	A KORPUSZNYELVÉSZETI MÓDSZER ELŐNYEI A FORDÍTÁSKUTATÁSBAN	89
4.1	Bevezetés	89
4.1.1	A korpusznyelvészeti módszer perspektívái az alkalmazott nyelvészeti kutatásban	89
4.1.2	A korpusznyelvészet, mint módszer	89
4.1.3	Korpusznyelvészet a fordításkutatásban	91
4.2	Eszközök, számítógépes alkalmazások	92
4.2.1	A korpuszok legalapvetőbb ismérvei	92
4.2.2	Az elektronikusan elérhető korpuszok	93

4.2.3	A fordításkutatás számára fontos korpuszok	97
4.2.4	Fordítástámogatás	100
4.2.5	Saját korpusz építése	102
4.2.6	A korpusztervezés fő szempontjai	103
4.2.7	Szövegválasztás	103
4.2.8	A korpuszok annotálása	104
4.2.9	Egyszerű korpuszalapú elemzések	105
III.	A KUTATÁS MÓDSZEREI	108
1	A VIZSGÁLAT TÁRGYA	108
2	A KUTATÁS FŐBB SZAKASZAI	109
2.1	Korpuszépítés	109
2.2	Annotálás	110
2.2.1	Kézi annotálás	110
2.2.2	Gépi annotálás	111
2.3	A lexikai kohézió vizsgálata korpusznyelvészeti módszerrel	112
2.3.1	A kísérleti kutatás	112
2.3.2	A Magyar Lexikai Kohéziós Projekt (Lexical Cohesion in Hungarian)	114
3	A KORPUSZ BEMUTATÁSA	115
3.1	A korpusz szerkezete	115
3.2	A korpusz összetétele	117
3.2.1	Az alkorpuszok tervezése	118
3.2.2	Autentikus magyar szövegek	119
3.2.3	Fordított magyar szövegek	120
3.3	A korpusz mérete	121
4	AZ ELEMZÉS BEMUTATÁSA	122
4.1	Az elemzéshez használt nyelvtechnológiai eszközök	122
4.1.1	Princeton WordNet	123
4.1.2	EuroWordNet és BalkaNet	124
4.1.3	HuWordNet	124
4.2	Szegmentálás és morfológiai elemzés	125
4.2.1	Bekezdésre és mondatra bontás	125
4.2.2	Tokenizálás, egyértelműsítés	126

4.3	A lexikai kohéziós kapcsolatok automatikus azonosítása	126
5	AUTOMATIKUS GÉPI ESZKÖZÖKKEL AZONOSÍTOTT LEXIKAI KOHÉZIÓS KAPCSOLATOK A MAGYAR LEXIKAI KOHÉZIÓS PROJEKT SORÁN	128
5.1.1	A lexikai kohéziós kapcsolatok realizációja a Magyar Lexikai Kohéziós Projektben	130
5.1.2	További példák az antonima kapcsolat realizációjára a Magyar Lexikai Kohéziós Projektben	136
5.1.3	További példák a meronima kapcsolat realizációjára a Magyar Lexikai Kohéziós Projektben	138
5.1.4	További példák a hiponima kapcsolat realizációjára a Magyar Lexikai Kohéziós Projektben	141
6	A MAGYAR LEXIKAI KOHÉZIÓS PROJEKT EREDMÉNYEI	144
6.1	Szövegtípus és nyelvpár szerinti összesítések	145
6.2	A shift, az eltolódás kimutatása	148
6.2.1	Az eltolódás kimutatása a közéleti szövegek esetében	149
6.2.2	Az eltolódás kimutatása az szépirodalmi szövegek esetében	149
6.2.3	Az eltolódás kimutatása a tudományos szövegek esetében	151
6.2.4	Az eltolódás kimutatása a vallási szövegek esetében	153
IV.	AZ EREDMÉNYEK ÖSSZEGZÉSE, KÖVETKEZTETÉSEK	155
	ÁBRÁK JEGYZÉKE	161
	TÁBLÁZATOK JEGYZÉKE	163
	IRODALOM	164
	NYÍLVÁNOS FORRÁSOK	178
	PUBLIKÁCIÓS LISTA	179

I. Bevezetés

1 Eredetiség nyilatkozat

EREDETISÉGI NYILATKOZAT

Ezúton kijelentem, hogy a *Fordított szövegek számítógépes összevetése* című doktori értekezést önállóan készítettem el, és nem használtam fel tanulmányokból, illetve értekezésekből jelöletlenül átvett idézeteket. Az értekezést más felsőoktatási intézményben, illetve az egyetem más karán nem nyújtottam be, az értekezésemben szerepeltetett – más szerzőktől származó – gondolatokat és idézeteket a tudományos munkáknak megfelelően jelöltem.

2011. december 2.

Seidl-Pécs Olívia

STATEMENT OF AUTHORSHIP

I hereby declare that the thesis titled *Fordított szövegek számítógépes összevetése (Computer-based Comparison of Translated Texts)* is my own work. It was solely written by me – no other article, study or thesis of mine is partially or fully represented here; all references to other academic works are cited in its original format and noted as per the exact source in full conformity and integrity with the academic requirements of Doctor of Philosophy, and has not been submitted to any other institution.

2 December 2011

Seidl-Pécs Olívia

2 Köszönetnyilvánítás

Köszönettel tartozom témavezetőmnek, Prószéky Gábornak, aki segítségemre volt a téma helyes fókuszának megtalálásában, és aki végigkísért tanácsaival az értekezés írása és az azt megelőző doktori képzés során. Köszönetet szeretnék mondani Klaudy Kingának, a Fordítástudományi doktori program vezetőjének, aki nagy hangsúlyt fektetett arra, hogy a doktori program keretében a tudományterület lehető legszélesebb palettájával ismertessen meg. Az általa javasolt szakirodalom feldolgozásával kialakíthattam naprakész tudományos szemléletemet. Hálával tartozom a Fordítástudományi Doktori Program oktatóinak, akik közül szeretném külön kifejezni köszönetemet Fóris Ágotának, Heltai Pálnak, Károly Krisztinának és Váradi Tamásnak. Az értekezés megírásához szükséges elméleti ismeretek és gyakorlati tapasztalatok fontos részét szereztem meg előadásaik, útmutatásaik, építő gondolataik, tanácsaik nyomán.

Köszönetet szeretnék mondani az elemzés automatizálását megvalósító program megírásáért Orosz Györgynek és Laki Lászlónak, a Pázmány Péter Katolikus Egyetem Információs Technológiai Kara két PhD-hallgatójának, akik munkája nélkül az elemzéseket nem tudtam volna ilyen nagy korpuszon elvégezni. Köszönettel tartozom az MTA Nyelvtudományi Intézet Nyelvtechnológiai Kutatócsoportja munkatársai közül Héja Enikőnek, Kuti Juditnak, Oravec Csabának, Pajzs Júliának, Perényi Dórának, Pintér Tibornak, Sass Bálintnak és Simon Eszternek, akik gyakorlati tapasztalataik átadásával, az eszközkészlethez és a szakirodalomhoz való hozzáférésem elősegítésével voltak hasznomra kutatásom korpusznyelvészeti hátterének megalapozásában. Köszönetet szeretnék mondani mindazoknak, akik az elemzéshez a fordított szövegeiket a rendelkezésemre bocsátották.

Végül, de nem utolsó sorban köszönöm férjem, Tibor, valamint gyermekeim és szüleim kitartóan bátorító segítségét, támogatásukat a kutatás befejezéséhez és az értekezés megírásához.

*"A problémákat nem új információk
segítségével oldjuk meg, hanem azáltal, hogy
rendszerbe foglaljuk azt, amit már régóta
tudunk."
(Ludwig Wittgenstein)*

3 Az értekezés témája és célja

Az értekezés nem irodalmi, hanem nyelvészeti szempontból foglalkozik a fordításokkal. Az értekezés alkalmazott nyelvészeti beágyazottságát az adja, hogy a kutatás fókuszában a fordítás eredményeképpen kapott szöveg áll, amit elsősorban a szövegnyelvészet és a korpusznyelvészet eszközeivel vizsgál. Az értekezés így témaválasztásánál fogva három tudományterület – a fordításkutatás, a szövegnyelvészet és a korpusznyelvészet – kutatási szférájának metszéspontjában helyezkedik el.

A fordított szövegek korpusznyelvészeti elemzése eredményeként manifesztálódó eltolódásjelenségek okait vizsgáló értekezésem csatlakozik ahhoz a Petőfi S. (2004) által propagált szövegnyelvészeti megközelítéshez, amely a szövegek komplexebb vizsgálatát tűzi célul maga elé. A korpuszokból kinyert tisztán statisztikai eredmények számszerű adatokkal bizonyítják, hogy a fordított szövegekre jellemző kohéziós eszközhasználat nem egyezik az autentikus szövegprodukciónak jellemző kohéziós eszközhasználattal. Igyekeztem az eredményközpontú megközelítés mellett a forrásnyelvi szöveg hatására bekövetkező célnyelvi szövegsajátosságok és eltolódások leírásához a funkció- és folyamatorientált jellemzőket is a dolgozat elméleti fejezeteiben részletesebben is számba venni. 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” (Károly 2007:13).

Kutatásom nagyobb mennyiségű nyelvi adat alapján végzett összehasonlító elemzés és így a kontrasztív elemzések (*Contrastive Analysis*) között a kontrasztív szövegnyelvészeti elemzések közé sorolható. A vizsgálat kontrasztív jellegét a többféle műfaji és szövegtipológiai csoportba sorolható magyar nyelven keletkezett autentikus és magyarra fordított célnyelvi szövegek szövegépítő sajátosságainak egybevetésének köszönheti. Az eredeti magyar és a fordítás eredményeként keletkezett magyar szövegek lexikai kohéziós mintázatának leírása az értekezést azon szövegnyelvészeti kutatások közé helyezi, amelyek sora a mondathatárokon túlmutató nagyobb nyelvi egységek nyelvészeti elemzését tűzi ki célul, azaz „a mondatokból létrejövő szöveg alkotását modellálja, annak szabályait és a nyelvi rendszerben meghatározó törvényszerűségeit kutatja” (Kocsány 2002:146).

A szövegnyelvészet kutatási területére sorolható a lexikai kohéziós mintázat, azaz az egyes nyelvek szövegszinten megjelenő tulajdonságának vizsgálata. A kontrasztív szövegnyelvészeti vizsgálat újszerűségét az az új megközelítés adja, amely összehasonlító elemzés segítségével a fordított szövegekre jellemző lexikai kohéziós eltolódást vizsgálja. Az elemzés a magyar nyelv esetében bizonyítani kívánja, hogy a fordítás során a célnyelvi szövegek kohéziós eszközhasználatában bekövetkező változások az eltolódásjelenségek sorába tartoznak, mivel a fordított szövegek kohéziós mintázata nem egyezik a célnyelven írt autentikus szövegek kohéziós mintázatával. A korábban kis szövegmintán, kézi annotálással végzett elemzést most nagyobb korpuszon, automatikus eszközök segítségével ismételt meg, ami egyértelműen bizonyítja a fordított szövegekre jellemző kohéziós mintázat egyezőségeit és különbözőségeit az autentikus célnyelven írt szövegek kohéziós mintázatához képest. Meggyőződésem szerint a fordítás során keletkező szövegek lexikai kohéziós sajátosságainak szövegnyelvészeti elemzése nagymértékben hozzájárul a fordított szövegekről alkotott tudásunk kiteljesítéséhez, s ez által további szempontokat kínál a fordításértékelés és a fordítóképzés számára. A kutatás végeredményeként célul tűzttem ki az autentikus magyar és a fordítás eredményeként létrejött célnyelvi magyar szövegek lexikai kohéziós tipológiájának megalkotását a vizsgált nyelvpárok, szövegtípusok és műfajok

bontásában, különös tekintettel az autentikus és a fordítási szövegprodukciónak megkülönböztető eltolódásokra.

Kutatásom azon korpuszalapú vizsgálatok közé sorolható, amelyek jellegzetessége az írott nyelv szövegépítő sajátosságainak feltárása, ezen belül a variabilitást befolyásoló szövegkörnyezeti tényezők és a nyelvhasználat szabályszerűségeinek leírása. Az elemzés főbb területei az univerzálé kutatásba illeszkedve a fordított és nem fordított szövegek összevetése és a nyelvpár specifikus fordítói magatartásra fókuszáló forrásnyelvi és célnyelvi magyar szövegek nyelvi jellemzőinek összehasonlítása. A korpusznyelvészeti technikák a fordításra vonatkozó hatékony általánosításokat tesznek lehetővé, többek között a fordítási univerzálék feltárását, amelyek a fordított szöveget az autentikus szövegprodukciónak megkülönböztető általános tulajdonságok (v.ö. Baker 1993, 1996).

Az elemzés – mint minden korpuszalapú fordításkutatás (corpus-based translation studies) – empirikusan közelíti meg a nyelv leírását, ragaszkodik a használati előfordulások elsődlegességéhez, mivel a korpuszban összegyűjtött szövegek a ténylegesen használt nyelv olyan megtestesülései, amelyek a nyelvész beavatkozása nélkül jöttek létre. E tulajdonságoknak köszönhetően a korpuszalapú fordításkutatás kiválóan alkalmas a fordítás eredményeként létrejött szövegek tulajdonságainak leírására, és a célnyelvi- és forrásnyelvi, valamint az autentikus- és fordított szövegek közötti különbségek megvilágítására (v.ö. Baker 1995).

4 Az értekezés felépítése

Az értekezés részei: *Bevezetés, Elméleti háttér, A kutatás leírása* és az *Összefoglalás*. Az elméleti háttér négy alfejezetre tagolódik: (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*.

5 A kutatás módszere

5.1 Előzmény

A jelen kutatást megelőzte egy három éves pilot projekt, amely a kézíleg annotált kisméretű német–magyar párhuzamos korpusz gépi lekérdezésével kimutatta a német-magyar nyelvpár esetében az autentikus és a fordított szövegekre jellemző lexikai kohéziós mintázatot.

5.1.1 Új megközelítés a kohézió-kutatásban: összehasonlító elemzés

Pilot kutatásom a fordítástudomány és kohézió-kutatás területének összekapcsolásával újszerű megközelítést kínált, melynek lényegét az összehasonlító elemzés, illetve a fordítás során a kohéziós elemek használatában bekövetkezett változások feltárása jelentette. Ez a fajta összehasonlító elemzés nem véletlenül nem valósulhatott meg a szövegkutatás korábbi időszakában, hiszen hagyományos eszközökkel végzett kohéziókutatás keretein belül egy ilyen összehasonlítás igen körülményes és időigényes.

Nem tartom túlzottnak azt a megállapítást sem, hogy a fordításokban realizált lexikai kohéziós elemek egybevetése az autentikus szövegprodukcióna jellemző megvalósulásukkal alapvetően az új technológia elterjedésének és hozzáférhetőségének köszönhetően valósulhat meg. A korpusznyelvészeti módszer új távlatokat nyit az alkalmazott nyelvészeti kutatások számára, aminek segítségével a csak felhasználói szintű számítógépes ismerettel rendelkező nyelvész-kutató számára sem irreális célkitűzés a szövegek, illetve szövegrészletek párhuzamosítása.

5.1.2 Szövegválasztás

Pilot kutatásomhoz hivatalos internetes honlapokról töltöttem le szövegeket, amelyeket Microsoft Word formátumban archiváltam. Ezek olyan leíró szerepet betöltő informatív szövegek, amelyek egy-egy intézményt, tájegységet, szervezetet stb. mutatnak be. Szövegválasztásomban három fő szempont érvényesült. (i) A kiválasztott szövegek stílusa hasonló a leíró jellegből fakadóan, ami kizárja az eltérő stílusú szövegekre jellemző, pragmatikai és stiláris okokra visszavezethető különbségeket a lexikai kohéziós elemek használatánál (v.ö. Szikszainé Nagy 1999: 69). Ez a szempont azért lényeges, mert „a kohézió csak részben tekinthető a szöveg inherens (bennfoglalt) tulajdonságának. A kohézió foka / mértéke szövegtípusonként más és más lehet“ (Vass 2002: 9). E tényező figyelembevételét a gyakorisági vizsgálatok, és azok eredményeinek összehasonlíthatósága indokolja. (ii) A szövegek német fordítását feltehetően profi fordítók készítették, hiszen ezek az információk komoly marketing érdekeket szolgálnak. A fordítások hasonló és jó minősége a garancia arra, hogy a vizsgált célnyelvi szövegek esetében a lexikai kohéziós elemek használatában tapasztalható eltérés nem a suta fordítói kompetenciával, hanem a nyelvek közötti, a kohéziós eszközök alkalmazásakor megfigyelhető eltéréssel magyarázható. (iii) A harmadik, ám nem elhanyagolható szempont, hogy a nyilvánosan hozzáférhető szövegek letöltésével időt takarítottam meg.

5.1.3 A Károly-féle taxonómia

Kutatásom elméleti háttéréül Károly Krisztina 2002-ben kidolgozott rendszerét választottam, amely a lexikai ismétlődés szövegszervező funkcióját vizsgálja. Taxonómiája a szöveg kohézióját biztosító lexikai egységként az olyan egységet definiálja, amelynek jelentése nem vezethető vissza alkotóelemeinek jelentésére.

Károly a kohéziós kapcsolat típusait vizsgálva két alapvető típust különböztet meg: a lexikai és a szöveg alapú relációkat. Az első csoportot is

további két alcsoportra osztja, és különbséget tesz azon esetek között, (i) amikor a lexikai egység változatlan formában ismétlődik (repetition), illetve (ii) amikor a lexikai egységnek csak az információtartalma ismétlődik meg teljes egészében vagy részben egy vele szemantikai kapcsolatban álló lexikai egység formájában (synonymy, opposites, hyponymy, meronymy). (iii) Elemzési modelljében külön kategóriát képviselnek a szöveg alapú relációk (instantial relation). Károly az ismétlés, a szinonima és az ellentét jellegét tekintve különbséget tesz egyszerű és derivációs alakok között. Derivációról akkor beszél, ha a lexikai egység képzett szó formájában ismétlődik (Károly 2002: 95-107).

A Károly-féle taxonómia tehát összesen kilenc lexikai kohéziós kategóriát különböztet meg a szövegfelszín szintaktikai-szemantikai hálójának vizsgálatakor. E kilenc kategória segítségével végeztem én is pilot-kutatásomat a szövegszerveződés szempontjából meghatározó szemantikai kapcsolatok feltárására.

5.1.4 A gépi feldolgozás előkészítése

A kiválasztott szövegek feldolgozását, azaz a kohéziós eszközök bejelölését a CLaRK System (Simov 2004) elemzést segítő gépi eszközzel végeztem. A CLaRK program jellegzetessége, más hasonló nyelvészeti elemzésre írt programokhoz hasonlóan, hogy lehetővé teszi a számítógépes feldolgozásra szánt szövegek strukturált kódolását. Ez annyit jelent, hogy míg az eredeti szöveg integritása megmarad, annotációként többletinformációt, azaz a nyelvészeti elemzés eredményeit vihetjük be a szövegbe (Prószéky, Kis 1999). Ilyen szövegbe ágyazott többletinformáció a jelen elemzés esetében a mondathatárok bejelölése, illetve az ismétlődő lexikai elemek kiválasztása, és a kohézió típusának megjelenítése (1. ábra).


```
<Sentence No="No15">A
<LexicalUnit "InstantialRelation" reference="H.1.1.">
Magyar Fordítóház Alapítvány
</LexicalUnit>
köszönettel
<LexicalUnit "SimpleRepetition" reference="H.1.2.">
fogad
</LexicalUnit>
adományokat: ösztöndíjakat,
<LexicalUnit "Hyponymy" reference="H.1.6.">
könyveket
</LexicalUnit>.</Sentence>
```

1. ÁBRA: A pilot projekt annotált szövegrészlete

Az eligazodásban és a szemléletességben segít, ha a CLaRK programban az eredeti szövegrészeket és az annotált információkat a képernyőn különböző színnel jelenítjük meg. Itt most a szemléltető ábrán eltérő színek helyett különböző betűtípusok segítenek az annotált információk közé ékelt elemzett szöveg elkülönítésében.

A vizsgálni kívánt kategóriákat a kutatás kezdetekor a CLaRK rendszerben definiálhatjuk. Kutatásunk célja érdekében tetszőleges számú és elnevezésű kategóriát határozhatunk meg, illetve rendelkezhetünk arról is, ha sorszámmal akarjuk ellátni őket, és/vagy valamilyen jellemző tulajdonságo(ka)t kötelező vagy választható jelleggel kívánunk hozzájuk rendelni. Így a konkrét elemzés során a kurzor mozgatásával már csak a legördülő ablakokból kell a kívánt értékeket kiválasztanunk.

5.1.5 Elemzés és annotálás

Az elemzés során a CLaRK rendszerbe beolvasott folyamatos szövegekben a mondathatárokat jelöltem meg, illetve a mondatokat sorszámmal láttam el (2. ábra). A mondathatárookra és mondatszámra vonatkozó többletinformáció a korpusznyelvészeti program számára is értelmezhető adat, ami elengedhetetlen feltétele az eredeti - és a fordított dokumentum mondatszintű párhuzamosíthatóságának. A mondatsorszám (No6, No7, ...) jelzi, hogy itt a magyar szöveg hatodik, hetedik, stb. mondatát látjuk (2. ÁBRA ábra), míg a

német szöveg megfelelő fordított mondatainak sorszámai mellett egy „d” betű található.

```
<Sentence No="No6">A házban jól felszerelt konyha található</Sentence>
<Sentence No="No7">A vendégek ösztöndíjuktól fedezik ellátásuk
költségeit.</Sentence>
<Sentence No="No8">A Fordítóházat a legjelesebb kortárs magyar
képzőművészek alkotásai díszítik.</Sentence>
<Sentence No="No9">A fordítóknak a Magyar Fordítóház Alapítványhoz kell
eljuttatniuk pályázatukat.</Sentence>
<Sentence No="No10">A fordítandó mű terjedelmétől függően 2-4-6-8 hetes
időtartamra szóló ösztöndíjat nyerhetnek el.</Sentence>
```

2. ÁBRA: A mondatok bejelölése és sorszámozása a pilot projektben

A következő lépésben az első mondatot elemeztem. Bejelöltem azokat a lexikai egységeket, amelyek a kohéziós kapcsolatok, a Károly-féle ismétlődési taxonómia szempontjából meghatározóak (3. ábra). Ezeket a későbbiekben ismétlődő lexikai egységeket (RLU = repeated lexical unit) hivatkozási számmal láttam el. E hivatkozási számok (H.1.1, H.1.2, H.1.3, stb.) utalnak arra, hogy a magyar szöveg első mondatának hányadik (1. 2. 3. stb.) eleméről van szó, illetve hasonló módon jelöltem meg a többi mondat ismétlődő lexikai elemeit is. A hivatkozási számok segítségével tudtam a későbbi ismétlődéseknél az előző mondatok lexikai elemeire visszautalni.

```
<Sentence No="No1">A
<RepeatedLexUnit RLU="H.1.1.">Magyar Fordítóház</RepeatedLexUnit>
1998. januári megnyitása óta
<RepeatedLexUnit RLU="H.1.2.">fogadja</RepeatedLexUnit> azokat a külföldi
<RepeatedLexUnit RLU="H.1.3.">fordítókat</RepeatedLexUnit>, akik
<RepeatedLexUnit RLU="H.1.4.">magyar</RepeatedLexUnit>
<RepeatedLexUnit RLU="H.1.5.">szépirodalmi</RepeatedLexUnit>, bölcséleti,
társadalomtudományi
<RepeatedLexUnit RLU="H.1.6.">műveket</RepeatedLexUnit>
<RepeatedLexUnit RLU="H.1.7.">fordítanak</RepeatedLexUnit> saját
<RepeatedLexUnit RLU="H.1.8.">nyelvükre</RepeatedLexUnit>.</Sentence>
```

3. ÁBRA: Az első mondat ismétlődő lexikai egységei a pilot projekt egyik szövegénél

Ezt követően az első mondatban kiválasztott lexikai egységek (lexical unit) ismétlődéseit kerestem meg az összes többi mondatban. Ezeket a kapcsolódás

jellegét tekintve elemeztem, és minden esetben bejelöltem a Károly-féle taxonómia 9 kategóriájának (egyszerű és derivációs ismétlés, egyszerű és derivációs szinonima, egyszerű és derivációs antonima, meronima, hiponima és szöveg-alapú relációk) megfelelő kohéziós kapcsolatot. A kohéziós kapcsolat meghatározását egy olyan referenciális szám követi, ami az első mondatban hivatkozási számmal ellátott ismétlődő egységekre utal vissza. Ennek a referenciális számnak a lekérdezés során lesz jelentősége.

Annak mintájára, ahogy az első mondat ismétlődő kapcsolatait meghatároztam, sorban elemeztem a szöveg többi mondatát és megkerestem mindegyiknek az ismétléses kapcsolatait az összes többi mondattal (1. ábra). Így jutottam el a teljes szöveg elemzéséhez, és minden mondatot mindegyik mondattal összevettem.

A pilot projekt során 3 teljes magyar szöveget és ezek fordításait elemeztem. A szövegek mondatszám 11 és 17 között váltakozott, ami az egyes szövegekhez kapcsolódó elemzések számát is meghatározta, ami 10 és 16 közötti elemzést jelentett szövegenként, eggyel kevesebbet, mint az adott szövegre jellemző mondatszám, hiszen egy elemzés legalább 2 szövegmondat lexikai elemeinek egybevetését jelenti. Összesítve 117 kiindulási mondat esetében vizsgáltam az adott mondat lexikai elemei és a többi szövegmondat lexikai elemei között fennálló kohéziós kapcsolatot.

5.1.6 Lekérdezések

Lekérdezéseimmel sikerült a kirajzolódó tendenciákat szemléltetni és a kiindulási hipotézisemet alátámasztani, miszerint a fordítási tevékenység befolyásolja a szövegfelszín strukturális összetartozását biztosító kohéziós elemek átváltását, illetve megőrzését. A kapott eredmények tanúsága szerint a lexikai kohéziós eszközök disztribúciójában tapasztalható különbségek együtt járnak a szövegek kohéziós sűrűségbeli differenciáltságával. Eredményeimet – amelyek a kutatás automatizálására és nagyobb szövegmennyiség vizsgálatára ösztönöztek – több nyilvános előadáson ismertettem valamint magyar és idegen nyelven írt tanulmányokban is publikáltam (Pécs 2007a, 2007b, 2007c, 2007d, 2009a, 2009b, 2011).

5.2 A kiválasztott módszer

5.2.1 *Miért a szöveg produkció, miért nem a percepció van a kutatás fókuszában?*

A leíró fordítástudomány (Descriptive Translation Studies) hagyományos Holmes-i hármass felosztásából kiindulva az értekezést az eredményközpontú leíró fordítástudományi (Product-oriented DTS) munkák közé lehet besorolni (Holmes 1972). Bár napjainkban a folyamatközpontú (Process-oriented DTS) és funkcióközpontú (Function-oriented DTS) fordítástudományi terület is nagy népszerűségnek örvend, a jelen kutatási kérdések megközelítését csak a fordítási folyamat eredményeként keletkezett célnyelvi szövegeket vizsgáló eredményközpontú leíró fordítástudományi módszer teszi lehetővé, mivel csak a célnyelvi szöveg alapján kaphatunk pontos képet a fordító végső döntéseiről (v.ö. Hatim, Mason 1990: 3).

A fordítás folyamatával egyrészt a fordítók munkakörülményeit, az adott társadalmi rendszerben való megítélését kutató szociológiai vizsgálatok, másrészt a fordítók fejében végbemenő folyamatokat feltáró pszichológiai vizsgálatok foglalkoznak. Ezek a vizsgálatok következtetéseiket a fordítókkal és megbízóikkal készített interjúk, kérdőíves felmérések alapján, illetve a fordítási folyamat megfigyeléséhez előre összeállított szempontrendszer szerint kitöltött jegyzőkönyvek összesítése alapján vonják le a döntéshozatal menetére és részleteire, a fordítói megakadásokra, hezitálásokra és önkorrekcióra vonatkozóan. További kedvelt kutatási lehetőséget kínál a fordítók által írt bevezetők, előszók, utószók, lábjegyzetek feldolgozása, illetve a fordítói memoárok kutatása.

Választott kutatási témám, amelynek célja a fordított és nem fordított (autentikus) célnyelvi szövegek szövegépítő sajátosságainak egybevetése két párhuzamos szövegkorpusz összehasonlítható elemzésén alapul, ami az eredményközpontú leíró fordítástudományi módszer keretei közé sorolható. Az összehasonlító vizsgálatok ezen típusa lehetővé teszi a fordított és nem fordított

célnyelvi szövegek esetében bizonyos szövegtulajdonságok különbségeinek kvantitatív módszerekkel mérhető relatív eloszlásának kimutatását. Ez a megközelítés a kutatás elemzési szakaszában alkalmazott számítógépes nyelvészet számára érdekes és megragadható feladatot jelent, a kutatás számára pedig nagyobb szövegtulajdonságok végzett, ezért jobban általánosítható eredményekkel kecsegtető kvantitatív mutatók kinyerését teszi lehetővé.

5.3 Újszerűség

Vizsgálatomat az alkalmazott nyelvészet számára új perspektívát kínáló nyelvtechnológiai alkalmazások (Human Language Technologies) segítették, amelyek „a nyelvelírás és a szoftvertchnológiai eszközök találkozásának” (Prószéky 2005) köszönhetően a szövegfeldolgozás nyelvi tartalom szerinti megvalósítását teszik lehetővé. A szövegek intelligens kezelése (fogalmazástámogatás, kereséstámogatás, megértéstámogatás, fordítástámogatás) iránt támasztott egyre növekvő igénynek köszönhetően a számítógépes nyelvészet napjainkban egyre nagyobb lendülettel fejlődő területté nőtte ki magát, ami egyben számtalan új gyakorlati nyelvtechnológiai alkalmazás megjelenését is eredményezi. Ezen alkalmazások megjelenése és elterjedése a nyelvtechnológiai iparágban (pl. morfológiai elemzés) korántsem biztosítja felhasználásuk elterjedését a nyelvészeti kutatásokban.

A jelen kutatás újszerűsége a kutatott témának, a lexikai kohézió vizsgálatának nyelvtechnológiai alkalmazások segítségével történő feltárásán alapszik. A kutatás során a következő lépések automatizálását lehetett nyelvtechnológiai eszközökkel megvalósítani: (1) a bekezdések bejelölése, (2) mondatra bontás, (3) tokenizálás, (4) morfológiai elemzés, (5) szófaji egyértelműsítés, (6) a főneveknek a bekezdésen belüli ismétlés relációinak megtalálása, (7) a főneveknek a bekezdésen belüli lexikai kohéziós relációinak feltárása a WordNet-élek mentén haladva. A számítógépes nyelvészeti eszközök alkalmazásának köszönhetően a következő területeken lehet a jelen 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, korpuszon 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 szövegtörzsek ellenére is.

5.4 Eredmények

A kutatás igen fontos eredménye, hogy autentikus és fordított magyar szövegek esetében feltárja szövegtípus és nyelvpár függvényében a célnyelvi magyar szövegekre jellemző lexikai kohéziós mintázatot. A vizsgálat összehasonlító jellegéből következően az eredmények között egyaránt megtalálható a vizsgált szövegfajták esetében az autentikus magyar szövegprodukción és a fordítás eredményeképpen létrejött magyar szövegekre jellemző lexikai kohéziós tipológia, illetve nyelvpár specifikus bontásban az

autentikus és fordítási szövegalkotásra jellemző lexikai kohéziós mintázatok különbségei.

Mivel az elemzést a hipotézisek megfogalmazásához elegendő támpontot szolgáltató pilot projekt vizsgálati korpuszához képest most lényegesen nagyobb szövegtörzson, nyelvtechnológiai alkalmazások segítségével ismételttem meg, a kutatás kimeneteként felállított lexikai kohéziós mátrix egyértelműen bizonyítja a fordított magyar szövegekre jellemző kohéziós mintázat egyezőségeinek és különbözőségeinek tendenciáit az autentikus célnyelven írt magyar szövegek kohéziós mintázatához képest. Meggyőződésem szerint a fordítás során keletkező szövegek lexikai kohéziós sajátosságainak szövegnyelvészeti elemzése nagymértékben hozzájárul a fordított szövegekről alkotott tudásunk kiteljesítéséhez, és ezáltal további szempontokat kínál a fordításértékelés és a fordítóképzés számára.

A kutatás végeredményeként célul tűztem ki az autentikus magyar és a fordítás eredményeként létrejött célnyelvi magyar szövegek lexikai kohéziós tipológiájának megalkotását a vizsgált nyelvpárok és szövegtípusok bontásában, különös tekintettel az autentikus és a fordítási szövegprodukción megkülönböztető eltolódásokra. A szövegfajták és a nyelvpárok ilyen széles skálájának vizsgálata a kutatásban alkalmazott nyelvtechnológiai módszereknek köszönhető. A magyar nyelv esetében a nyelvtechnológiai eszközök alkalmazásán alapuló lexikai kohéziós szövegjellemzők feltárását megvalósító vizsgálatom úttörő jellegű, amivel hozzájárul a nyelvléírás korpusznyelvészeti eszközökkel megközelíthető szempontrendszerének bővítéséhez, illetve a mondathatárokon túlmutató kohéziós kapcsolatok automatikus feltárásának módját bemutató kutatás a gépi fordítás számára kínálja a WordNet formalizmus alkalmazhatóságának bizonyítását.

A kutatás eredményei között kell említeni az összehasonlíthatóság feltételeit kielégítő összehasonlító korpusz megalkotását, amelynek autentikus és fordított magyar szövegeket tartalmazó alkorpuszai az egyes szövegfajták és nyelvpárok tekintetében kiegyensúlyozottak. Az autentikus és fordítási alkorpuszok szétválasztása már önmagában is külön eredménynek tekinthető, mivel a hazai korpuszépítési gyakorlatra nem volt eddig jellemző, hogy a magyar

szövegeket e kritériumok alapján külön alkorpuszokba csoportosítsa. Így például a magyar nyelv jelenleg ismert legnagyobb reprezentatív korpusza, a *Magyar Nemzeti Szövegtár* (<http://corpus.nytud.hu/mnsz/>) sem tesz különbséget a magyar szövegek között aszerint, hogy azok az elsődleges vagy a másodlagos szövegalkotás eredményeképpen keletkeztek-e.

6 Kutatási kérdések és hipotézisek

6.1 Kutatási kérdések

Az előzőekben leírt célok alapján a következő kutatási kérdéseket fogalmaztam meg:

- Kimutatható-e adott nyelvpárok között a fordítás következtében megváltozó lexikai kohéziós eszközhasználat esetében valamilyen törvényszerűség a kohéziós eltolódás jellegét illetően?
- Milyen mértékben egyezik az adott nyelvű autentikus szövegek kohéziós mintázata az azonos célnyelvi szövegekével?
- Azonosíthatók-e nyelvpár specifikus fordításnyelvi kohéziós eltolódások?

6.2 Kutatási hipotézisek

A vonatkozó szakirodalom tanulmányozását követően, illetve a pilot projekt tapasztalatainak és az eddig kapott elemzési eredményeim tükrében a következő hipotéziseket fogalmaztam meg:

- A fordított szövegekre jellemző kohéziós eszközhasználat nem egyezik az autentikus szövegprodukcióna jellemző kohéziós eszközhasználattal.
- A célnyelvi szövegek kohéziós eszközhasználatát a forrásnyelvi szöveg kohéziós eszközhasználata befolyásolja.

- A célnyelvi szövegek kohéziós eszközhasználatának eltolódása azonos szövegtípus esetén nyelvpár-függő.
- A célnyelvi szövegek kohéziós eszközhasználatának eltolódása fordítási tendencia.

7 A kutatás haszna

A kutatás interdiszciplináris jellegénél fogva három tudományterület eredményeihez járul hozzá. Ezek a fordítástudomány, a szövegnyelvészet és a korpusznyelvészet. A leíró fordítástudomány számára kiemelkedően fontos a célnyelvi szövegek jellemzőinek és szövegépítő sajátosságainak, ezen belül a fordítási szövegprodukciónak eltolódás (shift) jelenségeinek feltárása a célnyelvi szövegek különböző szintjein. A kutatás központi feladata a témát szűkebben szemlélve a kohéziós eszközök terén tapasztalható eltolódások feltárása, amit az autentikus szövegek jellemzőinek és a fordítási szövegtermékek sajátosságainak egybevetésével vizsgál.

A fordításkutatás számára a vizsgálat több új szempontot ad a célnyelvi szövegalkotás leírásához, mivel megvalósítja a célnyelvi szövegalkotásra jellemző kohéziós eszközhasználat leírását, hozzájárul a fordított szövegek nyelvi normáinak és a fordításízűség egyik okának feltárásához, illetve felhívja a figyelmet a mondathatáron túlmutató összefüggések fontosságára. A kutatás eredményei relevánsak a fordításoktatás számára is, mivel segíthetik a megfelelő kohéziós eszközhasználat elsajátítását és tudatosítását a fordításoktatásban, a kompetencia-elsajátítást az idegennyelv-tanulásban, az anyanyelvi tudatosítást a fordítóképzésben, illetve adatokkal alátámasztott szempontokat szolgáltatnak az értékelési kritériumok (hibatipológia) kidolgozásához. Az értekezés eredményei hasznosak továbbá a gépi fordítás számára is, mivel hozzájárulnak a párhuzamos nyelvi szerkezetek elemzéséhez és ezáltal egy hatékonyabb nyelvfeldolgozási modell lehetőségéhez, illetve az emberi tudás gép számára jobban értelmezhető leírásához. Mivel a gépi fordítás számára jelenleg komoly kihívást jelent az

egybefüggő szövegek minden egyes lexikai elemének a szövegkörnyezetnek megfelelő jelentésének kiválasztása, a kutatásban bemutatott WordNet formalizmus alkalmazása a mondathatárokon túlmutató kohéziós kapcsolatok azonosítására hatékony megoldást kínálhat ennek a problémának a megoldásához.

A kutatás a szövegnyelvészet számára is új adatokat szolgáltat, mivel most történt meg először ilyen nagy mennyiségű magyar nyelvű szöveg lexikai kohéziós elemzése és az autentikus és célnyelvi szövegprodukcóra jellemző kohéziós mintázat különbségeinek feltárása és leírása. Az értekezés támpontot nyújt a fordítási szövegprodukciónál szövegtípus szerinti szövegszemantikai jellemzőinek leírásához is. Kutatásom újszerűsége, hogy a magyar szövegek lexikai kohéziós vizsgálatánál a számítógépes nyelvészet legmodernebb eszközkészletét felhasználva az igen időigényes kézi annotálás helyett az automatikus kategória felismerést választottam, ami nagy mennyiségű szöveg hatékony elemzését tette lehetővé.

Az értekezés a korpusznyelvészet számára is fel tud mutatni releváns eredményeket, mivel a kutatás keretében létrehoztam egy 3.050.000 szövegszóból álló fordított magyar korpuszt (FMK). A FMK szövegtípus és forrásnyelv szerinti elrendezésben tartalmazza a magyar célnyelvi szövegeket. Megalkottam a célnyelvi magyar korpusz szövegeinek legfontosabb bibliográfiai adatait tartalmazó header annotációt, amely a későbbiek során adaptálható a korpusz további bővítésekor. A kutatás hasznai között kell említeni a WordNet formalizmus felhasználását a mondathatárokon túlmutató szövegszemantikai relációk feltárásában, ami egyrészt alkalmat adott a HuWordNet ontológia működésének nagyméretű szövegkorpuszon való tesztelésére, másrészt a kutatás a WordNet adatbázis relációinak felhasználásával a mondathatárokon túlmutató lexikai-szemantikai kapcsolatok meghatározására kínál automatikusan megvalósítható gépi eljárást.

II. Elméleti háttér

1 A fordított szöveg mint a tudományos vizsgálat tárgya

1.1 Bevezetés

Napjaink kommunikációval átszótt világának mindennapi működésében és működtetésében egyre hangsúlyosabb a kultúraközi kommunikáció jelentősége és szerepe. Erre a jelenségre irányította a figyelmet a közelmúltban lezajlott Interkulturális Párbeszéd Európai Éve (2008) is. A kultúraközi kommunikáció egyik meghatározó ismérve a fordított szövegek számának nagymértékű növekedése, ami a hétköznapi életet is jellemzi meghatározó arányaival, és ezáltal igen komoly hatással van az autentikus szövegprodukcóra és az anyanyelvhasználatra. A fordítástudomány területén végzett kontrasztív szövegnyelvészeti elemzések ennek következtében nagyon fontos szerephez jutnak az autentikus és a célnyelvi szövegprodukcó leírásánál és tudatosításánál. Az értekezés fordítástudományi elméleti háttérét tárgyaló fejezetben azokat a szempontokat gyűjtöttem össze, amelyek a szakirodalom alapján kiemelkedően fontosak a fordított szövegek vizsgálatával kapcsolatban.

1.2 A fordított szövegek vizsgálata

A fordításról való gondolkodás már több ezer éves múltra tekint vissza (Ciceró, Szt. Jeromos, Horatius, Quintilianus, Boéthius, Leonardo Bruni, Luther, Rotterdami Erasmus, Herder, Goethe, Schleiermacher, Arany János, Brassai Sámuel, Rajnis József, Batsányi János, Szász Károly ...), amire a XX. századig elsősorban az irodalmi megközelítés és a preskriptív normatív szemlélet volt a jellemző. A múlt század elején például Radó elméleti ismereteit és gyakorlati tapasztalatait összegezve fogalmazza meg a fordítóval, és a műfordítással szemben támasztott elvárásait, mely szerint legyen „költői dikciónk folyékony, sima, magyaros“ (Radó 1909: 133), vagy a „műfordítást, ép úgy mint a költészet bármely ágát, nem lehet és nem szabad

gyárilag úzni; hangulat kell a költő tolmácsolásához is, türelem és kitartás a kész munka csiszolásához“ (i. m.: 135).

Az irodalmi fordítások mellett egyre nagyobb arányban megjelenő szakfordítások következtében a XX. század közepétől megjelenik a nyelvészeti fordítástudomány. Az önálló fogalom- és terminuskészlet kialakításában, illetve az önálló tudományos diszciplína megteremtésében olyan szaktekintélyek játszottak úttörő szerepet, mint Fjodorov (1953), Jakobson (1959), Mounin (1963), Revzin & Rozencvejk (1964), Nida (1964), Catford (1965). Ebben az időszakban a fordításelmélet középpontjában a fordíthatóság problematikája (Fjodorov: 1953) és az ekvivalencia fogalma (Nida: 1964, Catford: 1965) áll.

Az 1970-es évektől már a hétköznapi kommunikációban is egyre nagyobb szerepet kapnak a fordított szövegek és a fordítóknak egyre határozottabb célja, hogy az eltérő nyelvi-, kommunikációs- és kulturális közegek képviselői közötti megértést elősegítve a forrásnyelvi szöveget a célnyelvi szöveg olvasóközönsége számára minél könnyebben feldolgozható formában közvetítsék. A szakirodalom az átvitel folyamat-, funkció- és eredmény orientált tudományos szintű vizsgálatakor (Holmes: 1972) számba veszi a forrásnyelvi szöveg célnyelvre történő átültetését befolyásoló nyelvi és nem nyelvi tényezőket, és törekszik a fordítás eredményeként létrejött produktum lehetőleg objektív elemzésére. Az addigi – a forrásnyelvi szöveg elsődlegességét hangsúlyozó – gyakorlat preskriptív megközelítését a deskriptív (Descriptive Translation Studies) – a fordítási folyamat jellemzőit leíró, az átvitel funkcionalitásait elemző és a célnyelvi szöveg jellegzetességeit feltáró – megközelítés (Holmes: 1972) váltja fel.

A fordításelméletben két fő irányzatot különböztethetünk meg, a normatív (előíró) és a deskriptív (leíró) irányzatot. A normatív irányzat képviselői azért foglalkoznak a fordítás kutatásával, hogy megmondják, hogyan kell fordítani. A deskriptív irányzat képviselői nem akarnak tanácsokat adni, nem értékelik a fordításokat, őket a fordítás folyamatának leírása érdekli. Nem az, hogy hogyan kell fordítani, hanem az, hogy hogyan fordítunk valójában? (Klaudy 2007:45)

A tudományos igényű tanulmányok nyitnak a társtudományok (szociológia, pszichológia, szövegnyelvészet, korpusznyelvészet) vizsgálati tárgyai és eszközei felé, ami hozzájárul a tudományág (Holmes 1972) interdiszciplináris és empirikus jellegének kialakulásához. Az 1970-es és 1980-as évektől kezdve a fordítástudomány elsődleges célja már a jelenségek leírása és a jelenségek ismétlődésének és szabályszerűségeinek megmagyarázása és előrejelzése.

A fordítás gyakorlati tevékenység. A fordítástudomány azonban, ha tudománynak tartja magát, nem állhat meg a gyakorlat leírásának vagy gyakorlati receptek keresésének szintjén, hanem a fordítás során tapasztalt jelenségeket valamilyen elméleti keretben kell értelmeznie. (Heltai 2009: 13)

1.3 A leíró fordításkutatás megjelenése

A fordítások számának ugrásszerű növekedése maga után vont a fordítás mibenlétével, elveivel, követelményeivel foglalkozó tudományos gondolkodás iránti egyre nagyobb igényt. A fordítástudomány holmesi belső felosztása (Holmes: 1972) elméleti, leíró és alkalmazott területekre megteremtette a tudományág részterületeinek elkülönülését a gyakorlatban, és az igényt az egyes területek kutatási kérdéseinek és feladatköreinek meghatározására.

A fordítástudomány szempontjából történelmi jelentőségű *The Name and Nature of Translation Studies* című cikkében Holmes a diszciplína hármas belső felosztása után a leíró fordítástudomány (Descriptive Translation Studies – DTS) számára a jelenségek leírása mellett az általános elvek kialakítását jelölte meg alapcélkitűzésként. A leíró fordítástudomány két fontos feladatköre tehát: (1) a fordítási tevékenységgel és a fordításokkal kapcsolatos tapasztalatok leírása és (2) a tapasztalatok alapján azon általános elvek meghatározása, amelyek segítségével a leírt jelenségeket meg lehet magyarázni és előre lehet jelezni (Holmes 1972: 71).

Ebben a cikkben Holmes a leíró fordítástudományt további három nagy területre osztja, és rögzíti az eredmény-, a folyamat- és a funkcióközpontú kutatások mibenlétét. A három kutatási irány közül a kutatásomat meghatározó

szövegnyelvészeti megközelítés szempontjából az eredményközpontú fordításkutatás bír nagy jelentőséggel, amit Holmes a következőképpen definiált:

Product-oriented DTS, that area of research which describes existing translations, has traditionally been an important area of academic research in translation studies. The starting point for this type of study is the description of individual translations, or text-focused translation description. A second phase is that of comparative translation description, in which comparative analyses are made of various translations of the same text, either in a single language or in various languages. (Holmes 1972: 71)¹

Az eredmény-, azaz szövegeközpontú – a létező fordítások leírásával és összevetésével foglalkozó – kutatásoknak köszönhető többek között a célnyelvi szövegalkotás leírásának több fontos szempontja. A létező fordítások leírásának kutatása akkor kapott nagyobb lendületet, amikor a számítástechnika fejlődésének köszönhetően már az asztali gépek kapacitása is lehetővé tette nagyobb méretű szövegek tárolását, illetve a szövegek elemzését segítő automatikus és félautomatikus eszközök alkalmazását. A célnyelvi szövegek sajátosságainak feltárása automatikusan indukálta a fordításokra jellemző univerzális szövegépítő tulajdonságok meghatározására való törekvést, sőt mára a fordítástudomány „egyik legjobban kutatott területe a fordított szövegek univerzális jellemzőinek (korpuszok segítségével történő) vizsgálata” (Heltai 2009: 22).

A fordítás eredményeképp kapott szövegek gépi elemzése új lendületet adott a fordítási univerzálék kutatásának. A fordítási univerzálék kérdése a célnyelvi megközelítés előtérbe kerülésével párhuzamosan vált központi témává a fordítástudományban. Ha ugyanis a célnyelvi szöveget nem a forrásnyelvi szöveg másodrangú (netán rontott) változatának fogjuk fel, hanem saját jogán is méltó kutatási tárgynak tartjuk, akkor azonnal felmerül a kérdés, vannak-e a fordítás eredményeképpen keletkezett szövegeknek olyan általános (univerzális) sajátosságai, amelyek megkülönböztetik őket az eredeti, nem fordítás eredményeképpen keletkezett szövegektől. (Klaudy 2007: 38)

1.4 A fordítási szituáció mint interkulturális tevékenység

A célnyelvi szövegalkotás kiemelkedően fontos feladata a kommunikáció és a megértés biztosítása. Számos kutatás kísérli meg a fordítás folyamatát, eredményét és funkcióját vizsgálva a fordítási tevékenység során szerepet játszó összes nyelvi és nem nyelvi tényezőt számba venni. Közöttük több kontrasztív és szövegnyelvészeti, pszicho- és szociolingvisztikai vizsgálat található. A legtöbb kutató a fordító legfontosabb feladatai közé sorolja a különböző kultúrák és nyelvközösségek közötti szellemi hídépítést (Hönig 1997: 18), illetve közvetítést. Hönig maga a fordító legfontosabb készségei között említi a forrásnyelvi szöveg alapján a célnyelvbe és a célnyelvi kultúrába belesimuló szöveg előállításának képességét (Hönig 1997: 26).

Az egységesülő Európában – ahol 2010-ben összesen 1,5 millió oldalt fordítottak le a Bizottság hivatalos fordítói és ahol a 2700 tolmács fejenként 153 napot tolmácsolt a tavalyi év során – időszerű minden, az interkulturális kommunikáció jellemzőit – így a célnyelvi szövegek sajátosságait – feltáró kutatás, mivel „a beszélő és hallgató társadalomban elfoglalt helyének, a köztük levő távolságnak, a saját közösségükhöz való viszonyuknak a megítélése erősen kultúrától függő tényező” (Szili 2006: 39).

A kultúrák hasonlóságai és különbözőségei felőli közelítés jellemzi Vermeer translációelméletét is, aki minden emberi cselekvést, így a szöveget is a kultúrába ágyazottan szemléli. Hangsúlyozza, hogy a szöveg maga is egy cselekedet végterméke, és ezért kapcsolatban van az őt alkotó személy viselkedésével és annak kulturális hátterével. Vermeer a fordítást nem a szavak és mondatok egyik nyelvről a másikra történő átkódolásaként definiálja, hanem komplex cselekvési folyamatként, ahol egy új szituációban, új funkcionális, új kulturális és új nyelvi feltételek között kell tudósítani a formailag maximálisan megőrzött szövegről. A fordítónak tehát a forrásnyelvi szöveget úgy kell a célnyelvi kultúrába ágyazottan megfeleltetnie, hogy az általa feltételezett szövegalkotói elképzelést szem előtt tartja, és lehetőleg a saját világával összehangolja (Vermeer 1994: 33–34).

Ismertek olyan tanulmányok is, amelyek egészen konkrétan vetik föl a fordított szövegek, a fordítási szituáció interkulturális beágyazottságát. Budin például a nemzetközi szakkommunikációt és a többnyelvű információ menedzsmentet meghatározó transzkulturális tudásátadást, valamint a globalizáció és a vállalati szféra nemzetközivé válását követően a fordítási piacon egyre nagyobb teret nyerő gazdasági kommunikációt, és a szó tágabb értelmében vett, a vállalati kommunikációt is magában foglaló szakszövegfordítást az interkulturális szakmai kommunikáció egyik formájának tekinti (Budin 2002: 81). Földes szerint pedig „tárgyszinten talán úgy lehetne meghatározni az interkulturális kommunikációt, hogy akkor áll fenn, ha a nyelvi kommunikációt a résztvevő felek legalább egyike interkulturálisnak tekinti és kommunikációs hozzáállását, viselkedését ennek megfelelően alakítja” (Földes 2007: 25). Ez a „kommunikációs hozzáállás” kétségtelenül jellemzi a fordító tevékenységét, és a fordítói magatartás fontos kritériuma.

1.5 A fordított szövegek hatása az anyanyelv használatra

A XXI. század elejére a termelő gazdaságban betöltött meghatározó szerepe mellett a fordítási szektor tevékenysége – az Európai Unió alapdokumentumaival és munkarendjével összhangban – a közigazgatási szférában is felerősödött. Az élet minden területére egyre inkább jellemző fordítási tevékenység, valamint a dokumentumok terjesztését nagymértékben felgyorsító internet használat következtében a társadalom egyre szélesebb rétegeihez egyre nagyobb számban jutnak el fordított dokumentumok. Napjainkra a hétköznapi életet fokozott mértékben hatja át a fordítási tevékenység formájában megjelenő interkulturális kommunikáció, aminek közvetett és közvetlen hatása az anyanyelvhasználatra egyre egyértelműbben kimutatható. Az anyanyelvhasználat terén tapasztalt eltolódások késztetnek egyre több kutatót a célnyelvi szövegalkotás anyanyelvre gyakorolt hatásának tudományos vizsgálatára.

Mint azt Klaudy már 2001-ben előrevetítette a feltáró és leíró célú kontrasztív fordításelemzések új perspektívát nyithatnak az anyanyelvhasználat tudatosításában, és ennek szükségszerűsége egyre sürgetőbb. A fordítandó szövegmennyiség

exponenciális növekedése láttán előre lehetett látni, hogy „ezt hivatásos fordítókkal elvégeztetni lehetetlen. A nyelvek közötti megfeleltetés a jövőben nem a hivatásos nyelvi közvetítők monopóliuma, hanem a művelt magyar állampolgár napi rutinja lesz” (Klaudy 2007/2001: 55). Ehhez kapcsolódva a fordítás útján keletkezett szövegek kontrasztív vizsgálata, illetve a fordításokra jellemző fordításnyelv sajátosságainak feltárása az alkalmazott nyelvészet kiemelkedően fontos kutatási területe, melynek eredményei többek között hozzásegíthetnek „a magyar nyelvleírás egy korszerűbb, funkcionális megközelítéséhez” (i. m.: 55).

Heltai a fordító nyelvi normáival foglalkozó tanulmányorozatának bevezetőjében foglalkozik a fordításnyelv problematikájával:

A fordításnyelv kifejezéshez erős negatív konnotációk járnak: a fordításnyelv nem felel meg a célnyelv megszokott normáinak, idegenszerű, nehezen érthető – szóval nem olyan, mint amit eredetileg az adott nyelven írtak. Azok, akiket aggaszt anyanyelvük állapota, netán „romlása”, gyakran a fordításnyelvet jelölik meg az egyik fő veszélyforrásnak. (Heltai 2004: 408)

Heltai a tanulmány folyamán kimondottan hangsúlyozza, hogy a leíró nyelvész feladatkörébe nem tartozik bele az egyes nyelvi változatok „szépségének” megítélése, sőt semmiféle követendő előírás támogatása, csupán az egymás mellett élő nyelvváltozatok feltárása, leírása. Ugyanakkor azt is elismeri, hogy a fordítónak „az átlagos művelt beszélőnél jobb és tudatosabb anyanyelvi kommunikatív kompetenciával kell rendelkeznie” (Heltai 2005b: 169).

A fordításnyelv sajátosságainak feltárására hívja fel Károly is a figyelmet, amikor hangsúlyozza, hogy „a kultúrák közötti közvetítés igen lényeges lesz a fordítások szövegszintű sajátosságainak tárgyalása során is” (Károly 2007: 30). A célnyelvi szövegek szövegépítő sajátosságainak leírásánál azonban figyelembe kell venni, hogy e szövegek a kultúrák közti közvetítés termékei, s így „... a szöveget nem önmagában vizsgáljuk, hanem felhasználója világról és társadalomról alkotott tudása függvényében” (i. m.: 18).

1.6 A célnyelvi szövegek keletkezését befolyásoló tényezők

A fordítástudomány feladatai közé tartozik a fordított szövegekre jellemző szövegalkotási sajátosságok feltárása, és a tapasztalatok felhasználása a fordítóképzésben, az anyanyelvhasználat tudatosításában, illetve a számítógépes nyelvészetben. Az információ áramlás felerősödésének következtében a hétköznapi élet számos területén megtöbbszöröződött a fordított szövegek mennyisége (vásárlói tájékoztatók, hírek, uniós dokumentumok...), melyeknél azonban gyakran az autentikus célnyelvi szövegekre jellemző szövegépítési sajátosságoktól eltérő szövegépítési tulajdonságokat figyelhetünk meg.

A fordítás egyik legfontosabb tulajdonsága, hogy másodlagos beszédtevékenység, ami azt jelenti, hogy a fordításban sokkal több tényező van befolyással a nyelvi eszközök kiválasztására, mint az egynyelvű beszédtevékenységben. Az elsődleges (egynyelvű) beszédtevékenységben a nyelvi eszközök kiválasztására nagyon leegyszerűsítve három tényező van hatással: (1) a nyelvi rendszer, amely az eszköztárat nyújtja, (2) a nyelvhasználati szabályok, amelyek szabályozzák a válogatást ebből az eszköztárból és (3) a nyelvhasználat tág értelemben vett kontextusa, amely pótlólagos kötöttségeket jelent a mindenkori történelmi, társadalmi, földrajzi stb. viszonyoknak megfelelően. A másodlagos (kétnyelvű) beszédtevékenységben a fenti három tényező minimum megduplázódik. A fordítói döntésekben egyaránt szerepet játszik a forrásnyelv és a célnyelv rendszere, nyelvhasználati szokásai és kontextuális kötöttségei. De még ennél is bonyolultabb a helyzet: ugyancsak befolyásolja a fordítói döntéseket mindezeknek a tényezőknek az egymáshoz való viszonya, illetve az az elképzelés, amit a fordító intuitíven vagy tudatosan kialakított a fenti három tényező viszonyáról, azaz a két nyelv rendszerében, a nyelvhasználati szokásokban és a történelmi, társadalmi viszonyokban megnyilvánuló hasonlóságokról és különbségekről. (Klaudy 2007: 190)

A fordítási szövegalkotás fontos jellemzői között kell említeni, hogy a fordító a fordítási folyamat során az egyik kultúrában keletkezett forrásnyelvi szöveget reprodukálva hozza létre a célkultúrába ágyazott célnyelvi szöveget. Szövegprodukciónak és szövegprodukciónak egyszerre vannak jelen (v.ö. Károly 2007: 47), és pedig két különböző kultúrához kapcsolódva. A fordítási szövegalkotás tehát a kultúraközi közvetítés olyan speciális esete, amikor a fordító mindkét kultúrában egyidejűleg aktívan cselekszik, de e cselekedetek végterméke csak a célnyelvi kultúrában ölt testet.

A fordítási folyamat forrás- és célnyelvspecifikus befolyásoltságát vizsgálva Klaudy a fordítói viselkedést leíró szerzőktől veszi kölcsön a „fordítási viselkedés” (translational behaviour) terminust (Klaudy 2007: 196). A „fordítási viselkedés” ebben az esetben azt a jelenséget jelöli, amikor az adott forrásnyelvű szöveg másképpen viselkedik, ha különböző célnyelvekre kell lefordítani (i. m.: 196).

Hatim és Mason azt hangsúlyozza, hogy mivel a fordító a két kultúrában egyidejűleg cselekszik, ezért a két nyelv ismerete mellett bikulturális látásmóddal is rendelkeznie kell. E bikulturális látásmód segíti megoldani a jelentésbeli eltérésekből fakadó átviteli nehézségeket (Hatim, Mason 1990: 223–224). Megfogalmazásukban a fordítási folyamat, s ennek következtében – meggyőződésem szerint a fordított szöveg is – egyik meghatározó jellemzőjét annak a kettősségnek köszönheti, hogy a fordítási szövegprodukciónak bemenete az olvasási folyamat szokásos kimenetét jelentő információ. Ez az információ az aktuális olvasásélmény egyedi végterméke. A célnyelvi szöveget meghatározó tényezők között így számon kell tartani – esetlegessége ellenére is – a fordító olvasását is, ahogy ez Bassnett cikkében a fordítások értékeléséről is megjelenik: „For translation is essentially an act that involves not only writing, but also reading and what we see when we read a translation are the traces of each individual translator's own subjective interpretation of the source text” (Bassnett 2003: 42).¹¹

A fordító olvasása, és ezáltal megértése mindig a szöveg egészére irányul, azaz a fordítói olvasás szövegszintű művelet (Klaudy 2007: 139). A fordítói olvasás kapcsán fontos hangsúlyozni azt a különbséget, ami az elsődleges és másodlagos szövegalkotásra ebből a szempontból jellemző. Az elsődleges szövegalkotás során a szövegépítő az egymással valamilyen logikai kapcsolatban álló gondolatai mentén

haladva, a szövegfolyam kisebb egységekből történő felépítésével teremti meg a szöveg egészét. Ezzel szemben a fordító a fordítási folyamat szövegprodukciónak szakaszában már a szöveg egészét ismerve önti azt új nyelvi formába. Míg az elsődleges szöveg alkotója a kisebb szövegrészekből apránként építi fel a teljes szöveget, addig a másodlagos szövegprodukciónak során a fordító a szöveg egészéből kiindulva képezi le az egyes szövegrészeket. Ennek következtében számolni kell azzal a ténnyel, hogy a fordítói szövegprodukciónak egészét befolyásolja a teljes célnyelvi szöveg ismerete, s adott esetekben nagy valószínűséggel kimutatható már a célnyelvi szövegek elején az adott forrásnyelvi szöveg hatása akár a lexikaválasztásban, akár szókapcsolatok előnyben részesítésében, akár jelentések megfeleltetésében (Péchy 2010: 76).

A fordított szövegek relativitása már a fordítói olvasásélmény egyedisége kapcsán is felvetődött. Ez a relativitás-gondolat azonban Neubertnél még hangsúlyosabban jelentkezik, aki a fordított szövegek relativitásával a forrásnyelvi szövegre gyakran jellemző eltérő értelmezési lehetőségek, illetve az egyfajta értelmezés esetén sem ritka különböző célnyelvi megfeleltetések jelenléte miatt foglalkozik (Neubert 1994: 97). A fordított szövegek relativitását Paepcke viszont a forrásnyelvi szöveg kulturális és történelmi beágyazottságából fakadó, szövegszinten nem feltétlenül megfogható értelmezéssel kapcsolatos különbségeknek tulajdonítja (Paepcke 1994: 106–108). A fordított szövegek relativitását e szempontok mellett tovább árnyalja a fordítók szociokulturális kompetenciája, ami már az elsődleges szövegalkotást is nagymértékben befolyásolja. A szociokulturális kompetenciához köthető relativitás következtében a fordított szövegek feldolgozásához szükséges erőfeszítés nagysága eltérő, amit részben magyarázhat a szövegek különböző kognitív környezete is.

A kognitív környezetbe tartozik sok minden, amit máshol „kontextus” néven emlegetnek. Az elnevezés nyilvánvalóan kiemeli azt a tényt, hogy nem a kontextus önmagában (például a beszélő és a hallgató közötti viszony) határozza meg a megnyilatkozás értelmezését, hanem a kontextus mentális reprezentációja. A kognitív környezet fogalmába beletartoznak a

kommunikátorok ún. enciklopédikus ismeretei, a közvetlenül megfigyelhető adatok és az előző megnyilatkozások ismerete is. (Heltai 2009: 14)

Anélkül, hogy a fordított szövegeknek az eredetijükhöz viszonyított minőségi megítélését fölvetném, bátran csatlakozni merek ahhoz a felfogáshoz, amely a forrásnyelvi és a célnyelvi szöveg kognitív környezete közötti különbséget vallja, hiszen könnyen belátható, hogy a „fordító kognitív környezet nem azonos a szerzőével, kivéve, ha saját művét fordítja” (Heltai 2009: 20).

1.7 A célnyelvi szövegalkotásra jellemző sajátosságok

Az adott nyelv grammatikai és mondatalkotási szabályai szerint helyes, ám az anyanyelvi beszélő „szövegről alkotott intuitív elképzeléseinek” mégsem megfelelő célnyelvi szövegek tulajdonságainak feltárása központi témája a fordítási szövegek kontrasztív elemzésének (v.ö. Papp 1972/2006: 122–123). A célnyelvi szövegek tulajdonságait vizsgálva Klaudy egyértelműen állást foglal amellett, hogy „vannak, illetve elkülöníthetőek szövegszintű fordítási problémák, vagy legalábbis bizonyos fordítási problémák jobban köthetők a szövegszinthez, mint mások” (Klaudy 2007: 138). A fordítási szövegalkotás problematikájára fókuszálva Klaudy az elemzés során a fordítási folyamat három szakaszát különíti el: a megértést, az átvitelt és a megformálást. Az átvitel szövegszintű vizsgálata kapcsán megállapítja, hogy a fordító nem tudja magát teljesen függetleníteni a forrásnyelvi szöveg hatásától, mivel a szövegalkotás során a forrásnyelvi szöveg szövegalkotási sajátosságai hatnak a szövegszerűség kritériumaiért felelős, egyes intuitív módon megteremtődő szövegszintű tulajdonságokra:

Aki azt állítja, hogy a fordítás csak a forrásnyelvi szöveg megértéséből (dekódolásból) és a célnyelvi szöveg megalkotásából (kódolásból) áll, annak számára nincs értelme forrásnyelvi szövegsajátosságok átviteléről beszélni. ... Klaudy 1987-ben „kvázi helyességről” beszél, ami azt jelenti, hogy a forrásnyelvre jellemző szövegépítési sajátosságok nem kerülnek át a

célnyelvbe, de megváltoztatják a célnyelvi szöveg mondatszerkesztési sajátosságainak eloszlását. (Klaudy 2007: 139)

Az intuitív módon megteremtődő szövegszintű tulajdonságok közé kell sorolni a szöveg lexikai kohéziójának megteremtését, hiszen ennek létrehozására a szövegalkotó nem szentel külön figyelmet. A kohéziós eszközök egyes nyelvekre jellemző sajátosságai eredményezik a különböző nyelveknél az egymástól eltérő kohéziós eszközhasználatot (Halliday, Hasan 1976).

Blum–Kulka az explicitációs hipotézis kapcsán írja le, hogy a célnyelvi szöveg minden esetben explicitebb, mint a forrásnyelvi szöveg – függetlenül attól, hogy a forrás- és célnyelv között milyen rendszerbeli különbségek vannak. Az explicititás tehát inherens módon benne rejlik a nyelvi közvetítés folyamatában. Ennek okait Blum–Kulka a nyelvek közötti különbségekre és a fordítás folyamatára vezeti vissza, konkrétan fogalmazva a kohéziós eszközök alkalmazására, a fordítás során végrehajtott átváltásokra, a nyelvek különböző stilisztikai preferenciáira és a fordítói értelmezés alaposágára. Mivel a kohéziós eszközök ismétlését az adott nyelv rendszere határozza meg, az egyes nyelvek rendszerbeli különbsége megmutatkozik a fordítás következtében tapasztalható kohéziós eltolódásban, azaz a forrásnyelvi szövegben és a célnyelvi szövegben alkalmazott kohéziós eszközök eltéréseiben. Ezt a kvantitatív módszerekkel vizsgálható eltolódást tovább árnyalják a különböző nyelvek stilisztikai preferenciái bizonyos kohéziós eszközök iránt (Blum–Kulka 1986).

A célnyelvi norma jellegzetes jegyeitől való eltérések okai után kutatva meg kell említeni a fordító forrásnyelven és célnyelven történő gondolatának szükségszerű interferálását, hiszen „interkulturális fordítást hajt végre, amely egy adott kultúrában létező fogalmaknak egy másik kultúrába való átültetését jelenti” (Simigné 2008: 126). Mint arról már korábban is szó volt, a fordító a két kultúrában egy időben cselekszik, és a két kultúrára jellemző, a világról való eltérő gondolkodás egyszerre van befolyással a keletkező célnyelvi szövegre. Frege 1891-ben úgy fogalmazott, hogy a „gondolatok világának megvan a képmása a mondatok, kifejezések, szavak, jelek világában” (Frege 2000/1891: 226). Az 1970-es évektől megjelenő kognitív nyelvészet feladatai között ugyanakkor már hangsúlyosan szerepel „a világ nyelvi képének rekonstrukciójával” való foglalkozás, amelyet

Bañcerowski-t idézve a kognitív nyelvészet „a nyelvben rögződött objektív valóság interpretációjaként értelmez” (Bañcerowski 1999: 82). Számos modern nyelvészeti kutatás kiindulópontjául szolgál az a tény, hogy a különböző nyelvek eltérő módon ragadják meg a valóságot (pl. metafora kutatás). Ez az eltérő interpretáció a fordított szövegek keletkezésénél sem elhanyagolható, mivel ugyanazon gondolatok manifesztációja az eltérő nyelvek nyelvi jelei segítségével eltérő módon történhet.

Ezen a ponton ismét hangsúlyozom a fordított szövegek tanulmányozásának jelentőségét a gondolatok nyelvi jelként való megformálásának kutatásában, hiszen „a szó segít a valóságból elkülöníteni valamit, ami aztán szervesen beépül kép- és fogalomrendszerünkbe” (Papp 1979/2006: 167). A fordítások szövegépítési sajátosságainak tanulmányozása során a kontrasztív elemzéseknek köszönhetően felszínre kerülhetnek az egyes nyelvek esetében a denotatív világ szegmensei és az adott nyelv nyelvi jelei és gondolati kategóriái közötti összefüggések:

A fordításkutatás nagyon alkalmas eszköz a nyelvi formától a gondolatig és a gondolattól a nyelvi formáig vezető út kutatására. Egy nyelven belül ugyanis a nyelvi formát nagyon nehéz a gondolathoz viszonyítani, mert a nyelvi forma születése elválaszthatatlan a gondolattól. A fordítás esetében az eredeti szöveg felfogható úgy, mint a gondolat sajátosan kódolt formája. (Klaudy 2008: 43)

1.7.1 Fordításnyelv – kvázi helyesség

A fordított szövegek szövegépítő sajátosságaira már korábban is történt utalás. Az autentikus szövegproduktiótól az anyanyelvi olvasó számára feltűnően különböző célnyelvi szövegekkel a hazai és a nemzetközi kutatások is egyre gyakrabban foglalkoznak a fordított szövegek arányának növekedésével összhangban. „A fordított szövegek eltérései a nem-fordított szövegektől disztribúciós jellegűek: a fordított szövegek nyelvtanilag helyesek, de bizonyos szerkezetek vagy szavak nagyobb vagy kisebb gyakorisággal fordulnak elő, mint a hasonló típusú nem fordított szövegekben” (Heltai 2009: 23). Ezt a jelenséget a szakirodalom hol kvázi-helyességnek (Papp 1972/2006, Klaudy 1987, Vehmas-Lehto 1989, Lindström 1996, Räisänen 2000, Aaltonen 2003, Schmidt 2005), hol fordításnyelvnek (Gellerstam

1986, Blum-Kulka 1986, Santos 1995, Tirkkonen-Condit 2002), hol harmadik kódnak (Frawley 1984) nevezi.

A mondatszinttel szemben a szöveg elsődlegességét hangsúlyozó kutatások az 1970-es évektől kezdve egyre gyakrabban foglalkoznak az autentikus- (primer) és a fordítási szövegprodukcióna (másodlagos szövegprodukcióna) jellemző szövegépítési sajátosságokkal. A fordított szövegek esetében gyakran megfigyelhető az elsődleges szövegalkotás sajátosságaitól eltérő, vagy részben eltérő jellemző tulajdonság, amit a szakirodalom kvázi-helyességnek (quasi-correctness) nevez. Ezeket a szövegszinten jelentkező különbségeket a mondatok szintjén általában nem is lehet észlelni és/vagy kimutatni, de a célnyelvi szövegek eltérései az autentikus szövegprodukciónaól mégis érzékelhetőek. A kvázi-helyesség tehát az adott célnyelv grammatikai szabályainak tökéletesen megfelelő, a szövegszinten érzékelhető eltérések miatt mégis idegenül ható fordítások minőségére utal. A szövegszinten érzékelhető leggyakoribb különbségek okai között egyes szerkezetek, lexikai egységek szokatlan disztribúcióját és a szokatlan retorikai szerkezeteket szokás említeni. A fordításnyelv (translationese, translated language) jelzővel tehát a természetes anyanyelvi szövegalkotástól eltérő, az autentikus célnyelvi szövegprodukcióna nem belesimuló, stilisztikai eltéréseket mutató fordított szövegeket szokás illetni (v.ö. Péch 2011).

A kvázi-helyes fordításokban a fordítók az összes kötelező átváltási műveletet elvégezték, de a különböző célnyelvi lehetőségek szövegszintű mérlegelése nélkül. Fordításaikon ezért még pótlólagos szövegszintű átváltási műveleteket kell végezni annak érdekében, hogy 1. a mondat elemeinek ugyanolyan legyen a kommunikatív súlya a célnyelvben, mint amilyen volt a forrásnyelvben, 2. a fordítás eredményeképpen létrejött célnyelvi szöveg egésze olyan legyen, mint amilyenek az eredetileg célnyelven fogalmazott szövegek szoktak lenni. (Klaudy 1987: 36)

A fordított szövegek kvázi-helyes voltának kiküszöbölése kapcsán Papp a mondathatárokon túlmutató összefüggések figyelembevételének szükségességét hangsúlyozza (Papp 1972/2006: 133). Lindström a turisztikai brosúrák szövegeiben található szavak szófaj szerinti előfordulási statisztikai alapján jellemzi finn–angol

vonatkozásban az angol autentikus broszúrák statisztikai jegyeitől eltérő finnből fordított szövegeket kvázi-helyesnek (Lindström: 1996).

The idea of quasi-correctness is a very useful notion to summarise and define features that are located on the fringe between correct and incorrect in translated texts; it can be applied to a whole range of phenomena that are present in this kind of text, which we usually have difficulty describing and explaining properly. I shall make use of the notion of quasi-correctness to describe and characterize translated texts that are formally correct but lack in qualitative aspect. (Schmidt 1995: 20)^{III}

A Klaudy-féle fordítási szövegalkotási gondolatmenetet folytatva Balaskó 2007-ben tűzi ki célul eredeti angol, fordított angol és fordított magyar szövegeket tartalmazó korpuszalapú vizsgálattal a fordított szövegekre jellemző, valamint az autentikus szövegalkotási sajátosságoktól eltérő harmadik kód jellegzetességeinek meghatározását. A fordításnyelv újradefiniálása során megkülönbözteti a fordítási univerzáléktól a célnyelvi jegyek célnyelvi normától eltérő használatát:

A fordításnyelv két komponensből tevődik össze: egyrészt azokból a jegyekből, amelyek a célnyelvi rendszerhez tartoznak ugyan, de használatuk valamilyen téren különbözik a célnyelvi normától, a tipikus célnyelvi jegyek használatától (eltérő gyakoriság, disztribúció, nyelvi minták). A harmadik kód másik komponensét azok a jelenségek alkotják, amelyeket a leíró fordítástudomány fordítási univerzáléknak nevez (pl. az explicitáció, a szimplifikáció, a normalizáció. (Balaskó 2007: 161–162)

1.7.2 A szövegszintű eltolódások okai

Az eltolódások okainak vizsgálatának feltárásával foglalkozik Catford (1965), aki a formális megfelelést megkülönbözteti a szövegekivivalenciától. Míg a formális megfelelés szerinte a nyelvi rendszerek különbözősége folytán nagyon ritka, addig a leginkább a mondat szintjén megvalósuló ekvivalencia viszony az egymásnak

megfelelő forrásnyelvi és célnyelvi egységek szituáció szempontjából releváns jegyeinek azonosságán alapul. A formális megfelelőktől való eltéréseket Catford eltolódásokként értelmezi és az eltérések okát magában a fordítás folyamatában, a forrásnyelvi szöveg célnyelvi szöveggé alakulásában látja. Szintbeli eltolódásokról (level shifts) beszél, ha az adott szövegben az adott nyelvi szinten (nyelvtani vagy lexikai szint) működő forrásnyelvi elemnek egy másik nyelvi szinthez tartozó célnyelvi elem lesz az ekvivalense. A kategóriabeli eltolódások (category shifts) között különbséget tesz nyelvtani szerkezet különbözősége, a szórend változása miatt bekövetkező strukturális eltolódások (structure shifts), a morfémák, szavak, tagmondatok szintjén végbemenő egységbeli eltolódások (unit shifts), az adott szófaj elemének más szófaj elemével történő fordításából adódó szófajbeli eltolódások (class shifts) és a fordító saját választásán alapuló kifejezésbeli eltolódások (intrasystem shifts, changes of term) között.

A fordítást befolyásoló szociokulturális kötöttségeket Toury (1995) három alapvető csoportba sorolja, amelyek (i) a nyelvi rendszerek sajátosságain alapuló abszolút érvénnyel bíró objektív szabályok, (ii) az adott közösség tagjai által elfogadott általános normák és (iii) az egyén szubjektív döntéseit tükröző jellegzetességek. A kötelező és az opcionális átváltásokból adódó eltolódások ebben a rendszerben a normavezérelt kategóriába kerülnek.

A szövegszintű eltolódások okai között három alapvető típust különböztethetünk meg: (i) a fordítói hibákat, (ii) azokat az eseteket, amikor a fordító meg kívánja könnyíteni a célnyelvi befogadó szövegfeldolgozását (iii) és a célnyelv és forrásnyelv rendszeréből adódó különbségeket. A fordítói megértésből, transzferből, szövegalkotásból vagy figyelmetlenségből adódó hibák miatti forrásnyelvi és célnyelvi különbségek jellegüknél fogva inkább a fordításoktatás és a fordítói kompetencia fejlesztés területére tartoznak. Ezeket a fordítók szövegértési, szövegalkotási, idegennyelvi és anyanyelvi kompetenciájának fejlesztésével, illetve a fordítás közben aktivált szövegszintű műveletek (grammatikai, lexikai és mondat szerkezeti választások) tudatosításának begyakorlásával, illetve koncentrációs gyakorlatok segítségével lehet kiküszöbölni. Az eltolódások stilisztikai, ideológiai vagy akár politikai okokra vezethetők vissza azokban a használatnyelvészeti esetekben, amikor a fordítók meg kívánják könnyíteni a

célnyelvi olvasók szövegfeldolgozását. A fordítási szakirodalom több kutatási terület kapcsán is behatóan tárgyalja ezeket az eseteket. Itt kell említést tenni többek között a műfajkutatásról, az ekvivalencia kutatásról, a fordítási univerzálék kutatásáról (explicitáció, implicitáció, egyszerűsítés, normalizáció), a reáliakutatásról és a relevancia elméletekről.

A fordítói hibák és a célnyelvi befogadók szövegfeldolgozásának megkönnyítése miatti eltérések előre pontosan nem meghatározható esetekben, azaz a fordító személyével és a fordítás körülményeivel kapcsolatba hozható opcionális esetekben következnek be, elsősorban a fordítás adott pillanatára jellemző szövegértelmezésnek és az átvitelrel kapcsolatos egyedi döntéseknek függvényében. A célnyelv és forrásnyelv nyelvi rendszerének különbözőségeire visszavezethető eltérések ezzel szemben viszont bizonyos mértékig előre megjósolhatóak, ha nem is száz százalékosan – hiszen ezekben a rendszernyelvészeti magyarázattal bíró esetekben is számolni kell az egyes nyelvpárok, illetve a különböző fordítók által előnyben részesített átváltási műveletekkel. Ezek közé az esetek közé kell sorolnunk a szövegfelszín szemantikai kapcsolatainak nyelvspecifikus voltára visszavezethető, a lexikai kohéziós elemek eloszlásában bekövetkező változásokat.

1.8 A fordítói kompetencia

A fordítói kompetencia leírásakor ide kívánczok kiindulásként az a már szinte közhelynek számító megállapítás, hogy a „nyelvi kompetencia szükséges, de nem elégséges feltétele a fordításnak” (Heltai: 2002: 112). A barcelonai PACTE kutató csoport 2003-as fordítói kompetencia modellje hat alkompetenciát különböztet meg, melyek között természetesen megtalálható a két nyelv ismeretének alkompetenciája is, a nyelvektől független enciklopédikus tudás alkompetenciája, a fordítással kapcsolatos ismeretek alkompetenciája, az eszközhasználatra vonatkozó instrumentális alkompetencia, a stratégiai kompetencia és a pszicho-fiziológiai kompetencia mellett (PACTE 2005).

Az eredeti és a fordítói szövegalkotás folyamat és produktum orientált megközelítései általában e két művelet különbségeire alapozva tárgyalják a két eltérő szövegalkotási folyamathoz szükséges kompetenciák mibenlétét. „There has to be a source text, which the translator then renders in his or her own way. The act of translation is therefore one of great subtlety, involving as it does reading and (re)writing skills” (Bassnett 2003: 42).^{IV} Napjainkra ugyanakkor egyre inkább jellemző az az álláspont, amely a fordított szöveget a forrásnyelvi szöveggel kommunikatív és funkcionális aspektusból is egyenértékű szövegként kezeli (Klaudy 2001). E felfogás szerint a célnyelvi szöveg önmagában is képes helyt állni a célnyelvi kontextusban, azaz az adott információt a célnyelvi normáknak megfelelő nyelvi, pszichológiai és társadalmi kontextusba helyezni. A célnyelvre vonatkozó nyelvi és szövegalkotási kompetenciákon túl tehát a fordítónak szüksége van még a forrásnyelvi kompetenciára is, a forrásnyelvi és célnyelvi kultúra vonatkozásában kulturális kompetenciára, szaktárgyi kompetenciára, kommunikatív kompetenciára és a fordítás megvalósításához szükséges átváltási és transzfer kompetenciákra (Klaudy 1999).

A Közös Európai Referenciakeret (KER) a kommunikatív nyelvi kompetencia összetevői között számon tartja az adott forrásnyelvre és célnyelvre vonatkoztatva a nyelvi, a szociolingvisztikai és a pragmatikai kompetenciákat.

A nyelvi kompetenciák körébe tartoznak a lexikai, fonológiai és mondattani ismeretek és készségek, valamint a nyelvi rendszer egyéb jellemzői, a különböző nyelvváltozatok szociolingvisztikai értékétől és pragmatikai funkciójától függetlenül. Az egyén kommunikatív nyelvi kompetenciájának emez alkotóeleme nem csupán az ismeretek terjedelmével és minőségével áll összefüggésben (pl. hogy milyen megkülönböztetésekre képes a fonetika területén, vagy milyen kiterjedt és árnyalt szókinccsel rendelkezik), hanem azok kognitív rendszerezésével és az ismeretek tárolásának módjával is (pl. különböző asszociatív hálókkal, amelyekben a beszélő a lexikai elemeket elhelyezi), akár csak az ismeretállomány kezelésével (aktiválás, előhívás és elérhetőség). A tudás rendszerezettsége és aktiválhatósága nemcsak

egyénenként változik, hanem egyazon nyelvhasználó esetében is. (Európai Tanács Nyelvpolitikai Osztály 2001/2005: 30–31)

A KER által definiált nyelvi kompetencia utal továbbá az írásbeli közvetítés általános nyelvi szintjére, a szókincs gazdagságára és használatának pontosságára, a nyelvtani pontosságra és a helyesírás pontosságára (i. m.: 57–58). A szociológiai kompetenciák szükségességét a KER szerint az indokolja, hogy a nyelvi közvetítés igen érzékeny a különböző kultúrákra jellemző „társadalmi szokásokra (pl. a generációk, nemek, társadalmi osztályok és csoportok közötti érintkezést szabályozó, normákként működő udvariassági szabályokra, vagy a közösségben gyakorolt alapvető rituálék nyelvi kódolására)” (i. m.: 31). A pragmatikai kompetenciák tekintetében a KER a nyelvi eszközök funkcionális használata kapcsán a megfelelő beszédaktus percepcióját és produkcióját, a rugalmasságot, a téma kifejtését és a nyelvi megfogalmazás precizitását hangsúlyozza, továbbá ide sorolja „a diskurzusra, a kohézióra és koherenciára vonatkozó” kompetenciákat, „a különböző szövegfajták és szövegtípusok” felismerésének képességét (i. m.: 31). A KER foglalkozik továbbá a stratégiai kompetenciák meglétével is, amelyek segítenek a végszó felismerésben és a következtetések levonásában, valamint a tervezés, a kompenzálás, a monitorozás és a javítás folyamatának végrehajtásában (i. m.: 58).

A fordítási kompetencia mérése során a kommunikatív nyelvi tevékenységekről szóló skálák alapján a KER szerint a következő kompetenciák meglétét kell ellenőrizni, illetve a különböző szintek szerint besorolni: (i) a forrásnyelvi olvasási szövegértési kompetenciák aktivizálása, (ii) a szöveg információtartalmának részletes megértéséhez szükséges forrásnyelvi kompetenciák aktivizálása, (iii) a forrás- és célnyelv rendszerére vonatkozó szociolingvisztikai ismeretek aktivizálása, (iv) a célnyelv nyelvi sémáinak aktivizálása és (v) a célnyelvi rendszerben a szövegalkotáshoz szükséges nyelvi kompetenciák aktivizálása.

A fordítói kompetenciák leírására és modellálására született számtalan tudományos igényű munka és elmélet bemutatása nem tartozik szorosan a tanulmány központi témájához. Tanulmányom szempontjából csupán annak a tisztázása a fontos, hogy a fordítóknak a nyelvi kompetencián belül külön kompetenciákra van szüksége ahhoz, hogy a célnyelvi szöveg megalkotásánál

megteremtődjenek a forrásnyelvi szöveg szövegépítő sajátosságai alapján a célnyelvi normáknak megfelelő célnyelvi szöveg szövegszerúségért felelős tulajdonságai. Erre vonatkozó iránymutatást tartalmazott a KER előzőekben ismertetett kompetenciamérést előíró skálája is, de megtaláljuk az erre való utalást Károly Kriszta *Szövegtan és fordítás* című munkájában is:

Ide tartoznak azok a részben tudatosan, részben ösztönösen vagy automatikusan végrehajtott szövegépítő feladatok (ezek pontos eloszlásáról még nincs elegendő ismeretünk), amelyek a szöveget kohézióvá, retorikailag/szerkezetileg (a műfajnak vagy szövegtípusnak megfelelően) jól formálttá, logikussá és mentálisan feldolgozhatóvá, vagyis értelmezhetővé, koherenssé teszik. (Károly 2007: 58)

A fordításoktatásra és a fordítóképzésre tehát fokozottan jellemző kell legyen a hatékony kommunikáció megteremtéséhez szükséges nyelvi viselkedésforma tudatosításáról és fejlesztéséről való gondolkodás, mivel ez már az anyanyelvi kommunikáció esetében is fontos kritériumként jelenik meg a szakirodalomban.

1.9 A norma és az ekvivalencia fogalma a fordításkutatásban

Bár korábban éppen a fordított szövegek keletkezését esetlegesen befolyásoló tényezőkről esett szó, mindenképpen említést kell tenni a kompetens hivatásos fordítók tevékenysége és az általuk készített minőségi fordítások nyomán kialakuló fordítási normáról, illetve a normavezéreltség kapcsán szintén gyakran felvetődő fordítási ekvivalencia fogalmáról.

1.9.1 A fordítási norma

A szakma által alkalmazandó, széles körben elfogadott módszerek és stratégiák a tekintéllyel, befolyással bíró kompetens fordítók tevékenységéhez köthetők. Ezek persze korántsem statikus, vagy abszolút érvényű elvárások, hanem mindig az adott kor, adott szociokulturális környezet igényeinek tesznek eleget. Másként fogalmazva

a normatív fordítási magatartás olyan normavezérelt stratégiákban nyilvánul meg, amelyeket a kompetens hivatásos fordítók adott köre az adott szociokulturális környezetben nagy valószínűséggel alkalmaz.

Míg Toury-nál (1995) a kompetencia az összes rendelkezésre álló megoldási lehetőséget, a performancia az ezek közül az adott pillanatban éppen aktuálisan megvalósuló megoldást jelenti, addig a fordítási norma az egy adott korban, az adott szociokulturális környezetben rendszeresen kiválasztott tipikus fordítói magatartást jelöli. Toury a forrásnyelvi szöveg és célnyelvi megfelelője közötti viszony leírására a forrásnyelvi kultúra normáit előnyben részesítő adekvátság (adequacy), a megfelelőség (appropriacy) és a célnyelv és a célnyelvi kultúra normáinak követését preferáló elfogadhatóság (acceptability) megkülönböztetését alkalmazza. E hármas felosztást figyelembe véve elmondhatjuk, hogy napjainkban előtérbe kerül a célnyelv követelményeinek való megfelelés elve. Egyes területeken, mint például a multinacionális cégek kommunikatív szempontból meghatározó dokumentumai esetében (pl. termékleírások, tenderdokumentációk) kifejezetten hangsúlyossá válik a célnyelvi olvasó elvárásainak való megfelelésre törekvés.

Chesterman normaelmélete (1993) az etikai, társadalmi és nyelvi szempontokat tömörítő szakmai (professional) normákat különbözteti meg a befogadók által megszabott elvárásalapú (expectancy) normáktól. Felfogásában a fordítói stratégiát – azaz a fordítás során felmerülő esetleges problémák megoldására szolgáló tudatos eljárást – a forrásnyelvi szöveg, a megbízás körülményei, a célnyelv és a célnyelvi közönség határozza meg. A kompetens fordítók által rendszeresen alkalmazott stratégiák idővel átalakulhatnak normatív törvényszerűséggé.

Heltai (2004: 414) a norma alkalmazásához való viszonyunkat úgy fogalmazza meg, hogy az adott szituációban az alternatív, önmagukban véve helyes formák közül egyet kitüntetünk, és általában azt használjuk. Ennek a norma szerinti, azaz szokásos nyelvhasználatnak kétségtelen előnye, hogy az ismétlődés következtében részben automatizálódik, és így megszokottsága révén hozzájárul a könnyebb feldolgozhatósághoz. „A nyelvi norma betartását ösztönözheti az a tény is, hogy a norma szerinti viselkedés a csoportidentitás jelzésére is szolgál.” (i. m.: 414). Míg egy nyelv grammatikai struktúráját kötelező érvényű szabályok határozzák meg, addig

Heltai a lexika, a mondat és a szövegszintű műveletek lehetséges alternatívái közötti választások esetében a normák érvényesülését tartja meghatározónak. A lexika vonatkozásában például nem lehet egyértelműségről beszélni, hiszen egy-egy szót adott kontextuális környezetben más-más jelentéssel kell fordítanunk. Mindent összegezve Heltai normaszemléletében is mérvadó szempontként a célnyelvi közösség igényeinek elsődlegessége fogalmazódik meg:

A betartandó normák minden nyelvváltozatban és minden kommunikációs helyzetben mások. A fordítónak tehát mindig azokhoz a normákhoz kell igazodnia, amelyek az adott kommunikációs helyzetben érvényesek. Az érvényességet a relevancia elve szabja meg: a fordításnak az adott célközönség, a fordítás tervezett használója számára az adott szituációban optimálisan relevánsnak kell lennie. (Heltai 2005b: 166–167)

Hönig (1997) is a szövegfunkciót szolgáló fordítási norma követésére tesz javaslatot még a szakszövegek esetében is, és ennek érdekében például a terminuskészletből a felhasználói csoport igényeinek megfelelő kifejezés-variáns kiválasztását tartja megfelelőnek. Hönig szerint a fordított szöveg érthetősége sem határozható meg abszolút kritériumok alapján, ezért minden szakszöveg esetében külön meg kell határozni az érthetőség biztosításának követelményeit, amit a kommunikációs cél érdekében a kommunikációs célcsoport igényeihez kell igazítani.

Wilss-nél (1996) ez a kérdés a fordítási nehézséggel kapcsolódik össze, miszerint a fordítónak kevésbé jelent nehézséget a szöveg fordításközpontú elemzése, mint sokkal inkább a forrásnyelvi szövegből a funkció-azonos célnyelvi szöveg létrehozásának folyamata. A fordítói tevékenységet meghatározó szabályokat Wilss a nyelvhasználat következő területein vizsgálja: nyelvi rendszer, felhasználói norma, egyéni stílus. A nyelvi rendszer esetében szerinte a norma a „langue” (lexikon, szóképzés, frazeológia, mondat- és szöveggrammatika) szintjén valósul meg. A felhasználói normán egyrészt az állandósult és szituációtól független, az adott nyelvi- és kulturális közösségre jellemző kifejezéseket érti. Felfogásában a felhasználói norma másrészt az olyan állandósult kifejezéseket takarja, amelyek használatuk gyakorisága miatt kvázi szabályértékűvé emelkedtek. Ezeket a

szituációfüggő, ismétlődő kifejezéseket tekinti Wills a kommunikációs kompetenciánk meghatározó elemeinek. Az egyéni stílus esetében pedig éppen a példaértékű kifejezésmódokkal szemben megjelenő nyelvformáló erők szabad játékát hangsúlyozza.

1.9.2 Az ekvivalencia

A fordítások normativitásának tárgyalása során gyakran előtérbe kerül az ekvivalencia kérdésköre is. A forrásnyelv és célnyelv különböző egységei között megvalósuló ekvivalencia viszony szintjét tekintve beszélhetünk ekvivalencia viszonyról (1) a szavak szintjén, (2) a szavak szintje fölötti, de még nem mondat szintű lexikai egységek (kollokációk, idiómák, állandósult szókapcsolatok) esetében, (3) a lexikai kategóriába nem sorolható (például számok, nemek miatti) nyelvtani ekvivalenciáról, (4) a szövegszintű ekvivalenciáról, ahol ekvivalencia viszony jöhet létre (i) a tematikus- és információ struktúra terén, (ii) a szövegkohézió terén és (iii) a szövegek különböző kommunikációs szituációkban történő használatát tükröző pragmatikai téren (v.ö. Baker 1992). A minőségi fordításokkal szemben támasztott elvárások alappillérei a célnyelvi szöveg (1) tartalmának, (2) hangulatának és stílusának lehető legteljesebb visszaadása, (3) a forrásnyelvi szöveg természetes és közérthető megfogalmazása és (4) a forrásnyelvi szöveg hatásához hasonló célnyelvi hatás kiváltása. E négy kritérium közül az üzenet tartalmának napjainkban tetten érhető elsőbbsége elkalauzol minket a különböző ekvivalencia felfogások áttekintéséhez.

Catford a *Linguistic Theory of Translation* című művében (1965) megkülönbözteti a formális megfelelést és a szövegek ekvivalenciát. A formális megfelelést azon célnyelvi kategóriák (lexikai egységek, grammatikai struktúrák, szerkezetek, szerkezeti elemek), esetében alkalmazza, amelyekről megállapítható, hogy a célnyelv rendszerében ugyanazt a helyet foglalják el, mint az a forrásnyelvi kategória a forrásnyelvben, amelynek fordításául szolgálnak. Catford szerint a formális megfelelés a nyelvi rendszerek különbözősége folytán nagyon ritka és sokkal inkább a mondat szintjén hozható létre ekvivalencia viszony. Ekkor viszont már szövegek ekvivalenciáról beszél, amely bármely célnyelvi egység (szöveg vagy

szövegrész) esetében megvalósul, amelyről megállapítható, hogy egyenértékű az adott forrásnyelvi egységgel (szöveggel vagy szövegrésszel), és a forrásnyelvi és célnyelvi ekvivalens egységek esetében megvalósul a szituáció szempontjából releváns jegyek azonossága.

A forrásnyelvi üzenet formájának és tartalmának lehető leghűbb visszaadására törekvő, a forrásnyelvi szövegre és kontextusra összpontosító formális ekvivalenciára épülő fordítás Nida (1964) ekvivalencia-felfogásának is részét képezi. Nida ezzel állítja szembe a célnyelvi befogadó kulturális közegének szövegalkotására jellemző elvárásokat szem előtt tartó, a célnyelvi olvasó esetében a forrásnyelvi befogadó eredeti szöveghez való viszonyával azonos hatás kiváltására törekvő fordítói magatartást. A dinamikusan ekvivalenciával jellemezhető fordítások a legtermészetesebb módon illeszkednek a célnyelvhez és a célnyelvi kultúrához, ezért a célnyelvi közönség számára természetesen hatnak. A természetes hatás érdekében a fordítók a grammatikai és lexikai elemek átváltása esetén a megfelelő stilisztikai elemek kiválasztására törekszenek.

Toury (1980, 1991, 1993) az ekvivalencia viszony kritériumainak feltárása helyett az adekvátságra és az elfogadhatóságra helyezi a hangsúlyt, mivel a forrásnyelvi és a célnyelvi szöveg közötti ekvivalenciát eleve adott dolognak tekinti. Véleménye szerint a fordítók tisztában vannak azokkal a tényezőkkel, amelyek az adott korban, célnyelvi kultúrában meghatározzák a célnyelvi szövegjegyek és a szövegek elfogadhatóságát. Ennek értelmében az a megközelítés, ami alapján az adott célnyelvi befogadó közösség különbséget tesz a megfelelő és a nem megfelelő fordítások között az ekvivalenciának egyfajta funkcionális jelentőséget tulajdonít.

Klaudy a tartalmi ekvivalencia azon eseteit tartja kutatásra érdekes témának, amelyeknél – mint ahogy erről a kvázi helyesség tárgyalásánál már szó esett – a fordítók jólformált célnyelvi mondatokat hoznak létre, de a célnyelvi olvasó mégsem érti meg ezeket a szövegeket a kommunikatív ekvivalencia hiánya miatt. Klaudy a kommunikatív ekvivalencia előfeltételeként tartja számon a tartalmi ekvivalencia meglétét és a célnyelvi szöveg befogadójának a forrásnyelvi szöveg befogodójával azonos reakcióját (Klaudy 1987: 37).

Wills (1996) a fordítási szabályokkal kapcsolatban két esetet különböztet meg a fordítói gyakorlatban. A fordító bizonyos esetekben olyan szöveghelyekkel találja

magát szemben, amelyet könnyen megért és így a célnyelvi megfelelőt is könnyedén előállítja. Ebben az esetben „prestabilizált ekvivalencia viszonyról” beszél. Ugyanakkor azoknál a szövegrészleteknél, amelyek fordítása nagy problémamegoldó képességet igényel, a fordító csak egy viszonylagos ekvivalencia megteremtésére számíthat.

Neubert (1994) a translációs folyamat viszonylagosságára hívja fel a figyelmet, mivel még a legtapasztaltabb és legrutinosabb fordítók munkái is némi ingadozást mutatnak. Ezt szerinte sokan tévesen a fordítói folyamat szubjektív jellegével magyarázzák, holott a viszonylagosság a fordítások forrásnyelvi szöveghez való viszonyából fakadó objektív adottság. Mivel a teljes szöveg meghatározza a grammatikai és lexikai választások lehetőségét, felelős az alsóbb szintek ekvivalenseinek relativizálásáért.

Reiss (1984) szerint az ekvivalencia szintjét a célnyelvi szöveg – sokszor a forrásnyelvi szövegétől eltérő – célja határozza meg, aminek következtében a fordítás lehet interlineáris, szó szerinti, filológiai és kommunikatív. Az interlineáris fordítás az egyes szavak izolált szó szerinti fordításával a szavak szintjén valósítja meg az ekvivalenciát, míg a szó szerinti fordításkor már a szavak mellett a mondatstruktúrák helyessége is előtérbe kerül. A még mindig forrásnyelvi szöveg orientáltságú filológiai fordítás az előbbieken túl a stílus tekintetében is törekszik az ekvivalenciára, míg a kommunikatív fordítás már elsődlegesen a célnyelvi befogadó igényeihez igazodik. E kategóriák mentén Reiss különbséget tesz a kontrasztív nyelvészeti (langue típusú) – két nyelv nyelvi jelei közötti megfeleltetésekre vonatkozó – és a fordítástudomány szöveg központú (parole típusú) ekvivalenciája között, mely utóbbi az adott szöveg nyelvi jeleinek a két nyelvközösség adott szociokulturális kontextusába ágyazott megfeleltetése. Az egy az egyhez megfeleltetésekről csak a gépi fordítás esetében beszél, mivel ez a humán fordítások esetében kizárt a nyelvközösségek szituatív nyelvhasználati különbségeiből adódó eltérések következtében (Reiss 1984).

Drei Folgerungen lassen sich aus dem Beispiel ableiten: 1. Die Möglichkeit, Äquivalenz zwischen Ausgangs- und Zieltext herzustellen, wird von den unterschiedlichen Sprachstrukturen beeinflusst und gesteuert und kann in

verschiedenen Sprachenpaaren zu unterschiedlichen Lösungen führen, 2. die Einschätzung von Äquivalenz ist auch eine Frage der Bewertung [...] 3. Äquivalenz ist keine 1:1-Entsprechung, zu der es keine Alternative gibt. Mehrere Übersetzungen können äquivalent sein. (i. m.: 166)^V

A fordítók döntéseit egyre kevésbé a szavakhoz és a szűkebb értelemben vett szöveghűséghez való ragaszkodás jellemzi, bár a szöveg információ-közvetítő funkciójának tiszteletben tartása továbbra is alapkövetelmény. Egyre inkább előtérbe kerül a kommunikatív hatékonyság igénye, ami a célnyelvi olvasó vagy hallgató szövegértési munkájának megkönnyítését tartja szem előtt. Heltai fordítás és relevancia kapcsolatáról írt cikkében egyrészt leszögezi, hogy a fordítás éppen a forrásnyelvi szöveghez való hasonlósága által válik relevánssá, másrészt viszont megállapítja, hogy mivel a másodlagos kommunikációs helyzetben a célnyelvi olvasó kognitív környezete eltér a forrásnyelvi olvasó kognitív környezetétől, ezért „a fordítás által közvetített üzenet nem lehet ekvivalens (ha az ekvivalenciát azonosságnak tekintjük), hanem csak hasonló lehet az eredeti üzenethez” (Heltai 2009: 16). Napjaink fordításértékelése, a forrásnyelvi szöveg primátusát és az ekvivalencia megvalósítását szem előtt tartó korábbi állásponttal szemben, talán már túlságosan is a szöveg minőségére koncentrált (Heltai 2005b). Így például ma már az interkulturális gazdasági kommunikációval, a szakszövegek fordításával foglalkozó tudományos megközelítés nem a kultúraspecifikus normák és hagyományok leírására helyezi a hangsúlyt. Ezzel szemben viszont divatossá vált az egységes tárgyalási kontextus, ami a különböző kultúrkörökből származó emberek interakciója és közös munkája során megjelenő eredményeket helyezi előtérbe (Kelz 2002).

Sager (1994) szövegalkotó- és szövegbefogadó stratégiákat különböztet meg (senderorientierte und empfängerorientierte Strategie), aszerint, hogy a fordítás megrendelője magának a szövegnek tulajdonít nagy jelentőséget, vagy inkább a célnyelvi befogadó érdekeit tartja szem előtt. Fontos azonban leszögezni, hogy bár egyrészt a teljes ekvivalencia illúziójával szakítva a fordított szövegek relativitásáról kell beszélni, másrészt viszont e célnyelvi szövegek esetében léteznek már feltárt univerzális szövegalkotási jellemzők.

1.10 A fordítási univerzálé

A nyelvészet és nyelvtan történetében komoly hagyománya van az emberi nyelv univerzális tulajdonságainak és ezáltal az ember intellektuális képességeinek meghatározására vonatkozó törekvésnek. Az univerzálék csoportjába a nyelv azon tulajdonságait szokás sorolni, amelyek nem csak egy nyelvre vagy nyelvcsoporthoz, hanem minden nyelvre, vagy legalábbis a nyelvek többségére jellemzőek. A kutatások kapcsán az univerzálékat fajtájuk szerint három nagyobb csoportba szokás sorolni, melyek (i) a nyelvi elemzési kategóriákat jelentő szubsztantív (főnév, igealak, ...), (ii) az elemkombinációk absztrakt szabályait tömörítő formai (nyelvtan) és (iii) a nyelv két tulajdonsága közötti állandó összefüggéseket jelölő implikációs (ha – akkor) univerzálék. A nyelvtan univerzális tulajdonságait kutatók kivétel nélkül szembesülnek előbb vagy utóbb a reprezentativitás kérdésével, és a leírni kívánt univerzálé relatív vagy abszolút jellegének problematikájával (vö. Crystal 1998: 112).

Az univerzálé kutatás hagyományaihoz igazodva fordítási univerzálénak azokat a nyelvi jegyeket tekintjük, amelyek a fordítás folyamatában aktuálisan szereplő nyelvpároktól függetlenül a fordított szövegeket jellemzik. A fordítástudományi szakirodalom a célnyelvi szövegek vonatkozásában a Laviosa által a *Routledge Encyclopedia of Translation Studies*-ban leírt ismérvek alapján öt ilyen alapvető univerzális tulajdonságot különböztet meg: (i) az egyszerűsítést és a forrásnyelvi szöveg ismétléseinek kerülését, (ii) az explicitációt, (iii) az úgynevezett „normahútságot” vagy konzervativizmust, (iv) a diskurzustranszfert és az interferencia törvényt, valamint (v) a célnyelvi lexikai elemek eltérő előfordulási gyakoriságát (Laviosa 2001: 288–291).

Az egyszerűsítés lehet lexikai (tágabb jelentéskörű kifejezések alkalmazása – pl.: *kifli/péksütemény* –, körülírás, átfogalmazás reáliák esetében), szintaktikai (bonyolult mondatstruktúrák egyszerűsítése) vagy stilisztikai (hosszú mondatok feldarabolása, körülírások lerövidítése, bonyolult kifejezőmód helyetti rövidebb szókapcsolatok). A forrásnyelvi szövegben előforduló szó szerinti ismétlések kerülése talán a célnyelvi szövegek egyik legjellegzetesebben szembeötlő tulajdonsága.

Már korábban is szó esett Blum-Kulka explicitációs hipotéziséről, mely szerint a fordítási folyamat – mint kommunikációs aktus – szükségképpen kohéziós és

koherenciális eltolódásokkal jár. Kohézió alatt itt a szövegelemek közötti nyílt/látható (grammatikai), azaz a nyelvspecifikus eszközök (markerek) által jelölt összefüggést érti, míg a koherencia a szövegelemek közötti belső tartalmi (szemantikai, pragmatikai), azaz az olvasóban zajló értelmezési folyamatok révén megnyilvánuló kapcsolat. Az explicitációs hipotézis szerint a szöveg szerveződését meghatározó kohézió-jelölők a célnyelvi szövegben gyakran eltolódnak és az esetleges fordítói betoldások következtében a célnyelvi szöveg terjedelme általában meghaladja a forrásnyelvi szöveg terjedelmét. A forrásnyelvi szöveg értelmezését segítő nyíltabb, világosabb kifejezésnek köszönhetően – azaz a nyelvi közvetítésekre általánosan jellemző stratégia következtében – a célnyelvi szöveg tartalma a forrásnyelvi szövegéhez képest explicitebb. Akkor beszélünk explicitációról, ha a forrásnyelvi szövegben megtalálható implicit információkat tartalmazó fontos szemantikai elemeket a befogadó nyelvben explicit módon kell azonosítani. A leggyakrabban alkalmazott explicitációs technikák: a megjegyzések közbevetése; a kötőszók, módosítószók, minőségjelzők, kiegészítések, magyarázatok, kulturális háttérinformációk betoldása; a kétértelmű információk pontosítása (v.ö. Blum-Kulka 1986).

A fordításokra jellemző normahúségről vagy konzervativizmusról a fordítói tevékenységet befolyásoló, a célnyelvre jellemző stilisztikai normákkal kapcsolatos fordítói feltételezések kapcsán szokás szót ejteni. Az olvashatóságot és közérthetőséget előtérbe helyező stratégia következtében felerősödik a szövegkonvenciókhoz való hűség, így többek között a hiányzó írásjelek pótlása; írásjelek módosítása; befejezetlen mondatok befejezése; logikusabb szövegtagolás megvalósítása; az elbeszélésmódban a hagyományosabb formákhoz való igazodás; az idejétmúlt kifejezések helyett modern kifejezések használata stb. (v.ö. Vanderauwera 1985). A fordítás során lejátszódó mentális folyamatok hatására a fordítók hajlamosak a forrásnyelvi szöveg sajátosságait átvinni a célnyelvi szövegbe. Viszont a forrásnyelvi szöveg sajátosságai minél jobban hatnak a diskurzusátvitel során a célnyelvi szövegre, annál inkább tetten lehet érní a célnyelvi szövegben az interferencia jeleit. A forrásnyelvi szöveg sajátosságainak helytelen átvitele a negatív transzfer (v.ö. Toury 1995). A célnyelvi szövegek lexikai elemeinek eltérő előfordulási gyakoriságát Baker a nyelvi közvetítés folyamatának alapvető tulajdonságai közé

sorolja, ami bizonyos esetekben a fordított szövegek egyes lexikai elemeinek szokatlan eloszlását is eredményezheti (v.ö. Baker 1993, 1996, Blum-Kulka, 1986).

ⁱ Az eredményközpontú leíró fordítástudomány - a fordításkutatásnak a meglévő fordításokat leíró részterülete - a tudományos igényű fordításkutatásnak mindig is fontos területe volt. Az ilyen típusú tanulmányok kiindulópontja az egyedi fordítások leírása, vagyis a fordítások szövegközpontú leírása. A második megközelítés az összehasonlító fordításleírás, amikor ugyanazon szöveg egyazon nyelven belüli, vagy különböző nyelvű fordításainak összehasonlító elemzését végzik el. (saját fordítás)

ⁱⁱ A fordítások szempontjából igen fontos, hogy a fordítási tevékenység nemcsak az írást, hanem az olvasást is magában foglalja. A fordítások olvasásakor a forrásnyelvi szövegnek az egyes fordítók egyéni értelmezése szerinti olvasatát látjuk. (saját fordítás)

ⁱⁱⁱ A kvázi-helyesség nagyon hasznos fogalom a fordított szövegek esetében a helyes és helytelen határán lévő jellegzetességek összefoglalására és meghatározására; az ilyen típusú szövegekre jellemző, általában nehezen leírható és megmagyarázható jelenségek széles skálájára vonatkoztatható. A kvázi-helyesség fogalmát fogom használni azon fordított szövegek leírására és jellemzésére, amelyek ugyan formálisan helyesek, de mégsem elégítenek ki bizonyos minőségi követelményeket... (saját fordítás)

^{iv} Szükség van egy forrásnyelvi szövegre, amit a fordító aztán a saját módszerével fordít le. A fordítás ezért nagyon bonyolult tevékenység, ami olvasási és (újra)írási szakmai jártasságot is feltételez. (saját fordítás)

^v A példa alapján három következtetést lehet levonni: 1. A forrás- és a célnyelvi szöveg közötti ekvivalencia megteremtését a nyelvi szerkezetek különbözősége befolyásolja és irányítja, és ez a különböző nyelvpárok esetében különböző eredményekre vezethet. 2. Az ekvivalencia az értékelési szempontok függvénye is [...] 3. Az ekvivalencia nem jelent 1:1-es megfelelést, aminek nincs alternatívája. Több fordítás is lehet egymással ekvivalens. (saját fordítás)

2 A kutatás szövegnyelvészeti beágyazottsága

2.1 A szövegkutatás megjelenése

„A modern nyelvészet szülötteként keletkező, ókori gyökerekből táplálkozó és sok tudományágtól megtermékenyített textológia egységbe foglalja a szövegek tanulmányozását. Forrásainak sokasága is egyik oka interdiszciplináris tudományággá fejlődésének” (Szikszainé Nagy 1999: 19). Bár a szöveggel való foglalkozás gyökerei az ókorba nyúlnak vissza (retorika, stilisztika, filológia, hermeneutika) (v.ö. de Beaugrande, Dressler 2000: 37–39), a szövegek tudományos igényű vizsgálatára csak a múlt század második felétől került sor, „amikor fölmerül az igény a szöveg mibenlétének tisztázására, a szöveg, illetve szövegek leírására. Csak ennek az igénynek az egyértelmű megfogalmazása után beszélhetünk szövegtanról mint elképzelt vagy ténylegesen létező tudományágról” (Tolcsvai Nagy 1994).

A Saussure-i értelmezéssel – miszerint a mondat az egyén produktumaként a parole szintjére tartozik – szakító generatív iskola már magát a mondatot helyezte a kutatások középpontjába, majd az egyes mondatok formalista és strukturalista szempontú leírását a mondatok fölötti szint szisztematikus vizsgálatának igénye váltotta föl. A szöveget mindinkább egységes egésznek, a nyelvhasználat egységének tekintő felfogás megerősödésével a szöveg és alkotómondatainak kapcsolatában a hangsúly arról, hogy a szöveg mondatokból áll, áttevődött a szöveg mondatok által történő realizációjára (Halliday, Hasan 1976). A mondat szintű elemzések mellett egyre népszerűbbek lettek a teljes szövegegységre kiterjedő vizsgálatok, elterjedtek a szöveg kommunikatív célját és az adott szöveggel kapcsolatos funkcionális szempontokat is integráló szövegszintű kutatások. Az a tény, hogy a szöveget ezentúl az emberi kommunikáció alapegységének tekintették, együtt járt a szövegek tudományos megalapozottságú vizsgálatának megjelenésével: „In view of the fact that language communication always takes place in the form of texts, domains such as native and foreign language instruction, translation, and so on, increasingly demand a text-theoretical foundation. The creation of a linguistic text theory is thus urgent not only for linguistics, but for society at large” (Isenberg 1976: 47). Sőt a

német nyelvterületen megjelent a nyelvészetet a szövegnyelvészettel azonosító irányzat is: „Linguistik ist Textlinguistik“ (Weinrich 1967: 109).

Kutatásra érdemes témává vált (i) maga a szöveg, a szöveg struktúráját, felszíni jegyeit kutató szövegtani vizsgálatok számára, (ii) megjelentek a szöveg keletkezésének körülményeit elemző beszédaktus-elméletek (iii) és a szöveg percepcióját, azaz a szöveg befogadóra gyakorolt hatását vizsgáló kognitív kutatások. A modern nyelvelméletek közül a szövegek vizsgálatával foglalkozik többek között a strukturalizmus, a kommunikáció elmélet, a pszicholingvisztika, a beszédaktus elmélet, a tartalomelemzés, a diskurzuselemzés, a modern retorika és a kognitív nyelvészet is. A szöveget elsődleges nyelvi jelnek tekintő irányzat szerint az emberi írott és szóbeli kommunikáció nem mondat-, hanem szövegszinten zajlik, és a nyelv alapegysége (v.ö. Schiewe 2009: 95), illetve a fordítás egyetlen lehetséges egysége is a szöveg (v.ö. Dressler 1973: 12), bár a kölcsönösen releváns szövegek maguk is összekapcsolódnak tartalmilag, formailag és funkcionálisan is a diskurzus keretében.

A szöveg tanulmányozásával foglalkozó tudományág elnevezései között megtaláljuk többek között a szövegnyelvészetet (text linguistics), a szöveg tudományt (science of texts), a szövegkutatást (Textforschung), a szövegelemzést (text analysis), a diskurzuselemzést (discourse analysis), a textológiát (Textologie) és a szövegtant (Textwissenschaft). Tanulmányomban a szöveggel foglalkozó tudományterület megjelölésére egységesen a magyar szakirodalomban nagyon gyakran használt szövegkutatás kifejezést (v.ö. Károly 2007: 16) használom.

2.2 A szövegkutatás feladata

A nyelvi szerkezeteken túllépő és már a teljes szövegegységre kiterjedő, a szövegek felépülésének okait és kritériumait, illetve a szövegek használatának módját és célját kutató vizsgálatok elterjedésével egy időben megkezdődött a szövegtan szerepének és feladatkörének, a szöveg nyelvészeti alapú tudományának meghatározása. Dressler (1973: 1) minden olyan kutatást a szövegnyelvészet tárgykörébe utal, ami a szöveget tekinti az írásbeli vagy szóbeli emberi kommunikáció alapegységének, vagy amelyik legalább a mondatgrammatika

határain túllépve mondatfolyamok vagy nagyobb szövegegységek vizsgálatával foglalkozik. A szövegtan legfontosabb feladatai között tartja számon annak meghatározását, hogy:

... a szöveg nyelvi összetevői hogyan kapcsolódnak össze: milyen nyelvtani kapcsolóelemek és jelentésbeli kapcsolóelemek tartják össze a szöveget, azoknak milyen formában milyen jelentésük van, hogyan egészülnek ki a beszédhelyzetből és a világról való tudásból (pl. tudáskeretekből, forgatókönyvekből származó) ismeretekkel, és azt, hogy e tényezők hogyan alakítanak ki a szövegben összefüggő jelentéshálózatot ... (Kugler, Tolcsvai 2000: 249–250)

A szövegről, annak funkciójáról, alaptermészetéről, szerkezeti tulajdonságairól és fajtáiról való tudományos gondolkodás feladatai közé tartozik tehát (i) a szövegszerűség kritériumainak feltárása, (ii) a szöveg szerkezeti, grammatikai, jelentéstani, pragmatikai sajátosságainak meghatározása (iii) a szövegalkotás folyamatának és a szövegalkotást befolyásoló tényezőknek a leírása, (iv) a szövegbefogadás folyamatának és a szövegbefogadást befolyásoló tényezőknek a leírása és (v) az egyes szövegtípusok meghatározása és tulajdonságainak leírása. Károly megfogalmazásában:

A szövegtan vizsgálati tárgya a szöveg, amely komplex nyelvi, kognitív és társadalmi jelenség: egyszerre nyilvánul meg produktumként (mint szerkezet) és folyamatként (mint művelet sor). A szöveg kutatásának célja ennek az összetett jelenségnek a szükségképpen több szempontú és megfelelően árnyalt leírása. (Károly 2007: 22)

A szövegalkotás, szövegszerűség, szövegértés, szövegfunkciók, szövegtípusok, stb. feltárása alapvetően a szövegek „működésének” leírásával foglalkozik az emberek közötti interakcióban, így többek között a fordítási szituációban is. A fordított szövegek szövegközpontú vizsgálatánál az utóbbi évtizedekben három vizsgálati témakör került előtérbe: (i) a Halliday és Hassan (1976, 1980) óta népszerű

kohéziós vizsgálatok, (ii) a swalesi (1990) hagyományt folytató műfajelemzések és (iii) a Dijk és Kintsch (1983) nyomdokait követő makrostrukturális tanulmányok (v.ö. Károly 2007: 33). Mivel a jelen tanulmány a lexikai kohéziós vizsgálatok sorát kívánja gazdagítani, ezért a szövegkutatás elméleti kereteit meghatározó megközelítések közül azokat a szakirodalmi hivatkozásokat gyűjtöttem össze, amelyek a célnyelvi szövegek lexikai kohéziós szövegépítési sajátosságainak feltárására irányuló vizsgálatok jelentőségét támasztják alá.

Kiefer (1979: 46) a szövegek nyelvészeti vizsgálatát és „grammatikai (fonológiai, morfológiai, szintaktikai és szemantikai)” leírását, valamint a mondatok egymásutánjából szöveget előállító fenomén megragadását tekinti a szövegtan legfőbb feladatának. Károly Sándor (1979: 55) ugyanakkor a szövegszerkesztés általános szabályainak, a „szövegtanilag jó mondat”, „jó mondatsor, jó szöveg fogalmának” tisztázását és a lineáris kohézióval való foglalkozást tekinti a szövegtan feladatainak. Szikszainé Nagynál (1999: 18) a szövegekkel foglalkozó tudományág kutatási céljai között szerepel „azoknak a feltételeknek, lehetőségeknek, általános szabályszerűségeknek, eszközöknek a feltárása, amelyek a szövegtanhoz kellene, illetve magának a szövegalkotás folyamatának kiderítése” és „a gyakorlati teendők szolgálata (a kommunikációs zavarok csökkentése vagy kiszűrése, a fordítás könnyítése stb.)”. Fehérnél már a szöveg kommunikációs aktusban betöltött szerepének hatékonysága kerül előtérbe:

A szöveget nem csupán részekből és viszonyokból álló összetett jelként, hanem olyan értelmes cselekvésként kell értelmezni, amelynek célját, kifejtettségét az adott szituációnak való megfelelés határozza meg. A fő kérdés immár nem az, hogy miként kapcsolódnak össze a szöveg összetevői, hanem az, hogy miként funkcionál a kommunikációs interakcióban és milyen tulajdonságai teszik ezt lehetővé és hatékonyá. (Fehér 1999: 474)

Petőfi S. János szövegdefiníciója a „legtágabb értelemben vett nyelvhasználatot” veszi alapul, és a világ éppen aktuális szegmentumára vonatkozó tudásunkat is a szöveg részeként értelmezi. Ez a szövegfelfogás „egy írott vagy nyomtatott fizikai manifesztációval rendelkező, dominánsan verbális, relációs

szemiotikai objektumra – jelölő-jelölt relációra – utal” (Petőfi S. 2004: 29). A szövegnyelvészet feladata e felfogás szerint „valamennyi olyan nyelvészeti módszerekkel kimutatható reláció elemzése és leírása, amely ’átlép’ a „szövegmondatok”-nak tekinthető kommunikátum-összetevők közötti határokon” (i. m.: 74). A napjainkban egyre hangsúlyosabbá váló szövegnyelvészeti irányzat önmaga szűkebb keretein túllépve, „széleskörű interdiszciplináris megalapozásra” törekszik, s a „szövegeket komplex jelekként kezeli, és vizsgálja azokat mind szintaktikai (vagy pontosabban: formai), mind szemantikai, mind pragmatikai nézőpontból” (i. m.: 86).

Petőfi S. János a szövegtani kutatások közé sorol minden olyan kutatást, „amelynek célja, hogy a szövegek formai és szemantikai architektonikája nyelvészeti és hermeneutikai elemzéséhez és leírásához adekvát elméleti keretet biztosítson” (i. m.: 73). Petőfi S. több évtizedes szövegkutatási tapasztalataira támaszkodva hangot ad továbbá annak a véleményének is, hogy a szövegtannal foglalkozó szemiotikai textológia és retorika tantárgyakat fel kellene venni az egyetemi oktatásba, ahol központi szerepet kapnának.

A leíró szövegkutatás kezdeti hármas tagozódását (grammatikai, szemantikai és pragmatikai vizsgálatok) mára a vizsgálati területek sokszínűsége váltotta fel: szövegpragmatika, szövegsemantika, szöveggrammatika, szövegszerkezetan, szövegstilisztika, szövegtipológia, szövegfonetika, szövegformatan, amit a szöveg alkotásának és befogadásának sajátosságainak vizsgálata egészít ki (Szikszainé Nagy 1999: 23).

2.3 A fordítások szövegnyelvészeti megközelítése

A fordítási szöveg elemzése az elméleti (fogalmi meghatározások, rendszerbe foglalás) vagy empirikus kutatások (megfigyelés, adatgyűjtés, elemzés és kísérletezés) révén elsősorban leíró jellegű következtetések levonására nyújt lehetőséget. A fordításkutatásban a szövegelemzés tárgya lehet maga (i) a forrásnyelvi szöveg (pl. szerző, közönség, funkció aspektusából történő vizsgálata), (ii) a célnyelvi és forrásnyelvi szövegek összehasonlító elemzése (pl. az adott

forrásnyelvi szöveg különböző fordításainak összevetése adott célnyelven vagy az adott forrásnyelvi szöveg különböző célnyelveken elkészült fordításainak összevetése), (iii) az adott nyelven keletkezett fordított és autentikus szövegproduktumok bizonyos szövegjegyeinek vagy bizonyos szövegjelenségek eloszlásának relatív különbségeit vizsgáló összevetés (comparable texts) és (iv) az annotált fordítások vizsgálata (TAP, retrospekció eredménye).

Míg a szöveggel való tudományos igényű foglalkozás kezdetben csak a saját keretein belüli feladatleírásra összpontosított, mára túlnőtt a diszciplína szabta kereteken. A szövegnyelvészet kutatási kérdései jelentőségükhöz mérten egyre több interdiszciplináris területen is megjelennek. A fordítástudományi kutatások közül például az eredményközpontú megközelítések számára nélkülözhetetlen a szövegek összehasonlító vizsgálata és elemzése és ennek jelentőségére egyre több szerző utal is munkájában. Szathmári (1996: 15) a szövegszintű kutatások szerepét a fordítástudomány szemszögéből vizsgálva megállapítja, hogy „a szövegtani oldal új összefüggésekre, megoldásokra, módszerekre hívhatja fel a fordítók figyelmét”. Károly (2007: 14) a szövegkutatás és fordítástudomány kutatási területeinek összekapcsolását azért tartja meghatározónak, mert „mind elméleti, mind módszertani szempontból jelentős elvekkkel és eredményekkel gazdagíthatja a fordítási szöveggel és a fordítási szövegalkotással kapcsolatos tudásunkat.”

Klaudy (2007: 139) értelmezésében a fordítástudomány a fordítási feladatok természetéből adódóan, azaz a mondathatárokon túlmutató szövegegységek „átkódolása” miatt nem nélkülözheti a szövegszintű vizsgálatokat. A diszciplína kutatási kérdései közé tartozik szerinte többek között annak megállapítása, hogy „Hogyan jelentkezik a szövegszint az átvitelben?”

Aki azt állítja, hogy a fordítás csak a forrásnyelvi szöveg megértéséből (dekódolásából) és a célnyelvi szöveg megalkotásából (kódolásából) áll, annak számára nincs értelme forrásnyelvi szövegsajátosságok átviteléről beszélni. A fordított szövegek elemzése viszont azt mutatja, hogy a fordítás eredményeként keletkezett szövegek mások, mint az autentikus célnyelvi szövegek, és éppen a szövegszint az, ahol a másság kimutatható (Klaudy 2007: 139).

A fordítástudományon belül a szövegek vizsgálatának két nagy csoportja különül el egymástól, annak függvényében, hogy a vizsgálatok fókuszában a forrásnyelvi vagy a célnyelvi szövegek tulajdonságainak leírása áll. A forrásnyelvi szövegre fókuszáló megközelítés elnevezése (source text oriented translation studies) eredetileg Toury-tól (1995) származik, aki azt az irodalmi fordítások célnyelvi normatív megközelítése kapcsán alkalmazta. A forrásnyelvi szöveget vizsgáló kutatások napjainkban a fordítóképzésben kapnak jelentős szerepet, s e tekintetben számba veszik a fordítás alapjául szolgáló forrásnyelvi szövegre jellemző szövegspecifikus (felépítés, lexikon, mondatszerkezeti sajátosságok, nem verbális elemek alkalmazása és aránya stb.) és nem szövegspecifikus (szövegalkotó, szövegalkotási intenció, elképzelt célközönség, alkalmazott kommunikációs csatorna stb.) tulajdonságokat (v.ö. Nord 2005). A német szakirodalom ugyanakkor szívesen és behatóan foglalkozik a fordító számára mindenképpen megoldandó feladatot jelentő, a forrásnyelvi szövegben tetten érhető defektusokkal is (v.ö. Snell-Hornby és szerzőtársai 1998: 147–151).

A fordítás kommunikatív és interaktív szerepének előtérbe kerülésével a figyelem egyre inkább a fordítás funkciójára és a befogadói elvárásokat és a célnyelvi kultúrának a szöveg egészével és az egyes szövegfajtákkal szemben támasztott követelményeit szem előtt tartó fordítói szándékokra irányult. Ezzel a megközelítéssel a hangsúly áthelyeződött az forrásnyelvi szövegről és a szövegprodukciónról a célnyelvi produkcióra. A fordító szöveget alkot és immár mint célnyelvi „szerző” jelenik meg, akinek át kell élnie a fogalmazási folyamat minden lépését és nehézségét, bár alkotói szabadsága korlátozott, mivel a fordítás valójában a forrásnyelvi szöveg által indukált célnyelvi szövegprodukción (Neubert és Shreve 1992). A fordított szöveg tehát egyedülálló intertextuális kapcsolatban áll a forrásnyelvi szöveggel és egyre fontosabb szerepet kap annak feltárása is, hogy miként őrizhető meg a forrásnyelvi szöveg műfaji identitása a célnyelven. Mivel a szövegtípus és a műfaj döntően a kommunikatív szándék és szituáció függvénye, az ekvivalencia központú megközelítést egyre gyakrabban váltja fel a funkcionális (a célnyelvi befogadó számára könnyebben feldolgozható szöveget előállító) fordítás.

A fordítástudományon belül elterjedt, a célnyelvi szövegre fókuszáló megközelítés (target text oriented translation studies) (v.ö. Toury 1995) a fordítási folyamat eredményeként, a forrásnyelvi szöveg alapján keletkezett célnyelvi szöveg vizsgálatától várja, hogy az elemző elegendő ismerethez jusson „(többek között) a fordítási szöveg jellegzetességeire vonatkozóan. A fordítási szöveg nyelvi és nem nyelvi (kognitív, illetve társadalmi kontextus által meghatározott) sajátosságainak leírása ezért elengedhetetlen részét képezi minden, a fordítások minőségének értékelését, a fordítók képzését és a fordítói kompetencia fejlesztését célul kitűző vállalkozásnak” (Károly 2007: 13).

A célnyelvi szövegre fókuszáló fordításkutatás a szövegtulajdonságok tipologizálása érdekében több fajta kutatási módszert is meghonosított, így a (i) forrásnyelvi és célnyelvi szövegek kontrasztív elemzését, (ii) az adott forrásnyelvi szöveg eltérő célnyelvi megfelelőinek összehasonlító elemzését, (iii) a célnyelven íródott autentikus és fordított szövegek szövegépítő sajátosságainak egybevetését, és (iv) a különböző nyelveken elérhető tematizált csoportokba sorolt fordítások témacsoporton belüli összehasonlítását (v.ö. Komissarov 2002).

2.4 Mit tekintünk szövegnek?

Az 1960-as évektől egyre nagyobb számban megjelenő szövegnyelvészeti tanulmányok eleinte a szöveget egyszerűen csak a mondatnál nagyobb egységként kezelték, majd az egymást követő jól formált mondatok sorozataként definiálták: „ein durch ununterbrochene pronominale Verkettung konstituiertes Nacheinander sprachlicher Einheiten.“ (Harweg 1968: 148) Csak amikor már bebizonyosodott, hogy az egymással szomszédos mondatok szintaktikai-szemantikai összekapcsolódása nem bizonyul a szövegszerűség elégséges kritériumának, tulajdonítottak a későbbi szöveg meghatározások a szöveg pragmatikai és funkcionális vetületének is kiemelkedően fontos szerepet, miszerint a nyelvi jelsor „egy tényleges vagy feltételezett kommunikáció-szituációban egy tényleges vagy feltételezett kommunikatív funkció betöltésére alkalmas” (Petőfi S. 1988: 43).

A szövegértékűség legfőbb kritériuma pragmatikai meghatározottságú: abban, hogy egy nyelvi jelet vagy nyelvi jelek egy halmazát szövegnek tekinthetünk-e, az játssza a leglényegesebb szerepet, hogy egy adott helyzetben be tudja-e tölteni egy teljes közlés szerepét. Egyes esetekben erre akár egy mondat is alkalmas lehet, például: *Fűre lépni tilos!*, bizonyos szituációkban egyetlen mondatértékű szó is közölhet teljes értékűen: *Ne!*, sőt akár egyetlen hangnyi közlés is önálló szöveg lehet, mint például *Ó!* indulatszavunk is. (H. Varga 2007: 13)

A szakirodalom egységesen az olyan közléseket tekinti szövegeknek, amelyek a szövegszerűség hét ismérvének (kohézió, koherencia, szándékoltság, elfogadhatóság, hírérték, helyzetszerűség, intertextualitás) eleget tesznek, mivel ezek bármelyikének hiányában a szöveg nem lesz kommunikatív értékű, azaz nem lesz alkalmas a közlési feladat végrehajtására. A szövegszerűség ismérvei alapjaiban határozzák meg a szöveg jellemzőit. Ezek mellett a szövegkommunikáció egészét a szöveg hatékonysága, hatásossága és helyénvalósága befolyásolja. A szöveg hatékonysága a kommunikációs interakcióban részt vevő szereplők erőfeszítésére, a szöveg hatásossága a kommunikációs cél megvalósulásának mértékére, míg a szöveg helyénvalósága a szövegszerűség ismérveinek adott helyzethez való igazodására utal. Petőfi S. (1990: 7), aki „a szöveget a nyelvhasználat és nem a nyelvi rendszer elemének” tekinti, kihangsúlyozza továbbá, hogy a szöveg „írott vagy nyomtatott fizikai manifesztációval” rendelkezik, „dominánsan verbális” jellegű szemiotikai objektum, melynek „domináns jelentéshordozó elemei lexikai elemek”.

A szöveg meghatározásának ma is ismert definícióinak sorát csak a szövegszerveződés ismérveinek feltárására irányuló vizsgálatok megjelenése és egy univerzálisan érvényes szöveggrammatika leírási kereteinek megteremtését szorgalmazó igény vonta maga után. A számos meghatározás közül a terjedelmi keretek miatt Szikszainé Nagy Irma *Leíró magyar szövegtanának* a szakirodalom legjelentősebb szövegdefinícióit összegző szöveg meghatározását idézem:

A szöveg nyelvi elemekből álló, túlnyomórészt szerves kapcsolódásukból felépülő olyan írásos vagy szóbeli közlés, amely a nyelvi kommunikáció

alapegységeként meghatározott beszédhelyzetben vagy szöveggörnyezetben valamely kommunikációs szerepnek tesz eleget. A szöveg alapvetően nyelvi jelek összekapcsolt sora, sőt lánc, a nyelvhasználat legmagasabb szintjén szerveződött legkomplexebb nyelvi jelcsoport, ennek következményeként egyrészt jelentéskomplexum, egységes egész alkot, másrészt szerkesztett, globális struktúrájú, tehát részekre tagolható. (Szikszainé Nagy 1999: 53)

2.5 A szövegszerűség kritériumai

A szövegszerűség, más néven textualitás kritériumainak meghatározása számos szövegnyelvészeti értekezés témája, melyek szerint a szöveg szöveggként való érzékelése egyrészt a szöveg alkotójának, másrészt a szöveg befogadójának szövegről alkotott elképzeléseitől függ. Kiefer (1979: 47) szövegkoherenciának nevezi mindazt, „ami a mondatokat szöveggé szervezi”, a szövegkoherenciát meghatározó elemeket pedig a „szövegkoherencia mutatóinak”.

Canisius és Knipf (1996: 25) a szövegszerűség meghatározásának három szintjét különbözteti meg: a pragmatikai, a szemantikai és a szintaktikai szinteket. A szerzők a szövegszerűség pragmatikai aspektusánál a befogadó szöveggel szemben támasztott elvárásait tartják mérvadónak, azaz az olyan nyelvi valóságon kívüli tényezőket, mint a szöveg szerzőjének szándéka, a szöveg keletkezésének és befogadásának konkrét körülményei, a szöveg érthetőségének mértéke a befogadó szempontjából, vagy a szöveg teljes kommunikációs környezete. A szövegszerűség szemantikai aspektusát ezzel szemben a szöveg témája jelenti, azaz az a tény, hogy a szövegelemek a közös kontextus és a közös szituáció tartja össze. A szövegelemek közötti kapcsolatok rendszere ugyanakkor már a szövegszerűség szintaktikai aspektusának része.

A szöveg értékű közlés jellemzői a tartalmi folyamatosság, a kommunikációs célhoz igazodó, teljesen és lezártan kifejtett egységes téma, a mondatok közötti tartalmi-logikai és grammatikai kapcsolat, az egymásra épülő szerkezeti egységek és a homogén stílus (Szikszainé Nagy 1999: 57).

A szövegértékűség a szöveget alapvetően meghatározó, azzal vele járó tulajdonság. Szöveggé: szövegmondatok láncává a mondatok sorát elsősorban a szövegalkotó teszi a szövegösszefüggés megvalósításával, azaz objektíve létező szövegtényezők miatt „értékelődik” egy nyelvi jelsor szövegnek: a változatlan téma, a koreferenciális (az azonos jeltárgyra vonatkozó szavak kapcsolata), az expliciten jelölt vagy impliciten jelzett tartalmi-logikai összefüggés stb. miatt. Viszont a befogadótól is függ egy mondat sor szöveggé történő értelmezése, az ő sokféle előismeretén múlik, hogy egy szöveghez tud-e jelentést rendelni. (Szikszainé Nagy 1999: 56)

A kommunikatív értékű közlésekre kivétel nélkül igaz a szövegszerűség hét ismérvének realizációja (De Beaugrande és Dressler 2000: 23), amelyek a szövegalkotót és a befogadót segítik, hogy „kapcsolatokat tudjon teremteni vagy azonosítani a szövegen belül, illetve a szöveg és az azt körülvevő emberi kontextus között, valamint, hogy el tudja dönteni, mely kapcsolódások relevánsak a kommunikáció szempontjából” (Károly 2007: 30). A szövegszerűség hét ismérve közül a kohézió és a koherencia szövegközpontú tulajdonságok, míg a szöveg segítségével megvalósuló kommunikációra és a kommunikációs aktus szereplőire vonatkozó másik öt ismerv (szándékolttság, elfogadhatóság, hírérték, helyetszerűség, intertextualitás) felhasználó-központú tulajdonság.

Halliday a szöveggé szerveződés kritériumai között strukturális és kohéziós elemeket különböztet meg. A strukturális elemek között a téma-réma viszonyt tartalmazó tematikus struktúrát valamint az új és régi információtartalmat megjelenítő információs struktúrát és fókusz jelöli meg. A kohéziós elemek négy csoportját különbözteti meg, így a referenciát, az ellipszist és helyettesítést, az összekapcsolást és a lexikai kohéziót (v.ö. Halliday 1985: 313).

A kohézió a szövegfelszín alkotóelemei, azaz az egyes szavak közötti kölcsönös grammatikai függőségeken alapuló kapcsolat. Ezek a különböző mondatokat, szövegrészeket összekötő, explicit módon megjelenő felszíni kapcsolatok meghatározóak a lexikai egységek jelentésének szempontjából. A koherencia ezzel szemben a fogalmak és viszonyok szövegfelszín alatt meghúzódó kölcsönösen releváns volta. A szövegben a fogalmak közötti tartalmi viszony (okság, lehetővé

tétel, indoklás, célzat) gyakran nem explicit, azaz nem mindent aktivizálunk felszíni kifejezések révén. A koherencia nem csupán a textualitás jellemzője, hanem egyben a percepció során lezajló kognitív folyamatok eredménye is, hisz a szöveg befogadója a szövegben foglaltak alapján aktivizálja a saját valóságismereteit, következtetéseket von le, koherenciaviszonyokat fedez fel, illetve hoz létre. E kognitív tényezők következtében minden egyes olvasó – legalábbis részben – eltérő értelmet tulajdoníthat ugyanazon szövegnek, illetve „különböző mennyiségű és minőségű viszonyok hozzátételére van szüksége ugyanazon szöveg megértéséhez” (Károly 2007: 31).

A szándékoltság a szöveg alkotójának kommunikatív célkitűzése, míg az elfogadhatóság a szöveg befogadója szempontjából értelmezi a szöveg használhatóságát és relevanciáját. A hírérték ugyanakkor a szövegben megnyilvánuló közlés tartalmának jellegére (várt/váratlan, ismert/ismeretlen információ) vonatkozik. Ebből a szempontból figyelembe kell venni a szöveg befogadójának feldolgozó kapacitását, mivel a nagy hírértékű szöveg feldolgozása több erőfeszítést igényel, míg a szélsőségesen alacsony hírértékű szöveg unalmas lehet. A helyzetszerűség az adott szituációban teszi a szöveget relevánssá, míg az intertextualitás az adott szöveg feldolgozásához szükséges korábbi szövegismeretekre utal.

Az eredeti szövegalkotáshoz képest komplexebb fordítási szövegprodukción során a nyelvi- és szövegalkotási kompetencia mellett kiemelt szerepet kapnak a kommunikációs aktus kultúra- és helyzetspecifikus vonatkozásai is. A fordítási szöveg, mint kommunikatív esemény, akkor képes elérni a célját, ha funkcionálisan és a szövegszerűség kritériumai szerint is adekvát, azaz ha a kommunikáció és a fordítás célja szempontjából is releváns és megfelel a célnyelv szövegalkotási elvárásainak.

A fordítási szövegalkotás funkcionális megközelítése a fordítás célja és funkciója függvényében dinamikusan változó mértékben tartja a fordítási tevékenységet produktív és reprodukív természetűnek. A fordítási szövegalkotásnak ez a dinamikus jellege megköveteli a fordítótól, hogy a forrásnyelvi szöveg és a fordítás célja alapján döntsön a fordítási stratégiáról (szó szerinti / szabad fordítás) és arról, hogy a forrásnyelvi szöveg mely aspektusait reprodukálja változtatás nélkül

és melyeken változtat a hatékony kommunikáció és a szövegszerűség kritériumainak való megfelelés érdekében. Minthogy a fordító produkciós és reprodukciós feladatai nem definiálhatóak általánosságban és statikusan – ezek a fordítás célja/funkciója és a fordítandó szöveg műfaja/típusa függvényében dinamikusan változnak. Ennek következtében az is a fordítási szituáció egyedi jellemzője, hogy a fordító az adott helyzetben a szövegszerűség ismérvei közül mely tulajdonságok reprodukciójáért és mely tulajdonságok produkciójáért felelős.

Klaudy (2006) például úgy vélekedik, hogy a fordítónak a szövegszerűség ismérvei közül a kohézió, az elfogadhatóság és az intertextualitás kapcsán van produktív feladata, míg a koherenciát illetően csak az átvitelért felelős, mivel a fogalmak közötti viszonyokat nem ő, hanem a forrásnyelvi szöveg szerzője teremti meg. Előfordulhat viszont, hogy a fogalmak közötti koherencia viszonyok egy az egyben történő reprodukciója eredményeként olyan célnyelvi szöveg jön létre, amelynek értelemszerkezetét a célnyelvi befogadó nem tudja feldolgozni és ami számára nem koherens, mivel a célnyelvi befogadó más háttértudással rendelkezik mint a forrásnyelvi szöveg szerzője és/vagy befogadója. Amennyiben a fordító ezt a veszélyt felismeri és képes a forrásnyelvi szöveg fogalmi és relációs hálóját a célnyelvi közönség számára értelmezhető módon átültetni a célnyelvi szövegbe, akkor valójában aktívan részt vesz a célnyelvi szöveg koherenciájának megteremtésében, vagyis magában a szövegprodukciónban.

Klaudy (i. m.) szerint a fordító feladata továbbá a forrásnyelvi szöveg intertextuális viszonyainak helyes dekódolása is, aminek alapján eldönti, hogy a fordítás milyen típusú célnyelvi szövegek közé fog majd beilleszkedni. A megfelelő szövegtípus megteremtéséhez szükséges célnyelvi eszközök kiválasztásához meghatározóak a fordító szövegtipológiai ismeretei is. A fordítás célja/funkciója és a szöveg típusa/műfaja mellett a két nyelv közelsége és távolsága is meghatározhatja a fordító produkciós és reprodukciós feladatainak arányát. Minél nagyobb a két nyelv és a két kultúra egymáshoz viszonyított távolsága, a fordítónak annál produktívabban kell közreműködnie abban, hogy a célnyelvi szöveg megfeleljen a célnyelvi befogadó szövegszerűséggel kapcsolatos elvárásainak, és ezáltal hatékonyan tölthesse be kommunikatív funkcióját.

3 A lexikai kohézió vizsgálata fordított szövegekben

3.1 Bevezetés

A fordítóknak, és a gépi fordítóprogramok készítőinek legfontosabb célkitűzése a célnyelvi kultúrától és nyelvezettől nem idegen célnyelvi szövegek előállítása (vö.: Hönig 1997: 26). A legtöbb fordítás esetében mégis éppen ezen a területen érhetjük tetten az idegenszerűséget. Az anyanyelvi olvasó számára a szövegszint jegyei a legárulkodóbbak, azaz a szövegszerűség meglétéért felelős jegyek valamelyikének hiánya, vagy azoknak az adott nyelven, az autentikus szövegprodukciónál keletkezett szövegekre jellemző disztribúciójától eltérő alkalmazása. Jelen fejezet a fordítástudomány és a szövegnyelvészet közös kutatási területén belül a szövegszerűség egy adott elemére, a lexikai kohézióra fókuszálva tekinti át e két tudományterület interdiszciplináris célkitűzéseit, eredményeit, távlatait.

3.2 A kohézió

Mint arról már az előző fejezetben szó esett, egy nyelvi jelsornak, illetve az egymás után álló mondatok sorának a szövegszerűség alapvető kritériumainak kell megfelelnie, hogy azokat a szövegalkotó és szöveg befogadó is szöveggé értelmezze. A szövegegység és szövegösszefüggés megteremtéséért felelősek a szövegen belül mind a makroszerkezeti, mind a mikroszerkezeti kapcsolódások, amely kapcsolódásokat lehet a megvalósulásuk szerinti szövegszint alapján és a kapcsolatot jellemző utalás iránya alapján csoportosítani. A szövegen belüli összefüggés-kapcsolatok közül a pragmatikai jellegű a koherencia, a szemantikai jellegű a kohézió, a szintaktikai jellegű a konnexitás, a szerkezeti szintű a kompozicionális konnexitás, a stilisztikai jellegű a stíluskohézió, a fonetikai-prozódia jellegű az akusztikai kohézió. A szövegen belüli összefüggés-kapcsolatok eszközei a különböző szövegszinteken eltérőek és a különböző szövegtípusok esetében az összekapcsoltság mértéke változik. A szövegen belüli kapcsolatot jellemző utalás iránya alapján megkülönböztetünk anaforikus, azaz visszafelé utaló, kataforikus, azaz

a szövegben előre utaló és deiktikus, azaz rámutató funkciójú kapcsolódásokat (v.ö. Szikszainé Nagy 1999: 57-72).

A szövegegység és szövegösszefüggés megteremtéséért felelős kapcsolódások egyike a szövegelemek szövegen belüli változatlan és koreferenciális (ugyanarra a valóságszegmensre vonatkozó) ismétlődése. Ezen ismétlődések részben az egyes mondatokon belül valósulnak meg, részben a szöveg egymástól akár kicsit távolabb elhelyezkedő mondatai között. A szakirodalom különbséget tesz a szöveg kisebb egységein belül megvalósuló lineáris kohézió és a tágabb szövegkörnyezetre, vagy akár a teljes szövegre jellemző globális kohézió között. A jelen tanulmány középpontjában a szöveg kisebb egységein belüli – az automatikus elemzések számára szövegszerkesztési és központosítási tulajdonságai révén könnyen megragadható (az egyik beljebb kezdéstől a következő beljebb kezdésig tartó) bekezdés szintű – szemantikai jellegű összefüggés-kapcsolatok révén megvalósuló kohéziós viszony leírása és a kutatás eredményei alapján ezen viszony további tulajdonságainak feltárása áll.

3.2.1 A kohézió meghatározása és a kohéziós eszközök

A szövegek mikroszerkezeti szintjén, az egymást követő mondatokban, szövegegységekben megvalósuló grammatikai-szemantikai összefüggés láncolatot biztosítja a lineáris kohézió. A lineáris kohézió megteremtéséért felelős nyelvi elemek rendszerbe foglalása a különböző szerzők kohézióról alkotott elképzeléseinek és a különböző nyelvek kohéziós preferenciáinak és eszköztárának függvénye. Mivel a kohéziókutatás hagyományai az angolszász szakirodalomban gyökereznek, a tanulmányban ezek bemutatása került előtérbe. A kohéziós taxonómiák egymásra épülnek, mivel az újabb rendszerek iránti igény mindig akkor merült fel, amikor az azt megelőző nem bizonyult teljesen kielégítőnek, részben az általa használt kategóriák elmosódottsága miatt, részben mert a taxonómia mégsem fedett le minden lexikai kategóriát.

A kohéziókutatás kezdeti szakaszában Harweg (1968) a helyettesítést tekintette a szövegszerveződés egyik műveletének. Elképzelése szerint egy kifejezés kohéziós vagy koherens viszonyt teremt az őt követő azonos értelmű, vagy ugyanarra

a dologra utaló másik kifejezéssel. A lexikai kohéziós vizsgálatok esetében leggyakrabban alkalmazott taxonómiák közé tartozik Halliday és Hasan rendszere (1976, 1989), amelyben a szemantikai természetű kohézió „a szöveg egy elemének a szemantikai viszonya egy másikhoz, amely elem az elsőnek az értelmezéséhez szükséges” (Halliday, Hasan 1976: 8). Ennél a megközelítésnél a kohézió a jelentés szintjén érvényesül, s a szöveg létrehozása szempontjából nem annyira a kohézió mértéke a lényeges, hanem az a tény, hogy egy elemet egy másik elem értelmez, vagyis magának a kohézióknak a jelenléte.

Halliday és Hasan (1976) angol nyelvre kidolgozott kohéziós taxonómiájában a kohéziós eszközök öt különböző típusát határozzák meg, melyek a referencia, az ellipsis, a helyettesítés, az összekapcsolás és a lexikai kohézió. A referencia olyan szemantikailag önmagában nem értelmezhető grammatikai elem, amelyik a szöveg egy másik elemére való utalásán keresztül nyer értelmet. Az utalások két nagy csoportja közül az exofora a szövegen kívüli valóságra, a beszédhelyzetre vagy annak egy elemére utal (exophoric/situational reference), míg az utalások másik csoportjába a szövegen belüli utalások tartoznak (endophoric/textual reference). Az endofora lehet továbbá előre utaló (katafora) vagy hátrafelé mutató (anafora). A referencia (i) utalhat személyes és birtokos névmások használatával személyre (personal reference) (ii) lehet a mutatószók és a határozott névelő által mutató (demonstrative reference) vagy (iii) az azonosságot, hasonlóságot, különbözőséget, mennyiségi vagy minőségi hasonlítást kifejező jelzők és határozók következtében hasonlító (comparative reference). A helyettesítést (substitution) és az ellipszist (ellipsis) ugyanazon kohéziós kapcsolat kétféle megnyilvánulásaként tartják számon. Míg a helyettesítés során a szöveg egy elemét egy másik helyettesíti, addig az ellipsis esetén a szöveg egyik eleme kimarad, azaz a magától értetődő szövegrész esetében zéró helyettesítés történik. Az angolban a behelyettesítést végző elem névszói, igei, vagy mellékmondati funkciót tölthet be és így az ellipsisnek is három fajtája létezik: névszói, igei és mondatellipsis. Az összekapcsolásban (conjunction) szerepet játszó kötőszavak önmaguk nem kohezívek, de kapcsolatot tudnak teremteni a szöveg strukturálisan nem kapcsolódó részei között. A kapcsolatok természetétől függően a kötőszavak öt csoportja különböztethető meg: (i) a kapcsoló (*and, besides*), (ii) az ellentétes (*but, however*), (iii) az okozati (*so, hence*), (iv) az

időhatározói (*then, next*) és a folytatólagos (*well, after all*) összekapcsolást biztosítók. A lexikai kohéziós kapcsolatokat a lexikaválasztás biztosítja, ami megvalósulhat egy lexikai elem ismétlése révén (*reiteration*) vagy az általában hasonló lexikai környezetben együtt előforduló, ám rendszerszintű szemantikai kapcsolattal nem rendelkező lexikai elemek asszociációja révén (*collocation*).

Hasan 1984-ben átdolgozta Halliday és Hasan taxonómiáját és Gutwinski (1976) közeli ismétlés kategóriái alapján újraértelmezte a szöveg lexikai egységei közötti különféle kohéziós relációkat, mivel nem tartotta kielégítőnek az 1976-os Halliday és Hasan taxonómiában a lexikai és grammatikai kohéziós kapcsolatok elkülönítését, illetve a pontatlanul definiált szemantikai kategóriák miatti elmosódott határokat. Az időközben lefolytatott kutatások cáfolták azt az eredetileg megfogalmazott hipotézist is, hogy a kohéziós elemek láncon belüli arányának növekedésével a szöveg koherensebb lesz. A lexikai kohéziós kapcsolatokat két alapvetően különböző kategóriába sorolta, így megkülönböztette az általános (*general*) kohéziós kapcsolatokat az adott szöveggörnyezetre jellemző eseti (*instantial*) lexikai kohéziós kapcsolatoktól. Az általános lexikai kohéziós kapcsolatok közé sorolta az ismétlést (*repetition*), az antonímiát (*antonymy*), az adott lexikai elemmel rokon értelmű elem szinonímia kapcsolatát (*synonymy*), konkrét – általános összefüggést jelölő haponímiát (*hyponymy*) és a rész – egész kapcsolatot jelölő meronímiát (*meronymy*). Az adott szövegre jellemző pillanatnyi relációk közül megkülönböztette az egyezést (*equivalence*), az elnevezést (*naming*) és a hasonlóságot (*semblance*). Hasan különbséget tett az ugyanarra a szövegelemre visszautaló szövegfüggő láncok (*identity chain*) és a nem szövegfüggő láncok (*similarity chain*) között. Ugyanígy a lexikai elemeket is két kategóriába sorolta aszerint, hogy azok részt vesznek vagy sem a láncalkotásban. A szövegszerveződés szempontjából meghatározó lexikai elemek (*relevant tokens*) részt vesznek a láncalkotásban, míg a szöveg témájához nem kapcsolódóak (*peripheral tokens*) nem alkotnak láncokat.

Halliday (1985) felfogásában a tematikus struktúra és a fókusz a szövegalkotásban fontos szerepet játszik. Az elbeszélő és érvelő szövegekben az adott téma a szöveg folyamán többnyire fókusz pozícióban van, míg más regiszterek esetében az előző tagmondat rémája képezi általában az aktuális tagmondat

témáját. A fókusz a legfontosabb információt tartalmazza és az információs fókusz a mondat végére kerül. Ezzel szemben a téma rendszerint a mondat elején dominál. Halliday a szövegalkotás másik fontos elemének a kohéziót tekinti, ami a jelentéstani kapcsolatok egy nagyon széles tartományának kódolását végzik, úgynevezett referenciális láncokat képez. A szövegszerkezet ugyan nem grammatikai, hanem szemantikai természetű, de a különböző referenciális láncok alkotó elemei más láncok elemeivel grammatikai viszonyban is állhatnak, ami a szövegkohéziót tovább erősíti. A kohézióra Halliday a szöveghez hasonlóan folyamatként tekint és megvalósulásának négy fajtáját különbözteti meg az angol nyelv esetében, amelyek (i) a referencia, (ii) az ellipszis és helyettesítés, (iii) az összekapcsolás és (iv) a lexikai kohézió. A lexikai kohéziót a szöveg lexikai egységei közötti különféle relációk hozzák létre, melynek fajtái az (1) Ismétlés (repetition) pontosan egyező, inflexiós és derivációs változatai (2) a szinonímia viszonyok és (3) a kollokációk (collocation). A szinonímia viszonyok között megkülönbözteti a referencia azonossággal rendelkező szinonimát (synonymy) és fölérendelést (superordinates), valamint a referencia azonossággal nem rendelkező, a konkrét – általános viszonyra utaló hiponímiát (hyponymy), a rész – egész kapcsolatra utaló meronímiát (meronymy), és antonímiát (antonymy). A kollokációk elemei gyakran együtt szerepelnek a szövegben, általában hasonló lexikai környezetben fordulnak elő, ám rendszerszintű szemantikai kapcsolat nincs az elemek között.

De Beaugrande és Dresslernél (2000) a kohézió, mint már említettük a szövegszerűség hét ismérvének egyikeként szerepel. De Beaugrand és Dressler felfogásában a kohézió grammatikai függőségeken alapul és azt tükrözi, hogy „a hallott vagy látott szavak között milyen kölcsönös összefüggések vannak egy adott szószorozaton belül” (De Beaugrand és Dressler 2000: 24). A közvetlenül a szövegfelszínen megragadható kohéziós eszközök tekintetében a szerzőpáros megkülönbözteti (1) a visszatérést – a már használt elemek és minták teljes vagy részleges (pl.: másik szóosztályban történő: *forog – forgó*) ismétlését – (2) a teljes szerkezetet új elemek beiktatásával ismétlő párhuzamot, (3) az adott szövegegység tartalmát más kifejezésekkel körülíró parafrázist, (4) a tartalmas elemeket tömörítési céllal helyettesítő szómásokat (névmások, igemások, hasonlók), (5) a szövegfelszín

bizonyos elemsorozatának részleges ismétlésén alapuló ellipszist, (6) és az igeidő, az igeszemlélet és a kötőszavak révén történő beillesztéseket.

A kohéziós kutatások további eredményeire támaszkodva dolgozta ki Hoey rendszerét (1991), aki szintén a lexikai relációk alapján mutatta ki a mondatok közötti összefüggéseket. Meghatározása szerint a mondatok között abban az esetben van kapcsolat, ha legalább három közös referenciapont, ismétlődő kapocs szerepel bennük, ami az egymástól távol eső mondatok esetében is megvalósulhat. A szöveg tartalma szempontjából a kapcsolódások megléte következtében elkülönülnek a tartalom szempontjából fontos információkat hordozó központi mondatok és a témához lazábban kapcsolódó mondatok. Hoey taxonómiája az ismétlések típusait tekintve megkülönbözteti a lexikai ismétléseket, a parafrázisokat és a szubsztitúciós kapcsolatokat jelölő nem lexikai ismétléseket. A lexikai ismétlések és a parafrázisok esetében különbséget tesz egyszerű és komplex esetek között.

Szikszaíné Nagy (1999) a szöveg jelentésbeli egységét a téma változatlanóságában és a szerkezeti egységek jelentésének egymásutánosságában látja, míg a szöveg jelentésbeli összetartó elemeként a kohéziót jelöli meg. A szöveg lineáris kohéziójának megteremtődésének okát a fogalmi kategóriák egymásba épülésében és a lexikai egységek jelentésbeli viszonyainak meglétében látja. A kohéziós eszközök között megkülönbözteti a korreferenciát, a mezőösszefüggést, az asszociációs mezőt, a mellérendelő kötőszavakat és az ellipszist. Az azonos valóságalemeire utaló, koreferenciális szövegelemek esetében szemantikai ekvivalenciáról beszél, ami megvalósulhat egyszerű ismétlés révén vagy szinonimákkal, hiperonimákkal, hiponimákkal, antonimákkal és proformákkal. Megfogalmazásában az azonos jelentéssztyálya tartozó lexikai elemek közös valóságvonatkozása által megteremtett tematikus kapcsolat a mezőösszefüggés, míg az asszociációs mező a szavak asszociációs erőterének következménye. A gondolatkötések gyakori és szokványos megnyilvánulásait látja a mellérendelésben, míg a szemantikai hiányon alapuló ellipszist pedig az egyik legerősebb szövegkapcsoló tulajdonságként tartja számon.

A szöveg strukturális elemeinek lineáris kapcsolódását Canisius és Knipf (1996: 68–69) a német szaknyelvi terminológiára hivatkozva konnektív kapcsolatnak nevezi, míg a mondat különböző elemei, a tagmondatok, a mondatok és a nagyobb

szövegegységek közötti oksági, temporális és lokális (stb.) kapcsolatokat megteremtő elemeket konnektoroknak. A kapcsolódásokat biztosító egy- vagy többszavas elemeket a szerzők a kapcsolódások típusa szerint öt nagyobb csoportba sorolják be, így megkülönböztetik a morfológiai (kötőszavas, névszói kapcsolatok és egyezőségen alapuló viszonyok), a szintaktikai (ellipszisen alapuló kapcsolat, mondaton belüli szórenden alapuló kapcsolat), a lexikai (deriváción és összetételen alapuló viszonyok, szinonima és hiperonima viszonyok), a stilisztikai (anaforikus kapcsolat és ismétlés) és a prozódiai (intonáción alapuló viszony) kapcsolódásokat.

A modern szövegnyelvészetben a kohézió gyakran a koherenciával együtt kerül említésre, mivel többek megítélése szerint „azonos jelenségre vonatkoznak a konnexió, kohézió és a koherencia fogalmak, amelyek (a szemiotika háromelemű jelfelfogása értelmében) a szöveghkapcsolódás grammatikai, szemantikai, illetőleg pragmatikai dimenzióira utalnak, ám ezeket nem elkülöníthető szintekhez, hanem egymással kölcsönviszonyban álló funkciókhoz kötik” (Fehér 1999: 469). A német szövegnyelvészeti szakirodalom a koherencia és kohézió két központi fogalmát a szűkebben vett konkrét értelmezés mellett szélesebb értelmezésben is használja, ahogy erről Csatár Péter és Haase Zsófia a két fogalom értelmezését összegző munkájában számot ad:

Ez a széles értelmezés azonban nem csak annyit jelent, hogy az újabb irányzatok újraértelmezik, és ennek során bővítik a koherenciához, illetve a kohézióhoz sorolandó jelenségek körét. Sokkal inkább arról van szó, hogy a koherencia és a kohézió fogalma több más szövegnyelvészeti szakkifejezés meghatározásában is egyre inkább szerephez jut. Vagyis olyan metaelméleti terminusként alkalmazzák őket a kutatók, amelyek arra szolgálnak, hogy a szövegnyelvészet által vizsgált jelenségeket jelölő elméleti terminusok jelentését magyarázzák. (Csatár P. és Haase Zs. 2011: 60)

3.2.2 A lexikai kohézió

A kohéziós eszközök közül a lexikai kohézió különös figyelmet érdemel szövegalkotó, szövegszervező szerepe miatt, ahogy azt a mondathatárokon

túlmutató szemantikai-szintaktikai kapcsolatokat vizsgáló, az előzőekben bemutatott ismétlésmodellek mindegyike hangsúlyozza (Halliday és Hasan 1976, 1989, Hasan 1984, Hoey 1991). A szövegek lexikai mintázatának feltérképezése, a lexikai elemek ismétlődéseinek és szemantikai kapcsolatainak feltárására irányuló kutatások a magyar szövegnyelvészeti szakirodalomban is a kezdetektől helyet kaptak.

Károly Sándor (1979: 53) az általa „szemantikai koherencia” névvel illetett jelenség esetében a mondaton belül, vagy nagyobb szövegegységben az „ugyanarra a valóságélemre, személyre, tárgyra” vonatkozó szavak „közös valóságra utaló” szerepének következményeként írja le a szöveg kohéziójának, folyamatosságának és egységének megteremtését. E szerint a „szöveg szemantikai összefogó elemei között mindazok a szótári, lexikai egységek szerepelnek, amelyek a szövegben megismétlődnek.” A szemantikai kapcsolatok, a szemantikai progresszió esetében ugyanakkor a szövegre jellemző többi törvényszerűséghez hasonlóan a szövegtípusok szerinti specializáltságot tartja érvényesnek (i. m.: 55).

Békési Imre (1979: 56–57) számára a szövegben az „azonos vagy egymáshoz közeli szemantikai elemek” ismétlődése teremti meg a szövegegység szerkezetét, amit szemantikai–gondolati–logikai jelzővel lehet minősíteni, „de megnevezhetjük egy negatív előjellel is: nem grammatikai.” E lényegileg nem grammatikai kapcsolatok esetében fontos tudnunk, milyen a „kapcsolás típusa, formája, alakja és módja”, illetve a kapcsolás „terjedelme, ereje, szorossága”, mivel Békési a konstrukciós forma felépítésének szempontjából megkülönböztet fő- és alkapcsolásokat.

Károly Krisztina felhívja a figyelmet arra, hogy a kohéziós eszközök nemcsak a mondatokat és mellékmondatokat összekapcsoló nyelvtani struktúrák azonosításához szükségesek, hanem a szövegelemek közötti mondathatárokon túlmutató kapcsolatok megértéséhez is (Károly 2002: 19)¹. Hoey (1991) taxonómiájából kiindulva Károly 2002-ben megalkotta a lexikai ismétlődés szövegszervező funkcióját vizsgáló rendszerét. A szövegfelszín kohézióért felelős elemeinek elsődleges funkcióját a kapcsolatok jelzésében és ezáltal a szöveg folyamatosságának biztosításában látja (i. m.: 65). Taxonómiájában a szöveg kohézióját az olyan lexikai egységek biztosítják, amelyeknek jelentése nem vezethető vissza alkotóelemeiknek jelentésére. Károlynál a kohéziós kapcsolatok két alapvető

típusát a lexikai- és a szövegalapú, pillanatnyi relációk jelentik. A lexikai relációk esetében formailag különbséget tesz egyszerű és derivált alakok között, míg tartalmilag megkülönbözteti, ha (i) a lexikai egység változatlan formában ismétlődik (ismétlés), illetve (ii) amikor a lexikai egységnek csak az információtartalma ismétlődik meg (szinonima, ellentét, hiponima, meronima) (i. m.: 103–106).

3.2.3 A lexikai egység

Petőfi S. (1988: 42) szövegfelfogásából kitűnik, hogy „célszerűnek látszik „dominánsan nyelvi szövegekről” beszélni, ahol a dominancia (intuitive) azt jelenti, hogy domináns információ-hordozóknak a vehiculum lexikai elemeit tekintjük.” Ezeknek az információ hordozó elemeknek a szemantikai-szintaktikai ismétléseit kívánják a különböző lexikai kohéziós taxonómiák megragadni.

Károly 2002-es ismétlés modellje a kohéziós elemzések megvalósításához szükségesnek tartja a szöveg kohézióját biztosító lexikai egység meghatározását, mivel a szövegek kohéziós kapcsolatait nem feltétlenül a helyesírási szavak szemantikai kapcsolata biztosítja. Károly lexikai egységként az olyan egységet definiálja, amelynek jelentése nem vezethető vissza alkotóelemeinek jelentésére, az egyes szavak mellett megjelennek a szavaknál nagyobb, a több szóból felépülő nyelvi egységek, a kollokációk és az állandósult szókapcsolatok is. Az elemzések alapjait szolgáló lexikai egységek lehetnek tehát (i) önálló szavak (pl.: *sun, books*), (ii) idiómák (pl.: *bitter pill*), vagy (iii) állandósult szókapcsolatok (pl.: *give up, National Theater*) (Károly 2002: 95).

3.2.4 A kohézió kutatás legismertebb képviselői

A lexikai kohézió vizsgálata a szövegkutatás egyik legvitatottabb és meghatározó irányzata, melynek tárgya elsősorban írott szövegek leíró, funkcionális elemzése. A szövegközpontú szemlélet az 1960-as évek végétől vált népszerűvé, mivel a mondatokból szöveggé szerveződött nyelvi formák kapcsán felvetődő kérdések nem mindegyikét lehetett már a mondatgrammatika keretein belül

megválaszolni (Tolcsvai Nagy 1994). A szöveget mindinkább egységes egésznek, a nyelvhasználat egységének tekintő felfogás szerint a szöveg és alkotómondatainak kapcsolatában a hangsúly már nem arra került, hogy a szöveg mondatokból áll, hanem inkább a szöveg mondatok által történő realizációjára (Halliday és Hasan 1976).

A szöveg jelentésbeli összefüggés-hálózatát kutató szövegszemantikai megközelítés a szöveg szempontjából releváns lexikai elemek ismétlődését vizsgálja. Halliday és Hasan meghatározásában „a kohézió a szöveg egy elemének a szemantikai viszonya egy másikhoz, amely elem az elsőnek az értelmezéséhez szükséges” (Halliday és Hasan 1976: 8). A szöveg kohéziós kapcsolatait feltáró tanulmányok a lexikai elemeket a szövegben betöltött szerepük szempontjából két csoportra osztják. A lexikai kohéziós kutatások középpontjában a centrális vagy releváns elemek állnak, amelyek a téma fő gondolatát viszik, míg a témához lazábban vagy egyáltalán nem kapcsolódó lexikai elemeket figyelmen kívül hagyják (Beaugrande és Dressler 2000). De Beaugrande és Dressler mellett számos kutató felismerte a kohézió meglétének jelentőségét a szövegprodukciónak szempontjából, és a szövegnyelvészeti területen belül mind a grammatikai, mind a lexikai kohézió kutatása fontos szerephez jutott.

A kohéziókutatás célja azoknak a nyelvi eszközöknek a vizsgálata, amelyek a szöveg felszínén töltenek be szövegösszetartó, szövegszervező funkciót. A nyelvi rendszerek és az egyes nyelvek és kultúrák nyelvhasználati és retorikai hagyományainak eltérései miatt a különböző nyelvek a kohéziót is eltérő minőségű és mennyiségű eszköz segítségével fejezik ki. (Károly 2007: 62)^{II}

A lexikai kohézió vizsgálata iránti érdeklődés növekedésével napjainkra egyre népszerűbbek lettek a témakörben folytatott korpusz alapú kutatások. A lexikai kohéziókutatás korpusznyelvészeti feltárásának növekvő igényét bizonyítja többek között a Benjamins kiadó gondozásában megjelent, kifejezetten a témakörben megvalósított kutatások feldolgozására szentelt gyűjteményes kötet. A *Lexical Cohesion and Corpus Linguistics*ben megjelentetett hat tanulmány közül egy a kohézió és a retorikai struktúra kapcsolatát, három a lexikai kohézió hangzó szövegekben

történő vizsgálatát, kettő pedig a tanulói korpuszokban megvalósuló realizációját kutatja.

Morley kutatása egyrészt a klasszikus megközelítés értelmében az azonos szemantikai mezőbe tartozó szavak közötti logikai kapcsolatokat vizsgálja, másrészt az újságcikkek főcímei és szövegtartalma közötti összefüggéseket. A vizsgálat kimutatja, hogy mivel a címek alapján a cikkek lexikai tartalma előrevetíthető, ezért ezeknek igen fontos a kohéziós szerepe a sajtószövegekben. Ugyanakkor Morley megállapítja azt is, hogy az újságcímek sokszor utalnak az argumentum során alkalmazott grammatikai struktúrákra és a lexikai kohézió hozzájárul a szöveg retorikai struktúrájának kialakításához. A szöveg tematikus pozíciójában egy-egy ellenérvet elindító lexikai elem sokszor sztereotipikusnak tűnik, ám ezek felismerése elsősorban a nyelvtanulók számára a szöveg értelmezésénél nagy segítséget jelenthet.

Nesi és Basturken (2006) a *British Academic Spoken English* (BASE: <http://www2.warwick.ac.uk/fac/soc/al/research/collect/base/>) és a *Michigan Corpus of Academic Spoken English* (MICASE: <http://quod.lib.umich.edu/m/micase/>) összesen 160 tudományos előadásából kialakított korpusz esetében vizsgálta a tudományos szövegekben előforduló lexikai kohéziós elemeket, több szavas kifejezéseket és a gyakori együttes szóelőfordulásokat. A korpuszban a WordSmith Tools cluster funkciójának segítségével azonosították a minimum tízszer megismétlődő 4 szavas együttes előfordulásokat, amelyek százalékos előfordulási gyakoriságát a *Longman Grammar of Spoken and Written English* (Biber et al. 1999) társalgásra és tudományos cikkekre jellemző adataival vetették egybe. Az eredmények ismeretében a szerzőpáros rávilágít a négyzavas ismétlődő előfordulások szövegépítő jelentőségére és a kohézió megteremtésében játszott szerepére.

Warren (2006) az ötcsillagos Hong Kong-i hotelekben felvett, a pályázók szakmai karrierje szempontjából nagy jelentőséggel bíró állásinterjúkat elemezte, ahol a felek által az aktuális szövegkörnyezetben az informativitás és a szöveg kohéziója szempontjából előnyben részesített lexikaválasztást vizsgálta. A vizsgált anyag a kétmillió szövegszavas *Hong Kong Corpus of Spoken English* (HKCSE: <http://rcpce.engl.polyu.edu.hk/HKCSE/>) anyanyelvi angol beszélők és Hong Kong-i

kínaiak párbeszédeit tartalmazó korpusz része, amit Warren prozódiaileg is annotált. A szógyakorisági listák, kulcsszó listák, konkordancia vizsgálatok és prozódiai elemzések eredményeként a kutatás végkövetkeztetései között szerepel, hogy a beszélők a lexikai kohéziós elemeket részesítik előnyben a nem lexikaiakkal szemben, ami összefüggésbe hozható azzal a beszédhelyzetben aktiválódó feltételezésükkel, hogy ezek lexikai kohéziós elemek az adott szituációs helyzetben informatívabbak.

Cheng (2006) a *Hong Kong Corpus of Spoken English* (HKCSE) részét képező SARS (Severe acute respiratory syndrome) krízissel foglalkozó szöveganyagot elemezte. A ConcGam^{III} programmal megtalált hat leggyakoribb lexikai elem kollokációs kapcsolatainak feltárása után megállapította, hogy az együttes előfordulások identifikációja segítségével árnyaltabb képet kaphatunk a szövegek koherenciájáról, mintha csak a szövegekben a lexikai kohéziót vizsgálnánk.

Flowerdew (2006) és Mahlberg (2006) kutatásai az angol nyelv tanítása szempontjából foglalkoznak a kohézióval. Flowerdew főként a tanulói korpusz hibaelemzéseinek tanulságai alapján vonja le következtetéseit a kínai diákok kohéziós hibáival kapcsolatban. Mahlberg inkább a kohézió oktatásának nehézségeit veszi számba az angol mint idegennyelv oktatásban és a kohézió témakörének az idegennyelv oktatásban való integrálására a korpusz alapú módszert tartja a legmegfelelőbbnek.

3.3 A lexikai kohézió kutatása és a fordítás

3.3.1 *Miért érdekes számunkra a lexikai kohézió vizsgálata fordított szövegekben?*

Vélekedésem szerint az átvitel következtében a célnyelvi szövegek lexikai kohéziós mintázatában bekövetkező változások meghatározó jellemzői a fordításnyelvnek. Értekezésem központi gondolatai közé tartozik annak az állításnak a bizonyítása, hogy a fordítás során a célnyelvi szöveg kohéziós eszközhasználatában bekövetkező változások az eltolódásjelenségek sorába tartoznak. Feltevésem szerint a fordított szövegek kohéziós mintázata nem egyezik sem a kiindulásukként szolgáló forrásnyelvi szövegek, sem a célnyelven írt autentikus szövegek kohéziós

mintázatával. Ezért tartom a lexikai kohézió vizsgálatát különösen érdekesnek a fordításkutatás szempontjából, amit a következő négy szerző argumentumai is alátámasztanak. (i) Hoey (1991) véleménye szerint a kohézió inkább lexikai, mint grammatikai relációk eredménye, így a szövegszerveződés szempontjából a lexikai kohézió kitüntetett jelentőséggel bír. (ii) Hatim és Mason (1990) megállapítják, hogy a lexikai kohézió a szövegalkotó szándékos lexikai választásán múlik, szemben a többnyire az egyes nyelvsziszterek jól formáltsági előírásait kielégítő grammatikai kohézióval. (iii) Halliday és Hasan (1976) szerint – mivel a kohézió nyelvi jelenség – az egyes nyelvekre jellemző különböző sajátosságok következtében az eltérő nyelvek kohéziós eszközhasználata sem fedheti száz százalékban egymást. (iv) Végül, de nem utolsón ide kívánczik Blum-Kulka (1986) azon megállapítása, mely szerint ha a különböző nyelvek eltérő módon tolerálják a kohéziós elemek ismétlését, akkor a fordításra eltolódás lesz jellemző ebben a tekintetben is.

... cohesion, or sequential connectivity of surface elements – are much more likely to be language-specific or text-specific. There are many possible cohesive devices capable of reﬂaying, say, a given relationship between propositions. And in a given language, some are likely to be preferred options. (Hatim, Mason 1990: 195)^{IV}

A fordított szövegek jellemzőit kutató Balaskó (2005: 111–112) szerint „számos olyan jelenség fedezhető fel, amely köztesnyelvnek minősíthető annak ellenére, hogy a fordítók – a nyelvtanulóktól eltérően – általában kompetens és tapasztalt használói a célnyelvnek. Köztesnyelvi jelenségek azonban nemcsak az idegen nyelvre, hanem az anyanyelvre történő fordításnál is megjelennek, ami azt a feltételezést támasztja alá, hogy bár az idegen nyelvi / célnyelvi kompetencia hiánya elősegíti a köztesnyelvi formák megjelenését, nem okolható egyedül azok keletkezéséért.” A fordítás során keletkezett szövegek tulajdonságai Balaskó (2005: 113) szerint kimutathatóak, mivel „ha a fordításnyelv önálló nyelvi rendszer, akkor megkülönböztethető más nyelvi rendszerektől, azaz felismerhető.”

3.3.2 Lexikai kohézió és fordítás

Az eddigi tudományos eredmények ismeretében megállapítható, hogy már az egy nyelven belüli szövegprodukciónak esetében is nagy jelentőséggel bír a kohéziós eszközhasználat megléte és a használata milyensége. Ahogy Klaudy fogalmaz „anyanyelvünkön nem tudatosítjuk azt, hogy mondatszerkesztési műveleteinknek milyen szövegszintű következményei vannak” (2006: 142), ezzel szemben a fordítások esetében a célnyelvi szövegek helyes kohéziós eszközhasználatára nagyobb gondot kell fordítani. Bár a fordítástudomány jelenlegi álláspontja a kohézió átvitelével kapcsolatban az, hogy „a forrásnyelvi szöveg kohéziós elemeit a fordítónak nem kell átvinnie, csupán helyesen dekódolnia” (Klaudy 2006: 140), a fordítástudomány mégis egyre nagyobb figyelmet szentel a lexikai kohézióknak mind a leíró fordításkutatás deskriptív elemzéseiben, mind a fordításoktatás preskriptív követelményrendszerének kialakításánál. Ma már a szakirodalom is hangsúlyozza azt a megállapítást, hogy a fordítások minőségének lényegi eleme a kohézió megléte, illetve helyes átvitele.

Heltai például a fordító feladatai közé sorolja, „a természetesnek vett „nyelvhelyességi normákon” kívül (Heltai 2005b: 169) „az optimális szövegkohézió” biztosítását (Heltai 2005a: 36), továbbá „a szöveg megformálása során elsősorban a [...] szövegszintű normákra, konvenciókra, illetve disztribúciós mintákra” való ügyelést. A célnyelvi normák részét képezi tehát mondat és szövegszinten is a megfelelő kohéziós eszközhasználat is (Heltai 2005b: 169).

A fordított szövegekben a kohézió megteremtésének funkcionális jelentőségére utal Klaudy a kvázi-helyesség fogalom bevezetésével, mely szerint „a fordított magyar szövegek kohéziója csökken, az olvasó késve azonosítja a mondatok szerkezetét, nem világos a hangsúlyos és hangsúlytalan részek viszonya”. A kvázi-helyes szövegek esetében a megszokott célnyelvi szövegalkotási normáktól való eltérés miatt „az olvasó számára sokkal nagyobb megterhelést jelent a mondanivaló kihámozása a szokatlan nyelvi formából” (Klaudy 1987: 7), illetve az olvasók az „egész szöveget mégis elutasítják, mert nem felel meg a helyesen szerkesztett szövegről alkotott intuitív elképzeléseiknek” (Klaudy 2004: 389).

Szabó Mihály a szlovákiai kétnyelvű célközösség számára készülő, többnyire nem önálló magyar nyelvű szövegtermékeket vizsgálva sürgeti a fordítások értékeléséhez elengedhetetlen „objektív és egységes szempontrendszer” kialakítását. Szabó Mihály szerint szövegszinten a kohézió is meghatározó elemnek kell lennie a „retorikai cél; a műfaj; a grammatikai és lexikai regiszter – a szaknyelv; a tematikus rend” mellett (Szabó Mihály 2003: 58).

A fordításoktatás számára kidolgozott Dróth-féle hibatipológia formatív értékelési szempontrendszerének öt kritériuma között is szerepel a lexikai kohézió. A hibák súlyozásának rendszerével Dróth azt szorgalmazza, hogy a más célnyelvi normák figyelembevételéhez hasonlóan megteremtődjön „az anyanyelvre fordított szöveg idegenszerűség nélküli, a célnyelvi igényeknek és hagyományoknak megfelelő nyelvi megformálásához szükséges készségek kialakítása, fejlesztése”, s ezek között a „megfelelő kohéziós eszközök alkalmazásának képessége.” (Dróth 2002: 95) A fordítás eredményeként keletkezett célnyelvi szövegek osztályozásakor figyelembe vett kritériumok elsődlegesen arra összpontosítanak, hogy a célnyelvi szövegek mennyiben felelnek meg a célnyelvi normák követelményrendszerének, mivel a „fordítások formatív értékelésekor már nem a forrásnyelvi szöveget vetjük egybe a célnyelvi szöveggel, hanem a célnyelvi szöveget elemezzük.” (i. m.: 98)

A feltáró kontrasztív fordításelemzés új perspektívát nyithat az anyanyelv tudatosításában. Ez egyre sürgetőbb a Klaudy által 2001-ben előrevetített és azóta mindinkább jellemző folyamat miatt. A fordítandó szövegmennyiség exponenciális növekedése következtében az összes fordítást „...hivatásos fordítókkal elvégeztetni lehetetlen. A nyelvek közötti megfeleltetés a jövőben nem a hivatásos nyelvi közvetítők monopóliuma, hanem a művelt magyar állampolgár napi rutinja lesz” (2007: 55). Ezért is tartom a fordítás útján keletkezett szövegek kontrasztív vizsgálatát, illetve a fordításokra jellemző fordításnyelv sajátosságainak feltárását az alkalmazott nyelvészet kiemelkedően fontos kutatási területének. A kutatási eredmények többek között hozzásegíthetnek „a magyar nyelvleírás egy korszerűbb, funkcionális megközelítéséhez” (Klaudy 2007: 55).

3.3.3 *Leíró fordításkutatás*

A szövegalkotással a fordítási tevékenység tükrében foglalkozó munkák közül említést érdemel Heltai *A fordító és a nyelvi normák* című tanulmányorozata, ahol a szerző bár a „kohéziós eszközök használatát” a „szövegalkotási normák közé” sorolja (Heltai 2004: 417) mégis megállapítja, „hogyan a szövegalkotási normák nem fogalmazhatók meg olyan egyértelműen és explicit módon, mint a morfológiai és szintaktikai normák”.

A szövegalkotási normák közé sorolhatunk részben olyan jelenségekkel kapcsolatos normatív elvárásokat, amelyeket hagyományosan a szintaktika körébe szoktunk sorolni, így például a tematikus tagolás és a fókusz kérdéseit. Ide sorolhatjuk a kohéziós eszközök használatát is. Normatív jellegűnek tarthatjuk azt a követelményt, hogy a szövegben a referencia nyomon követhető legyen. A szövegnormák, szemben a nyelvtani normákkal, gyakran statisztikai jellegűek, az egyes szerkezetek, lexikai vagy frazeológiai egységek különböző szövegtípusokban való disztribúciójával kapcsolatosak, és ebből következőleg többnyire nem nyilvánvalóak. (Heltai 2004: 417)

Heltai a szövegalkotási normák nehezen megragadható volta ellenére a fordító feladatai közé sorolja „a természetesnek vett „nyelvhelyességi normákon” kívül (Heltai 2005b: 169) „az optimális szöveg kohézió” biztosítását (Heltai 2005a: 36), illetve „a szöveg megformálása során elsősorban a [...] szövegszintű normákra, konvenciókra, illetve disztribúciós mintákra” való ügyelést, melynek részét képezi a megfelelő kohéziós eszközhasználat mondat és szövegszinten (Heltai 2005b: 169).

A múlt század második felében tudományosan elfogadottá vált az a felismerés, hogy az egymást követő mondatok sorából csak a szövegalkotó tudatos tevékenysége következtében alakulhat ki szöveg: a „szövegmondatok láncává a mondatok sorát elsősorban a szövegalkotó teszi a szövegösszefüggés megvalósításával, azaz objektíve létező szövegtényezők miatt „értékelődik” egy nyelvi jelsor szövegnek” (Szikszainé Nagy 1999: 56). A tudatos szövegszerveződés

feltételeinek vizsgálata mellett felvetődött a szövegértékűség kritériumainak meghatározása iránti igény.

Sziksainé Nagy (1999: 58–72) a szövegösszetartó erő meglétében látja a szöveggé szerveződés kritériumát. Szerinte a szövegösszetartó erő különböző fajtáinak működése eredményezi a textualitást. Sziksainé Nagy a szöveg szintjei szerint különbözteti meg a szövegösszefüggést eredményező szövegszervező erő fajtáit, így a pragmatikai szintű összefüggést megteremtő koherenciát, a szemantikai viszonyért felelős kohéziót, a szintaktikai szintű kapcsolatért felelős konnexitást, a szerkezeti szintű kapcsolatokat létrehozó kompozicionális konnexitást, a stilisztikai szintű összefüggéseket megteremtő stíluskohéziót, a hangzó szövegek fonetikai prozódia jellegű akusztikai kohézióját, az írott szövegeket jellemző formai-tipográfiai kötöttségeket megteremtő vizuális egybetartozást, és a nyílt, vagy rejtett utalások által megvalósuló intertextuális kohéziót.

Petőfi S. (2004: 29–30) meghatározásában azon szemiotikai objektum elégti ki a textualitással szemben támasztott követelményeket, amely „egy adott vagy feltételezett kommunikációs helyzetben, egy adott vagy feltételezett funkciónak (intenciónak) eleget tevő, összefüggő és teljes tényállás-konfigurációt juttat kifejezésre, egy összefüggő és teljes megformáltsággal rendelkezőnek tekinthető dominánsan verbális objektum formájában.” E szemlélet szerint a szöveg esetében hangsúlyossá válik továbbá a szöveg jelentéshordozó elemeinek továbbra is dominánsan lexikális jellege, az írásos vagy hangzó manifesztáció, és a szöveg autonóm teljes, vagy részszoveg jellege.

3.3.4 Fordításoktatás

A fordításoktatás számára a megfelelő lexikai kohéziós eszközök célnyelvi használatának helyes elsajátíttatása az előző pontban leírtakból következően alapcélkitűzésként jelenik meg. A fordítóképzés célja – mint ahogy azt Dróth tantárgyleírásában olvashatjuk – „az anyanyelvre fordított szöveg idegenszerűségek nélküli, a célnyelvi igényeknek és hagyományoknak megfelelő nyelvi megformálásához szükséges készségek kialakítása, fejlesztése.” (Dróth 2002: 95)

Szabó Mihály a szlovákiai kétnyelvű célközönség számára készülő, többnyire nem önálló magyar nyelvű szövegtermékeket vizsgálva sürgeti a fordítások értékeléséhez elengedhetetlen „objektív és egységes szempontrendszer” kialakítását, melynek szövegszinten a „retorikai cél; műfaj; a grammatikai és lexikai regiszter - szaknyelv; tematikus rend” mellett a kohézió is meghatározó elemnek kell lennie (Szabó Mihály 2003: 58).

A „megfelelő kohéziós eszközök alkalmazásának képessége” megtalálható a fordítóképzés során folyamatosan finomítandó fordítási kompetencia részei között, s ez a terület nemcsak a készségek elsajátításakor kap figyelmet, hanem iránymutató szándékkal a Dróth-féle ötszintű hibatipológia formatív értékelési szempontrendszerének kritériumai között is szerepel (Dróth 2002: 95). Ezen értékelési kritériumok viszont a fordítás eredményeként keletkezett célnyelvi szövegeket aszerint osztályozzák, hogy mennyiben felelnek meg a célnyelvi normák követelményrendszerének, mivel a „fordítások formatív értékelésekor már nem a forrásnyelvi szöveget vetjük egybe a célnyelvi szöveggel, hanem a célnyelvi szöveget elemezzük” (Dróth 2002: 98).

^I Eredeti szövegrészlet: “... the means of cohesion do not only refer to the grammatical structuring of sentences and clauses, but they also include the actual intersentential relationship that are formed between the elements of the text ...”

... a kohézió jelentése nemcsak a mondatok és tagmondatok nyelvtani szerkezetére vonatkozik, hanem magában foglalja a szöveg egyes részei között fennálló mondatok közötti kapcsolatokat is ...

^{II} Saját javítás. Az eredeti idézet 2. mondata: “A nyelvi rendszerek és az egyes nyelvek és kultúrák nyelvhasználati és retorikai hagyományainak eltérései miatt a különböző nyelvek, kohéziót is eltérő minőségű és mennyiségű eszköz segítségével fejezik ki.”

^{III} A ConcGram© Chris Greaves (Department of English, The Hong Kong Polytechnic University) által fejlesztett keresőmotor

^{IV} ... a kohézió, vagy a felszíni elemek egymás utáni kapcsolódása sokkal inkább nyelv-specifikus, mint szöveg-specifikus jelenség. Sok lehetséges kohéziós eszköz van, ami képes a mondatrészek közötti adott viszony meghatározására. Egy adott nyelv esetében némelyik elem választása előnyt élvez. (saját fordítás)

4 A korpusznyelvészeti módszer előnyei a fordításkutatásban

4.1 Bevezetés

Az utóbbi évtizedek robbanásszerű számítástechnikai fejlődésének köszönhetően számos új eszköz, számítógépes alkalmazás segíti a fordítástudomány területére tartozó alkalmazott nyelvészeti kutatásokat. Ezek az eszközök alkalmasak több szöveg együttes vizsgálatára, a forrásnyelvi és célnyelvi szövegek összehasonlíthatóságát megkönnyítő szövegpárhuzamosításra, gyors, pontos és hatékony gépi lekérdezések elvégzésére.

A korpuszalapú fordításkutatás (corpus-based translation studies) Mona Baker (1993, 1995) iskolateremtő munkásságának köszönhetően terjedt el. Baker hangsúlyozta, hogy a korpusban összegyűjtött szövegek a ténylegesen használt nyelv olyan megtestesülései, amelyek a nyelvész beavatkozása nélkül jöttek létre. A korpuszalapú megközelítés legfőbb erénye ugyanakkor az, hogy empirikusan közelíti meg a nyelv leírását, ragaszkodik a használati előfordulások elsődlegességéhez. E tulajdonságoknak köszönhetően a korpuszalapú fordításkutatás kiválóan alkalmas a fordítás eredményeként létrejött szövegek tulajdonságainak leírására, és a célnyelvi és forrásnyelvi, valamint az autentikus és fordított szövegek közötti különbségek megvilágítására (Baker 1995).

4.1.1 *A korpusznyelvészeti módszer perspektívái az alkalmazott nyelvészeti kutatásban*

4.1.2 *A korpusznyelvészet, mint módszer*

A múlt század közepétől a szövegek véletlenszerű gyűjteményeit és archívumait – a számítógépes tárolókapacitás robbanásszerű növekedésének köszönhetően – a szövegek mérete, fajtája, forrása vagy más meghatározó kritériuma alapján végzett tervszerű adatgyűjtés eredményeként létrejött szövegbankok váltották fel. A modern korpuszok a gépileg olvasható, explicit módon megfogalmazott szempontok szerint összeállított szöveg(-részlet)ek gyűjteményei,

melyek a nyelv empirikus vizsgálatát valódi nyelvhasználati minták alapján teszik lehetővé (v.ö. McEnery, Wilson 1996: 1). Mivel a korpuszban tárolt szövegek a nyelv természetes megvalósulásai, ezért a nyelv adott állapotát vagy egy változatát testesítik meg és ezáltal alkalmasak a nyelvészeti kutatás számára. A *Corpus Linguistics* antológia a következőképpen határozza meg a korpusznyelvészeti módszerek az alkalmazott nyelvészeti kutatásban betöltött szerepét:

Corpus linguistics today is often understood as being a relatively new approach in linguistics that has to do with the empirical study of “real life” language use with the help of computers and electronic corpora. In the first instance, a “corpus” is simply any collection of written or spoken texts. However, when the term is employed with reference to modern linguistics, it tends to bear a number of connotations, among them machine readable form, sampling and representativeness, finite size, and the idea that a corpus constitutes a standard reference for the language variety it represents. (Lüdeling, A., Kytö, M. 2008: 5.)¹

A korpusznyelvészet legfontosabb területei közé tartoznak a leíró nyelvészeti, a lexikográfiai, a szövegnyelvészeti, az alkalmazott nyelvészeti és a fordítástudományi vizsgálatok. A leíró nyelvészet az adott nyelv nyelvhasználatára jellemző lexikai/grammatikai és nem-nyelvi asszociációs mintákat kutatja, azaz azt, hogy milyen nyelvi jelenségek milyen lexikai/grammatikai asszociációkkal társulnak. A lexikográfia számára az egyes szavak különböző jelentéseinek reprezentációja, adott regiszterre jellemző gyakorisága és kollokációs mintázata az érdekes. A szövegnyelvészet az adott regiszterre és szövegtípusra jellemző szövegszintű nyelvi eszközök (referencia, ismétlés) leírásával foglalkozik. Az alkalmazott nyelvészetben belül a korpuszalapú kutatások a szaknyelvkutatásban és a szaknyelvoktatásban, az anyanyelv és idegen nyelv elsajátításának vizsgálatakor, az anyanyelv és idegen nyelv tanításban, a stilisztikában, a fordításkutatásban és a fordító képzésben terjedtek el.

4.1.3 *Korpusznyelvészet a fordításkutatásban*

A korpuszban összegyűjtött szövegek fordítástudományi vizsgálatának (corpus-based translation studies) előnyére Mona Baker hívta fel először a figyelmet (Baker 1993). Baker a korpuszalapú vizsgálatok mellett szóló legfőbb érvként azt emelte ki, hogy a korpuszban összegyűjtött szövegek a ténylegesen használt nyelv megtestesülései, illetve a nyelvi viselkedés természetes megvalósulásai, melyek a nyelvész beavatkozása nélkül jöttek létre. A korpuszokban található szövegek jellemzésénél Sinclair is hangsúlyosan említi, hogy azok mind valódi nyelvhasználati minták, azaz olyan természetes nyelvi szövegek, amelyek jellemzőek a nyelv adott állapotára vagy egy-egy változatára: „[corpus is] a collection of naturally-occurring language text, chosen to characterize a state or a variety of a language” (Sinclair 1991: 171).ⁱⁱ Sinclair máshol a szöveggyűjtés szempontjait számba véve azt emeli ki, hogy a szövegek lehetnek teljes szövegek, vagy szövegrészletek, illetve hogy a korpusz összeállítója explicit módon megfogalmazott nyelvészeti szempontok szerinti és nem véletlenszerű adatgyűjtés eredményeként hozza létre az adott nyelv vagy nyelvváltozat mintáját megtestesítő korpuszt: „a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language” (Sinclair 1995:).ⁱⁱⁱ

A korpuszalapú fordításkutatás ismérvei között is fontos tehát számon tartani, hogy empirikusan közelíti meg a nyelv leírását, ragaszkodik a hiteles használati előfordulások elsődlegességéhez, azaz a beszélt vagy írott nyelvhasználatot nem izolált példákkal szemlélteti. A korpuszalapú fordításkutatás vizsgálatának fókuszában egyrészt az autentikus (nem fordított) és fordított szövegek összevetése, másrészt a forrásnyelvi és célnyelvi szövegek nyelvi jellemzőinek leírása áll. Az előbbi esetben az egynyelvű korpuszokra jellemző lexikai, szintaktikai és szövegszintű nyelvi jelenségek vizsgálatai a szövegek olyan egyedi sajátosságaira mutatnak rá, amelyek kialakulása sem a forrásnyelv, sem a célnyelv hatásával nem magyarázhatóak és nagy valószínűséggel a fordítások univerzális tulajdonságai közé sorolhatóak. Ezzel szemben az utóbbi esetben a kétnyelvű párhuzamos korpuszok segítségével a kutató a nyelvpár specifikus fordítói magatartást, a gyakorlott fordítók fordítási problémákra adott megoldási javaslatait tudja vizsgálni. A valós nyelvi

előfordulások vizsgálatának terjedő népszerűségéről árulkodik a korpuszalapú fordításkutatások 1990-es évektől egyre növekvő száma, melyek közül terjedelmi keretek miatt itt csak a legjelentősebbek említésére szorítkozom: Baker (1993, 1995, 1996), Laviosa (1997, 1998, 2002), Kennedy (1998), Tymoczko (1998) Stubbs (2001), Kenny (2001), Olohan (2004), Teubert és Cermáková (2007).

A korpuszalapú fordítástudomány népszerűségének jövőbeni növekedését vetíti előre Tymoczko az e kutatási irány által megvalósítható távlati lehetőségek számbavételével (Tymoczko 1998). A fordítástudomány igényeinek megfelelő technológiák alkalmazása és adaptálása mellett Tymoczko kiemeli a korpuszok felhasználásának köszönhetően kutathatóvá váló témák közül a nyelvészeti és kulturális megközelítés integrálását a fordítástudományban és az ideológiák fordításokra gyakorolt befolyásának vizsgálatát. A kulturális vonatkozású és nyelvhelyességi kérdések tisztázását segítő nagy referencia korpuszok kapcsán Stewart is megemlíti (Stewart 2000), hogy a fordító ezeket a korpuszokat az idegen nyelvre történő fordítás során a hagyományos szótáraknál sokkal jobban tudja hasznosítani a nagy autentikus nyelvi mintának köszönhetően a nyelvi és kulturális korlátok leküzdésére. Tymoczko továbbá annak a reményének is hangot ad, hogy a korpuszalapú fordítástudománynak köszönhetően nyitva áll a lehetőség a fordítás elméleti és gyakorlati területének egymáshoz közelítése, sőt akár újraegyesítése előtt is (Tymoczko 1998).

4.2 Eszközök, számítógépes alkalmazások

4.2.1 A korpuszok legalapvetőbb ismérvei

Az 1960-as évektől megjelenő korpusznyelvészet (corpus linguistics) az a „nyelvészeti irányzat, mely a nyelv és nyelvhasználat vizsgálatát speciális módszerek és számítógépes programok segítségével korpuszra alapozva végzi” (Szirmai 2005: 170). A korpuszok (corpora) maguk olyan a “nyelvészeti vizsgálatok céljából, bizonyos szempontok alapján összeválogatott írott vagy beszélt nyelvi szövegek” (Szirmai 2005: 170), amelyek elektronikus formában állnak rendelkezésünkre.

A nyelvész korpuszokkal kapcsolatos tevékenységét Kennedy négy nagyobb csoportba sorolja (Kennedy 1998), annak megfelelően, hogy a nyelvészeti tevékenység középpontjában (i) a korpuszépítés (számítógépes tárolásra és feldolgozásra alkalmas szövegek gyűjtése, és ezekből adatbázis építése), (ii) a nyelvléírás (lexikon és/vagy nyelvtan leírása adott nyelv esetében), (iii) a statisztikán alapuló elemzés (az adathalmaznak a számítógépes nyelvészet statisztikai módszereivel történő elemzése), vagy (iv) valamely a hétköznapi életben is elterjedt alkalmazás (lexikográfia, nyelv- és/vagy fordítás oktatás és tanulás, stb.) áll.

4.2.2 Az elektronikusan elérhető korpuszok

Mivel a korpuszépítéssel és a nyelvészeti elemzés céljából történő saját korpusz összeállításával bővebben a 3. pontban foglalkozom, most a nyelvléírás és a fordítástudományi vizsgálatok számára jelenleg is elérhető on-line és off-line korpuszokat tekintem át. Vizsgálhatunk írott vagy beszélt nyelvi korpuszokat, mely utóbbinál mindenképpen szükséges a beszélt nyelvi anyag átírása az élőbeszéd jellemzőinek (szünet, hangerősség, hanglejtés, hangsúly, stb.) rögzítésével. A természetes nyelvfeldolgozás (Natural Language Processing, vagy NLP) utóbbi évtizedekben tapasztalható robbanásszerű fejlődésének köszönhetően a nem elemzett korpuszok mellett egyre több nyelven válnak elérhetővé az automatikus, vagy félautomatikus eszközök segítségével elemzettek is. (i) A szegmentálás, tokenizálás során azonosításra kerülnek a bekezdés-, mondat- és szóhatárok, (ii) a lemmatizáló program az azonos szótókból származó alakokat szűri ki, (iii) a lexikai elemek meghatározásához nélkülözhetetlen a morfológiai elemzés, jelentés egyértelműsítés (word sense disambiguation), és tulajdonnév felismerés, (iv) míg a mondat szerkezeti egységeinek azonosítását, és a közöttük lévő összefüggések feltárását a szintaktikai elemzés (parsing) valósítja meg.

Az írott nyelvi on-line és off-line elérhető korpuszokat többek között a bennük található szövegek nyelve alapján szokás csoportosítani. A világ első elektronikus korpusza az 1964-ben W. N. Francis és H. Kucera által Brown University Standard Sample of Present-Day American English néven publikált *Brown Corpus* volt (<http://khnt.hit.uib.no/icame/manuals/brown/INDEX.HTM>). Ezt számos egynyelvű

korpusz (monolingual corpora) követte, melyek közül a teljesség igénye nélkül a fordítástudomány számára legfontosabbakat említtem.

A legnagyobb méretű, angol nyelvű szövegeket tartalmazó korpuszok közül a *British National Corpus* (www.natcorp.ox.ac.uk) első verziója 1991 és 1994 között készült el. A 100 millió szövegszót tartalmazó gyűjtemény 90%-ban írott, 10%-ban beszélt nyelvi szövegeket tartalmaz. A korpusz alkotói a késő XX. századi modern brit nyelv széles körű reprezentálását tűzték célul maguk elé, ezért a szövegek az általános nyelvi kép megvalósítása érdekében különböző témaköröket, stílusokat, műfajokat és regisztereket képviselnek, illetve a szerzők szerinti kiegyensúlyozottság biztosításának érdekében teljes szövegeket és rövidebb szövegrészleteket is tartalmaznak. A BNC első verzióját 2001-ben követte a *BNC World*, majd 2007-ben a *TEI* szabvány (The Text Encoding Initiative) standardizált annotációját integráló *BNC XML Edition*.

Mark Davies 2008-ban indította el a *Brigham Young University*-n a legnagyobb szabadon hozzáférhető angol nyelvű korpusz, a *Corpus of Contemporary American English* (www.americancorpus.org) építését, melyet azóta is évente egyszer vagy kétszer frissítenek minden évben további 20 millió szövegszóval. Az utolsó frissítés a honlap tanúsága szerint 2011. márciusában történt. A mára már több mint 425 millió szövegszót tartalmazó *Corpus of Contemporary American English-t* (COCA) három nagy részkorpusz alkotja: (i) a British National Corpus, (ii) a 100 millió szövegszót tartalmazó *TIME Corpus* (1920-2000) és (iii) a 400 millió szövegszót tartalmazó *Corpus of Historical American English* (COHA; 1810-2009). A korpuszban egyenlő megoszlásban található beszélt nyelvi, irodalmi és tudományos szövegek, valamint népszerű folyóirat- és újságcikkek teljes listája lekérdezhető mind tartalom, mind évszám vagy műfaj szerinti csoportosításban. A COCA keresője lehetővé teszi a teljes korpuszban vagy az egyes részkorpuszokban szavak, mondatok, lemmák, szófajok, kollokációk és ezek mindenfajta kombinációjának lekérdezését.

A legnagyobb méretű, német nyelvű szövegeket tartalmazó *TIGER Project* (www.ims.uni-stuttgart.de/projekte/TIGER/) célkitűzése volt egy nyelvészeti elemzett német sajtószöveg-gyűjtemény létrehozása. A projekt első szakaszában a 700 000 szövegszót tartalmazó 40 000 mondatot, míg a másodikban további 200 000 szövegszót tartalmazó 10 000 mondatot annotáltak félautomatikus gépi elemzéssel.

A jelenleg publikált TIGER Treebank (Version 2.1) a Frankfurter Rundschau szövegeiből létrehozott (900 000 token, 50 000 mondat), morfológiailag elemzett és lemmatizált német sajtókorpusz.

A számos magyar egynyelvű korpusz közül a *Magyar Nemzeti Szövegtár* (<http://corpus.nytud.hu/mnsz/>) építésére 1998 és 2005 között került sor az MTA Nyelvtudományi Intézetének Korpusznyelvészeti Osztályán. A projekt célja egy 100 millió szavas szövegtörzs létrehozása volt, ami a mai magyar nyelv jellegzetes megnyilvánulásait tartalmazza. A korpusz jelenleg 187,6 millió szövegszót tartalmaz, amelyek az öt kárpátmedencei regionális nyelvváltozat (magyarországi, szlovákiai, kárpátaljai, erdélyi, vajdasági) és az öt stílusréteget (sajtó, szépirodalom, tudományos, hivatalos, személyes) mutatják be. A korpuszban található szövegek elemzése és az elemzések egyértelműsítése automatikus gépi eszközökkel történik, melynek kimenete a szótő, a szófaj és a szó morfológiai elemzése.

A *Szeged Treebank 2.0* (www.inf.u-szeged.hu/projectdirs/hlt/) keletkezése az 1999-es indulással szintén az ezredfordulóhoz köthető. A korpusz legfontosabb ismérve, hogy morfoszintaktikailag elemzett és kézzel egyértelműsített természetes nyelvi szövegeket tartalmazó 1 millió szövegszavas magyar adatbázis. A kézzel egyértelműsített annotálásnak köszönhetően a Szeged Treebank különböző verziói megbízható tanulási és tesztelési referencia-adatbázist jelentenek a magyar természetesnyelv-feldolgozással foglalkozó kutatások számítógépes tanulóalgoritmusai számára (Nyelv- és Beszédtechnológiai Platform 2010: 17).

A többnyelvű korpuszokat (multilingual corpora) funkciójuk és összeállításuk szempontja szerint alapvetően két nagy csoportra oszthatjuk: összehasonlító és párhuzamos korpuszokra. Ilyen többnyelvű korpusz például a *MLCC Multilingual and Parallel Corpora* (www.elda.org/catalogue/en/text/W0023.html). A többnyelvű korpuszok tárgyalásánál meg kell említeni a fordítási korpuszokat is, amiket a szakirodalom speciális funkcionalitásuk és felhasználóik meghatározott köre miatt szokott hol az összehasonlító, hol a párhuzamos korpuszokon belül elkülöníteni (v.ö. Lüdeling, A., Kytö, M. 2008: 272).

Az összehasonlító korpuszok (comparable corpora) kettő vagy több különböző nyelven előállított egynyelvű korpuszt tartalmaznak, azaz csak autentikus (nem fordított) szövegek gyűjteményei. Az összehasonlító korpuszokat alkotó egynyelvű

korpuszok szövegei valamilyen kritérium – mint például a szövegek keletkezési kora, a szövegtípus, a tárgy, a forma ... – alapján válnak összevethetővé. Az összehasonlító korpuszok összeállításakor sokszor meglévő egynyelvű korpuszokat használnak fel. Kutatásuk különösen a ritkán vagy szinte soha le nem fordított szövegfajták esetében – mint például autentikus beszélt nyelvi szövegek, rádió adások, turisztikai szövegek, céges szórólapok – nyújt támpontot az alkalmazott nyelvészeti kutatások számára.

A párhuzamos korpuszok (parallel corpora) ezzel szemben az autentikus szövegek mellett ezek fordítását is tartalmazzák. A napjainkban már több nyelvpárra elkészített párhuzamos korpuszok lehetnek egy- vagy kétirányúak, mondat vagy szó szinten párhuzamosítottak. A *Parallel Corpora in Uppsala* (<http://xml.coverpages.org/etap-over.html>) például csak eredeti svéd szövegekből, és azok fordításaiból áll, míg az *English-Norwegian Parallel Corpus* tartalmaz angol és norvég eredeti szövegeket, és azok fordítását is a másik nyelven (<http://www.helsinki.fi/varieng/CoRD/corpora/ENPC/>). A fordítástudomány számára különösen érdekesek a mondatillesztéssel illetve szó szinten is párhuzamosított korpuszok, mint az *Aligned Hansards of the 36th Parliament of Canada* (<http://www.isi.edu/natural-language/download/hansard/>), és a *European Parliament Proceedings 1996-2001* (<http://www.statmt.org/europarl/>). Az EUR-Lex (<http://eur-lex.europa.eu/hu/tools/about.htm>) a 23 hivatalos nyelven ingyenes hozzáférést biztosít az Európai Unió joganyagához és más dokumentumokhoz. Jelenleg 2 815 000 dokumentumot tartalmaz, amit az új csatlakozó államoknak a csatlakozás előtt le kell fordíttatnia a saját nemzeti nyelvére.

Ismertek olyan párhuzamos korpuszok is, amelyek valamely nyelv autentikus szövegeinek több nyelvre való fordítását is tartalmazzák. Ilyen többnyelvű párhuzamos korpusz például a különböző nyelvekre fordított angol szövegeket tartalmazó *International Corpus of English* (www.ucl.ac.uk/english-usage/ice/). Ezen elv alapján az adott nyelv szövegeinek több különböző nyelvre történő fordítását tartalmazza a *Pedant Corpus* (Danielsson, Ridings 1996), az *Oslo Multilingual Corpus* (Johansson 1997) és a *Triptic Corpus* (Paulussen 1999).

A fordítási korpusz (translational corpus) megnevezést jelenleg a szakirodalom két különböző felfogás alapján használja. 1995-ben Mona Baker a fordítási korpusz nevet csak az adott nyelvre fordított szövegek gyűjteményére értette, míg Johansson

1998-ban ugyanezt az megnevezést gyakorlatilag a párhuzamos korpusz megnevezés helyett, azaz forrásnyelvi szövegeket és azok célnyelvi fordításait tartalmazó szöveggyűjteményekre alkalmazta (v.ö. Lüdeling, A., Kytö, M. 2008: 277). A Baker-i felfogás értelmében, azaz a különböző nyelvekről csak angol nyelvre fordított szövegek gyűjteményeként állították össze a mintegy 10 millió szövegszót tartalmazó *The Translational English Corpus*-t (TEC: <http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>) a Manchester-i Egyetem Fordító és Tolmács Központjában.

A tanulmány korpuszok ismertetésével foglalkozó fejezete a terjedelmi keretek miatt a kutatási témába illő írott nyelvi korpuszokra összpontosít és nem tér ki a beszélt nyelv elsődlegessége következtében kiemelkedően fontos kutatási terepet nyújtó beszélt nyelvi korpuszok (Spoken Language Corpora) részletes bemutatására. A beszélt nyelvi korpuszok esetében mindenképpen meg kell említeni a nyelvtechnológiai szoftverekkel való feldolgozhatóság érdekében valamely beszédfelismerő alkalmazás szükségességét a lejegyzés automatizálásához. A beszélt nyelvi korpuszok közül említést érdemel az *EUR-ACCOR* korpusz (<http://www.cstr.ed.ac.uk/research/projects/artic/accor.html>), ami nyelvenként 5-10 beszélőtől tartalmaz szövegeket.

4.2.3 A fordításkutatás számára fontos korpuszok

A fordítási korpuszok különösen alkalmasak a fordított szövegek sajátosságainak vizsgálatára, egyes szövegpárok, illetve egy-egy lexikai elem vagy kifejezés különböző fordításainak tanulmányozására. Az egynyelvű korpuszokkal történő egybevetésük eredményeként magyarázatot kaphatunk a fordításízűség szövegszinten kimutatható okaira, illetve feltárhatjuk az adott nyelvpárra és fordítási irányra jellemző fordítási mintázatot.

Ha a forrásnyelvi és a célnyelvi szöveg is rendelkezésre áll, akkor szegmenspárok, fordítási egységek hozhatók létre. A szöveg párhuzamosítása történhet manuálisan vagy automatikusan, ahol a szinkronizált szövegpár mindkét tagja mondatra vagy mondatnak tekintett részekre szegmentált és a szegmensek meg vannak egymásnak feleltetve. A szöveg párhuzamosítása során a mondat az a

legkisebb nyelvi elem, amelynek szintjén egy szövegpár még pontosan szinkronizálható, de létezik bekezdésszintű szinkronizálás is. A szegmentálás automatizálásának nehézségei között mindenképpen meg kell említeni, hogy a szegmenshatárok (mondat/pont) definiálása a számítógépes feldolgozás számára egyáltalán nem triviális feladat. Ezért a párhuzamosítás folyamatában fontos támpontot jelentenek a számok, időpontok, mennyiségek, tulajdon nevek (v.ö. Prószycki, Kis 1999: 263–266). A szinkronizálásnak köszönhetően létrehozott, az autentikus szövegeket és ezek fordításait tartalmazó párhuzamos korpuszok speciális használati lehetőségük miatt fontosak a fordítástudomány számára, mivel alkalmasak a fordítási szövegminták tanulmányozására (v.ö. Johansson 1997), és a fordítási univerzálék feltárására (v.ö. Baker 1995). Ezek a korpuszok kulcsfontosságúak továbbá a nem angol korpusznyelvészet (ázsiai nyelvek, kis nyelvek) számára.

McEnery és Xiao (McEnery, Xiao 2007: 18) a párhuzamos és összehasonlító korpuszok előnyeként kiemeli, hogy (i) általuk olyan új szempontokra figyel fel a komparatív nyelvészet, amelyek nem merülnek föl az egynyelvű korpuszok vizsgálatakor; (ii) kibővítik az egyes nyelvek kulturális és nyelvi jellemzőinek különbségeire vonatkozó ismereteinket; és (iii) megvilágítják a célnyelvi és forrásnyelvi szövegek, illetve az autentikus és fordított szövegek közötti különbségeket. E korpuszok alkalmazása a fordítástudományon kívül számos területen hasznos, így a lexikográfiában, nyelvoktatásban is.

Klaudy (2005) arra hívja fel a fordításkutatók figyelmét, hogy az 1990-es évek közepén a szövegkorpuszok felfedezése a tudományág számára nagyban elősegítette, hogy a fordítástudomány, mint diszciplína normatív (előíró) tudományból, deskriptív (leíró) tudománnyá tudjon átalakulni. E szemléletváltásnak köszönhetően a kutatók a fordított szöveget már nem rontott szöveggé, hanem önmagában is érdekes kutatási tárgyként kezdték kezelni. A korpuszalapú fordításkutatás elterjedésének tehát mindenképpen pozitív hozadéka a forrásnyelvi szöveg primátusának megszűnése, és a célnyelvi szöveg kutathatóságának előtérbe kerülése. Klaudy továbbá azt is hangsúlyozza, hogy a párhuzamos korpuszok megmutatják a rutinos fordítók megoldásait, míg a többnyelvű korpuszok

segítségével a célnyelv elemeit természetesebb környezetben lehet vizsgálni, feltárhatók a jellemző lexikai kapcsolódások és szerkezeti minták.

Baker is a fordított szövegek lexikai jellegzetességeinek feltárását emeli ki a korpuszok szerepével kapcsolatban (Baker 1995). A korpuszalapú vizsgálatok segítenek abban, hogy a kutató empirikus adatokkal támassza alá, mely szavak fordulnak elő gyakrabban, mely szavaknak más a jellegzetes környezete a fordított szövegekben, mint az azonos nyelv autentikus szövegeiben. Továbbá vizsgálható az is, hogy az autentikus nyelvhasználat mely szavai nem fordulnak elő a fordított szövegekben, illetve hogy az idegen szavak használatában hogyan tér el a fordított szövegek és az azonos nyelvű autentikus szövegek szóhasználata. Baker munkássága során többször is felvetődik az a kérdés, hogy vannak-e a fordított szövegeknek olyan általános tulajdonságai, amelyek megkülönböztetik azokat az autentikus szövegproduktiótól, azaz beszélhetünk-e fordítási univerzálékról? (Baker 1993, 1996) Felvetésében éppen az intuíció és a manuális elemzések alapján feltételezhető fordítási univerzálék létezését kell a korpuszok segítségével bizonyítani vagy cáfolni.

Károly a korpusznyelvészet és fordításkutatás összekapcsolódásnak előnyét az írott vagy beszélt nyelv szabályszerűségeinek vizsgálatában látja, ami segít a variabilitást befolyásoló, szövegkörnyezeti tényezők és a nyelvhasználat szabályszerűségeinek feltárásában (Károly 2002). Az univerzálékutatás számára a forrásnyelvi és célnyelvi szöveg jellemzőinek összevetése segíthet a nyelvpár specifikus fordítási magatartás feltárásában. Károly a korpuszalapú kutatások előnyei között tartja számon, hogy segítségükkel kimutatható az egyes regiszterekre jellemző grammatika, hogy segítségükkel igazolható vagy kijavítható az anyanyelvi beszélők intuíciója, hogy segítségükkel kimutatható, mely nyelvi jelenséget használnak gyakrabban vagy ritkábban a nyelvtanulók, mint az anyanyelvi beszélők, hogy megmutatják a forrásnyelv és célnyelv asszimétrikus jellegét. Károly ugyanakkor bírálja is a korpuszalapú kutatásokat, mivel azok csupán a szöveg felszíni jelenségeit (gyakoriság, kvantitatív adatok) vizsgálják és az összehasonlítások alapjául az anyanyelvi beszélők performanciája szolgál. A korpuszalapú kutatások hátránya még, hogy figyelmen kívül hagyják az egyéni sajátosságokat, az egyedi előfordulásokat, bár éppen előnyös tulajdonságuk is egyben, hogy a gyakran és jellemzően előforduló esetekkel foglalkoznak, kizárva a statisztikailag nem szignifikáns eseteket.

A fordítástudományban vizsgált számos korpuszalapú kutatás közül a teljesség igénye nélkül felvillantott kutatási kérdések jól bizonyítják, hogy milyen szerteágazóan széleskörű a vizsgálatok spektruma. Kenny például a GEPCOLT korpusz (German-English Parallel Corpus of Literary Texts) segítségével a fordításokra jellemző lexikai kreativitást vizsgálta (Kenny 1998, 2001), míg Holmes és Laviosa a TEC korpuszt (The Translation English Corpus) használta fel a fordított szövegek általános tulajdonságainak (a magyarázó betoldások miatti explicitégségi szint növekedése; a bonyolult mondatok feloldásának köszönhető egyszerűsítés; normalizáció) feltárásához (v.ö. Holmes 1972, Laviosa 2002). Gyakori tárgya a fordítástudományi kutatásoknak annak vizsgálata, hogy bizonyos nyelvi jellegzetességek eloszlása más a fordított szövegekben, mint az autentikus szövegekben. Blum-Kulka ilyen indíttatásból vizsgálta a kohéziós eszközök eloszlását a fordított szövegekben (Blum-Kulka 1986), Johansson pedig az explicitációt az angol és norvég párhuzamos szövegekben (v.ö. Johansson 1995). Hansen-Schirra, Neumann és Steiner német-angol, illetve angol-német relációban vizsgálta a fordítások explicitégségi szintjének növekedése és a kohéziós eszközök eloszlásának változása közötti kapcsolatot (Hansen-Schirra, Neumann, Steiner 2006). Baker és Olohan kutatta a fordító stílusát (v.ö. Baker 2000, Olohan 2004), Mauranen az interkulturális jellemzőket (Mauranen 1996), Doherty német-angol nyelvpár esetében az információs struktúrát (Doherty 1999), míg Hasselgárd a tematikus szerkezetet (Hasselgárd, Oksefjell 1999).

4.2.4 Fordítástámogatás

A korpuszok hasznosítása kapcsán a fordításkutatásban ha csak röviden is, de mindenképpen szót kell ejteni a fordítás gépi támogatásáról – ide értve az elektronikus szótárakat, a terminológiai adatbázisokat és terminológia kezelő rendszereket, a fordítómemóriát és az előfordító rendszerek használatát, illetve magát a gépi fordítást is. A fordítást támogató eszközöket működésük alapján – aszerint, hogy igényelnek-e emberi interakciót – három nagy csoportba szokás besorolni, melyek (i) a teljesen automatikus gépi fordítás (fully automatic machine translation), (ii) az emberi elő- és/vagy utószerkesztéssel segített gépi fordítás

(human aided machine translation) és (iii) az emberi fordítás gépi támogatása (machine aided human translation) (v.ö. Prószéky, Kis 1999).

A teljesen automatikus gépi fordítórendszerek olyan nyelvtechnológiai alkalmazások, amelyek a természetes nyelvek közötti fordítás folyamatának automatizálására törekszenek. A szabályalapú rendszerek (rule based machine translation) beépített szótárakra és nyelvészeti ismeretekre épülnek, a forrásnyelvet több szinten elemzik, és az elemzés kimenetét szabályokkal képezik le a célnyelven. A fordítás történhet közvetlenül a két nyelv között (transzfer) vagy közvetítőnyelven (interlingua) keresztül. Ezzel szemben a korpuszalapú (statisztikai) rendszerek (statistics based machine translation) nagyméretű beépített háttér adatbázisok segítségével mondatonként vagy szavanként feleltetik meg a szövegeket. A fordító rendszer ebből a háttéradatbázisból keresi ki az adott forrásnyelvi mondathoz legjobban hasonlító minta célnyelvi megfelelőjét. Ez a módszer ma igen népszerű, mert nem igényli az adott nyelv ismeretét, bár a fordítás minősége éppen emiatt romlik általában.

Napjaink automatikus gépi fordítórendszerei elsősorban behatárolt témájú szakmai szövegek fordítására alkalmasak, használatuk során felmerülnek stilisztikai és nyelvtani hibák és az egyértelműsítésben elkövetett hibák. E rendszerek hátrányai között szokás említeni, hogy túl következetesen használják a terminológiát, és gyakoriak a fordítási klisék.

Az emberi fordítást támogató gépi eszközök a fordításhoz kapcsolódó konkrét részfeladatokat látják el. Az elektronikus szótárak esetében meg szokás különböztetni a közvetlen felhasználású szótárakat – amelyekben a felhasználó maga keres – a közvetett felhasználású szótáraktól – amelyek nyelvtechnológiai rendszerek belső adatbázisaként szolgálnak és amelyhez ezért a felhasználó nem fér hozzá. A fordítástámogató szoftverek egy csoportját CAT-eszközökként (computer-assisted translation, computer-aided translation tools) szokás említeni. A CAT-eszközök három olyan fontos alrendszert tartalmaznak, mint a fordítómemória (translation memory), a terminológiai adatbázis (terminology management) és a szövegpárhuzamosító eszköz (alignment tool). Komoly szakirodalommal és szakmai érdeklődéssel bír a fordítástudományon belül a fordítómemória és a terminológiai adatbázis kutatása az olyan nyelvterületeken (mint például a német), ahol a vállalati

szférában is komoly erőforrásokat szentelnek a terminológiai konzisztencia megteremtésére és a konzisztens fordítások megvalósítására. A CAT-eszközök alkalmazása a következetes terminushasználat megteremtése mellett elsősorban nagyobb volumenű, hasonló tartalmú, ismétlődő fordítási megbízások esetén segíti a fordítás folyamatban a feldolgozáshoz szükséges idő csökkentését és ezáltal a költségoptimalizálást.

4.2.5 Saját korpusz építése

A fordításkutatók körében egyre népszerűbb az egyetemi kutatóhelyek által létrehozott, és elektronikusan elérhető korpuszok helyett a saját kutatási célokat jobban kiszolgáló, egyénileg épített korpuszok vizsgálata. A korpuszépités maga több egymásra épülő részfeladatra bontható, kezdve a tervezés és szöveggyűjtés fázisával, amit a tárolás követ, majd végül az annotálás, azaz a nyelvészeti elemzés eredményeit tartalmazó többletinformáció feltüntetése.

A korpusztervezés kapcsán két alapvető szempont merül fel, a reprezentativitás igénye és az osztályozás, ami a korpuszok összehasonlíthatóságát biztosítja. Mivel a korpusz tervezőjének elsődleges célja, hogy garantálja kutatási eredményeinek érvényességét és megbízhatóságát (v.ö. Kennedy 1998), az eredmények általánosíthatósága érdekében ügyelnie kell a korpusz reprezentativitására. A reprezentatív korpusz a mindennapi nyelvhasználatnak megfelelő arányban tartalmazza a szövegtípust, a műfajt, a szövegfunkciót, a szövegalkotó neme, a térbeli és időbeli lefedettség, stb. szempontjából különböző jellegű szövegeket, amiért a szakirodalom kiegyensúlyozott (well-balanced) korpusznak is hívja. A korpusz reprezentativitása összefügg a korpusz méretével és a mintavétel meghatározásával is.

A korpusz méretét a szövegszavak számával szokás jelölni. Bár a számítástechnika tárolási kapacitásának ugrásszerű növekedése következtében egymás után születnek a több millió szavas korpuszok, a kutatási céltól és a lekérdezéshez rendelkezésre álló eszközöktől függően sok esetben előnyösebb a hipotézisek vizsgálatát kisebb korpuszokon megkezdeni.

4.2.6 A korpusztervezés fő szempontjai

A korpusz méretének meghatározása mellett a tervezésénél dönteni kell arról, hogy a korpusz a teljes- vagy a részszövegek gyűjteményét tartalmazza-e, illetve hogy statikus- vagy dinamikus korpuszt (monitor korpusz) kívánunk-e létrehozni (v.ö. Kennedy 1998). Míg a teljes szövegekből felépített korpusz esetében nehezebb megszerezni a szerzői jogi hozzájárulást, addig a részszövegek gyűjtése azonos korpuszméret mellett több szerző szövegrészletének tárolását teszi lehetővé. A szerzői jogi hozzájárulás szükséges mind az elektronikus formában történő tároláshoz, mind a szövegek későbbi kutatásban való felhasználásához. A szövegek letöltését, ma már megkönnyíti számos elektronikusan elérhető adatbázis, mint a *Digitális Irodalmi Akadémia*, *The database of translations of Hungarian literary works* (<http://translations.bookfinder.hu/indexn.htm>), *Magyar Elektronikus Könyvtár*, *The Online Books Page* (University of Pennsylvania), *Electronic Text Collections in Western European Literature*, *La biblioteca telematica*, *Literatur im Netz*, stb.

4.2.7 Szövegválasztás

A statikus korpusz változatlan formában tartalmazza a meghatározott szempontok alapján válogatott szövegeket, míg a dinamikus korpusz a belső arányok megtartása mellett folyamatosan bővül, ezért alkalmas a nyelv változásának nyomon követésére. A szövegválasztásnál több szempontot kell a korpusz összeállítójának figyelembe vennie aszerint, hogy kutatása során milyen nyelvű, autentikus, vagy fordított szövegeket kíván vizsgálni, és a vizsgált szövegekkel milyen időperiódust szeretne felölelni.

A kiválasztott szövegeknek illeszkedniük kell a vizsgálat tárgyához, ami irányulhat a forrásnyelvi vagy a célnyelvi szöveg vizsgálatára, illetve lehet összehasonlító elemzés. A vizsgálandó szövegek gyűjtését befolyásolja továbbá a kutatási eredmények tervezett felhasználása, ami történhet a leíró fordítástudományban, közvetlenül az oktatásban, vagy valamely alkalmazott nyelvészeti területen, mint például a szótárkészítésben. A szövegválasztást befolyásolják még a szöveg csak elemzéssel meghatározható belső tulajdonságai, és

a szövegalkotás jellemzőiként számon tartott külső tulajdonságok. A szövegalkotás külső tulajdonságai közé szokás sorolni a nyelvi összetételt; a szöveg szinkrón vagy diakrón idő periódusba való tartozását; az olyan tematikus szempontokat, mint a szerző, a terület, a közzététel helye és a befogadó; a szövegtípust és a szövegek beszélt vagy írott nyelvi voltát.

4.2.8 A korpuszok annotálása

A számítógépek teljesítményének növekedése következtében egyre nagyobb méretű szövegbázisok, korpuszok létrejöttével párhuzamosan fejlődött a szöveg nyelvi tartalom szerinti feldolgozása és intelligens kezelése, illetve az automatikus gépi tanuló eljárások kifejlesztése. Az ember és a számítógép közötti természetes nyelven történő kommunikációt megvalósítani kívánó és az ehhez szükséges számítógépes programok megalkotásával foglalkozó nyelvtechnológiának köszönhetőek a nyelvi elemző rendszerek. Minden nyelvi elemző rendszer esetében két fő cél érvényesül, ami a lefedettség (recall – minél több nyelvi elem azonosítása illetve kódolása) és a pontosság maximalizálása (precision – az azonosítás vagy kódolás hibamentessége). A nyelvtechnológia célja olyan nyelvi elemzők létrehozása, amelyek minél több találat azonosítására képesek minél kevesebb tévedés mellett. Itt kell említést tenni arról a nehezen feloldható ellentmondásról, hogy a találatok számának növeléséhez megengedőbb rendszerre van szükség, míg a tévedések csökkentését szigorításokkal lehet elérni. A nyelvtechnológia elemző legoptimálisabb találati listájának beállítása tehát e két ellentétes művelet aprólékos hangolásának eredményéből fakad.

A nyelvtechnológia feladatok esetében is a gépi fordításnál már említett két irányzatot különböztethetjük meg, (i) a produktív szabályokat alkalmazó megközelítést és (ii) a megtanult minta alapú statisztikai módszert, mely utóbbi a példatárakból, korpuszokból automatikusan kivont mintákat alkalmazza más nyelvi megnyilvánulások feldolgozására. A program a mintákhoz valószínűségeket rendel, majd az együttes valószínűség maximalizálásával illeszti a mintákat. A mai modern nyelvtechnológiai alkalmazásokban legtöbbször keverednek a szabályalapú és a statisztikai eszközök, míg az alsóbb szinteken (beszédfelismerés, szófaji elemzés) a

statisztikai megoldások dominálnak, addig a felsőbb szinteken (parsing, szemantika) a szabályalapú módszerek a jellemzőbbek.

Mivel a számítógépes elemző rendszerek számára a természetes nyelvi megnyilatkozások csak karakterek sorozataiként jelennek meg, ezért szükség van a homogén karaktorsorozatok vonatkozásában a gép számára néhány alapvető támpont definiálására. Ilyen alapvető szövegtulajdonság megadása a szöveg kisebb egységekre – bekezdésekre, mondatokra, szavakra – történő tagolása, a szegmentálás. Az olyan ragozó típusú nyelvek feldolgozása esetében, mint amilyen például a magyar nyelv elengedhetetlen a morfológiai elemzés. A morfológiai elemző felismeri a szavak egyes alakjait, a ragozott formákat, a képzett szavakat és az összetett szavakat. A ragozó nyelvek esetében nem lehet a nyelv szókincsének gépi reprezentációját a nyelv összes szóalakjának szótárban való elhelyezésével megvalósítani, hanem a morfémáknak és azok jellemző tulajdonságainak felsorolását a szótárban a szóalakok felépítéséhez szükséges szabályok megadásával kell kiegészíteni. A toldalékolt szavak gépi reprezentációja így a minimális egységek (tő és toldalékok) által hordozott információk összességéből tevődik össze. A magyar morfológiai elemző, a HUMor (High-speed Unification MORpology) 1991-ben készült el és több mint kétmilliárd helyes magyar szóalak elemzésére képes.

A morfológiailag elemzett szövegek esetében a homonim szóalakok (pl.: *vár*, *sír*) közül a szövegek további feldolgozása megkívánja a szófaji vagy morfoszintaktikai egyértelműsítést, ami során az adott szóalak adott kontextusba illő elemzését választja ki a rendszer relatív gyakoriságon és szóhármás vizsgálatokon alapuló lekérdezések segítségével. A lemmatizálás során az azonos szótőből származó szóalakokat tudja az elemző rendszer az adott nyelv szavait tartalmazó szótár szavainak megfeleltetni. A magasabb szintű gépi műveletek (pl. fordítástámogatás) feltételezik már a statisztikai és grammatikai modellek alapján végrehajtott szintaktikai elemzést (parsing) is.

4.2.9 Egyszerű korpuszalapú elemzések

A korpuszok elemzéséhez használt számítógépes programokra az információ-lekérdezés pontossága, gyorsasága, és jó minősége jellemző (Szirmai 2005). A ma

rendelkezésre álló eszközökkel házilagosan is könnyen lekérdezhető statisztikák, kulcsszó keresések, konkordanciák segítenek a korpusz lexikális jellemzőinek feltérképezésében, ami további kutatási hipotézisekre inspirálhat.

A szószám (token) a szóközzel határolt szavakat mutatja, azaz az összes előforduló szövegszót, függetlenül attól, hogy ugyanaz a szó hányszor szerepel az egész korpuszban vagy az egyes szövegegységekben. A szóalak/típus (type) a duplikátumok kiszűrése után a korpuszban szereplő különböző szótári szavak számáról tájékoztat. A type/token arány a korpuszra jellemző lexikai változatosságról ad számot. Mivel a szöveghosszal nő az ismétlések száma, az összehasonlíthatóság érdekében standardizálták a type/token arány számítását, és a szöveg ezer tokenes egységeiben mért eredmények átlagát vetítik a teljes szövegre. Készíthetünk továbbá kis-és nagybetűs betűgyakorisági listát, alfabetikus elrendezésű vagy gyakoriság szerinti szólistát, vagy vizsgálhatjuk az átlagos szó- és mondathosszt. Ez utóbbi a mondathatár felismerés miatt nehezen automatizálható, és így utólagos kézi javítást igényel.

A statisztikák készítésénél tapasztalható gondok közül meg kell még említeni az egymástól független jelentésű, de azonos hangalakú homonim szavakat (pl.: *Az ország egyik leglátogatottabb műemléke az egri vár. / Peti Katira vár.*), a ragozott alakok szótári töre való visszavezetésének, lemmatizálásának problematikáját (pl.: *alma, almát, almás*), a több szavas kifejezések egy szótári egységként való kezelésének nehézségét (pl.: *házi feladat*), az azonos szó különböző írásmódjainak egységes kezelését (pl.: *szőlő – szöllő*), és a kategória különbségek felismerését, mint például a kis- és nagy betűs írásmód hatására módosuló jelentést (pl. *attila* = alkalmi öltöny, *Attila* = ffi név).

A kulcsszó kereső program a vizsgált szöveg szólistáját hasonlítja egy hosszabb szöveg szólistájához, és a rövidebb listában gyakrabban előforduló szavakat jelöli meg kulcsszavakként. A klaszter vagy N-gram a szövegben szereplő több egységből álló szerkezetek vizsgálatát segíti elő. Az n helyére kerülő szám határozza meg, hogy a szövegben szereplő szavakat hányas egységekre bontjuk. Például a „*Pisti minden délben eszik levest.*” mondatot 3-gram esetében a következő egységekre bonthatjuk: *Pisti minden délben, minden délben eszik, és délben eszik levest.* A konkordancia (concordance) az „adott szó vagy kifejezés szövegben szereplő összes előfordulását

szövegkörnyezetében bemutató lista” (Szirmai 2005). Az eredményeket a csomóponttól jobbra vagy balra levő első/második.../ötödik szó alapján listáztathatjuk ki a programmal. A konkordanciára Sinclair, mint a korpusznyelvészet egyik legfontosabb vizsgálati eszközére tekint, mivel a szöveg mintázataira hívja fel a kutatók figyelmét (v.ö. Sinclair 1991).

^I A korpusznyelvészetet manapság gyakran tekintik a nyelvészet egy viszonylag új ágának, amely a „hétköznapi élet” nyelvének empirikus tanulmányozásával foglalkozik számítógépek és elektronikus korpuszok segítségével. Első megközelítésben a „korpusz” leírt, vagy beszélt nyelvi szövegek egyszerű gyűjteménye. Abban az esetben viszont, ha ezt a kifejezést a modern nyelvészettel összefüggésben használjuk, akkor számos további ismérvvvel is rendelkezik, mint az elektronikus olvashatóság, a mintavétel és reprezentativitás, a végleges méret, valamint az a szándék, hogy a korpusz az általa képviselt nyelvi választék hiteles referenciája. (saját fordítás)

^{II} [a korpusz] adott nyelv természetesen előforduló szövegeinek gyűjteménye, amelyek egy nyelv adott állapotára, vagy változataira jellemzőek. (saját fordítás)

^{III} olyan nyelvi anyagok gyűjteménye, amiket meghatározott nyelvészeti kritériumok alapján választottak ki és rendeztek össze, azzal a céllal, hogy az adott nyelv mintái legyenek. (saját fordítás)

III. A kutatás módszerei

1 A vizsgálat tárgya

Az elméleti fejezetekben vázolt szakirodalmi megközelítések alapján a fordítások minőségének lényegi eleme a kohézió megléte és a célnyelv rendszerének – mind fajtáját mind mennyiségét tekintve – megfelelő kohéziós eszköztár alkalmazása. A most bemutatásra kerülő kutatás a fordítás során a kohéziós elemek használatában bekövetkezett változások feltárására fókuszál. A fordítástudomány és kohéziókutatás területeinek összekapcsolásával az általam kínált új megközelítés meghatározó jellemzője az összehasonlító elemzés. Kutatásom során az autentikus magyar szövegekre és a fordított, azaz célnyelvi magyar szövegekre jellemző kohéziós eszközhasználatot vizsgálom, ezek egybevetésével a fordítási tevékenység és a forrásnyelvi szöveg hatására bekövetkező eltolódásokat kívánom a magyar szövegek lexikai kohéziós elemeinek disztribúciójában feltárni.

Az ilyen fajta összehasonlító elemzésre természetesen nem véletlenül nem került sor a szövegkutatás korábbi periódusaiban, hiszen a hagyományos eszközzel végzett lexikai kohéziókutatás keretein belül egy ilyen típusú összehasonlítás igen időigényes lett volna, mivel a hagyományos eszközzel végzett elemzés egy-egy szó kapcsolatait a szóból kiinduló, s az adott lexikai egység ismétléses kapcsolataihoz érkező vonalak segítségével jelöli (v.ö. Károly 2002: 80). Egyszerűen belátható, hogy a jelen vizsgálat alapjául szolgáló nagyobb korpuszhoz hasonló szövegkorpuszok esetén az ilyen típusú vizsgálat milyen ember feletti feladatot jelentene.

Az összehasonlító elemzés eredményeinek általánosíthatósága érdekében mind az autentikus magyar szövegprodukción, mind a célnyelvi magyar szövegprodukción reprezentáló korpusz összeállításánál törekedtem különböző szövegtípusú szövegek vizsgálatára. A célnyelvi magyar szövegeket tartalmazó korpusz összeállításánál további fontos szempontnak tekintettem, hogy a fordított magyar szövegek forrásnyelvi eredetijei különböző forrásnyelveket reprezentáljanak.

2 A kutatás főbb szakaszai

2.1 Korpuszépítés

Az elemzett szövegek a szövegprodukciónban részt vevők száma és egymáshoz való viszonya alapján mind monológikus szövegek, kommunikációs funkciójuk szerint pedig az elbeszélő, a leíró és az érvelő szövegtípusokba sorolhatóak. A narratív vagy elbeszélő szövegek fordulatokban gazdag, élményszerű szemléletességgel, gyakran leíró elemek beleépítésével tájékoztatnak eseményekről, történésekről és cselekvésekről. A szövegek felépítése lineáris időbeliségre vagy a téma térbeliségre épülő (rész-egész/közel-távol/koncentrikus) kifejtésére épül és az elbeszélés során az elbeszélő nézőpontja meghatározó szerepet kap. A deskriptív vagy leíró szövegek ezzel szemben tárgyilagosan ismertetik és mutatják be a dolgokat, tényállásokat és körülményeket, a leíró személyét és érzéseit háttérbe szorítva adnak jellemzést vagy leírást. Az érvelő szövegek központi témája a bizonyos jelenségekkel, személyekkel, körülményekkel kapcsolatos állásfoglalás, s ennek indoklására, igazolására és a meggyőzésre törekednek. Az argumentatív vagy érvelő szövegek az állítások közötti logikai rendre – az ok-okozati kapcsolatra, a szövegben bemutatott állítások bizonyítására vagy cáfolatára épülnek (v.ö. H. Varga 2007: 27–34).

A korpusz szövegei a szövegtípusok kommunikációs színterei szerint közéleti (hivatalos), egyházi, tudományos és szépirodalmi besorolást kaphatnak. A vizsgált szövegek a hordozó közeg vagy csatorna szerint mind írásbeli megnyilatkozások, melyekre átgondolt, tudatos felépítés, igényes, választékos mondatfűzés jellemző. A szövegek igényesen megszerkesztett volta teszi lehetővé a kohéziós elemek vizsgálatát, hiszen az írásbeli kommunikáció során 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.

A szövegeket részben nyilvános internetes oldalakról, adatbázisokból töltöttem le, részben barátok, ismerősök bocsátották fordításaikat nyelvészeti elemzés céljából a rendelkezésemre. A mindenki számára legálisan elérhető internetes honlapokról letöltött szövegek esetében is megtakarítottam a szerzői jogi kérdések tisztázását és

a begépelés fáradtságát. Az elemzés számára a szövegeket Microsoft Word formátumban archiváltam. Az archivált és az elemzés különböző szakaszaiban keletkezett szövegek elrendezésére a számítógépes dokumentumok tárolásánál bevált folderes struktúrát választottam, ahol az egyes foldereken belül a kiválasztott paraméterek alapján további egymással azonos szinten elhelyezkedő, illetve egymásba ágyazott alfolderek definiálására is lehetőség nyílik. Ez a folderstruktúra az emberi keresés számára is könnyen átláthatóvá teszi a szöveges dokumentumok nagyszámú halmazát.

2.2 Annotálás

A nyelvészeti elemzés számára szükséges jelenségek megjelölése a szövegben metanyelvi többletinformáció segítségével történik. Erre a célra kiválóan alkalmasak az úgynevezett jelölő (Mark-up) nyelvek, mint a HTML, SGML, XML (Hyper Text Markup Language, Standard Generalized Markup Language, Extensible Markup Language). A humán tudományok esetében általában a számítógépes feldolgozásra szánt digitális dokumentumokat SGML- vagy XML-alapú, platformfüggetlen kódolással látják el. A jelölőelemek (tagek) szabadon bővíthetők, de használatuknak szigorú szintaxisa van, amit a TEI (The Text Encoding Initiative) ajánlásai szabályoznak.

2.2.1 Kézi annotálás

Az elemzett és alkorpuszokba rendezett szövegeket kézi bevitel során láttam el header annotációval, ami az adott elektronikus fájl bibliográfiai leírását, azaz a szövegfájlban expliciten nem feltétlenül megjelenő tulajdonságokat tartalmazza. A headerben feltüntettem a fájlformátumot, ami a teljes kutatási anyag esetében három féle lehet: az eredetileg archivált Word formátum vagy az OpenOffice program segítségével kialakított bemeneti XML formátum, illetve az elemzés végeredményeként kapott, a lexikai kohéziós kapcsolatokat tartalmazó kimeneti XML formátum (WORD, source_XML, finite_XML). Mivel egyelőre csak az elemzett

alkorpuszt láttam el header annotációval, így a fájlformátum mező értéke minden esetben *finite_XML*.

A header tartalmazza a szövegfájl címét, szerzőjét, a publikálásra vonatkozó adatokat, valamint a szöveg elérhetőségét tartalmazó információt (4. ábra). A szöveg elérhetősége esetében az interneten nyilvánosan elérhető szövegeknél az elérés útvonalát tartalmazó webes linket adtam meg, míg az ismerősöktől kapott szövegek esetében a közvetítő személy monogramját tüntettem fel. A header tartalmazza továbbá a szövegek kommunikációs színtérének típusára és a forrásnyelvére vonatkozó információt. A szövegeket a kommunikációs színtér típusa szerint négy kategóriába soroltam be: irodalmi, tudományos, vallásos és közéleti (LIT, SCIENCE, RELIG, EU). A forrásnyelv szempontjából megkülönböztettem autentikus és célnyelvi magyar szövegeket (ORIG, TRAD). A fordítás eredményeként létrejött magyar szövegek esetében jelöltem a forrásnyelvet, ami a korpuszban jelenleg archivált szövegek esetében lehet angol, francia, német, olasz, latin és görög (ENG, FR, GER, IT, LAT, GREEK).

```
<Header>
  <formatDesc>
    <format><!-- ...      --></format>
    <!ATTLIST format   (WORD | source_XML | finite_XML)>
  </formatDesc>
  <fileDesc>
    <title><!-- ...      --></title>
    <author><!-- ...     --></author>
    <publication><!-- ... --></publication>
    <source><!-- ...     --></source>
  </fileDesc>
  <textDesc>
    <genre><!-- ...      --></genre>
    <!ATTLIST genre    (LIT | SCIENCE | RELIG | EU)>
    <language><!-- ...   --></language>
    <!ATTLIST language (ORIG | TRAD)>
    <TRAD type="(ENG | FR | GER | IT | LAT | GREEK)">
  </textDesc>
</Header>
```

4. ÁBRA: Az LC-HU korpusz header annotációja

2.2.2 Gépi annotálás

Az elemzésre kiválasztott szövegek feldolgozását, azaz a kohéziós eszközök bejelölését egy erre a feladatra készített korpusznyelvészeti program segítségével végeztem el. A nyelvészeti elemzést végző programok jellegzetessége, hogy lehetővé

teszik a számítógépes feldolgozásra szánt szövegek strukturált kódolását. Ez annyit jelent, hogy míg az eredeti szöveg integritása megmarad, addig annotációként többletinformációt, például a nyelvészeti elemzés eredményeit vihetjük be a szövegbe (Prószéky, Kis 1999). Az annotáció tehát a dokumentumok tartalmi elemeinek megjelölésére szolgál a szövegbe ágyazott metaadatok segítségével. Ilyen szövegbe ágyazott többletinformáció például a jelen elemzés esetében a bekezdések, a mondathatárok, a főnévi csoportok és az ismétlődő lexikai elemek bejelölése, illetve a kohézió típusának megjelenítése. Az eligazodásban és a szemléletességben segít, ha az annotált információkat a képernyőn különböző színekkel jelenítjük meg (5. ábra).

```

5073 <sentence><origin>Az ismeretek közvetítése külföldiek számára sokféle formában és szinten történhet,
· országon belüli és külföldi helyszíneken egyaránt. </origin><token><stem="az">tag="[DET]">Az </
· token><token><stem="ismeret">tag="[FN][PL][NOM]">ismeretek</token><token><stem="közvetítés">tag="[FN]
· [Pse3][NOM]">közvetítése</token><token><stem="külföldi">tag="[FN][PL][NOM]">külföldiek</token><token>
· stem="számára">tag="[NU][Pse3]">számára</token><token><stem="sokféle">tag="[MN][NOM]">sokféle</
· token><token><stem="forma">tag="[FN][NE]">formában</token><token><stem="és">tag="[KOT]">és</
· token><token><stem="szint">tag="[FN][SU.P]">szinten</token><token><stem="történik">tag="[IGE][HAT]
· [e3]">történhet</token><token><stem=",">tag="[PUNCT]">,</token><token><stem="ország">tag="[FN]
· [SU.P]">országon</token><token><stem="belüli">tag="[MN][NOM]">belüli</token><token><stem="és">
· tag="[KOT]">és</token><token><stem="külföldi">tag="[MN][NOM]">külföldi</token><token><stem="helyszín">
· tag="[FN][PL][SU.P]">helyszíneken</token><token><stem="egyaránt">tag="[HA]">egyaránt</token><token>
· stem=".">tag="[PUNCT]">.</token></sentence>¶

```

5. ÁBRA: Az LC-HU korpusz annotált szövegrészlete

2.3 A lexikai kohézió vizsgálata korpusznyelvészeti módszerrel

2.3.1 A kísérleti kutatás

A kutatás előzményeként bemutatott pilot projekt (ld. I. 4.1 fejezet) lexikai kohéziós vizsgálatát kisméretű, kézi címkézéssel annotált korpuszon végeztem. A vizsgálat előrehaladtával feltérképeztem a számomra is elérhető nyelvtechnológiai eszközök körét és az elemzésre alkalmas dokumentumok elérhetőségének paramétereit. A számítógépes eszköz segítségével végzett lekérdezések eredményei ismeretében a pilot projekt mintegy konklúziójaként fogalmaztam meg a fordításnyelvre jellemző és az autentikus szövegalkotási szokásoktól eltérő kohéziós mintázatra vonatkozó hipotéziseimet.

Az Magyar Lexikai Kohéziós Projekt (Lexical Cohesion in Hungarian) során a kohéziós elemek identifikációja nyelvtechnológiai eszközök összekapcsolásával

történt, ami a hipotézisalkotás szempontjából meghatározó pilot projekt szövegállományánál összehasonlíthatatlan mértékben nagyobb szövegtípus elemzését tette lehetővé. Mivel az LC-Hu korpusz mérete nem teszi lehetővé, hogy a lexikai kohéziós kapcsolatok bejelölését minden szövegfájl esetében ellenőrizsem, az elemzési lánc eredményeinek validitását a korpuszban található szövegek egy szűkebb szegmensének adatai alapján kívántam tesztelni.

A tesztadatok meghatározásánál abból a már elfogadott téziséből indultam ki (ld. II. 3.2.2. fejezet), mely szerint az azonos nyelven, a nyelv fejlődésének adott korszakában egymáshoz közeli időintervallumban keletkezett különböző szövegtípusokba sorolható szövegdokumentumok esetében a szövegtípusra jellemző szövegépítő sajátosságok között a lexikai kohézió esetében is megállapítható a szövegtípusok szerinti specializáltság (v.ö. Károly 1979: 55). Az elemzések kimeneteként kapott statisztikai eredményeket ezért először az autentikus magyar szövegek kisebb korpuszán ellenőriztem. Ennek a vizsgálatnak eredményeként a különböző szövegtípusba tartozó autentikus magyar szövegek egymással összehasonlítható nagyságú szövegtípus szerinti alkorpuszai esetében a várakozásoknak megfelelő, az egyes szövegtípusok szövegépítő tulajdonságainak különbözőségét tükröző kohéziós mintázatbeli eltéréseket mutató eredményeket kaptam.

A tesztadatok kinyerésére összeállított kísérleti korpusz tartalmazza az LC-HU korpusz összes szövegtípus kategóriájával azonos alkorpuszokat, tehát tartalmaz közéleti (EU), szépirodalmi (IROD), tudományos (TUD) és vallásos (VAL) szövegeket, alkorpuszonként 71.000 és 76.000 tokenszám közötti méretben. A kohéziós kapcsolatok bejelölése itt is a későbbiekben bemutatott magyar WordNet (Prószéky, Mihaltz 2008, Mihaltz 2005) élek mentén történ (ld. III. 4.3.), megkülönböztetve a következő kategóriákat: ismétlés (R), szinonima (S), az ellentét két fajtája (A), hiponima (H), meronima (M) és a be in state és a domén relációk (category domain, region domain). Mivel a relációk közül a be in state, és a region domain kategóriák számszerű előfordulása elenyésző (még a teljes korpusz esetében sem érik el vagy csak éppen a tokenszámmal viszonyított 0,01%-os előfordulási gyakoriságot), ezért az előfordulásukra vonatkozó adatok közlésétől ebben a fejezetben teljes mértékben eltekintettem.

A statisztikai eredményeket összegző táblázat soraiban az egyes szövegtípusok esetében az oszlopokban feltüntetett kategóriák esetében a tokenszámban kifejezett eredményeket közöltem, illetve megadtam az alkorpusz össztokenszámahoz viszonyítva az egyes kategóriákban mért elemszámok előfordulásának arányát.

TESZT AUTENTIKUS		TOKENEK	R	S	A	H	M	C_D	ÖSSZES
EU	ÖSSZES	71 427	5 235	455	80	1 751	143	43	7 708
	%		7,33%	0,64%	0,11%	2,45%	0,20%	0,06%	10,79%
IROD	ÖSSZES	71 244	6 342	348	159	1 141	97	89	8 182
	%		8,90%	0,49%	0,22%	1,60%	0,14%	0,12%	11,48%
TUD	ÖSSZES	75 930	12 613	490	126	1 886	144	80	15 380
	%		16,61%	0,65%	0,17%	2,48%	0,19%	0,11%	20,26%
VAL	ÖSSZES	71 038	12 894	563	479	1 580	648	127	16 295
	%		18,15%	0,79%	0,67%	2,22%	0,91%	0,18%	22,94%

1. TÁBLÁZAT : Az autentikus magyar szövegeken mért tesztadatok

Az 1. TÁBLÁZAT utolsó oszlopában (ÖSSZES) a lexikai relációk előfordulási eseteit összegezve az alkorpuszra jellemző lexikai relációk összmennyiségét és azok előfordulásának az össztokenszámmal viszonyított arányát emeltem ki. Mivel az egyes szövegtípusok esetében a várakozásoknak megfelelően jól érzékelhetőek a lexikai elemek eloszlásbeli gyakoriságai közötti különbségek, ezért következtetésként elmondhatjuk, hogy az autentikus magyar szövegek tesztkorpuszának eredményei megnyugtatóan igazolják az automatikus gépi eszközök összekapcsolásával kivitelezett kutatás mérési eredményeit.

2.3.2 A Magyar Lexikai Kohéziós Projekt (Lexical Cohesion in Hungarian)

A következő fejezetekben bemutatott Magyar Lexikai Kohéziós Projekt (Lexical Cohesion in Hungarian) előkészületeinek és elméleti kereteinek ismertetése már részben rávilágított arra a tényre, hogy a projekt előkészítése és a kutatás pontos megtervezése több éves előkészítő munkát igényelt. A tervezés során kikristályosodott, hogy a szövegszerveződés pontosan mely tulajdonságainak mérésére nyílik reális lehetőség automatikus gépi eszközök segítségével. A rendelkezésre álló eszközök működési paraméterei pedig meghatározták a

vizsgálható szövegek nyelvi tulajdonságát. A Magyar Lexikai Kohéziós Projekt során használt automatikus gépi eszközök kizárólag magyar szövegek elemzését tették lehetővé. Mivel a kohéziós kapcsolatok zöme a névszói csoportok relációjában valósul meg, a kutatás kereteit az elemzési idő és az elemzési eredmény optimalizálása érdekében a névszói csoportok kohéziós kapcsolatainak identifikációjára korlátoztam.

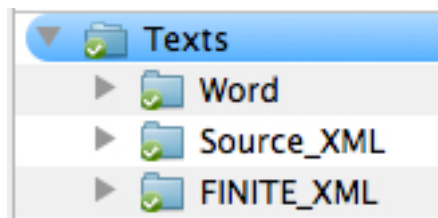
Bár kutatásom alapcélkitűzése az autentikus és a fordítás útján keletkezett magyar szövegek szövegépítő sajátosságainak összehasonlítása révén a fordítás következtében a célnyelvi magyar szövegek kohéziós elemeinek disztribúciójában bekövetkezett változások nyelvpár specifikus jellemzőinek feltárására irányult, mégsem elégedtem meg a kézi annotáció számára elérhető kisméretű szövegtörzsből kinyert elemzési eredmények realizálásával. Az automatikus eszközökkel segített elemzés megvalósítására több tényező is ösztönzött, melyek közül legfontosabbnak emelném ki azt a feltételezést, miszerint ha az automatikus eszközökkel segített kutatás eredményei alapján kimutatható az egymástól keletkezésük körülményeiben különböző szövegállományok alapvető szövegsemantikai szerveződésének különbözősége, akkor a kutatáson túlmutatva a számítógépes nyelvészet számára mindenképpen megfontolandó szempontrendszerként kínál a szövegek ilyen jellegű elemzése és vizsgálata. Tekintettel arra, hogy a gépi fordítás számára például az egyik legkomolyabb kihívást az egybefüggő gondolatmenetek minden elemének a szövegtörzsetbe adekvát módon beilleszthető fordításának megtalálása jelenti, meggyőződésem szerint a szöveg bekezdéseiben belüli lexikai kohéziós kapcsolatok automatikus felismerése hozzájárulhat ennek a problémának a kezeléséhez.

3 A korpusz bemutatása

3.1 A korpusz szerkezete

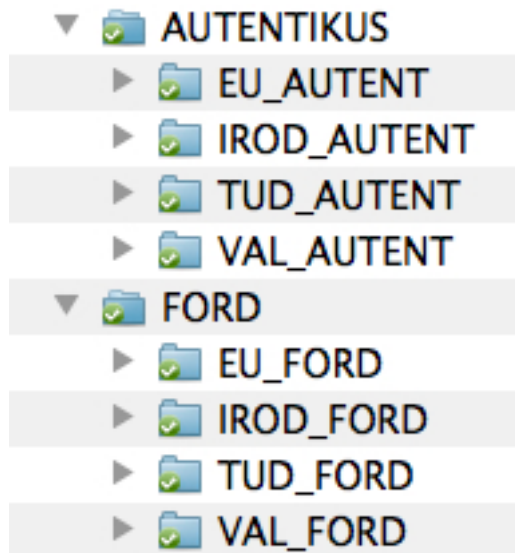
Az elektronikus formában tárolt szövegfájlok archiválására és elrendezésére a számítógépes dokumentumok rendszerezésénél ismert folderes struktúrát választottam, ami az egy szinten egymással több párhuzamos folder kialakításával és

a folderek egymásba ágyazásával az emberi keresés számára is könnyen átlátható elrendezést tesz lehetővé. A legfelső szinten a korpuszban tárolt szövegeket a dokumentum formátuma szerint osztottam három különböző folderbe. Így elkülönítettem az eredetileg archivált Word dokumentumokat és az OpenOffice program segítségével átalakított bemeneti XML fájlokat, illetve az elemzés végeredményeként kapott, a lexikai kohéziós kapcsolatokat tartalmazó kimeneti XML dokumentumokat (WORD, Source_XML, FINITE_XML). A szövegfájlok fájlformátum szerinti elkülönítése (6. ábra) alapvetően csak az emberi keresés és munkavégzés szempontjából érdekes, mivel az egyes fájl típusok expliciten kifejezett tulajdonságok alapján is könnyen megkülönböztethetőek egymástól.



6. ÁBRA: Az LC-HU korpusz legfelső folderszintje

Mindhárom fájlformátum szerinti alfolderben a szövegfájlokat egységesen tovább diverzifikáltam autentikus magyar és célnyelvi magyar mivoltuk alapján (AUTENTIKUS, FORDÍTOTT). A folderstruktúra következő szintjén – továbbra is teljesen egységesen az ezen a szinten található összes folderben – alakítottam ki a kommunikációs színtér típusa szerinti négy kategóriának megfelelően, a közéleti, az irodalmi, a tudományos és a vallásos (EU, IROD, TUD, VAL) alfordereket (7. ábra).



7. ÁBRA: Az LC-HU korpusz folderstuktúrája

Az autentikus szövegprodukciónál esetében a kommunikációs szintér típusa szerinti folderek tartalmazzák már az adott kategóriába tartozó szövegfájlokat. A fordítás eredményeként keletkezett szövegeket tartalmazó szubfolderek esetében viszont a kommunikációs szintér típusa szerinti négy kategórián belül az adott szövegtípus esetében a fordított szövegek eredetijeinek forrásnyelvei alapján még kialakítottam további alkategóriákat. Ezen a szinten már nem törekedtem az egységes struktúra továbbvitelére, hanem az alkategóriákat kizárólag az határozta meg, hogy az egyes szövegtípusok esetében milyen nyelvről fordított szövegeket tudtam a nyelvészeti elemzés számára a korpuszban archiválni. Jelenleg a korpusz egészét tekintve, de nem egységesen minden szövegtípus esetében a következő forrásnyelvekről fordított szövegek állnak rendelkezésemre: angol, francia, német, olasz, latin, görög. A szövegtípusokon belül a forrásnyelveket jelölő folderek tartalmazzák az egyes szövegfájlokat.

3.2 A korpusz összetétele

A korpusz építésének legfontosabb kritériuma az autentikus és célnyelvi magyar szövegprodukciónál reprezentálása volt. A szövegek gyűjtésekor elsődleges szempontként szerepelt, hogy a szövegek elektronikus fájlok formájában álljanak rendelkezésre, hiszen ilyen nagy mennyiségű szöveg esetén a dokumentumok

gépileg olvasható formába való konvertálása rengeteg időráfordítást jelentene. A dokumentumok gyűjtésének második meghatározó kritériuma a nyilvános elérhetőség volt a szerzői jogi problémák kiküszöbölése érdekében.

A kutatás számára mindenképpen nagy előnyt jelentett, hogy „az internet sokkal demokratikusabb publikálási forma, és ezáltal sokkal szélesebb körű mintavételt tesz lehetővé, mint a hagyományos nyomtatott médium” (Prószéky, Olasz, Váradi 2006: 1054). A szövegek nagy részének internetes elérhetősége ugyan kétségkívül gazdagította például az összehasonlításban szereplő szövegek forrásnyelvek szerinti diverzitását, de másrészt ennek ellenére éppen a szövegek nyilvános elérésének korlátai határozták meg legfőképpen a korpusz összetételének lehetőségeit.

3.2.1 Az alkorpuszok tervezése

Már utalás szintjén jeleztem, hogy a projekt tervezése során egyértelműen kirajzolódott, hogy a jelen kutatás lefolytatásához mely nyelvtechnológiai eszközök alkalmazása került számomra reálisan elérhető közelségbe. A rendelkezésre álló nyelvtechnológiai eszközök köre viszont csak magyar nyelvű szövegek automatikus elemzését tette lehetővé. Az eredetileg több nyelvpár esetében párhuzamos szövegek vizsgálatára tervezett kutatást így az eszközök szabta kereteknek megfelelően módosítottam és a vizsgálat fókuszába az autentikus és fordítás során keletkezett szövegprodukciónak szövegépítő sajátosságainak különbségét állítottam. Az alkorpuszokba gyűjtött szövegek esetében nagyon fontos szempontnak tekintettem ezért, 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.

Az autentikus magyar és a fordított magyar szövegeket tartalmazó alkorpuszok a szövegtípusok kommunikációs színterei szerint a következő alkorpuszokat tartalmazzák: közéleti, irodalmi, vallásos és tudományos. 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 ezért az Európai Unióval kapcsolatos szövegeket válogattam, az szépirodalmi alkorpuszba alapvetően regényeket és novellákat kerestem. A vallási szövegek esetében a hitbeli kérdésekkel foglalkozó egyházi szövegek közül válogattam. Ezt a szövegtípust azért tartottam a vizsgálat 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 esetében történelmi, filozófiai, nyelvészeti és pszichológiai témájú tudományos szövegeket archiváltam.

3.2.2 Autentikus magyar szövegek

A korpuszban található közéleti autentikus magyar szövegeket részben a „A 2011. évi magyar EU-elnökség” honlapjáról (www.mfa.gov.hu/kum/hu/bal/eu/2011_eu_elnokseg/), részben az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár (<http://mek.oszk.hu/>) archívumából töltöttem le. A kiválasztott dokumentumok esetében ügyeltem arra, hogy az általuk közvetített információ tartalma tekintetében ne jogszabályi felsorolások, hanem egybefüggő gondolatokat tartalmazó szöveges állományok legyenek. Az szépirodalmi alkorpuszba sorolható szövegeket az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából töltöttem le.

A vallásos témájú magyar nyelven írt szövegeket magyar papok nyilvános fórumokon közzétett publikációiból, az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából és a Magyar Katolikus Egyház hivatalos honlapján (<http://www.katolikus.hu/>) keresztül elérhető dokumentumok közül választottam ki. A letöltött dokumentumok között találhatóak elektronikus formában elérhető könyvek, a püspöki konferencia körlevelei és iránymutató írásai. A tudományos szövegtípusba sorolható dokumentumok között az autentikus magyar szövegeket tartalmazó alkorpuszban történelmi, filozófiai, nyelvészeti és pszichológiai témájú tudományos szövegeket archiváltam különféle tudományos honlapokról.

3.2.3 Fordított magyar szövegek

Eredeti elképzelésem szerint a fordított magyar korpusz az autentikus magyar korpusz kritériumaival azonos kritériumok alapján épített, a különböző forrásnyelvű kiindulási szövegek fordításaként keletkezett célnyelvi magyar szövegek lehetőleg minél többféle alkorpuszát tartalmazza. A fordítás eredményeként keletkezett célnyelvi magyar szövegek esetében sajnos bármennyire szerettem is volna, nem állt módomban a korpusz szövegeinek eredetijeiként számon tartott összes forrásnyelv vonatkozásában minden szövegtípus tekintetében elérhető magyar fordításokat találni. Jobb híján a célnyelvi magyar alkorpusz építésénél a már felkutatott, vagy ismerősöktől megkapott dokumentumok archiválását és elemzését tartottam követendő szempontnak. Ez az eljárás annál is célravezetőbbnek tűnt, mivel a jelenlegi célnyelvi szövegállomány összegyűjtése is hosszabb kutatómunka eredménye volt. Az adatok összehasonlíthatósága érdekében a szövegek forrásnyelvenkénti és szövegtípusonkénti voltát figyelembe véve igyekeztem a célnyelvi szövegek tokenszámban kifejezhető szövegtípusonkénti nagyságát a megfelelő autentikus magyar szövegek tokenszámban kifejezhető szövegtípusonkénti nagyságához igazítani.

A célnyelvi magyar korpusz esetében az angolból fordított közéleti szövegek nagy részét az Európai Unió hivatalos weboldaláról (http://europa.eu/index_hu.htm), egy részüket pedig az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából töltöttem le. Más weboldalak nem tartalmaztak hasonló témában és ellenőrizhetően fordítás eredményeként létrejött dokumentumállományt. Az szépirodalmi kategóriába tartozó szövegek között az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából töltöttem le angol, francia, német, olasz dokumentumokat.

A vallásos témájú fordított szövegek esetében a Vatikán honlapjáról (<http://www.vatican.va/>) töltöttem le a latinból fordított, nyilvánosan elérhető pápai enciklikákat, a németből fordított Anselm Grün könyvek magyar fordításait a fordító bocsátotta a rendelkezésemre. Az angol nyelvű vallásos témájú műveket az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából töltöttem le. A görögből fordított Újszövetség is nyilvánosan elérhető (<http://www.biblia->

tarsulat.hu/biblia/parhuzam.htm). A tudományos tárgyú szövegek esetében a szövegek egy részét az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából vagy az Educatio Társadalmi Szolgáltató Nonprofit Kft. által üzemeltetett és működtetett oldalról (<http://www.tankonyvtar.hu/>) töltöttem le, illetve részben kaptam ismerősök által készített fordításokat is.

3.3 A korpusz mérete

A korpusz jelenlegi méretét mint azt már korábban is említettem a számomra digitális formátumban elérhető dokumentumok száma határozta meg. Ez persze nem azt jelenti, hogy bizonyos nyelvek relációjában bizonyos szövegtípusok esetében nem lett volna könnyű tovább bővíteni a korpuszban található dokumentumok anyagát. Mivel azonban a szövegtípus szerinti alkorpuszok esetében a forrásnyelvek zöménél erre nemigen kínálkozott lehetőség, az adatok összevethetősége érdekében e tanulmány számára nem tartottam indokoltnak folytatni a szöveggyűjtést. Meggyőződésem szerint a korpuszban archivált anyagok összetétele és mennyisége alapján a Bevezetőben megfogalmazott hipotézisek (l. 5.2) bizonyítására a jelen kutatás alapjául szolgáló korpuszállomány megnyugtató méretű és összetételű és a korpuszból kinyert eredmények egyértelműen alátámasztják a tézisek állításait.

A korpusz méretének pontosabb bemutatását a szakmai gyakorlatban általában alkalmazott tokenszám megadásával kívánom teljesíteni. Az összes alkorpusz tokenszámban megadott méretének feltüntetésére terjedelmi keretek miatt most nem szeretnék külön kitérni, mivel ezek az egybevetések alkalmával jutnak csak nagyobb szerephez. A Magyar Lexikai Kohéziós Projekt teljes szövegállományának méretét az autentikus és a célnyelvi magyar alkorpuszokra vonatkozó szövegtípusonkénti adatok megadásával szemléltetem.

AUTENTIKUS	TOKENSZÁM
EU	249 315
IROD	255 699
TUD	257 442
VAL	245 009
ÖSSZES	1 007 465

2. TÁBLÁZAT: Az autentikus magyar szövegeket tartalmazó alkorpusz összetétele

FORDÍTOTT	TOKENSZÁM
EU	249 420
IROD	1 070 126
TUD	853 869
VAL	873 597
ÖSSZES	3 047 012

3. TÁBLÁZAT: A célnyelvi magyar szövegeket tartalmazó alkorpusz összetétele

Az itt bemutatott táblázatokból (2. TÁBLÁZAT és 3. TÁBLÁZAT) kiolvasható, hogy az autentikus magyar szövegeket tartalmazó alkorpusz közel 1.000.000 tokenszámú szöveget tartalmaz, míg a célnyelvi magyar korpusz közel 3.050.000 tokenszámú szöveget tartalmaz. Összességében megállapítható, hogy az autentikus és a célnyelvi szövegprodukciónak szövegépítő sajátosságait a lexikai kohézió magyar nyelvű szövegekben történő megvalósulásán keresztül feltáró Magyar Lexikai Kohéziós Projekt a jelen szakaszban közel 4.000.000 tokenszámú szöveg elemzéseinek eredményein keresztül szemlélteti. 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.

4 Az elemzés bemutatása

4.1 Az elemzéshez használt nyelvtechnológiai eszközök

A számítógéppel segített elemzések ismert előnyei közé tartoznak a relatív gyors megvalósíthatóság, az adatok és eredmények gépi tárolása, valamint az elemzési kategóriák megbízható és következetes bejelölése (Péché 2007). Ez utóbbi tulajdonság különösen a nagy szövegtörzseken lefuttatott elemzések esetében bír meghatározó jelentőséggel, mivel a nyelvtechnológiai eszköz az azonos kategória elemeket mindig azonos metaadatot tartalmazó címkével látja el. A számítógéppel

segített elemzések tehát nagymértékben hozzájárulnak a kutatások megbízhatóságához (reliability), hiszen a kutatást megismételve az automatikus elemzések során mindig ugyanazt az eredményt kapjuk.

A kohéziós elemek identifikációját az Magyar Lexikai Kohéziós Projekt (Lexical Cohesion in Hungarian) során a projekt céljainak megfelelően és érdekében összekapcsolt nyelvtechnológiai eszközök segítették. A kiindulási korpuszban archivált Word dokumentumokból első lépésben az OpenOffice program segítségével a bekezdéseket és mondathatárokat megtartó XML dokumentumokat készítettem. Az így kapott XML szövegfájlok szolgálták az Orosz György és Laki László által (PPKE ITK) az Magyar Lexikai Kohéziós Projekt célkitűzéseinek megvalósítása érdekében kifejlesztett elemző (Laki 2011, Novák, Orosz, Indig 2011) bemenetét. Az elemző először a mondathatárokat és a tokeneket azonosítja, majd a HuMor program segítségével végzi el a morfológiai annotációt és egyértelműsítést (Prószéky 1995) és végül az egyes bekezdéseken belül megkeresi a főnevek más mondatokban történő ismétléses előfordulásait, illetve a magyar WordNet Program (v.ö. Prószéky, Mihaltz 2008, Mihaltz 2005) élei mentén azonosítja a mondathatárokon túlmutató – szinonímia, antonímia, meronímia, hiponímia vagy domén reláción alapuló – szemantikai kapcsolatokat.

4.1.1 Princeton WordNet

A Princeton WordNet (<http://wordnet.princeton.edu/>) az 1990-ben a Princeton Egyetem Kognitív Tudományok Laboratóriumának professzora, George A. Miller által az angol nyelv teljes szókészletének gépileg olvasható adatbázisba való rendezésére kezdeményezett projektje. Az azóta is folyamatosan bővülő ontológia a szavak és fogalmak felsorolását és a közöttük levő viszonyok leírását tartalmazza. A természetes nyelv elemeinek és elemkapcsolatainak, azaz az ember mentális lexikonját modelláló, a számítógépes eszközök számára is értelmezhető rendszere a szóalakokat és a jelentéseket különféle lexikai relációk mentén kapcsolja össze. A kognitív mezők egyezőségei alapján a főneveket, igéket, mellékneveket és névszókat szófaji kategóriákként csoportosító szóhálózat fogalmi-szemantikai, illetve lexikai kapcsolatait a *synset* relációk, amelyek esetében a Princeton WordNetben

megkülönböztették a szinonímia, antonímia, hiponímia, meronímia, troponímia viszonyokat. A WordNet formalizmus alapvető jellemzője, hogy a synset alapján nyilvántartott szóhalmazok olyan azonos szófajhoz tartozó szavakat vagy szókapcsolatokat tartalmaznak, amelyek jelentése megközelítőleg azonos. Ez a felfogás és eljárás teljesen új feltételeket jelent az eddigi szótárstruktúra ABC rendben nyilvántartott lexikográfiai gyakorlatával szemben.

4.1.2 EuroWordNet és BalkaNet

A Princeton WordNet ontológiai célkitűzései alapján kezdték meg 1996–99 között az Európai Közösség által támogatott EuroWordNet (<http://www.illc.uva.nl/EuroWordNet/>) projekt fejlesztési munkálatait, ami nyolc nemzeti nyelv – az angol, a holland, a spanyol, az olasz, a francia, a német, a cseh, és az észt – esetében jelentette egy ontológikus adatbázis kiépítését. 2001 és 2004 között a BalkaNet keretein belül a WordNet projekthez tartozó ontológia építés további nyelvek – a görög, a román, a szerb, a török, a bolgár, a koreai, a kínai, a hindi és a tamil – adatbázisaival bővült.

4.1.3 HuWordNet

A 2005 és 2007 között fejlesztett magyar WordNet (HuWN) a magyar nyelv lexikai-szemantikai kapcsolatait tartalmazó gépileg olvasható adatbázis, a szavak jelentései mentén azonosítható lexikai relációk (pl. szinonima, antonima, hiponima, meronima) hálója. Mivel a Magyar Lexikai Kohéziós Projekt során ezidáig kizárólag a főnevek kohéziós kapcsolatainak feltárása valósult meg, a jelen tanulmány is csak a HuWordNeten belüli főnévi relációk bemutatására szorítkozik. A magyar főnévi WordNet adatbázisban szereplő főnevek hozzárendelése az angol nyelv 1.6-os verziójú WordNet synsetjeihez automatizált gépi módszerekkel történt. A feldolgozáshoz egy elektronikus magyar-angol kétnyelvű szótár főnév állománya és az egynyelvű magyar Értelmező Kéziszótár számítógéppel feldolgozható strukturális

és szemantikai információi alapján történt. Az eredmények validálása az adatbázis egy részkorpuszának kézi ellenőrzésével történt (v.ö. Miháltz 2003).

A HuWN projekt a főnévi synsetek esetében öt fő relációtípust használ, amelyek a szinonímia, az antonímia, a hiponímia, a merionímia és a domén relációk. A szinonímia mint reláció esetében az egy synsetbe tartozó főnevek jelentései közel azonosak (pl. autó – kocsi) és maguk a szinonimahalmazok rendkívül fontos szerepet játszanak a wordnetek felépítésében. Az antonima reláció a lexikális főnévi párok közötti fogalmi ellentétet jelöli (pl. *férj – feleség*). A hiponima viszony a generikus fogalmak és az azonos fogalmi körhöz tartozó specifikus egyedek közötti kapcsolatra utal (pl. *virág – rózsza*). A meronima viszony a rész-egész relációkat jelöli (pl. *fej – száj*). A domén relációknak három fajtája a category, a region és az usage. A category domain tartalmi/tematikus/szemantikai kapcsolatokat fejez ki valamilyen tevékenység vagy tárgyterület és a hozzá kapcsolódó fogalmak, eszközök, tevékenységek között (pl. *síelő – felcsatol, sícipő, lesiklás*). A region domain synsetek földrajzi területek térbeli egységeinek és a hozzájuk kapcsolódó fogalmi, kulturális, történelmi, szervezeti, stb. elemeknek a jelölésére szolgálnak (pl. *kilenc lyukú híd – Hortobágy*). Az usage domain synsetekkel jelölhető nyelvhasználati kategóriák (pl. költői, szleng) vizsgálata nem tartozott a jelen kutatás célkitűzései közé. A be in state reláció egy melléknévi synsetet csatol egy főnévhez, a főnév által leírt állapotban (pl. *éber – éberség*).

4.2 Szegmentálás és morfológiai elemzés

4.2.1 Bekezdésre és mondatra bontás

A lexikai kohéziós elemek identifikációja céljából annotált korpuszban a szegmentálás első alapvető lépése a téma tagolódásának vizuális jelölésére szolgáló bekezdések gépileg olvasható metainformáció segítségével (tag) való bejelölése. A vizuálisan az egyik bentebb kezdéstől a másik bentebb kezdésig tartó, míg akusztikusan a beszédben szünet tartásával jelölt bekezdések meghatározása a szövegben azért elengedhetetlen a lexikai kohéziós kutatások számára, mert a bekezdések a prózai művek legnagyobb mikroszerkezeti egységei, azaz a szövegek

szemantikai-logikai szegmentumai és így egyben a kohéziós kapcsolatok realizációjának színterei.

A mondatok, illetve a tagmondatok annotálása a bekezdéseken belül viszont azért jelentős, mert a lexikai kohéziós relációk a mondathatárokon túlmutató szövegsemantikai kapcsolatok. Az automatikus gépi eszköz számára tehát az egyes bekezdéseken belül található különböző mondatok lexikai elemei közötti kapcsolatok identifikációját kellett definiálni.

4.2.2 Tokenizálás, egyértelműsítés

A szegmentálás következő lépése során a mondatokon belül a HuMor morfológiai elemző (Prószéky 1995) a szavakat tövekre és azok toldalékaira bontja szét. A HuMor segítségével megtörténik a szótövek és toldalékok morfoszintaktikai jegyeinek kódolt leírása (Morphosyntactic Description), ami tartalmazza a szófajkódokat is. Az egyértelműsítés során a program azon szóalakok esetében, amelyek több kategóriába is besorolhatóak lennének, a konkrét szöveggörnyezet segítségével dönti el az aktuális szófaji, vagy morfológiai kategóriát. Az Orosz és Laki által készített elemző (Laki 2011, Novák, Orosz, Indig 2011) feladata ezen a szinten tehát az egy bekezdésen belül található főnévek felkínálása az ismétlésrelációk megtalálására és a magyar WordNet adatbázis számára.

4.3 A lexikai kohéziós kapcsolatok automatikus azonosítása

Az elemző az utolsó lépésben azonosítja bekezdésenként az egyes főneveknek a többi mondatban történő ismétléses előfordulását és feltárja a WordNet kategóriák alapján azonosítható lexikai kohéziós kapcsolatokat. A WordNet adatbázis szerepe tehát annak a vizsgálata, hogy a bekezdésben található bármely főnév esetében megtalálható-e bármiféle, az adatbázisban nyilvántartott reláció mentén azonos bekezdés többi mondatában. Az elemző egyrészt az egyes paragrafusok végén összegzi a kapcsolat típusa szerint a bekezdésben talált lexikai-szemantikai

kapcsolatok számát, másrészt a szöveg végén az egyes bekezdések után található végeredményeket összegezve adja meg a főneveknek az adott dokumentumban a bekezdéseken belül azonosított kohéziós ismétléses kapcsolatait.

5 Automatikusan gépi eszközökkel azonosított lexikai kohéziós kapcsolatok a Magyar Lexikai Kohéziós Projekt során

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 vallási 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 már csak mennyiségi jellemzőjük alapján is kiinduló pontjai lehetnek a jelen kutatás folytatásaként egy a kohéziós kötések szövegtípusonkénti nyelvi realizációját feltáró szövegnyelvészeti vizsgálatnak.

AUTENTIKUS	TOKENEK	ÖSSZES LEXIKAI KOHÉZIÓS KAPCSOLAT
EU	249 315	27 142
IROD	255 699	30 998
TUD	257 442	58 414
VAL	245 009	53 348
ÖSSZES	1 007 465	169 902

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

FORDÍTOTT	TOKENEK	ÖSSZES LEXIKAI KOHÉZIÓS KAPCSOLAT
EU	249 420	26 107
IROD	1 070 126	111 305
TUD	853 869	181 762
VAL	873 597	172 197
ÖSSZES	3 047 012	491 371

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

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

A lexikai kohéziós kapcsolatok azonosítása során az elemző bejelölte az ismétléseket, a szinonimákat, az antonimákat, a meronimákat, a hiponimákat, a category domain relációkat, a be in state relációkat és a region domain relációkat, melyek közül a jelen tanulmány keretein belül három kategóriát emeltem ki a szövegtípus specifikus realizációk részletesebb bemutatására. A kohéziós kategóriánként és szövegtípusonként bemutatott nyelvi példák elemzése nem tartozik a jelen dolgozat célkitűzései közé, csak a kohéziós kapcsolatok realizációjának gazdagságát és az adott korpuszra jellemző előfordulásait hivatott érzékeltetni. Az itt példaként bemutatott kohéziós kapcsolatok nagy része többször, vagy akár igen nagy számban is megismétlődik a korpusz során. A 4. TÁBLÁZAT és az 5. TÁBLÁZAT is alátámasztja, hogy a jelen értekezés a kohéziós kapcsolatok előfordulásának nagy száma miatt nem törekedhetett azok hiánytalan felsorolására, ez egy jövőbeni kutatás kiindulópontja lehet.

A következő alfejezetekben egyrészt szeretném érzékeltetni néhány szövegminta segítségével 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. A további alfejezetekben másrészt a korpuszban szereplő kohéziós kapcsolatok gazdagságának érzékeltetésére további példákat mutatok a terjedelmi keretek miatt a konkrét szövegrészletek mellőzésével. Itt ugyancsak a terjedelmi keretek miatt az antonima, a meronima és a hiponima kapcsolatoknak a Magyar Lexikai Kohéziós Projektben történő szövegtípus szerinti realizációjára szorítok. A felsorolt három kategóriára egyrészt a három kategóriára jellemző nyelvi példa nagy száma miatt,

másrészt a három kategória nyelvészeti vizsgálat számára érdekes volta miatt esett a választásom.

5.1.1 *A lexikai kohéziós kapcsolatok realizációja a Magyar Lexikai Kohéziós Projektben*

Az első példa (8. ábra) *Az unió kapujában*¹ című szöveg közéleti szövegtípusba tartozó eredeti Word-dokumentum részlete:

– *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 eredménytelenül támadnak a **maffiák**. Nagyon sokszor előfordult a **közpénzek felhasználásakor**, az állami megrendelésekkor, hogy valószínűsíthető volt: a **maffiák** vitték el **ezeket a pénzeket**.*

8. ÁBRA: Részlet *Az unió kapujában* című dokumentumból

Az elemzés első lépésében történik a szöveg szegmentálása és morfológiai elemzése, amit a bekezdés 2. mondatának elemzésén keresztül szemléltet a 9. ábra. A jelen tanulmány szempontjából viszont az igazán érdekes az elemzés következő lépése, amikor a bekezdésen belüli lexikai kohéziós kapcsolatokat azonosítja a program. A jelen mű fekete-fehér és lineáris ábrázolás módja sajnos nem tükrözi az elemző kimeneti dokumentumaként kapott színes annotációkat tartalmazó xml fájlok szemléletes struktúráját.

```

<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>

```

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

A 10. ábra lineárisan szemlélteti a példaként hozott közéleti szövegtípushoz tartozó szövegrészletben realizálódó kohéziós kapcsolatokat. 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.

"repetition":	"Erre"	>	"ez"
"repetition":	"Erre"	>	"ez"
"repetition":	"Erre"	>	"ebben"
"repetition":	"Erre"	>	"ez"
"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"	>	"ez"
"repetition":	"ez"	>	"ezeket"
"synonym":	"háború"	>	"küzdelemben"
"repetition":	"maffiának"	>	"maffiák"
"repetition":	"maffiának"	>	"maffiák"
"repetition":	"szüksége"	>	"szüksége"
"repetition":	"arra"	>	"arra"
"repetition":	"arra"	>	"arra"
"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"

10. ÁBRA: A 8. ábra szövegrészletében található kohéziós kapcsolatok

Egy másik illusztrációként szolgáló szövegrészlet a szépirodalmi szövegtípusba tartozó Ballai László *A kitagadott*¹¹ című regényéből (11. ábra) származik. Mivel a 11. ábra szövegrészletében a kohéziós kapcsolatok száma sokkal kevesebb, mint a 8. ábra szövegrészletében, így ez a példa (12. ábra) talán jobban szemlélteti a félkövér betűtípussal ki is emelt kohéziós kapcsolatok azonosítását. Itt kell még egyszer felhívni a figyelmet arra, hogy a jelen értekezés keretein belül csak a névszói elemek lexikai kohéziós kapcsolatainak beazonosítására került sor.

A 12. ábra esetében jól látszik, hogy az *étterembe* lexikai elemnek csak olyan ismétléses kapcsolata van, amit az azonos lexikai elem biztosít (repetition). Az adott szövegrészletben hasonlóan viselkednek a *konyháról*, a *vacsorára*, a *menüt* és a *kacsa* elemek is. A *szószokat* lexikai elem esetében több meronima-kapcsolatot azonosított az elemző az adott szövegrészletben, melyek: az *ételek*, az *ételeiket*, a *fogásoktól* és az *ételekig*. Az *ételek* lexikai egység esetében viszont találunk szó szerinti ismétléseket (*ételeiket*, *ételekig*) és szinonima-kapcsolatokat (*fogásoktól*) is.

– Na majd egyszer elmegyünk egy kínai **étterembe** – mondta a bácsi, és megállt a lépcsőfordulóban. – Nagyon érdekes, biztos fogod szeretni. Tudod, nem mindegy, hogy kivel eszik kínait az ember. Először Stockholmban jártam kínai **étteremben**, de egy olyan pasassal, akinek fogalma sem volt a kínai **konyháról**. Össze-vissza locsoltuk a tányérunkba az édes, a savanyú, a csípős **szószokat**, és a tetejébe még kenyeret is kértünk. Sokáig azt hittem, a kínaiak moslékot főznek. Végül Londonban egy kínai hívott meg **vacsorára**. Életem egyik legnagyobb gasztronómiai élménye volt. Elmeséltem neki stockholmi **vacsorámat**. Finoman mosolygott. „Általában a helyes arány megtalálásán múlnak a dolgok” - mondta, és adott néhány tanácsot. Aszerint válasszam meg a társaságomat, hogy mit akarok enni. Legjobb persze, ha egyedül megyek vacsorázni, mert így tudom gondtalanul átadni magamat az **ételek** élvezetének. Csak olyan kínai **étterembe** menjek be, ami tele van kínaiakkal. Lehetőleg **menüt** válasszak, mert a legjobban a kínaiak tudják, hogy milyen sorrendben érdemes fogyasztani az **ételeiket**, milyen meglepetésekkel szegélyezzék a könnyű **fogásoktól** a csípős **ételekig** vezető utat. Végül, **kacsa** mindig legyen a **menüben**, mert a **kacsát** egyetlen **konyha** sem tudja olyan porhanyósra sütni, mint a kínai.

11. ÁBRA: Szövegrészlet A kitagadott című regényből

"repetition":	"étterembe"	>	"étterembe"
"repetition":	"étterembe"	>	"étterembe"
"repetition":	"konyháról"	>	"konyha"
"meronym":	"szószokat"	>	"ételek"
"meronym":	"szószokat"	>	"ételeiket"
"meronym":	"szószokat"	>	"fogásoktól"
"meronym":	"szószokat"	>	"ételekig"
"repetition":	"vacsorára"	>	"vacsorámat"
"repetition":	"étterembe"	>	"étterembe"
"repetition":	"ételek"	>	"ételeiket"
"synonym":	"ételek"	>	"fogásoktól"
"repetition":	"ételek"	>	"ételekig"
"repetition":	"menüt"	>	"menüben"
"synonym":	"ételeiket"	>	"fogásoktól"
"repetition":	"ételeiket"	>	"ételekig"
"repetition":	"kacsa"	>	"kacsát"

12. ÁBRA: : A 11. ábra szövegrészletében található kohéziós kapcsolatok

A tudományos szövegtípust képviseli *A magyar történelem tizedei*^{III} című Bonfini mű magyar fordítása Kulcsár Péter fordításában. A 13. ábra szövegrészletéhez kapcsolódó lexikai kohéziós kapcsolatrendszer félkövér

betűtípussal kiemelt példái azt szemléltetik (14. ábra), amikor egy lexikai elem egy korábban előforduló lexikai elem ismétléses kapcsolata, majd később a szövegben az adott elemnek is megtalálható egy másik lexikai elem által biztosított ismétléses kapcsolata: a *zsinatot* elem hiponima kapcsolatban van a *gyűlést* lexikai elemmel, majd a *gyűlést* elem hiponima kapcsolataként jelenik meg a *gyülekezet* kifejezés.

Theodosius parancsára Ephesusban **zsinatot** tartanak, amelyen **Dioscorus** elnökletével Eutychest felmentik, Flaviánt elítélik. Ezzel szemben **Theodosius** halála után Leó Marcianus **utasítására** és védnöksége alatt Calchedonba hívatott össze **gyűlést**, amelyet egyetemesnek mondanak, mivel 630 szent **atya** vett részt rajta, amelyen a szentséges **gyülekezet** megerősítette, hogy **Krisztusban** kettős természet van, és ugyanazt az egyetlen **Krisztust** kell istennek és embernek hinni. Aztán valamennyi **atya** egyetértésével nyomban kiátkozták a manicheus fertőzés szerzőit, és **könyveiket** nyilvánosan elégették; **Dioscorus** dölőfjét megtaposták, és az eretnekséget eltörölték.

13. ÁBRA: Szövegrészlet A magyar történelem tizedei műből

"repetition":	"Theodosius"	>	"Theodosius"
"synonym":	"parancsára"	>	"utasítására"
"hyponym":	"zsinatot"	>	"gyűlést"
"repetition":	"Dioscorus"	>	"Dioscorus"
"hyponym":	"gyűlést"	>	"gyülekezet"
"repetition":	"atya"	>	"atya"
"hyponym":	"atya"	>	"szerzőt"
"repetition":	"Krisztusban"	>	"Krisztust"

14. ÁBRA: A 13. ábra szövegrészletében található kohéziós kapcsolatok

Az utolsó példa a *Caritas in veritate*^{IV} kezdetű enciklika szövegrészlete. Ebben a szövegrészletben egyrészt arra láthatunk példát, hogy az elemző a névmásokat is vizsgálja a kohéziós kapcsolatok szempontjából (*azt, arra*), másrészt példát találunk arra a jelenségre, amikor egy lexikai elem két különböző lexikai elem ismétléses kapcsolataként is megjelenik, mint például a *kívánalmakat – követelményeit* és a *szüksége – követelményeit* szinonima kapcsolatok esetében. A szövegrészletben ebben az esetben is félkövér betűtípus emeli ki a kohéziós kapcsolatokat biztosító lexikai elemeket.

VI. Pál a Populorum progressióban olyan piacgazdasági modell létrehozására hívta fel a figyelmet, amely legalább tendenciájában képes minden népet befogadni, nem csak a kellőképpen fejletteket. Szorgalmazta, hogy tegyenek **erőfeszítéseket** a **mindenki** számára emberibb **világ** felépítéséért, az olyan **világért**, amelyben „**mindenki** ad és kap anélkül, hogy az egyik fél haladása akadályozná a másik **fejlődését**”. [94] A pápa ilyen módon világméretűvé terjesztette ki a Rerum novarumban foglalt **kívánalmakat** és **törekvéseket**, amely enciklika az ipari forradalom következtében született, s korát mindenképp megelőzve, elsőként fogalmazta meg **azt** a **gondolatot**, hogy a polgári rendnek öngazgatásához **szüksége** van az állam osztó célzatú beavatkozására is. Ez a szemlélet ma a piacok és a társadalmak nyitottá válásának **folyamatai** miatt nemcsak válságban van, hanem elégtelen is **arra**, hogy kielégítse a teljes **értelemben** vett emberi gazdaság **követelményeit**. Amit az Egyház társadalmi tanítása a maga emberképe és társadalomképe alapján mindig is vallott, **azt** ma a globalizációra jellemző dinamika is megköveteli.

15. ÁBRA: Szövegrészlete a Caritas in veritate kezdetű enciklikából

"synonym":	"erőfeszítéseket"	>	"törekvéseket"
"repetition":	"mindenki"	>	"mindenki"
"repetition":	"világ"	>	"világért"
"hyponym":	"fejlődését"	>	"folyamatai"
"synonym":	"kívánalmakat"	>	"követelményeit"
"repetition":	"azt"	>	"arra"
"repetition":	"azt"	>	"azt"
"hyponym":	"értelemben"	>	"gondolatot"
"synonym":	"szüksége"	>	"követelményeit"
"repetition":	"arra"	>	"azt"

16. ÁBRA: A 15. ábra szövegrészletében található kohéziós kapcsolatok

Bár a nyelvészeti elemzés számára e szövegminták tanulmányozása további csemegéket tartogatna, most a sor a már említett terjedelmi keretek miatt csak a Magyar Lexikai Kohéziós Projektben található antonima-, meronima- és hiponima-kapcsolatok további realizációinak felsorolásszerű bemutatásával folytatódik. Ezek a minták csak a kohéziós kapcsolatok jelen korpuszban megtalálható sokszínű megvalósulását hivatottak alátámasztani.

5.1.2 További példa az antonima-kapcsolat realizációjára a Magyar Lexikai Kohéziós Projektben

AUTENTIKUS ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

"antonym":	"fogyást"	>	"növekedésükkel"
"antonym":	"kérdésre"	>	"válasz"
"antonym":	"egyetértésre"	>	"nézetkülönbségeket"
"antonym":	"veszélyek"	>	"biztonsága"
"antonym":	"erők"	>	"gyengeségünk"
"antonym":	"előnyei"	>	"hátrány"
"antonym":	"háború"	>	"béke"
"antonym":	"csőd"	>	"sikereket"
"antonym":	"rendet"	>	"zűrzavar"
"antonym":	"jövőbe"	>	"múltunkra"
"antonym":	"jövendő"	>	"múltunkra"

17. ÁBRA: Példák Az unió kapujában című szövegből

FORDÍTOTT ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

"antonym":	"asszony"	>	"férfihoz"
"antonym":	"nők"	>	"ember"
"antonym":	"versengés"	>	"együttműködés"
"antonym":	"türelmetlenség"	>	"tolerancia"
"antonym":	"kisebbségek"	>	"többséggel"

18. ÁBRA: Példák a Kisebbségi adattár^V című szövegből

"antonym":	"munkavállalók"	>	"munkáltatók"
"antonym":	"alkalmazott"	>	"munkáltató"
"antonym":	"jövedelem"	>	"kiadásokat"

19. ÁBRA: Példák az IP-07-752 azonosítójú EU-s dokumentum^{VI} szövegéből

"antonym":	"férfiak"	>	"nők"
"antonym":	"lányok"	>	"fiúknál"

20. ÁBRA: Példák az IP-07-1115 azonosítójú EU-s dokumentum^{VII} szövegéből

AUTENTIKUS ALKORPUSZ – SZÉPIRODALMI SZÖVEGTÍPUS

"antonym":	"bácsi"	>	"néni"
"antonym":	"ember"	>	"nők"
"antonym":	"férje"	>	"felesége"
"antonym":	"apám"	>	"Édesanyám"
"antonym":	"rendben"	>	"zavarba"
"antonym":	"fogyásnak"	>	"növekedését"
"antonym":	"siker"	>	"kudarcc"
"antonym":	"békén"	>	"háború"
"antonym":	"múltjához"	>	"jövőjéhez"

21. ÁBRA: Példák A kitagadott című regényből

FORDÍTOTT ALKORPUSZ – SZÉPIRODALMI SZÖVEGTÍPUS

"antonym":	"élőket"	>	"holtakat"
"antonym":	"anyjánál"	>	"apja"
"antonym":	"férfi"	>	"asszonyok"
"antonym":	"szülei"	>	"gyermekemből"
"antonym":	"szegényt"	>	"gazdagot"
"antonym":	"múltra"	>	"jövőt"
"antonym":	"gyűlölettel"	>	"szeretet"
"antonym":	"rendbe"	>	"zűrzavarában"
"antonym":	"kérdésre"	>	"választ"
"antonym":	"örömet"	>	"gyötrelmeket"
"antonym":	"örömet"	>	"fájdalom"
"antonym":	"világosság"	>	"sötétbe"

22. ÁBRA: Példák a Tájfun^{VIII} című antológiából

AUTENTIKUS ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

"antonym":	"nő"	>	"ember"
"antonym":	"lányaim"	>	"fiú"
"antonym":	"melegben"	>	"hidegben"
"antonym":	"vége"	>	"elején"
"antonym":	"fogyásával"	>	"növekedésével"
"antonym":	"kérdésre"	>	"választ"
"antonym":	"tökéletessége"	>	"tökéletlensége"
"antonym":	"békét"	>	"háború"
"antonym":	"kapitalizmus"	>	"szocializmus"
"antonym":	"ugyanaz"	>	"más"
"antonym":	"anód"	>	"katód"
"antonym":	"zűrzavarában"	>	"rendet"

23. ÁBRA: Példák a Négy könyv^{IX} című gyűjteményből

FORDÍTOTT ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

"antonym":	"rendnél"	>	"rendetlenségtől"
"antonym":	"háborút"	>	"békével"
"antonym":	"erényeid"	>	"tökéletlenségemmel"
"antonym":	"szeretetet"	>	"gyűlöletet"
"antonym":	"élőkhöz"	>	"halottak"
"antonym":	"utódának"	>	"őseikként"
"antonym":	"apját"	>	"anyját"
"antonym":	"mások"	>	"ugyanazon"
"antonym":	"asszonyokat"	>	"férfiakhoz"
"antonym":	"gyalázatban"	>	"tisztelettel"
"antonym":	"bátorságát"	>	"félelemtől"
"antonym":	"szenvéde"	>	"örömeiben"

24. ÁBRA: Példák A magyar történelem tizedei című műből

AUTENTIKUS ALKORPUSZ – VALLÁSOS SZÖVEGTÍPUS

"antonym":	"születések"	>	"haláláról"
"antonym":	"termékenység"	>	"terméketlenség"
"antonym":	"gyermek"	>	"szülőknél"
"antonym":	"férfi"	>	"nő"
"antonym":	"férj"	>	"feleség"
"antonym":	"emberek"	>	"nők"
"antonym":	"fiatalok"	>	"felnőtt"
"antonym":	"fiatalok"	>	"öreg"
"antonym":	"anya"	>	"apák"
"antonym":	"csökkenését"	>	"növekedését"
"antonym":	"mérséklődése"	>	"növekedése"
"antonym":	"munkanélküliség"	>	"állásukat"

25. ÁBRA: Példák A boldogabb családokért^X körlevélből

FORDÍTOTT ALKORPUSZ – VALLÁSOS SZÖVEGTÍPUS

"antonym":	"férfiak"	>	"nők"
"antonym":	"más"	>	"ugyanaz"
"antonym":	"embert"	>	"nőnek"
"antonym":	"gyermek"	>	"szülő"
"antonym":	"apa"	>	"anya"
"antonym":	"élők"	>	"holtak"
"antonym":	"világosság"	>	"sötétség"
"antonym":	"vereséget"	>	"győzelmet"
"antonym":	"egyenlőtlenségek"	>	"egyenlőségével"
"antonym":	"kezdet"	>	"vég"
"antonym":	"választ"	>	"kérdéseire"

26. ÁBRA: : Példák a Gaudium et spes^{XI} kezdetű lelkipásztori konstitúcióból

5.1.3 További példák a meronima-kapcsolat realizációjára a Magyar Lexikai Kohéziós Projektben

AUTENTIKUS ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

"meronym":	"április"	>	"tavaszt"
"meronym":	"éveken"	>	"évtizednek"
"meronym":	"hetekben"	>	"hónapon" /
"meronym":	"hónaptól"	>	"év"
"meronym":	"polgárok"	>	"országok"
"meronym":	"polgárai"	>	"nemzetekkel"
"meronym":	"Budapestre"	>	"Magyarországra"
"meronym":	"Erdély"	>	"Románia"
"meronym":	"Mallorcán"	>	"Spanyolországban"
"meronym":	"szintig"	>	"épületet"
"meronym":	"gyerekek"	>	"családjában"

27. ÁBRA: Példák Az unió kapujában című szövegből

FORDÍTOTT ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

"meronym":	"alprogramja"	>	"Program"
"meronym":	"szöveg"	>	"kiadványainak"
"meronym":	"feltétele"	>	"Egyezmény"
"meronym":	"személyiséget"	>	"személyek"
"meronym":	"részt"	>	"egésznek"
"meronym":	"cikkelyek"	>	"dokumentumoknál"
"meronym":	"Bulgária"	>	"Európában"
"meronym":	"Görögország"	>	"Európában"
"meronym":	"Bulgáriában"	>	"Európában"
"meronym":	"Szlovákiában"	>	"Európában"
"meronym":	"Montenegró"	>	"Szerbia"

28. ÁBRA: Példák a Kisebbségi adattár című szövegből

AUTENTIKUS ALKORPUSZ – SZÉPIRODALMI SZÖVEGTÍPUS

"meronym":	"polgárokkal"	>	"országokat"
"meronym":	"Coventry"	>	"Angliában"
"meronym":	"légkör"	>	"földön"
"meronym":	"padlásra"	>	"házakból"
"meronym":	"könyvtárban"	>	"házak"
"meronym":	"keze"	>	"ember"
"meronym":	"tenyeréről"	>	"kezeből"
"meronym":	"kezeből"	>	"karomat"
"meronym":	"lába"	>	"asztalról"
"meronym":	"kengyelből"	>	"nyergében"
"meronym":	"hegyét"	>	"kardot"
"meronym":	"órát"	>	"napot"

29. ÁBRA: Példák A kitagadott című regényből

FORDÍTOTT ALKORPUSZ – SZÉPIRODALMI SZÖVEGTÍPUS

"meronym":	"hét"	>	"hónap"
"meronym":	"órákor"	>	"nappal"
"meronym":	"napig"	>	"hét"
"meronym":	"birtokról"	>	"föld"
"meronym":	"birtokról"	>	"ország"
"meronym":	"felhőjét"	>	"ég"
"meronym":	"fenekére"	>	"tó"
"meronym":	"alakulatához"	>	"hadseregébe"
"meronym":	"csata"	>	"háborúkbán"
"meronym":	"gyerekek"	>	"családhoz"
"meronym":	"szemem"	>	"arca"
"meronym":	"kezére"	>	"embereknek"
"meronym":	"betűjére"	>	"ábécé"

30. ÁBRA: Példák a Tájfun című antológiából

AUTENTIKUS ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

"meronym":	"sejtmagban"	>	"sejtekben"
"meronym":	"sejtjei"	>	"szervezet"
"meronym":	"szövetekben"	>	"szervezet"
"meronym":	"éveiben"	>	"évtizedekben"
"meronym":	"keze"	>	"ember"
"meronym":	"réz"	>	"bronz"
"meronym":	"vasat"	>	"acél"
"meronym":	"agyagból"	>	"tégla"
"meronym":	"apszissá"	>	"templommal"
"meronym":	"pillérről"	>	"templommal"
"meronym":	"fa"	>	"erdők"
"meronym":	"Nílus"	>	"Egyiptom"

31. ÁBRA: Példák a Négy könyv című gyűjteményből

FORDÍTOTT ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

"meronym":	"arcot"	>	"emberek"
"meronym":	"lábát"	>	"testének"
"meronym":	"szeme"	>	"arcot"
"meronym":	"kezébe"	>	"embert"
"meronym":	"szülőktől"	>	"családjáról"
"meronym":	"Milánó"	>	"Itáliába"
"meronym":	"Rómában"	>	"Itáliába"
"meronym":	"Cellájában"	>	"kolostort"
"meronym":	"oszlop"	>	"templomot"
"meronym":	"falai"	>	"épületet"
"meronym":	"tetejéről"	>	"épületet"
"meronym":	"oszlop"	>	"szentélyt"

32. ÁBRA: Példák A magyar történelem tizedei című műből

AUTENTIKUS ALKORPUSZ – VALLÁSOS SZÖVEGTÍPUS

"meronym":	"polgárai"	>	"Országában"
"meronym":	"testvéreiknek"	>	"családok"
"repetition":	"házasságot"	>	"házasság"
"meronym":	"házasfelek"	>	"házaspárok"
"meronym":	"gyermeküket"	>	"családok"
"meronym":	"szülők"	>	"családok"
"meronym":	"fejéből"	>	"teste"
"meronym":	"lábából"	>	"teste"
"meronym":	"részük"	>	"egységet"
"meronym":	"karodra"	>	"emberek"
"meronym":	"keresztények"	>	"egyház"
"meronym":	"éveken"	>	"évtizedeken"

33. ÁBRA: Példák A boldogabb családokért körlevélből

FORDÍTOTT ALKORPUSZ – VALLÁSOS SZÖVEGTÍPUS

"meronym":	"múltra"	>	"életét"
"meronym":	"órát"	>	"napja"
"meronym":	"harcba"	>	"háború"
"meronym":	"polgárai"	>	"népei"
"meronym":	"polgárait"	>	"nemzet"
"meronym":	"testvéreinket"	>	"család"
"meronym":	"gyermekeinek"	>	"család"
"meronym":	"kezed"	>	"ember"
"meronym":	"részük"	>	"egészségben"
"meronym":	"jellegzetességei"	>	"természetébe"
"meronym":	"személyiségének"	>	"Lelke"
"meronym":	"keresztények"	>	"Egyház"

34. ÁBRA: Példák a Gaudium et spes kezdetű lelkipásztori konstitúcióból

5.1.4 További példák a hiponima-kapcsolat realizációjára a Magyar Lexikai Kohéziós Projektben

AUTENTIKUS ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

"hyponym":	"Uniónak"	>	"Amerika"
"hyponym":	"Európa"	>	"földrész"
"hyponym":	"kontinens"	>	"Európa"
"hyponym":	"integráció"	>	"unió"
"hyponym":	"hazánk"	>	"országainak"
"hyponym":	"tagállamokban"	>	"országok"
"hyponym":	"kísérletei"	>	"lehetőségét"
"hyponym":	"próbálkozás"	>	"lehetőségét"
"hyponym":	"kiadások"	>	"veszteségét"
"hyponym":	"kisvállalat"	>	"cégek"
"hyponym":	"kisvállalat"	>	"társaságokat"

35. ÁBRA: Példák Az unió kapujában című szövegből

FORDÍTOTT ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

"hyponym":	"munka"	>	"termék"
"hyponym":	"vállalkozások"	>	"munka"
"hyponym":	"munka"	>	"munkahely"
"hyponym":	"munkavállalónak"	>	"dolgozók"
"hyponym":	"költségek"	>	"kiadásait"
"hyponym":	"egészsége"	>	"jólléte"
"hyponym":	"nyereségességét"	>	"állapot"
"hyponym":	"gyakorlat"	>	"tevékenységeket"
"hyponym":	"tagállamra"	>	"ország"

36. ÁBRA: Példák az IP-07-752 azonosítójú EU-s dokumentum szövegéből

AUTENTIKUS ALKORPUSZ – SZÉPIRODALMI SZÖVEGTÍPUS

"hyponym":	"könyvtárban"	>	"helyiségben"
"hyponym":	"áruházban"	>	"üzletből"
"hyponym":	"világon"	>	"területére"
"hyponym":	"emberek"	>	"valaki"
"hyponym":	"halott"	>	"valaki"
"hyponym":	"könyvet"	>	"szakasza"
"hyponym":	"szakasza"	>	"zene"
"hyponym":	"szag"	>	"érzés"
"hyponym":	"illata"	>	"szag"
"hyponym":	"ingeket"	>	"ruhában"
"hyponym":	"Duna"	>	"folyónál"
"hyponym":	"Volga"	>	"folyónál"

37. ÁBRA: Példák A kitagadott című regényből

FORDÍTOTT ALKORPUSZ – SZÉPIRODALMI SZÖVEGTÍPUS

"hyponym":	"dologról"	>	"gondra"
"hyponym":	"kötelességemnek"	>	"dolgában"
"hyponym":	"adósság"	>	"tartozásait"
"hyponym":	"arcképén"	>	"képét"
"hyponym":	"özvegy"	>	"asszony"
"hyponym":	"család"	>	"nemzetsége"
"hyponym":	"királyságban"	>	"birtokán"
"hyponym":	"kastélyba"	>	"Házában"
"hyponym":	"nagybőjtre"	>	"idővel"
"hyponym":	"napján"	>	"idővel"
"hyponym":	"napot"	>	"idejében"
"hyponym":	"nap"	>	"csillagomnak"

38. ÁBRA: Példák a Tájfun című antológiából

AUTENTIKUS ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

"hyponym":	"öntöttvas"	>	"vas"
"hyponym":	"vas"	>	"fémekhez"
"hyponym":	"golyók"	>	"töltetét"
"hyponym":	"pajzsok"	>	"páncélok"
"hyponym":	"könyv"	>	"termékeinek"
"hyponym":	"nyomtatás"	>	"írás"
"hyponym":	"tenger"	>	"víz"
"hyponym":	"óceánra"	>	"víz"
"hyponym":	"óceán"	>	"víz"
"hyponym":	"hajóépítést"	>	"építésével"
"hyponym":	"találmányokban"	>	"alkotás"
"hyponym":	"művészetben"	>	"alkotás"

39. ÁBRA: Példák a Négy könyv című gyűjteményből

FORDÍTOTT ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

"hyponym":	"győzelemtől"	>	"sikerrel"
"hyponym":	"püspök"	>	"papjának"
"hyponym":	"papjának"	>	"lelkészt"
"hyponym":	"képében"	>	"arcot"
"hyponym":	"palástot"	>	"ruházatával"
"hyponym":	"köpenye"	>	"ruhája"
"hyponym":	"ütközetből"	>	"nézeteltérésének"
"hyponym":	"problémáinak"	>	"nehézségével"
"hyponym":	"halálból"	>	"vége"
"hyponym":	"idő"	>	"pillanatra"
"hyponym":	"meghívását"	>	"utasítására"
"hyponym":	"dolog"	>	"feladatot"

40. ÁBRA: Példák A magyar történelem tizedei című műből

AUTENTIKUS ALKORPUSZ – VALLÁSOS SZÖVEGTÍPUS

"hyponym":	"ember"	>	"egyén"
"hyponym":	"ember"	>	"személy"
"hyponym":	"jóléte"	>	"boldogulásának"
"hyponym":	"irányból"	>	"helyzetétől"
"hyponym":	"világából"	>	"területen"
"hyponym":	"kötelék"	>	"kapcsolat"
"hyponym":	"közösségére"	>	"kapcsolat"
"hyponym":	"tanítás"	>	"elvek"
"hyponym":	"civilizációval"	>	"társadalommal"
"hyponym":	"születések"	>	"változások"
"hyponym":	"változások"	>	"átalakulása"
"hyponym":	"vagyon"	>	"tulajdon"

41. ÁBRA: Példák A boldogabb családokért körlevélből

FORDÍTOTT ALKORPUSZ – VALLÁSOS SZÖVEGTÍPUS

"hyponym":	"értelmét"	>	"célja"
"hyponym":	"életének"	>	"léte"
"hyponym":	"jelenléte"	>	"léte"
"hyponym":	"alaptörvényével"	>	"törvény"
"hyponym":	"szentnek"	>	"személy"
"hyponym":	"emberek"	>	"személy"
"hyponym":	"ura"	>	"emberek"
"hyponym":	"család"	>	"egységét"
"hyponym":	"küldetés"	>	"feladatok"
"hyponym":	"fejlődést"	>	"folyamatát"
"hyponym":	"segítség"	>	"tevékenységnek"
"hyponym":	"összhangban"	>	"rendben"

42. ÁBRA: Példák a Gaudium et spes kezdetű lelkipásztori konstitúcióból

6 A Magyar Lexikai Kohéziós Projekt eredményei

Az előző fejezetben bemutatott Orosz–Laki-féle elemző (Laki 2011, Novák, Orosz, Indig 2011) által dokumentumonként összegzett végeredményeket az autentikus és a célnyelvi szövegprodukción külön kezelő Excel táblázatokban összegeztem. A táblázatokban a szövegeket az egyes szövegtípusok és az egyes forrásnyelvek függvényében külön tudtam kezelni. A táblázatok sorai tartalmazzák az egyes szövegekre, míg az oszlopai az egyes lexikai kohéziós kategóriákban a szövegre vonatkoztatott eredményeket. A táblázatokban a szövegek, a lexikai kohéziós kategóriák és a szövegeket tartalmazó folderek elnevezései a számítógépes dokumentumok kezeléséből ismert ékezet és *space* nélküli leírás szerint szerepelnek.

XX_sz.	SZÖVEG	TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
1	Ballai	48 187	8 649	312	615	1 569	511	150	7	0
2	Fekete	83 292	5 588	656	166	2 069	199	29	2	0
3	Sánta	71 244	6 342	348	159	1 141	97	89	6	0
	ÖSSZES	202 723	20 579	1 316	940	4 779	807	268	15	0
	%		10,15%	0,65%	0,46%	2,36%	0,40%	0,13%	0,01%	0,00%

6. TÁBLÁZAT: Részlet az eredményeket archiváló táblázatból

A táblázat oszlopaiban található adatok tehát a tokenszám (TOKENEK), az ismétlések száma (R), a szinonimák száma (S), az antonimák száma (A), a meronimák száma (M), a hiponimák száma (H), a *category domain* relációk száma (C_D), a *be in state* relációk száma (B_I_S) és a *region domain* relációk száma (R_D). Az egyes szövegtípusokra az egyes forrásnyelvek esetében jellemző adatokat az azonos kategóriába tartozó szövegek adatai alatt összegeztem (ÖSSZES), majd egy következő sorban az adott mennyiségnek az adott szövegtípus esetében vizsgált összes szövegmennyiségben mért százalékos előfordulási gyakoriságát adom meg két tizedesvessző pontossággal (%).

6.1 Szövegtípus és nyelvpár szerinti összesítések

Az előző pontban bemutatott táblázati struktúra előnyei közé tartozik, hogy az alkorpuszonkénti eredményeket szövegtípus és nyelvpár szerinti összesítések szerint tárolja. Ebben a fejezetben mutatom be táblázatos formában elrendezve a forrásnyelvek szerint összesített eredményeket, amelyek a részkorpuszok részben eltérő méretei miatt nem hasonlíthatók minden esetben össze. Az azonos méretű részkorpuszok összehasonlító adatait majd a következő alfejezetben ismertetem.

AUTENTIKUS		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
EU	ÖSSZES	249 315	19 018	1 766	249	5 471	459	164	3	5
	%		7,63%	0,71%	0,10%	2,19%	0,18%	0,07%	0,00%	0,00%
IRODALOM	ÖSSZES	255 699	22 094	1 427	956	5 223	941	334	19	0
	%		8,64%	0,56%	0,37%	2,04%	0,37%	0,13%	0,01%	0,00%
TUDOMÁNYOS	ÖSSZES	257 442	45 317	1 717	241	8 632	1 691	731	73	4
	%		17,60%	0,67%	0,09%	3,35%	0,66%	0,28%	0,03%	0,00%
VALLÁSI	ÖSSZES	245 009	41 262	2 886	789	6 760	1 269	358	21	3
	%		16,84%	1,18%	0,32%	2,76%	0,52%	0,15%	0,01%	0,00%

7. TÁBLÁZAT: Az autentikus magyar szövegek kohéziós elemeinek száma

Az 7. TÁBLÁZAT az autentikus magyar szövegkorpusz kohéziós elemeinek számát tartalmazza közéleti (EU), szépirodalmi (IROD), tudományos (TUD) és vallásos (VAL) szövegtípusonkénti bontásban. A táblázat sorai tartalmazzák a szövegtípusokra, míg oszlopai a kohéziós kapcsolatokra jellemző adatokat: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma (M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D).

FORDÍTOTT ANGOL		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
ANGOL EU	ÖSSZES	249 420	21 370	735	240	2 603	198	957	4	0
	%		8,57%	0,29%	0,10%	1,04%	0,08%	0,38%	0,00%	0,00%
ANGOL IRODALOM	ÖSSZES	278 909	23 443	1 538	575	4 169	1 077	138	17	1
	%		8,41%	0,55%	0,21%	1,49%	0,39%	0,05%	0,01%	0,00%
ANGOL TUDOMÁNYOS	ÖSSZES	187 334	49 107	818	140	3 325	213	70	10	0
	%		26,21%	0,44%	0,07%	1,77%	0,11%	0,04%	0,01%	0,00%
ANGOL VALLÁSI	ÖSSZES	169 363	13 705	1 226	250	2 555	282	167	5	0
	%		8,09%	0,72%	0,15%	1,51%	0,17%	0,10%	0,00%	0,00%

8. TÁBLÁZAT: Az angolból fordított célnyelvi magyar szövegek kohéziós elemeinek száma

A 8. TÁBLÁZAT az angol nyelvből fordított célnyelvi magyar szövegtörzs kohéziós elemeinek számát tartalmazza közéleti (EU), szépirodalmi (IROD), tudományos (TUD) és vallásos (VAL) szövegtípusonkénti bontásban. A táblázat sorai tartalmazzák a szövegtípusokra, míg oszlopai a kohéziós kapcsolatokra – tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma (M), hiponimák száma (H), category domain relációk száma (C_D), be in state relációk száma (B_I_S) és region domain relációk száma (R_D) – jellemző adatokat.

FORDÍTOTT FRANCIA		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
FRANCIA IRODALOM	ÖSSZES	257 731	12 392	897	396	2 617	1 460	222	14	0
	%		4,81%	0,35%	0,15%	1,02%	0,57%	0,09%	0,01%	0,00%
FRANCIA TUDOMÁNYOS	ÖSSZES	280 259	33 988	1 807	464	6 006	768	477	12	1
	%		12,13%	0,64%	0,17%	2,14%	0,27%	0,17%	0,00%	0,00%

9. TÁBLÁZAT: A franciából fordított célnyelvi magyar szövegek kohéziós elemeinek száma

A 9. TÁBLÁZAT a francia nyelvből fordított célnyelvi magyar szövegtörzs kohéziós elemeinek számát tartalmazza szépirodalmi (IROD) és tudományos (TUD) szövegtípusonkénti bontásban. A táblázat sorai tartalmazzák a szövegtípusokra, míg oszlopai a kohéziós kapcsolatokra jellemző adatokat: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma

(M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D).

FORDÍTOTT NÉMET		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
NÉMET IRODALOM	ÖSSZES	260 803	14 916	909	514	2 493	708	158	9	0
	%		5,72%	0,35%	0,20%	0,96%	0,27%	0,06%	0,00%	0,00%
NÉMET TUDOMÁNYOS	ÖSSZES	257 079	31 373	2 002	618	8 452	3 995	1 065	83	0
	%		12,20%	0,78%	0,24%	3,29%	1,55%	0,41%	0,03%	0,00%
NÉMET VALLÁSI	ÖSSZES	259 514	52 878	2 318	1 147	6 781	642	271	37	0
	%		20,38%	0,89%	0,44%	2,61%	0,25%	0,10%	0,01%	0,00%

10. TÁBLÁZAT: A németből fordított célnyelvi magyar szövegek kohéziós elemeinek száma

A 10. TÁBLÁZAT a német nyelvből fordított célnyelvi magyar szövegkorpusz kohéziós elemeinek számát tartalmazza szépirodalmi (IROD), tudományos (TUD) és vallásos (VAL) szövegtípusonkénti bontásban. A táblázat sorai tartalmazzák a szövegtípusokra, míg oszlopai a kohéziós kapcsolatokra jellemző adatokat: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma (M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D).

FORDÍTOTT OLASZ		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
OLASZ IRODALOM	ÖSSZES	272 683	31 918	2 172	1 586	6 010	798	136	22	0
	%		11,71%	0,80%	0,58%	2,20%	0,29%	0,05%	0,01%	0,00%
OLASZ TUDOMÁNYOS	ÖSSZES	129 197	25 488	2 411	494	6 174	1 495	848	58	0
	%		19,73%	1,87%	0,38%	4,78%	1,16%	0,66%	0,04%	0,00%

11. TÁBLÁZAT: Az olaszból fordított célnyelvi magyar szövegek kohéziós elemeinek száma

A 11. TÁBLÁZAT az olasz nyelvből fordított célnyelvi magyar szövegkorpusz kohéziós elemeinek számát tartalmazza szépirodalmi (IROD) és tudományos (TUD) szövegtípusonkénti bontásban. A táblázat sorai tartalmazzák a szövegtípusokra, míg oszlopai a kohéziós kapcsolatokra jellemző adatokat: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma

(M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D).

FORDÍTOTT VALLÁSI		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
LATIN VALLÁSI	ÖSSZES	269 713	37 236	4 488	530	11 471	1 062	674	27	0
	%		13,81%	1,66%	0,20%	4,25%	0,39%	0,25%	0,01%	0,00%
GÖRÖG VALLÁSI	ÖSSZES	175 007	28 532	1 812	702	2 517	547	268	67	0
	%		16,30%	1,04%	0,40%	1,44%	0,31%	0,15%	0,04%	0,00%

12. TÁBLÁZAT: A latinból és görögből fordított célnyelvi magyar szövegek kohéziós elemeinek száma

A 12. TÁBLÁZAT a latin és görög nyelvből fordított vallásos (VAL) témájú célnyelvi magyar szövegek kohéziós elemeinek számát tartalmazza. A táblázat oszlopai a kohéziós kapcsolatokra jellemző adatokat tartalmazzák: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma (M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D).

Az itt bemutatott táblázatok mindegyike alátámasztja és a százalékos értékeknek köszönhetően jól szemlélteti az autentikus magyar szövegeket tartalmazó tesztkorpuszon már bemutatott jelenséget, miszerint az azonos forrásnyelvi szövegek esetében a célnyelvi korpuszban a különböző szövegtípusoknál a kohéziós elemek gyakorisága és eloszlásának az aránya különböző. Az eredmények alapján továbbá az is jól megfigyelhető, hogy mind az autentikus magyar szövegprodukciónál, mind a célnyelvi magyar szövegprodukciónál elhanyagolható mind a *region domain*, mind a *be in state* relációk száma, mivel a két kategóriára jellemző adatok egyik részkorpusz esetében sem haladják meg a 0,1%-os előfordulási gyakoriságot.

6.2 A shift, az eltolódás kimutatása

Az előző fejezetben bemutatott forrásnyelvek szerint összesített eredményeket most az adatok könnyebb egybevetethetősége, illetve az eltolódásjelenség kimatathatóságának érdekében szövegtípusonkénti bontásban vizsgálom.

6.2.1 Az eltolódás kimutatása a közéleti szövegek esetében

A közéleti szövegek esetében csak autentikus magyar szövegek valamint az angolból fordított célnyelvi magyar szövegek adatait lehet a rendelkezésre álló szövegállomány alapján összehasonlítani.

EU		TOKENSZÁM	R	S	A	H	M	C_D	B_I_S	R_D
AUTENTIKUS	ÖSSZES	249 315	19 018	1 766	249	5 471	459	164	3	5
	%		7,63%	0,71%	0,10%	2,19%	0,18%	0,07%	0,00%	0,00%
ANGOL	ÖSSZES	249 420	21 370	735	240	2 603	198	957	4	0
	%		8,57%	0,29%	0,10%	1,04%	0,08%	0,38%	0,00%	0,00%

13. TÁBLÁZAT: A közéleti szövegtípusra jellemző kohéziós elemek eloszlása

A 13. TÁBLÁZAT az autentikus magyar szövegek és angolból fordított közéleti (EU) témájú célnyelvi magyar szövegek kohéziós elemeinek számát és eloszlásának arányait tartalmazza. A táblázat oszlopai a kohéziós kapcsolatokra jellemző adatokat tartalmazzák: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma (M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D). A lekérdezések tanúsága szerint a közéleti szövegek tekintetében közel azonos tokenszámú részkorpuszok esetében az angolból fordított célnyelvi magyar szövegek kohéziós elemeinek száma és eloszlásának aránya eltér az autentikus magyar szövegek esetében tapasztalható kohéziós eszközhasználatától. Míg az autentikus magyar közéleti szövegekre kevesebb ismétlés, ezzel szemben arányaiban nagyobb számú szinonima és hiponima kapcsolat a jellemző, addig az angolból fordított célnyelvi közéleti magyar szövegek esetében magasabb az ismétlések száma és kisebb a szinonima és hiponima kapcsolatok száma.

6.2.2 Az eltolódás kimutatása az szépirodalmi szövegek esetében

Az szépirodalmi szövegek esetében az autentikus magyar szövegekre valamint az angolból, franciából, németből és olaszról fordított célnyelvi magyar szövegekre jellemző adatokat lehet a rendelkezésre álló szövegállomány alapján összehasonlítani.

AUTENTIKUS		TOKENSZÁM	R	S	A	H	M	C_D	B_I_S	R_D
IRODALOM	ÖSSZES	255 699	22 094	1 427	956	5 223	941	334	19	0
	%		8,64%	0,56%	0,37%	2,04%	0,37%	0,13%	0,01%	0,00%

14. TÁBLÁZAT: A szépirodalmi szövegtípusra jellemző kohéziós eszközhasználat az autentikus magyar szövegek esetében

FORDÍTOTT IRODALOM		TOKENSZÁM	R	S	A	H	M	C_D	B_I_S	R_D
ANGOL IRODALOM	ÖSSZES	278 909	23 443	1 538	575	4 169	1 077	138	17	1
	%		8,41%	0,55%	0,21%	1,49%	0,39%	0,05%	0,01%	0,00%
FRANCIA IRODALOM	ÖSSZES	257 731	12 392	897	396	2 617	1 460	222	14	0
	%		4,81%	0,35%	0,15%	1,02%	0,57%	0,09%	0,01%	0,00%
NÉMET IRODALOM	ÖSSZES	260 803	14 916	909	514	2 493	708	158	9	0
	%		5,72%	0,35%	0,20%	0,96%	0,27%	0,06%	0,00%	0,00%
OLASZ IRODALOM	ÖSSZES	272 683	31 918	2 172	1 586	6 010	798	136	22	0
	%		11,71%	0,80%	0,58%	2,20%	0,29%	0,05%	0,01%	0,00%

15. TÁBLÁZAT: A szépirodalmi szövegtípusra jellemző kohéziós eszközhasználat a célnyelvi magyar szövegek esetében

A 14. TÁBLÁZAT az autentikus magyar szépirodalmi szövegek a 15. TÁBLÁZAT pedig a célnyelvi – angolból, franciából, németből és olaszból fordított – szépirodalmi magyar szövegek kohéziós elemeinek számát és eloszlásának arányait tartalmazza. A táblázat oszlopai a kohéziós kapcsolatokra jellemző adatokat tartalmazzák: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma (M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D). A lekérdezések tanúsága szerint az szépirodalmi szövegek tekintetében a hasonló tokenszámú részkorpuszok esetében a célnyelvi magyar szövegek kohéziós elemeinek száma és eloszlásának aránya mindegyik forrásnyelv szerinti részkorpusz esetében eltér az autentikus magyar szövegek esetében tapasztalható kohéziós eszközhasználattól. Az adatok azt a feltételezést is alátámasztják, hogy az egyes forrásnyelvek különbözőképpen befolyásolják a célnyelvi magyar szépirodalmi szövegekben a kohéziós elemek disztribúcióját. Az autentikus magyar szépirodalmi szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva az angolból fordított célnyelvi szépirodalmi magyar szövegek esetében nagyobb eltérés a hiponima viszonyok valamint a *category domain* relációk esetében tapasztalható. A franciából és a németből fordított célnyelvi szépirodalmi magyar szövegek esetében

az autentikus magyar szépirodalmi szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva mindkét részkorpusz esetében (bár a forrásnyelvek viszonylatában is különböző mértékekben) csökken az ismétlések, szinonima, antonima, és hiponima kapcsolatok számaránya. Ezzel szemben az olaszból fordított célnyelvi szépirodalmi magyar szövegekre az autentikus magyar szépirodalmi szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva arányaiban nagyobb számú ismétlés, szinonima, antonima és hiponima kapcsolat a jellemző, viszont kevésbé jellemző a meronima kapcsolat valamint a *category domain* relációk száma.

6.2.3 Az eltolódás kimutatása a tudományos szövegek esetében

A tudományos szövegek esetében az autentikus magyar szövegekre valamint az angolból, franciából, németből és olaszból fordított célnyelvi magyar szövegekre jellemző adatokat lehet a rendelkezésre álló szövegállomány alapján összehasonlítani. A 14. TÁBLÁZAT az autentikus magyar tudományos szövegek a 15. TÁBLÁZAT pedig a célnyelvi – angolból, franciából, németből és olaszból fordított – tudományos magyar szövegek kohéziós elemeinek számát és eloszlásának arányait tartalmazza. A táblázat oszlopai a kohéziós kapcsolatokra jellemző adatokat tartalmazzák: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma (M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D).

AUTENTIKUS		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
TUDOMÁNYOS	ÖSSZES	257 442	45 317	1 717	241	8 632	1 691	731	73	4
	%		17,60%	0,67%	0,09%	3,35%	0,66%	0,28%	0,03%	0,00%

16. TÁBLÁZAT: A tudományos szövegtípusra jellemző kohéziós eszközhasználat az autentikus magyar szövegek esetében

FORDÍTOTT TUDOMÁNYOS		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
ANGOL TUDOMÁNYOS	ÖSSZES	187334	49107	818	140	3325	213	70	10	0
	%		26,21%	0,44%	0,07%	1,77%	0,11%	0,04%	0,01%	0,00%
FRANCIA TUDOMÁNYOS	ÖSSZES	280259	33988	1807	464	6006	768	477	12	1
	%		12,13%	0,64%	0,17%	2,14%	0,27%	0,17%	0,00%	0,00%
NÉMET TUDOMÁNYOS	ÖSSZES	257079	31373	2002	618	8452	3995	1065	83	0
	%		12,20%	0,78%	0,24%	3,29%	1,55%	0,41%	0,03%	0,00%
OLASZ TUDOMÁNYOS	ÖSSZES	129197	25488	2411	494	6174	1495	848	58	0
	%		19,73%	1,87%	0,38%	4,78%	1,16%	0,66%	0,04%	0,00%

17. TÁBLÁZAT: A tudományos szövegtípusra jellemző kohéziós eszközhasználat a célnyelvi magyar szövegek esetében

A lekérdezések tanúsága szerint a tudományos szövegtípus tekintetében a részben hasonló, részben különböző tokenszámú részkorpuszok esetében a célnyelvi magyar szövegek kohéziós elemeinek száma és eloszlásának aránya mindegyik forrásnyelv szerinti részkorpusz esetében eltér az autentikus magyar szövegek esetében tapasztalható kohéziós eszközhasználatától. Az eltérő tokenszámú részkorpuszok esetében az össztokenszámra vetített előfordulási arányok szolgálhatnak az összehasonlítás alapjául. Az adatok azt a feltételezést is alátámasztják, hogy az egyes forrásnyelvek különbözőképpen befolyásolják a célnyelvi magyar tudományos szövegekben a kohéziós elemek disztribúcióját.

Az angolból fordított célnyelvi tudományos magyar szövegek esetében nagyobb az ismétlések, viszont kisebb a szinonima-, antonima-, hiponima- és meronima-kapcsolatok valamint a category domain relációk számaránya az autentikus magyar tudományos szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva. A franciából fordított célnyelvi tudományos szövegek esetében az autentikus magyar tudományos szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva csökken az ismétlések, az antonima, a hiponima és a meronima kapcsolatok valamint a category domain relációk

számaránya. A németből fordított célnyelvi tudományos szövegek esetében az autentikus magyar tudományos szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva csökken az ismétlések és a hiponima kapcsolatok számaránya, viszont növekszik a szinonima, az antonima és a meronima kapcsolatok valamint a category domain relációk számaránya. Ezzel szemben az olaszból fordított célnyelvi tudományos magyar szövegekre az autentikus magyar tudományos szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva arányaiban nagyobb számú ismétlés, szinonima-, antonima-, hiponima- és meronima-kapcsolat valamint category domain reláció a jellemző.

6.2.4 Az eltolódás kimutatása a vallási szövegek esetében

A vallási szövegek esetében az autentikus magyar szövegekre valamint az angolból, németből, latinból és görögből fordított célnyelvi magyar szövegekre jellemző adatokat lehet a rendelkezésre álló szövegállomány alapján összehasonlítani.

AUTENTIKUS		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
VALLÁSI	ÖSSZES	245 009	41 262	2 886	789	6 760	1 269	358	21	3
	%		16,84%	1,18%	0,32%	2,76%	0,52%	0,15%	0,01%	0,00%

18. TÁBLÁZAT: A vallásos szövegtípusra jellemző kohéziós eszközhasználat az autentikus magyar szövegek esetében

FORDÍTOTT VALLÁSI		TOKENEK	R	S	A	H	M	C_D	B_I_S	R_D
ANGOL VALLÁSI	ÖSSZES	169 363	13 705	1 226	250	2 555	282	167	5	0
	%		8,09%	0,72%	0,15%	1,51%	0,17%	0,10%	0,00%	0,00%
NÉMET VALLÁSI	ÖSSZES	259 514	52 878	2 318	1 147	6 781	642	271	37	0
	%		20,38%	0,89%	0,44%	2,61%	0,25%	0,10%	0,01%	0,00%
LATIN VALLÁSI	ÖSSZES	269 713	37 236	4 488	530	11 471	1 062	674	27	0
	%		13,81%	1,66%	0,20%	4,25%	0,39%	0,25%	0,01%	0,00%
GÖRÖG VALLÁSI	ÖSSZES	175 007	28 532	1 812	702	2 517	547	268	67	0
	%		16,30%	1,04%	0,40%	1,44%	0,31%	0,15%	0,04%	0,00%

19. TÁBLÁZAT: A vallásos szövegtípusra jellemző kohéziós eszközhasználat a célnyelvi magyar szövegek esetében

A 18. TÁBLÁZAT az autentikus magyar vallási szövegek a 19. TÁBLÁZAT pedig a célnyelvi – angolból, franciából, németből és olaszból fordított – vallásos magyar szövegek kohéziós elemeinek számát és eloszlásának arányait tartalmazza. A táblázat oszlopai a kohéziós kapcsolatokra jellemző adatokat tartalmazzák: tokenszám (TOKENEK), ismétlések száma (R), szinonimák száma (S), antonimák száma (A), meronimák száma (M), hiponimák száma (H), *category domain* relációk száma (C_D), *be in state* relációk száma (B_I_S) és *region domain* relációk száma (R_D).

A lekérdezések tanúsága szerint a vallásos szövegtípus tekintetében a részben hasonló, részben különböző tokenszámú részkorpuszok esetében a célnyelvi magyar szövegek kohéziós elemeinek száma és eloszlásának aránya mindegyik forrásnyelv szerinti részkorpusz esetében eltér az autentikus magyar vallási szövegek esetében tapasztalható kohéziós eszközhasználatától. Az eltérő tokenszámú részkorpuszok esetében az össztokenszámmra vetített előfordulási arányok szolgálhatnak az összehasonlítás alapjául. Az adatok azt a feltételezést is alátámasztják, hogy az egyes forrásnyelvek különbözőképpen befolyásolják a célnyelvi magyar vallási szövegekben a kohéziós elemek disztribúcióját.

Az angolból fordított célnyelvi vallásos magyar szövegek esetében arányaiban csökken az ismétlések, a szinonima-, antonima-, hiponima- és meronima-kapcsolatok valamint a *category domain* relációk száma az autentikus magyar vallási szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva. A németből fordított célnyelvi vallási szövegek esetében az autentikus magyar vallási szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva növekszik az ismétlések és az antonima kapcsolatok számaránya, viszont csökken a szinonima-, a hiponima- és a meronima-kapcsolatok valamint a *category domain* relációk számaránya. A latinból fordított célnyelvi vallási szövegek esetében az autentikus magyar vallási szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva csökken az ismétlések és a meronima kapcsolatok számaránya, viszont növekszik a szinonima-, az antonima- és a hiponima-kapcsolatok valamint a *category domain* relációk számaránya. A görögből fordított célnyelvi vallásos magyar szövegekre az autentikus magyar vallási szövegekre jellemző kohéziós kapcsolatok számarányához viszonyítva arányaiban nagyobb számú antonima kapcsolat, viszont kevesebb ismétlés, szinonima, hiponima és meronima kapcsolat, valamint a *category domain* reláció a jellemző.

IV. Az eredmények összegzése, következtetések

Az értekezés alkalmazott nyelvészeti megközelítésből foglalkozik a fordításokkal, a célnyelvi szövegek szövegépítő sajátosságainak feltárásával. A szövegek komplexebb vizsgálatát tűzi célul maga elé, amit egyrészt a szövegnyelvészet és a korpusznyelvészet eszközeivel kvantitatív módszerekkel vizsgál és 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 szöveggörnyezeti tényezők és a nyelvhasználat szabályszerűségeinek feltárásán keresztül. Az elemzés főbb területei az autentikus és célnyelvi szövegek kohéziós eszközhasználatának összevetése, a kohéziós eszközhasználat szempontjából a nyelvpár specifikus fordítói magatartás feltárása, a forrásnyelvi és célnyelvi magyar szövegek szövegkohéziós eltolódásainak megragadása és a fordított szöveget az autentikus szövegprodukciónól megkülönböztető kohéziós eszközhasználatra jellemző fordítási univerzálé feltárása. 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 alapján végeztem, mintegy 4 millió szövegszót tartalmazó, többféle műfaji és szövegtipológiai csoportba sorolható magyar nyelven keletkezett autentikus és magyarra fordított célnyelvi szöveg mondathatárokra túlmutató szövegépítő sajátosságait elemeztem. Az elemzett elbeszélő, leíró és érvelő szövegek a közéleti (hivatalos), egyházi, tudományos és szépirodalmi szövegtípusokba tartoznak. A Magyar Lexikai Kohéziós Projektben a kohéziós elemek identifikációját a projekt céljainak megfelelően és érdekében Orosz György és Laki László által (PPKE ITK) összekapcsolt nyelvtechnológiai eszközök (Laki 2011, Novák, Orosz, Indig 2011) automatizálták,

melyek közül a magyar WordNet Program (Prószéky, Mihaltz 2008) segítette a szemantikai kapcsolatok feltárását.

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. Elképzeléseim 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 előző fejezetben 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 szövegprodukcióna jellemző kohéziós eszközhasználattal.

A III. 5.2 fejezetben bemutattam, hogy az autentikus magyar és a célnyelvi magyar szövegeknél a vizsgált négy szövegtípus esetében az autentikus magyar szövegprodukcióna 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.

A III. 5.2 fejezetben 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.

A III. 5.2 fejezetben 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.

ⁱ Molnár P. (szerk.) 2000. *Az unió kapujában*. Budapest: Mundus Magyar Egyetemi Kiadó. Elérhető: mek.oszk.hu/01900/01989

ⁱⁱ Ballai L. 2000. A kitagadott. Elérhető: www.vasaros.com/art/ballai/

ⁱⁱⁱ Bonfini, A. 1581. *Rerum Ungaricarum decades*. In: S. Varga K. (szerk.) 1995. *A magyar történelem tizedei*. Budapest: Balassi Kiadó. Ford: Kulcsár P. Elérhető: <http://www.tankonyvtar.hu/tortenelem/magyar-tortenelem-080903-5>

^{iv} XVI. Benedek pápa. 2009. *Caritas in veritate kezdetű enciklikája*.

Elérhető: http://www.vatican.va/holy_father/benedict_xvi/encyclicals/documents/hf_ben-xvi_enc_20090629_caritas-in-veritate_hu.html

- ^v Bíró A.-M. (szerk.) 2002. *Kisebbségi Adattár*. Budapest: Teleki László Alapítvány.
Elérhető: mek.oszk.hu/05900/05981/05981.pdf
- ^{vi} Európai Munkahelyi Biztonsági és Egészségvédelmi Ügynökség. 2007. 06. 04. *Legyen könnyebb a teher – a váz- és izomrendszeri zavarok kezelése*.
Elérhető: http://osha.europa.eu/hu/news/news_article_EW2007_Launch
- ^{vii} Európai Unió sajtóügynökség. 2007. 07.18. A Bizottság fellép a férfiak és nők közötti fizetéskülönbség csökkentése érdekében.
Elérhető:
<http://europa.eu/rapid/pressReleasesAction.do?reference=IP/07/1115&format=HTML&aged=1&language=HU&guiLanguage=fr>
- ^{viii} Osztoivits L. (szerk.) 1968. *Tájfűn. Klasszikus angol kisregények*. Budapest: Európa Kiadó.
Elérhető: <http://mek.niif.hu/03600/03689/html/02.htm>
- ^{ix} Németh L. 1988. *Négy könyv*. Budapest: Szépirodalmi Kvk.
Elérhető: mek.oszk.hu/01000/01014/
- ^x A Magyar Katolikus Püspöki Konferencia körlevele. 1999. *A boldogabb családokért*. Budapest. Elérhető: <http://uj.katolikus.hu/nyomtat-konyvtar.php?h=157>
- ^{xi} II. Vatikáni Zsinat. 1965. *Gaudium et spes* kezdetű lelkipásztori konstitúciója. Az Egyház a mai világban.
Elérhető: http://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19651207_gaudium-et-spes_hu.html

Ábrák jegyzéke

1. ÁBRA:	A PILOT PROJEKT ANNOTÁLT SZÖVEGRÉSZLETE	17
2. ÁBRA:	A MONDATOK BEJELÖLÉSE ÉS SORSZÁMOZÁSA A PILOT PROJEKTBEN	18
3. ÁBRA:	AZ ELSŐ MONDAT ISMÉTLŐDŐ LEXIKAI EGYSÉGEI A PILOT PROJEKT EGYIK SZÖVEGÉNÉL	18
4. ÁBRA:	AZ LC-HU KORPUSZ HEADER ANNOTÁCIÓJA	111
5. ÁBRA:	AZ LC-HU KORPUSZ ANNOTÁLT SZÖVEGRÉSZLETE	112
6. ÁBRA:	AZ LC-HU KORPUSZ LEGFELSŐ FOLDERSZINTJE	116
7. ÁBRA:	AZ LC-HU KORPUSZ FOLDERSTUKTÚRÁJA	117
8. ÁBRA:	RÉSZLET AZ UNIÓ KAPUJÁBAN CÍMŰ DOKUMENTUMBÓL	130
9. ÁBRA:	PÉLDA AZ ELŐZŐ BEKEZDÉS EGYIK MONDATÁNAK MORFOLÓGIAI ANNOTÁLÁSÁRA	131
10. ÁBRA:	A 8. ÁBRA SZÖVEGRÉSZLETÉBEN TALÁLHATÓ KOHÉZIÓS KAPCSOLATOK	132
11. ÁBRA:	SZÖVEGRÉSZLET A KITAGADOTT CÍMŰ REGÉNYBŐL	133
12. ÁBRA:	A 11. ÁBRA SZÖVEGRÉSZLETÉBEN TALÁLHATÓ KOHÉZIÓS KAPCSOLATOK	133
13. ÁBRA:	SZÖVEGRÉSZLET A MAGYAR TÖRTÉNELEM TIZEDEI MŰBŐL	134
14. ÁBRA:	A 13. ÁBRA SZÖVEGRÉSZLETÉBEN TALÁLHATÓ KOHÉZIÓS KAPCSOLATOK	134
15. ÁBRA:	SZÖVEGRÉSZLETE A CARITAS IN VERITATE KEZDETŰ ENCIKLIKÁBÓL	135
16. ÁBRA:	A 15. ÁBRA SZÖVEGRÉSZLETÉBEN TALÁLHATÓ KOHÉZIÓS KAPCSOLATOK	135
17. ÁBRA:	PÉLDÁK AZ UNIÓ KAPUJÁBAN CÍMŰ SZÖVEGBŐL	136
18. ÁBRA:	PÉLDÁK A KISEBBSÉGI ADATTÁR CÍMŰ SZÖVEGBŐL	136
19. ÁBRA:	PÉLDÁK AZ IP-07-752 AZONOSÍTÓJÚ EU-S DOKUMENTUM SZÖVEGÉBŐL	136
20. ÁBRA:	PÉLDÁK AZ IP-07-1115 AZONOSÍTÓJÚ EU-S DOKUMENTUM SZÖVEGÉBŐL	136
21. ÁBRA:	PÉLDÁK A KITAGADOTT CÍMŰ REGÉNYBŐL	136
22. ÁBRA:	PÉLDÁK A TÁJFUN CÍMŰ ANTOLÓGIÁBÓL	137
23. ÁBRA:	PÉLDÁK A NÉGY KÖNYV CÍMŰ GYŰJTEMÉNYBŐL	137
24. ÁBRA:	PÉLDÁK A MAGYAR TÖRTÉNELEM TIZEDEI CÍMŰ MŰBŐL	137
25. ÁBRA:	PÉLDÁK A BOLDOGABB CSALÁDOKÉRT KÖRLEVÉLBŐL	138
26. ÁBRA:	PÉLDÁK A GAUDIUM ET SPES KEZDETŰ LELKIPÁSZTORI KONSTITÚCIÓBÓL	138
27. ÁBRA:	PÉLDÁK AZ UNIÓ KAPUJÁBAN CÍMŰ SZÖVEGBŐL	138
28. ÁBRA:	PÉLDÁK A KISEBBSÉGI ADATTÁR CÍMŰ SZÖVEGBŐL	139
29. ÁBRA:	PÉLDÁK A KITAGADOTT CÍMŰ REGÉNYBŐL	139
30. ÁBRA:	PÉLDÁK A TÁJFUN CÍMŰ ANTOLÓGIÁBÓL	139
31. ÁBRA:	PÉLDÁK A NÉGY KÖNYV CÍMŰ GYŰJTEMÉNYBŐL	140
32. ÁBRA:	PÉLDÁK A MAGYAR TÖRTÉNELEM TIZEDEI CÍMŰ MŰBŐL	140
33. ÁBRA:	PÉLDÁK A BOLDOGABB CSALÁDOKÉRT KÖRLEVÉLBŐL	140
34. ÁBRA:	PÉLDÁK A GAUDIUM ET SPES KEZDETŰ LELKIPÁSZTORI KONSTITÚCIÓBÓL	141

35. ÁBRA: PÉLDÁK AZ UNIÓ KAPUJÁBAN CÍMŰ SZÖVEGBŐL	141
36. ÁBRA: PÉLDÁK AZ IP-07-752 AZONOSÍTÓJÚ EU-S DOKUMENTUM SZÖVEGÉBŐL	141
37. ÁBRA: PÉLDÁK A KITAGADOTT CÍMŰ REGÉNYBŐL	142
38. ÁBRA: PÉLDÁK A TÁJFUN CÍMŰ ANTOLÓGIÁBŐL	142
39. ÁBRA: PÉLDÁK A NÉGY KÖNYV CÍMŰ GYŰJTEMÉNYBŐL	142
40. ÁBRA: PÉLDÁK A MAGYAR TÖRTÉNELEM TIZEDEI CÍMŰ MŰBŐL	143
41. ÁBRA: PÉLDÁK A BOLDOGABB CSALÁDOKÉRT KÖRLEVÉLBŐL	143
42. ÁBRA: PÉLDÁK A GAUDIUM ET SPES KEZDETŰ LELKIPÁSZTORI KONSTITÚCIÓBŐL	143

Táblázatok jegyzéke

1. TÁBLÁZAT :	AZ AUTENTIKUS MAGYAR SZÖVEGEKEN MÉRT TESZTADATOK	114
2. TÁBLÁZAT:	AZ AUTENTIKUS MAGYAR SZÖVEGEKET TARTALMAZÓ ALKORPUSZ ÖSSZETÉTELE	122
3. TÁBLÁZAT:	A CÉLNYELVI MAGYAR SZÖVEGEKET TARTALMAZÓ ALKORPUSZ ÖSSZETÉTELE	122
4. TÁBLÁZAT :	AZ AUTENTIKUS MAGYAR SZÖVEGEKET TARTALMAZÓ ALKORPUSZBAN FELTÁRT LEXIKAI KOHÉZIÓS KAPCSOLATOK SZÁMA	128
5. TÁBLÁZAT:	A FORDÍTOTT MAGYAR SZÖVEGEKET TARTALMAZÓ ALKORPUSZBAN FELTÁRT LEXIKAI KOHÉZIÓS KAPCSOLATOK SZÁMA	128
6. TÁBLÁZAT:	RÉSZLET AZ EREDMÉNYEKET ARCHIVÁLÓ TÁBLÁZATBÓL	144
7. TÁBLÁZAT:	AZ AUTENTIKUS MAGYAR SZÖVEGEK KOHÉZIÓS ELEMEINEK SZÁMA	145
8. TÁBLÁZAT:	AZ ANGOLBÓL FORDÍTOTT CÉLNYELVI MAGYAR SZÖVEGEK KOHÉZIÓS ELEMEINEK SZÁMA	146
9. TÁBLÁZAT:	A FRANCIÁBÓL FORDÍTOTT CÉLNYELVI MAGYAR SZÖVEGEK KOHÉZIÓS ELEMEINEK SZÁMA	146
10. TÁBLÁZAT:	A NÉMETBŐL FORDÍTOTT CÉLNYELVI MAGYAR SZÖVEGEK KOHÉZIÓS ELEMEINEK SZÁMA	147
11. TÁBLÁZAT:	AZ OLASZBÓL FORDÍTOTT CÉLNYELVI MAGYAR SZÖVEGEK KOHÉZIÓS ELEMEINEK SZÁMA	147
12- TÁBLÁZAT:	A LATINBÓL ÉS GÖRÖGBŐL FORDÍTOTT CÉLNYELVI MAGYAR SZÖVEGEK KOHÉZIÓS ELEMEINEK SZÁMA	148
13. TÁBLÁZAT:	A KÖZÉLETI SZÖVEGTÍPUSRA JELLEMZŐ KOHÉZIÓS ELEMEL ELOSZLÁSA	149
14. TÁBLÁZAT:	A SZÉPIRODALMI SZÖVEGTÍPUSRA JELLEMZŐ KOHÉZIÓS ESZKÖZHASZNÁLAT AZ AUTENTIKUS MAGYAR SZÖVEGEK ESETÉBEN	150
15. TÁBLÁZAT:	A SZÉPIRODALMI SZÖVEGTÍPUSRA JELLEMZŐ KOHÉZIÓS ESZKÖZHASZNÁLAT A CÉLNYELVI MAGYAR SZÖVEGEK ESETÉBEN	150
16. TÁBLÁZAT:	A TUDOMÁNYOS SZÖVEGTÍPUSRA JELLEMZŐ KOHÉZIÓS ESZKÖZHASZNÁLAT AZ AUTENTIKUS MAGYAR SZÖVEGEK ESETÉBEN	152
17. TÁBLÁZAT:	A TUDOMÁNYOS SZÖVEGTÍPUSRA JELLEMZŐ KOHÉZIÓS ESZKÖZHASZNÁLAT A CÉLNYELVI MAGYAR SZÖVEGEK ESETÉBEN	152
18. TÁBLÁZAT:	A VALLÁSOS SZÖVEGTÍPUSRA JELLEMZŐ KOHÉZIÓS ESZKÖZHASZNÁLAT AZ AUTENTIKUS MAGYAR SZÖVEGEK ESETÉBEN	153
19. TÁBLÁZAT:	A VALLÁSOS SZÖVEGTÍPUSRA JELLEMZŐ KOHÉZIÓS ESZKÖZHASZNÁLAT A CÉLNYELVI MAGYAR SZÖVEGEK ESETÉBEN	153

Irodalom

- Aaltonen, R. 2003. *From Quasi-Correctness to Quasi-Neutrality? A critical study of eleven translations of Vieraskynä columns published in Helsingin Sanomat in 2003*. HTÁBLÁZATelsinki. Doctoral Thesis.
- Aligned Hansards of the 36th Parliament of Canada*. Online elérés:
<http://www.isi.edu/natural-language/download/hansard/>
- Baker, M. 1992. *In Other Words. A coursebook on translation*. London: Routledge.
- Baker, M. 1993. Corpus Linguistics and Translation Studies. Implications and Applications. In: Baker, M., Francis, G., Tognini-Bonelli, E. (eds). 1993. *Text and Technology: In Honour of John Sinclair*. Amsterdam & Philadelphia: John Benjamins Publishing Company. 233-250.
- Baker, M. 1995. Corpora in Translation Studies. An Overview and Some Suggestions for Future Research. *Target*. 7/2. 223–243.
- Baker, M. 1996. Corpus-based Translation Studies: The Challenges that Lie Ahead. In: Harold, S. (ed). 1996. *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Baker, M. 2000. Towards a methodology for investigating the style of a literary translator. *Target*. 12/2. 241–266.
- Balaskó M. 2005. *Korpusnyelvészeti vizsgálatok és fordításnyelvi minták (angol és magyar tudományos szövegek anyaga alapján)*. PhD értekezés
- Balaskó M. 2007. A fordításnyelvről, avagy a flamand szónyeg láthatatlan szálairól. In: Heltai P. (szerk.) 2007. *Nyelvi modernizáció. Szaknyelv, fordítás, terminológia. A XVI. MANYE Kongresszus előadásai*. Gödöllő. 2006. április 10–12. (A MANYE Kongresszusok előadásai 3.) Pécs–Gödöllő: MANYE–Szent István Egyetem. 159–166.
- Bañcerowski J. 1999. A kognitív nyelvészet alapelvei. *Magyar Nyelvőr*. 123. évf. 1. szám. 78–87.
- Bassnett, S. 2003. Judging translation. *Wasafiri*. 18/40. 42–44.

- Beaugrand, R. d., Dressler, W. 2000. *Bevezetés a szövegnyelvészetbe*. Budapest: Corvina.
- Békési I. 1979. A kapcsolás problematikája. In: Kornya L., Czellerné Farkas M. (szerk.) 1996. *Magyar nyelvészeti cikk- és tanulmánygyűjtemény szakfordítóknak. II. Nyelvművelési és szakmai nyelvhasználat*. Debrecen: Kossuth Lajos Tudományegyetem Idegennyelvi Központ. 56–58.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Blum-Kulka, Sh. 1986. Shift of Cohesion and Coherence in Translation. In: Blum-Kulka, Sh. – Hause, J. (eds.) 1986. *Interlingual and Intercultural Communication*. Tübingen: Narr. 17–35.
- British National Corpus*. Online elérés: www.natcorp.ox.ac.uk
- Brown Corpus*. Online elérés: <http://khnt.hit.uib.no/icame/manuals/brown/INDEX.HTM>
- Budin, G. 2002. Wissensmanagement in der Translation. In: Best, J., Kalina, S. (Hrsg.) 2002. *Übersetzen und Dolmetschen*. Tübingen, Basel: Franke. 74–84.
- Canisius, P., Knipf, E. 1996. *Textgrammatik: Ein Einführung. Ein Lehr- und Arbeitsbuch für Anfänger*. Budapest: Nemzeti Tankönyv Kiadó.
- Catford, J. C. 1965. *A Linguistic Theory of Translation*. London: Oxford University Press.
- Cheng, W. 2006. Describing the extended meanings of lexical cohesion in a corpus of SARS spoken discourse. In: Flowerdew, J., Mahlberg, M. 2006. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 65–84.
- Corpus of Contemporary American English*. Online elérés: www.americancorpus.org
- Crystal, D. 1998. *A nyelv enciklopédiája*. Budapest: Osiris.
- Csatár P. Haase Zs. 2011. Koherencia és kohézió — Reflexiók e két fogalom értelmezésére a német nyelvű szakirodalom alapján. In: Dobi E. (szerk.) *A szövegösszefüggés elméleti és gyakorlati megközelítési módjai. Poliglott terminológiai és fogalmi áttekintés*. [Officina 16.] (Petőfi S. János 80. születésnapjára). Debrecen: Debreceni Egyetemi Kiadó. 57–81.

- Danielsson, P., Ridings, D. 1996. *Pedant: Parallel Texts in Göteborg. (Research Reports from the Department of Swedish, GU-ISS-96-2.)* Göteborg University: Spraúkdata.
- Digitális Irodalmi Akadémia.* Online elérés: <http://pim.hu/object.a76a23fd-564b-4b04-8a9b-5a1688afb4fa.ivy>
- Doherty, M. 1999. Clefts in Translations between English and German. *Target* 11/2. Amsterdam & Philadelphia: John Benjamins Publishing Company. 289–315
- Dressler, W. U. 1973. *Einführung in die Textlinguistik.* Tübingen: Max Niemeyer.
- Dróth J. 2002. Az értékelés mint rendszeres visszajelzés a fordítás oktatásában. In: Feketéné S. M. (szerk.) *Porta Lingua. 2002. Szaknyelvoktatásunk az EU kapujában.* Debrecen: Debreceni Egyetem. 94–100.
- Electronic Text Collections in Western European Literature.* Online elérés: http://wess.lib.byu.edu/index.php/Main_Page
- English-Norwegian Parallel Corpus.* Online elérés: <http://www.helsinki.fi/varieng/CoRD/corpora/ENPC/>
- EUR-ACCOR.* Online elérés: <http://www.cstr.ed.ac.uk/research/projects/artic/accor.html>
- EUR-Lex.* Online elérés: <http://eur-lex.europa.eu/hu/tools/about.htm>
- European Parliament Proceedings 1996-2001.* Online elérés: <http://www.statmt.org/euoparl/>
- Európai Tanács Nyelvpolitikai Osztály. 2001/2005. *Nyelvvizsgák szintillesztése a Közös Európai Referenciakerethez.* Budapest: Nyelvvizsgát Akkreditáló Testület. PH Nyelvvizsgáztatási Akkreditációs Központ.
- Fehér E. 1999. A magyar nyelvészeti szövegkutatás irányai, tárgykörei. *Magyar Nyelvőr.* 123. évf. 4. szám. 464–478.
- Fjodorov, A. V. 1953. *Vvegyenyije v tyeoriju perevoda.* Moszkva: Izd. lityeraturi na inosztrannih jazikah.
- Flowerdew, J. 2006. Use of signalling nouns in a learner corpus. In: Flowerdew, J., Mahlberg, M. 2006. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics.* Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 85–102.

- Földes Cs. 2007. „Interkulturális kommunikáció”: koncepciók, módszerek, kérdőjelek. *Fordítástudomány*. IX. évf. 1. szám. 14–39.
- Frawley, W. 1984. Prolegomenon to a Theory of Translation. In: Frawley, W. (eds.) *Translation: Literary, Linguistic, and Philosophical Perspectives*. London - Toronto: Associated University Press.
- Frege, G. 2000/1891. Fogalom és tárgy In: Máté A. (szerk) 2000. *Logikai vizsgálódások*. Budapest: Osiris.
- Gellerstam, M. 1986. Translationese in Swedish Novels Translated from English. In: Wollin, L., Lindquist, K. (eds.): *Translation Studies in Scandinavia*. Malmö: Liber Förlag. 88–95.
- Gutwinski, W. 1976. *Cohesion in Literary Texts : A Study of Some Grammatical and Lexical Features of English Discourse*. The Hague: Mouton.
- Halliday, M. A. K. 1985. *An Introduction to Functional Grammar*. London, Boston: Edward Arnold. 287–318.
- Halliday, M. A. K., Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Hansen-Schirra, S., Neumann, S., Vela, M. 2006. Multi-dimensional Annotation and Alignment in an English-German Translation Corpus. In: *Proceedings of the workshop on Multi-dimensional Markup in Natural Language Processing (NLPXML-2006) held at EACL, Trento, Italy, April 2006*. 35–42.
- Harweg, R. 1968. *Pronomina und Textkonstitution*. München: Fink.
- Hasan, R. 1984. Coherence and Cohesive Harmony. In: Flood, J. (Ed.): *Understanding reading comprehension: Cognition, language and the structure of prose*. Delaware: International Reading Association. 181–219.
- Hasselgaúrd, H., Oksefjell, S. (eds.) 1999. *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi.
- Hatim, B., Mason, I. 1990. *Discourse and the Translator*. London: Longman.
- Heltai P. 2002. A kontrasztív nyelvészet a fordítás oktatásában. In: Feketéné S. M. (szerk.) *Porta Lingua. 2002. Szaknyelvoktatásunk az EU kapujában*. Debrecen: Debreceni Egyetem. 112–116.
- Heltai P. 2004. A fordító és a nyelvi normák I. *Magyar Nyelvőr*. 128. évf. 4. szám. 407–434.

- Heltai P. 2005a. A fordító és a nyelvi normák II. *Magyar Nyelvőr*. 129. évf. 1. szám. 30–58.
- Heltai P. 2005b. A fordító és a nyelvi normák III. *Magyar Nyelvőr*. 129. évf. 2. szám. 165–172.
- Heltai P. 2009. Fordítás, relevancia, feldolgozás... *Modern Nyelvoktatás*. XV. évf. 1-2. szám. 13–28.
- Hoey, M. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.
- Holmes, J. 1972. The Name and Nature of Translation Studies. In: Holmes, J. 1988. *Translated!*. Amsterdam: Rodopi. 67–80.
- Hönig, H. 1997. *Konstruktives Übersetzen*. Tübingen: Stauffenburg.
- H. Varga Gy. (szerk.) 2007. *A verbális üzenet. Segédanyag a Nyelvi kommunikáció c. tárgyhoz*. Eger: Eszterházy Károly Főiskola.
- International Corpus of English*. Online elérés: www.ucl.ac.uk/english-usage/ice/
- Isenberg, H. 1976. Einige Grundbegriffe für eine linguistische Texttheorie. In Danes, F., Viehweger, D. (eds.) 1976. *Probleme der Textgrammatik*. Berlin: Akademie-Verlag. 47-145.
- Jakobson, R. 1959. On Linguistic Aspects of Translation. In: Brower, R. A. (ed.) *On Translation*. Cambridge: Harvard University Press. 232–239.
- Johansson, S. 1995. The approach of the Text Encoding Initiative to the encoding of spoken discourse. In: Leech, G., Myers, G., Thomas, J. (eds.) *Spoken English on Computer*. Harlow: Longman. 82-98.
- Johansson, S. 1997. Using the English-Norwegian Parallel Corpus – a Corpus for Contrastive Analysis and Translation Studies. In: Lewandowska-Tomasczyk, B., Melia, J. (eds.) 1997. *PALC '97. Practical Applications in Language Corpora*. Lodz: University of Lodz. 282–286.
- Károly K. 2002. *Lexical repetition in text. A study of the text-organizing function of lexical repetition in foreign language argumentative discourse*. Frankfurt am Main: Peter Lang.
- Károly K. 2007. *Szövegtan és fordítás*. Budapest: Akadémia Kiadó.
- Károly S. 1979. A szöveg és a jelentés szerepe a kommunikációs szemléletű nyelvészeti törekvésekben. In: Kornya L., Czellerné Farkas M. (szerk.) 1996. *Magyar nyelvészeti cikk- és tanulmánygyűjtemény szakfordítóknak. II.*

- Nyelvművelési és szakmai nyelvhasználat.* Debrecen: Kossuth Lajos Tudományegyetem Idegennyelvi Központ. 52–55.
- Kelz, H. P. 2002. Interkulturelle Kommunikation und translatorische Prozesse in der Wirtschaft. In: Best J., Kalina, S. (eds). 2002. *Übersetzen und Dolmetschen. Eine Orientierungshilfe.* Tübingen: Francke Verlag Tübingen und Basel
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics.* Harlow: Addison Wesley Longman Ltd.
- Kenny, D. 1998. Corpora in Translation Studies. In: Baker, M. (ed.) 1998. *The Routledge Encyclopedia of Translation Studies.* London: Routledge. 50–53.
- Kenny, D. 2001. *Lexis and Creativity in Translation. A Corpus-based Study.* Manchester: St. Jerome Publishing.
- Kiefer F. 1979. Szövegelmélet – szöveggrammatika –szövegnyelvészet. In: Kornya L., Czellerné Farkas M. (szerk.) 1996. *Magyar nyelvészeti cikk- és tanulmánygyűjtemény szakfordítóknak. II. Nyelvművelési és szakmai nyelvhasználat.* Debrecen: Kossuth Lajos Tudományegyetem Idegennyelvi Központ. 46–51.
- Klaudy K. 1987. Fordítás és aktuális tagolás. *Nyelvtudományi értekezések.* 123. Budapest: Akadémiai Kiadó.
- Klaudy K. 1999. *Bevezetés a fordítás gyakorlatába.* Budapest: Scholastica.
- Klaudy K. 2001. Mit tehet a fordítástudomány a magyar nyelv „korszerűsítéséért”? In: Klaudy K. (szerk.) 2007. *Nyelv és fordítás. Válogatott fordítástudományi tanulmányok.* Budapest: Tinta Könyvkiadó. 50–56.
- Klaudy K. 2004. A kommunikatív szakaszhatárok eltűnése a magyarra fordított európai uniós szövegekben. *Magyar Nyelvőr.* 128. évf. 4. szám. 389–407.
- Klaudy K. 2005. Párhuzamos korpuszok felhasználása a fordításkutatásban. In: Lanstyák I., Vančoné Kremmer, I. 2005. *Nyelvészetről - változatosan. Segédkönyv egyetemisták és a nyelvészet iránt érdeklődők számára.* Dunaszerdahely: Gramma Nyelvi Iroda. 153-185.
- Klaudy K. 2006. Szövegszintű műveletek a fordításban. In: Klaudy K. 2007. *Nyelv és fordítás. Válogatott fordítástudományi tanulmányok.* Budapest: Tinta Könyvkiadó. 138-144.

- Klaudy K. 2007. *Nyelv és fordítás. Válogatott fordítástudományi tanulmányok.* Budapest: Tinta Könyvkiadó.
- Klaudy K. 2008. Tizenöt tézis a fordítástudományról. In: Sárdi Cs. (szerk.) 2008. *Kommunikáció az információs technológia korszakában. A XVII. MANYE Kongresszus előadásai.* Siófok. 2007. április 19–21. (A MANYE Kongresszusok előadásai 4.) Pécs–Székesfehérvár: MANYE–Kodolányi János Főiskola. 41–50.
- Kocsány P. 2002. *Szöveg, szövegtípus, jelentés: a mondás mint szövegtípus.* Budapest: Akadémia Kiadó.
- Komissarov, V. N. 2002. *Sovremennoe perevodovedenie: uchebnoe posobie.* Moszkva: ÉTS.
- Kugler N., Tolcsvai N.G. 2000. *Nyelvi fogalmak kisszótára.* Budapest: Korona Kiadó. 249–250.
- La biblioteca telematica.* Online elérés: <http://www.liberliber.it/biblioteca/>
- 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.
- Laviosa, S. 1997. How Comparable Can 'Comparable Corpora' Be? *Target.* 9/2. 289–319.
- Laviosa, S. 1998. The Corpus-based Approach: A New Paradigm in Translation Studies. *Meta.* 43/4. 474–479.
- Laviosa, S. 2001. Universals of translation. In: Baker, M., Malmkjær, K. (eds.) 2001. *Routledge Encyclopedia of Translation Studies.* London: Routledge. 288–291.
- Laviosa, S. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications.* Amsterdam and Atlanta: Rodopi.
- Literatur im Netz.* Online elérés: <http://literatur-im-netz.dla-marbach.de/>
- Lindström, M. 1996. *The register of English travel brochures, with special reference to the quasi-correctness of translated brochures.* Magisterarbeit. Helsinki: Helsingin yliopisto.
- Lüdeling, A., Kytö, M. (eds.) 2008. *Corpus Linguistics An International Handbook. Volume 1.* Berlin-New York: Walter de Gruyter.
- Magyar Elektronikus Könyvtár.* Online elérés: <http://mek.oszk.hu/>

- Magyar Nemzeti Szövegtár*. Online elérés: <http://corpus.nytud.hu/mnsz/>
- Mahlberg, M. 2006. Lexical cohesion: Corpus linguistic theory and its application in English language teaching. In: Flowerdew, J., Mahlberg, M. 2006. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 103–123.
- Mauranen, A. 1996. Discourse Competence — Evidence from Thematic Development in Native and Non-Native Texts. In: Eija, V., Mauranen, A. (eds.) 1996. *Academic Writing. Intercultural and textual issues*. Amsterdam & Philadelphia: John Benjamins Publishing Company. 195–230.
- McEnery, T., Wilson, A. 2001. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, A., Xiao, Z. 2007. Parallel and comparable corpora: What is happening? In: Rogers, M., Anderman, G. (eds.). 2007. *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters. 18–32.
- Miháltz M. 2003. Magyar főnévi WordNet-ontológia létrehozása automatikus módszerekkel. In: Alexin Z., Csendes D. (szerk.) *Az 1. Magyar Számítógépes Nyelvészeti Konferencia előadásai*. Szeged: SZTE. 153–159.
- Miháltz M. 2005. Magyar EuroWordNet projekt: bemutatás és helyzetjelentés. In: Alexin Z., Csendes D. (szerk.) *A 3. Magyar Számítógépes Nyelvészeti Konferencia előadásai*. Szeged: SZTE. 68–78.
- MLCC Multilingual and Parallel Corpora*. Online elérés: www.elda.org/catalogue/en/text/W0023.html
- Mounin, G. 1963. *Les problèmes théoriques de la traduction*. Paris: Gallimard.
- Nesi, H., Basturkmen, H. 2006. Lexical bundles and discourse signalling in academic lectures. In: Flowerdew, J., Mahlberg, M. 2006. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 23–44.
- Neubert, A., Shreve, G. 1992. *Translation as Text*. Kent, Ohio: Kent State University Press.
- Neubert, A. 1994. Translatorische Relativität. Snell–Hornby, M. (Hrsg.) 1994. *Übersetzungswissenschaft. Eine Neuorientierung*. Tübingen, Basel: Francke Verlag. 85–105.

- Nida, E. 1964. *Toward a Science of Translating*. Leiden: Brill.
- Nord, Ch. 2005. *Text Analysis in Translation: Theory, Methodology, and Didactic Application of a Model for Translation-oriented Text Analysis*. Amsterdam: Rodopi.
- 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.
- Nyelv- és Beszédtechnológiai Platform. 2010. *Megvalósítási terv*. Budapest: MTA Nyelvtudományi Intézet.
- Olohan, M. 2004. *Introducing Corpora in Translation Studies*. London, New York: Routledge.
- Oslo Multilingual Corpus*. Online elérés: <http://www.hf.uio.no/ilos/forskning/prosjekter/sprik/english/corpus/>
- PACTE. 2005. Investigating Translation Competence: Conceptual and Methodological Issues. *Meta*. 50/2. 609–619
- Paepcke, F. 1994. Textverstehen – Textübersetzen - Übersetzungskritik. Snell-Hornby, M. (Hrsg.) 1994. *Übersetzungswissenschaft. Eine Neuorientierung*. Tübingen, Basel: Francke Verlag. 106–132.
- Papp F. 1972/2006. A mondatnál hosszabb szövegegységek végső szerkesztése, avagy az idegen nyelvű beszéd kvázi-helyessége. In: Klaudy K. (szerk.) 2006. *Papp Ferenc Olvasókönyv. Papp Ferenc válogatott nyelvészeti tanulmányai*. Budapest: Tinta Kiadó. 122–135.
- Papp F. 1979/2006. Szavak–Jelentés. In: Klaudy K. (szerk.) 2006. *Papp Ferenc Olvasókönyv. Papp Ferenc válogatott nyelvészeti tanulmányai*. Budapest: Tinta Kiadó. 153–168.
- Parallel Corpora in Uppsala*. Online elérés: <http://xml.coverpages.org/etap-over.html>
- Paulussen, H. 1999. *A Corpus-based Contrastive Analysis of English on/up, Dutch on and French sur within a Cognitive Framework*. Unpublished PhD thesis. University of Ghent.
- Péché O. 2007a. A lexikai kohézió vizsgálata a fordításkutatásban – új távlatok a korpusznyelvészeti módszernek köszönhetően. *Fordítástudomány*. IX. évf. 1. szám. 79–96.

- Péché O. 2007b. A lexikai kohézió vizsgálata a fordításkutatásban egy korpusznyelvészeti eszközzel végzett kutatás tükrében. In: Váradi T. (szerk.) 2007. *I. Alkalmazott Nyelvészeti Doktorandusz Konferencia kötete*. Budapest: MTA Nyelvtudományi Intézet. 114–127.
- Péché O. 2007c. Die Erforschung der lexikalischen Kohäsion mit korpuslinguistischen Mitteln. In: Schmitt, P. A. / Jüngst, H. E. (Hrsg.): *Translationsqualität. Leipziger Studien zur angewandten Linguistik und Translatologie Vol. 5*. Peter Lang Publishing Group.
- Péché O. 2007d. Korpusznyelvészeti eszköz alkalmazása a lexikai kohézió vizsgálatánál. In: Geccsó T., Sárdi Cs. (szerk.) *Nyelvelmélet – Nyelvhasználat*. Budapest: Tinta Könyvkiadó. 222–230.
- 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: Váradi 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 találkozása a fordításban. Doktori kutatások Klaudy Kinga tiszteletére*. Budapest: ELTE Eötvös Kiadó. 73–84.
- Péché 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. 621–628.
- Petőfi S. J. 1988. A szöveg mint interdiszciplináris kutatási objektum. In: Kornya L., Czellerné Farkas M. (szerk.) 1996. *Magyar nyelvészeti cikk- és tanulmánygyűjtemény szakfordítóknak. II. Nyelvművelési és szakmai*

- nyelvhasználat*. Debrecen: Kossuth Lajos Tudományegyetem Idegennyelvi Központ. 42–45.
- Petőfi S. J. 1990. *Szöveg, szövegtan, műelemzés*. Budapest: Országos Pedagógiai Intézet.
- Petőfi S. J. 2004. *A szöveg mint komplex jel*. Budapest: Akadémia Kiadó.
- Prószéky 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.
- Prószéky G., Kis B. 1999. *Számítógéppel – emberi nyelven. Intelligens szövegkezelés számítógéppel*. Bicske: Szak Kiadó
- Prószéky G. 2005. *Nyelvtechnológia (és) alkalmazásai*. Budapest: Aranykönyv Kiadó.
- Prószéky G., Olasz G., Váradi T. 2006. Nyelvtechnológia. In: Kiefer F. (szerk.) *Magyar nyelv*. Budapest: Akadémiai Kiadó. 1038–1072.
- Prószéky G., Miháltz M. 2008. Magyar WordNet: az első magyar lexikális szemantikai adatbázis. *Magyar Terminológia*. Vol. 1. No. 1. 43–57.
- Radó A. 1909. *A fordítás művészete*. Budapest: Franklin Társulat.
- Räisänen, N. 2000. *Quasi-correctness and the balance of power in Finnish translations of informative texts produced in the European Union*. Helsinki. Doctoral thesis.
- Reiss, K. 1984. Adäquatheit und Äquivalenz. In: Wilss, W. and Thome, G. (eds.) 1984. *Translation Theory and its Implementation in the Teaching of Translation and Interpreting*. Tübingen: Narr. 161–178.
- Sager, J. C. 1994. Die Übersetzung im Kommunikationsprozess: der Übersetzer in der Industrie. In: Snell-Hornby M. (ed.) 1994. *Übersetzungswissenschaft. Eine Neuorientierung*. Tübingen und Basel: Francke Verlag.
- Santos, D. 1995. On grammatical translationese. In: Koskeniemi, K. (ed.) *Short papers presented at the Tenth Scandinavian Conference on Computational Linguistics*. Helsinki. 59–66.
- Schiewe, J. 2009. Satz – Text – Diskurs. Überlegungen zum Status und zu den Aufgaben der Textlinguistik. In: *Linguistics. Germanic and Romance Studies*. Vilnius University Publishing House. 94–103.

- Schmidt, M. A. 2005. *How do you do it anyway? A longitudinal Study of Three Transator Students Translating from Russian into Swedish*. Stockholm: Almqvist and Wiksell International.
- Simigné F. S. 2008. Fordítás a feminizmus korában. In: Sárdi Cs. (szerk.) 2008. *Kommunikáció az információs technológia korszakában. A XVII. MANYE Kongresszus előadásai*. Siófok. 2007. április 19–21. (A MANYE Kongresszusok előadásai 4.) Pécs–Székesfehérvár: MANYE–Kodolányi János Főiskola. 125–130.
- Simov, K. 2004. *CLaRK System - an XML Based System For Corpora Development*. (Version 3.0). Online elérés: <http://www.bultreebank.org/clark/>
- Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. 1995. From theory to practice. In: Leech, G., Myers, G., Thomas, J. (eds.) *Spoken English on Computer*. Harlow: Longman. 99–112.
- Snell-Hornby, M., Hönlig H.G., Kussmaul, P., Scmitt, P. A. 1998. *Handbuch Translation*. Tübingen: Stauffenburg.
- Stewart, D. 2000. Conventionality, creativity, and translated text: The implications of electronic corpora in translation. In: Olohan, M. (ed.) *Intercultural faultlines. Research models in translation studies: Textual and cognitive aspects*. Manchester: St. Jerome. 73–91
- Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Szabó Mihály G. 2003. A Szlovákiai magyar szakfordítások minőségének javításáról és az objektív fordításkritika megteremtésének feltételeiről. *Társadalomtudományi Szemle*. V. évf. 4. szám. 55–68.
- Szathmári I. 1996. Nyelvtudomány, stilisztika, fordításelmélet és – gyakorlat. In: Kornya L., Czellerné F. M. (szerk.) 1996. *Magyar nyelvészeti cikk- és tanulmánygyűjtemény szakfordítóknak. II. Nyelvművelés és szakmai nyelvhasználat*. Debrecen: Kossuth Lajos Tudományegyetem Idegennyelvi Központ. 12–16.
- Szeged Treebank 2.0*. Online elérés: www.inf.u-szeged.hu/projectdirs/hlt/
- Szikszaíné Nagy I. 1999. *Leíró magyar szövegtan*. Budapest: Osiris.
- Szili K. 2006. Nyelvbe rejtett üzenetek. A nyelv és a kultúra viszonyáról általában és a magyar nyelv kapcsán. *Magyar Nyelv*. 102. évf. 1. szám. 33–43.

- Szirmai M. 2005. *Bevezetés a korpusznyelvészetbe. A korpusznyelvészet alkalmazása az anyanyelv és az idegen nyelv tanulásában és tanításában*. Budapest: Tinta Könyvkiadó.
- Teubert, W., Cermáková, A. 2007. *Corpus Linguistics: A Short Introduction*. New York: Continuum.
- The database of translations of Hungarian literary works*. Online elérés: <http://translations.bookfinder.hu/indexn.htm>
- The Online Books Page*. Online elérés: <http://onlinebooks.library.upenn.edu/>
- TEI. The Text Encoding Initiative*. Online elérés: <http://www.tei-c.org/index.xml>
- The Translational English Corpus-t (TEC)*. Online elérés: <http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>
- Tirkkonen-Condit, S. 2002. Translationese — a myth or an empirical fact? A study into the linguistic identifiability of translated language. In: *Target* 14/2. Amsterdam & Philadelphia: John Benjamins Publishing Company. 207–220.
- TIGER Project*. Online elérés: www.ims.uni-stuttgart.de/projekte/TIGER/
- Tolcsvai Nagy G. 1994. *A szövegtan kialakulása*. Online elérés: <http://www.jgytf.u-szeged.hu/~vass/szemm112.htm>
- Toury, G. 1995. *Descriptive Translation Studies and beyond*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Triptic Corpus*. Online elérés: http://universal.elra.info/product_info.php?cPath=42_43&products_id=465
- Tymoczko, M. 1998. Computerized Corpora and the Future of Translation Studies. *Meta* 43/4. 652–660.
- Vanderauwera, R. 1985. *Dutch Novels Translated into English: The Transformation of a 'Minority' Literature*. Amsterdam: Rodolpi.
- Vehmas-Lehto, I. 1989. *Quasi-correctness. A Critical Study of Finnish Translations of Russian Journalistic Texts*. Helsinki: Neuvostoliittoinstituutti.
- Vermeer, H. J. 1994. Übersetzen als kultureller Transfer. In: Snell-Hornby, M. (Hrsg.) 1994. *Übersetzungswissenschaft. Eine Neuorientierung*. Tübingen, Basel: Francke Verlag. 30–53.
- Warren, M. 2006. Cohesive chains and speakers' choice of prominence. In: Flowerdew, J., Mahlberg, M. 2006. *Lexical Cohesion and Corpus Linguistics*.

Benjamins Current Topics. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 45–64.

Weinrich, H. 1967. Syntax als Dialektik (Bochumer Diskussion). In: *Poetica* 1. 109-126.

Wilss, W. 1996. *Übersetzungsunterricht. Eine Einführung*. Tübingen: Gunter Narr Verlag.

Wittgenstein, L. 2004/1963. *Logikai-filozófiai értekezés*. Budapest: Atlantisz Kiadó.

WordSmith Tools. Online elérés:

<http://www.lexically.net/LexicalAnalysisSoftware/index.html>

Nyilvános források

Az Európai Unió portálja. Elérhető: http://europa.eu/index_hu.htm

Az Európai Unió Tanácsának magyar elnöksége. Elérhető: <http://www.eu2011.hu/hu>

Biblia Tarsulat. Elérhető: <http://www.biblia-tarsulat.hu/biblia/parhuzam.htm>

Educatio Társadalmi Szolgáltató Nonprofit Kft. Elérhető:

<http://www.tankonyvtar.hu/>

Jel Kiadó - Anselm Grün művei. Elérhető:

<http://www.prosperod.hu/katalogus/kereso/search.html?kotestipus=0&szerzo=Anselm%20Gr%FCn>

Magyar Katolikus Egyház. Elérhető: <http://www.katolikus.hu/>

Magyar Kurír. Elérhető: <http://www.magyarkurir.hu/>

Országos Széchényi Könyvtár Magyar Elektronikus Könyvtára. Elérhető:

<http://mek.oszk.hu/>

Vatikán. Elérhető: <http://www.vatican.va/>

Publikációs lista

Seidl-Pécs Olívia – Publikációk

2005

Joanna Best und Sylvia Kalina (Hrsg.): Übersetzen und Dolmetschen: eine Orientierungshilfe. Tübingen, Basel: Franke. 2002. Fordítástudomány VII. évf. 1. szám. 125–127. (recenzió)

Beszámoló a tavaszi fordítói konferenciáról. *Modern Nyelvoktatás* XI. évf. 2-3. szám. 119–120. (konferencia-beszámoló)

Az Akadémia Kiadó II. Szótárnapja. *Fordítástudomány* VII. évf. 2. szám. 119–121. (konferencia-beszámoló)

A magyar fordítók és tolmácsok napja. *Alkalmazott Nyelvtudomány* V. évf. 1-2. szám. 177–179. (konferencia-beszámoló)

2006

Szirmai Mónika: Bevezetés a korpusznyelvészetbe, Budapest: Tinta Könyvkiadó, 2005. *Fordítástudomány* VIII. évf. 1. szám. 123–124. (recenzió)

„Antal László és a mai magyar nyelvtudomány“. *Alkalmazott Nyelvtudomány* VI. évf. 1-2. szám. 219–222. (konferencia-beszámoló)

2007

A lexikai kohézió vizsgálata a fordításkutatásban - új távlatok a korpusznyelvészeti módszerek köszönhetően. *Fordítástudomány* IX. évf. 1. szám 79–96. (tanulmány)

Die Erforschung der lexikalischen Kohäsion mit korpuslinguistischen Mitteln. In: Schmitt, P. A., Jüngst, H. E. (Hrsg.): *Translationsqualität*. Leipziger Studien zur

angewandten Linguistik und Translatologie Vol. 5. Peter Lang Publishing Group.
(tanulmány)

A lexikai kohézió vizsgálata a fordításkutatásban egy korpusznyelvészeti eszközzel végzett kutatás tükrében. In: Váradi T. (szerk.) 2007. I. Alkalmazott Nyelvészeti Doktorandusz Konferencia kötet. 114–127. *Online elérés:*
<http://www.nytud.hu/alknyelvdok07/proceedings07/Pech.pdf> (tanulmány)

Korpusznyelvészeti eszköz alkalmazása a lexikai kohézió vizsgálatánál. In: Geccső T., Sárdi Cs. (szerk.) *Nyelvelmélet – Nyelvhasználat*. Budapest: Tinta Könyvkiadó. 222–230. (tanulmány)

2009

Miért érdekes a lexikai kohéziós eszközhasználat vizsgálata a fordításkutatás számára? In: Menyhárt J., Presinszky K. (szerk.) 2009. *Fordítás – kétnyelvűség (Nyelvészeti tanulmányok a fordítás elméletéről és gyakorlatáról)* Nyitra: Konstantin Filozófus Egyetem Közép-európai Tanulmányok Kara. 27–38. (tanulmány)

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. (tanulmány)

Analysing lexical cohesion in translation research using corpus linguistic methods. In: Váradi T. (szerk.) 2009. *Válogatás az I. Alkalmazott Nyelvészeti Doktorandusz Konferencia előadásaiból*. Budapest: MTA Nyelvtudományi Intézet. 103–116. (tanulmány)

2010

A volna – lenne előfordulási aránya fordított szövegekben. Szövegfeldolgozó eszközök alkalmazása a fordításkutatásban. In: Geccső T., Sárdi Cs. (szerk.) 2010. *Új módszerek az alkalmazott nyelvészeti kutatásban*. Budapest: Tinta Könyvkiadó. 238–243. (tanulmány)

Korpusznyelvészeti módszer a fordításkutatásban. In: Zimányi Á. (szerk.) 2010. A tudomány nyelve – A nyelv tudománya. Alkalmazott nyelvészeti kutatások a magyar nyelv évében. Eger: Eszterházy Károly Főiskola. 208–216. (tanulmány)

Szemponatok a fordított szövegek sajátosságainak leírásához. In: Károly K., Fóris Á. (szerk.) Nyelvek találkozása a fordításban. Doktori kutatások Klaudy Kinga tiszteletére. Budapest: ELTE Eötvös Kiadó. 73–84. (tanulmány)

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. (tanulmány)

Bozsik Gy., Seidl-Péché O. 2011. Fordítás- és tolmácsoláskutatás: Tegnap – Ma – Holnap. Beszámoló az Első Innsbrucki Fordítás- és Tolmácsolástudományi Konferenciáról. Fordítástudomány. 2011. XIII. évf. 1. szám. 115–120. (konferencia-beszámoló)

Seidl-Péché O., Tamás D. 2011. 5th Terminology Summit. Terminology 2010: Quality Matters – Terminológia 2010: Minőségügyi kérdések, Budapest, 2010. november 12-13. Magyar Terminológia Vol. 4. No. 1. 114-119. (konferencia-beszámoló)

2012

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. (tanulmány) (megjelenés alatt)

A témakörben tartott saját nyilvános előadások

2006

A lexikai kohézió vizsgálata korpusznyelvészeti eszközökkel. Elhangzott: VIII.

Fordítástudomány PhD Konferencia. ELTE BTK FTK.

Erforschung der lexikalischen Kohäsion mit korpuslinguistischen Mitteln. Elhangzott:

VIII. Leipziger Internationale Konferenz zu Grundfragen der Translatologie.

Universität Leipzig, Institut für Angewandte Linguistik und Translatologie.

A lexikai kohézió vizsgálata korpusznyelvészeti módszerrel. Elhangzott: Nyelvelmélet és nyelvhasználat Konferencia. ELTE-KJF.

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

2007

A lexikai kohézió vizsgálata a fordításkutatásban egy korpusznyelvészeti eszközzel

végzett kutatás tükrében. Elhangzott: I. Alkalmazott Nyelvészeti Doktorandusz

Konferencia. MTA Alkalmazott Nyelvészeti Munkabizottsága-MTA Nyelvtudományi Intézete.

2008

Hozzájárulás a szövegnyelvészet interkulturális megközelítéséhez. A lexikai kohézió

vizsgálata fordított szövegeknél. Elhangzott: Nyelv-Kultúra-Tudományok

Konferencia. Pápai Református Gimnázium.

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

2009

Korpusznyelvészeti módszer a fordításkutatásban. A tudomány nyelve – a nyelv tudománya Konferencia. Elhangzott: XIX. Magyar Alkalmazott Nyelvészeti Kongresszus.

Szövegfeldolgozó eszközök és alkalmazásai a fordítás kutatásban. Elhangzott: Új módszerek az alkalmazott nyelvészeti kutatásban Konferencia. ELTE–KJF.

Emlékezés Pápai Vilmára. Pápai Vilma szakmai munkássága. Elhangzott: A Papp Ferenc Baráti Kör 2009. decemberi ülésén. ELTE BTK FTK.

2010

Változik-e az opcionális volna : lenne arány a fordított szövegekben? Elhangzott: IV. Alkalmazott Nyelvészeti Doktorandusz Konferencia. MTA Alkalmazott Nyelvészeti Munkabizottsága-MTA Nyelvtudományi Intézete. (poszter szekció)

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.

Szemponatok a fordított szövegek sajátosságainak leírásához. Elhangzott: Nyelvek találkozása a fordításban. Doktori kutatások Klaudy Kinga tiszteletére. Budapest: ELTE.

2011

Warum wird die Verwendung lexikalischer Kohäsionselemente bei der translatorischen Tätigkeit beeinflusst? Elhangzott: 1. TRANSLATA-Konferenz. Institut für Translationswissenschaft der Leopold-Franzens Universität Innsbruck.