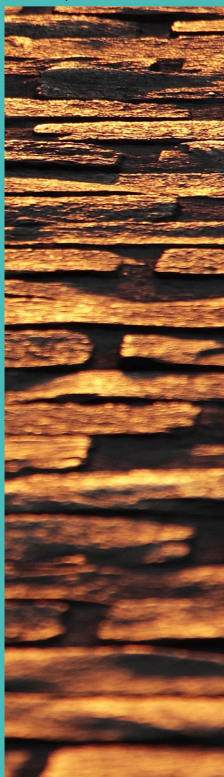




A FORDÍTÁSOK LEXIKAI KOHÉZIÓS MINTÁZATA

Seidl-Pécs Olívia

ELTE BTK Fordító- és Tolmácsképző Tanszék



FORDÍTÁSTUDOMÁNYI ÉRTEKEZÉSEK IV.
A FORDÍTÁSOK LEXIKAI KOHÉZIÓS MINTÁZATA
FORDÍTOTT ÉS AUTENTIKUS MAGYAR SZÖVEGEK
KORPUSZNYELVÉSZETI VIZSGÁLATA

Szakmai lektor:

Kóbor Márta

Édesanyámnak

SEIDL-PÉCH OLÍVIA

**A FORDÍTÁSOK
LEXIKAI KOHÉZIÓS MINTÁZATA**

**FORDÍTOTT ÉS AUTENTIKUS MAGYAR SZÖVEGEK
KORPUSZNYELVÉSZETI VIZSGÁLATA**

ELTE BTK
Fordító- és Tolmácsképző Tanszék

2024

ISBN 978-963-489-735-4 (online)

ISSN 3057-8043 (online)

DOI: <https://doi.org/10.21862/fordtudert4>

© Seidl-Péché Olívia, 2024

Kiadó: ELTE BTK Fordító- és Tolmácsképző Tanszék

Olvasószerkesztő: Hargitai Ildikó

A borító forrása: pixabay.com

Tördelőszerkesztő: Szabó Panna

Sorozatszerkesztő: Robin Edina

TARTALOM

| | |
|--|----|
| I. BEVEZETÉS | 7 |
| II. A FORDÍTOTT SZÖVEG | 13 |
| 1. Bevezetés | 13 |
| 2. A fordított szövegek vizsgálata | 14 |
| 3. A fordított szövegek hatása az anyanyelvhasználatra | 18 |
| 4. A fordítói kompetencia | 26 |
| 5. Norma és ekvivalencia a fordításkutatásban | 28 |
| 6. A fordítási univerzálé | 33 |
| III. SZÖVEGNYELVÉSZETI MEGKÖZELÍTÉS | 35 |
| 1. Bevezetés | 35 |
| 2. A szövegkutatás feladata | 36 |
| 3. A fordítások szövegnyelvészeti megközelítése | 38 |
| 4. A szöveg meghatározása | 40 |
| 5. A szövegszerűség kritériumai | 41 |
| IV. A LEXIKAI KOHÉZIÓ VIZSGÁLATA | 45 |
| 1. Bevezetés | 45 |
| 2. A kohéziókutatás legismertebb képviselői | 45 |
| 3. A kohézió | 47 |
| 4. A lexikai kohézió vizsgálata fordított szövegekben | 53 |
| V. A KORPUSZNYELVÉSZETI MÓDSZER | 57 |
| 1. Bevezetés | 57 |
| 2. A korpusznyelvészet mint módszer | 57 |
| 3. Korpusznyelvészet a fordításkutatásban | 58 |
| 4. A korpuszok legalapvetőbb ismérvei | 59 |
| 5. A WordNet-formalizmus | 69 |
| VI. A pilot kutatás..... | 71 |
| 1. Bevezetés | 71 |
| 2. A pilot kutatás szövegtörzse | 71 |

| | |
|---|-----|
| 3. A Károly-féle taxonómia | 72 |
| 4. A gépi feldolgozás előkészítése | 73 |
| 5. Annotálás és elemzés | 73 |
| 6. A pilot kutatás eredményei | 75 |
| | |
| VII. A MAGYAR LEXIKAI KOHÉZIÓS PROJEKT | 77 |
| 1. Bevezetés | 77 |
| 2. A Magyar Lexikai Kohéziós Projekt kutatási kérdései és hipotézisei | 78 |
| 3. A Magyar Lexikai Kohéziós Projekt szövegkorpusza | 80 |
| 4. A Magyar Lexikai Kohéziós Projekt annotálása | 86 |
| 5. A Magyar Lexikai Kohéziós Projekt elemzése | 87 |
| 6. A Magyar Lexikai Kohéziós Projekt eredményei | 97 |
| | |
| VIII. ÖSSZEGZÉS | 109 |
| | |
| Irodalom | 113 |
| Források | 127 |
| Függelékek | 129 |

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

I. Bevezetés

A fordításkutatás kezdeteitől fogva központi helyet foglal el a fordított szöveg jellemzőinek feltárása. A téma nagyrészt annak a mindennapi tapasztalatnak köszönhetően került középpontba, amely szerint a fordítás eredményeként létrejövő szövegek lényegesen eltérnek az azonos nyelven, de nem fordítás révén keletkezett szövegektől. A kezdetben csupán intuíciókra épülő megállapítást mára már több kutatás is igazolta konkrét adatokkal (pl. Blum-Kulka 1986, Lindström 1996, Räisänen 2000), de időszerűsége miatt a kutatási terület a gépi fordítás megjelenése után is kiemelt figyelmet kap (vö. Riley és mtsai. 2020, Zhang és Toral 2019).

Számos ezredfordulót követő kutatásnak is az a célja, hogy mélyebben megvilágítsa a fordított szövegek sajátosságait (pl. Chou és mtsai. 2023, Götz 2018, Károly és mtsai. 2022a, 2022b, Kenny 2001, Klaudy 2004, 2006, Klenk 2019, Laviosa 2001, Robin 2018). Ezek a kutatások arra törekednek, hogy azonosítsák azokat a nyelvi tényezőket, amelyek miatt a fordított szövegek különböznek az autentikus szövegektől. A fordítás során végbemenő nyelvi kódváltás számos esetben strukturális változásokat eredményez, amelyek befolyásolják a szövegek koherenciáját, kohézióját és stilisztikai jellemzőit.

Annak ellenére, hogy számos kutatás foglalkozott a fordítási szövegalkotás leírásával, mindmáig csak részleges eredmények állnak rendelkezésünkre. Egy-egy vizsgálat általában csak korlátozott nyelvi irányokban és/vagy szövegtípusok esetében tudja a kiválasztott nyelvi jellemzőt az autentikus és a fordítási szövegalkotás szempontjából leírni, – ahogy ezt például Károly (2021) a tudományos szakfordítás empirikus kutatásának eredményeit áttekintő tanulmányában bemutatja.

A fordítási szövegalkotás leírásával foglalkozó eddigi kutatások eredményei arra utalnak, hogy a fordított szövegek sajátos jellegzetességei nem csupán a nyelvi struktúrák szintjén jelentkeznek, hanem mélyebb, szemiotikai és pragmatikai szinteken is. A fordítások gyakran tartalmaznak olyan nyelvi elemeket és mintákat, amelyek az adott nyelvre nem, vagy másmilyen gyakorisággal és/vagy eloszlással jellemzőek, és a fordításokra jellemző különbségek jelentős hatással lehetnek a szövegek befogadására és értelmezésére.

A jelen mű nyelvészeti megközelítésben foglalkozik a fordításokkal, fókuszra a fordított szövegekre jellemző lexikai kohéziós mintázat feltárása, illetve annak vizsgálata, hogy a magyar nyelv esetében különbözik-e az autentikus és a fordított szövegek lexikai kohéziós mintázata. A bemutatott alkalmazott nyelvészeti kutatás úttörő jellegű, olyan területen dolgoz ki nyelvtechnológiai alkalmazásokkal támogatott elemzési módszert nagy-

mintájú szövegvizsgálathoz, amely eddig ismeretlen volt mind a hazai, mind a nemzetközi szakirodalomban. A lexikai ismétlés szövegszervező funkciójának vizsgálata jelen esetben a korpusznyelvészet eszközeivel valósul meg, a manuális elemzésre megalkotott taxonómiát integrálva a gépi elemzés folyamatába. A tárgyalt téma tehát három tudományterület – a fordításkutatás, a szövegnyelvészet és a korpusznyelvészet – kutatási területének metszéspontjában helyezkedik el. Az integrált megközelítés lehetővé teszi a nagymintájú szövegvizsgálat során a lexikai kohéziós elemek és a köztük levő kapcsolatok automatikus, következetes és pontos azonosítását. A kidolgozott módszertan új perspektívákat nyit a nyelvtechnológia számára a szövegelemzés területén.

A mű központi témájául választott célnyelvi szövegprodukciónak leírása, 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) A jelen, a fordított szövegekre jellemző eltolódásjelenségeket vizsgáló kutatás a lexikai kohéziós mintázatot korpusznyelvészeti módszerrel tárja fel és csatlakozik ahhoz a Petőfi S. (2004) által javasolt szövegnyelvészeti megközelítéshez, amely a szövegek komplexebb vizsgálatát tűzi ki célul.

A fordítási folyamathoz kapcsolódó lexikai eltolódásjelenségek vizsgálata ugyanakkor más szerzők értelmezésében a fordítási folyamat különböző fázisaiban is külön figyelmet érdemelhet. Klenk (2019) szerint érdemes a lektorálás folyamán keletkező lexikai eltolódásokat elemezni, mivel „ezen lexikai eltolódások vizsgálata a lektori kompetenciára és stratégiára nézve fontos adatokkal szolgálhat” (Klenk 2019:32). Szlávik (2023a) a neurális gépi fordítás következtében a lexikai ismétlés különféle típusainak szövegen belüli szerveződésében bekövetkező eltolódásokat kutatja, és megvizsgálja az eltolódások utószerkesztés során megvalósuló módosulásait.

A könyvben bemutatott lexikakohézió-kutatás nagyobb mennyiségű nyelvi adat alapján végzett összehasonlító elemzés és így 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 szöveg szövegépítő sajátosságainak egybevetése adja. Az autentikus és a fordítás eredményeként keletkezett magyar szövegek lexikai kohéziós mintázatának leírása a mondathatáron túlmutató nagyobb nyelvi egységek nyelvészeti elemzésével valósul meg, „a mondatokból létrejövő szöveg alkotását modellálja, annak szabályait és a nyelvi rendszerben meghatározó törvényszerűségeit kutatja” (Kocsány 2002:146).

A lexikai kohéziós mintázat vizsgálata a szövegnyelvészet kutatási területéhez tartozik. Az elemzés a magyar nyelv esetében bizonyítani kívánja, hogy a fordított szövegek kohéziós eszközhasználatában bekövetkező változások az eltolódásjelenségek sorába tartoznak, mivel a fordított szövegek kohéziós mintázata nem egyezik meg a magyar autentikus szövegek kohéziós mintázatával. A VI. fejezetben bemutatott kis szövegmintán, kézi annotálással végzett pilot elemzés, valamint a VII. fejezetben bemutatott nagyobb korpuszon (Magyar Lexikai Kohéziós Projekt), automatikus eszközök segítségével ki-

vitelezett vizsgálat rámutat az autentikus és a fordított magyar szövegekre jellemző kohéziós mintázatok különbözőségeire. A jelen vizsgálat tehát a fordítás során keletkező szövegek lexikai kohéziós sajátosságainak szövegnyelvészeti elemzésével nagymértékben hozzájárul a fordított szövegekről alkotott tudásunk kiteljesítéséhez, s ez által további szempontokat kínál a fordításértékelés és a fordítóképzés számára.

A kutatás – bár nem alapcélkitűzése – vizsgálati fókusza miatt kapcsolódik a műfajkutatás területéhez is. „A műfaj fogalmát nemcsak az irodalmi szövegfajtákra, hanem annál tágabb értelemben használjuk. Minden olyan szövegfajta műfajnak tekintünk, amely bizonyos kommunikatív célok megvalósítására adott szituációban és diskurzusközösségben meghatározott szerkezeti, stílári és tartalmi tulajdonságokkal konkrét olvasóközönség számára jön létre” (Károly 2007:128). A műfajkutatás tehát a szövegnyelvészetben belül elsősorban a különböző műfajokra/szövegtípusokra jellemző retorikai, szerkezeti, tartalmi és nyelvi sajátosságokat kutatja, leírja a szövegtípusokhoz társítható szerkezeti mintázatokat, tipikus elemeket és szerveződési elveket, ezen belül vizsgálja az egyes szövegtípusokra jellemző kohéziót és koherenciát is. Mivel a bemutatott vizsgálat több szövegtípus esetében vizsgálja az autentikus és a fordított magyar szövegekre jellemző lexikai kohéziós mintázatot, így eredményei hozzájárulnak a műfajkutatás vizsgálati eredményeihez is.

A kutatás a korpuszalapú vizsgálatok közé sorolható, jellegzetessége az írott nyelv szövegépítő sajátosságainak feltárása, ezen belül a variabilitást befolyásoló szöveggörnyezeti tényezők és a nyelvhasználat szabályszerűségeinek leírása. Betagozódik a korpusz-nyelvészeti technikákkal folytatott kutatások sorába, melyek lehetővé teszik a fordított szöveget az autentikus szövegproduktiótól megkülönböztető általános tulajdonságok, többek között a fordítási univerzálék feltárását (vö. Baker 1993, 1996, Robin 2023). Az elemzés fókusza az univerzálékutatásba illeszkedő fordított és nem fordított kontrasztív szövegelemzés, valamint a nyelvpárspecifikus fordítói magatartást vizsgáló, a magyar forrásnyelvi és célnyelvi szövegjellemzőket egybevető összehasonlítás.

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

A leíró fordítástudomány (*Descriptive Translation Studies*) hagyományos holmesi hármas felosztásából kiindulva a jelen kutatást az eredményközpontú leíró fordítástudományi (*Product-oriented DTS*) munkák közé lehet sorolni (Holmes 1972). Bár napjainkban a folyamatközpontú (*Process-oriented DTS*) és funkcióközpontú (*Function-oriented DTS*) fordítástudományi terület is nagy népszerűségnek örvend, a jelen kutatásban vizsgált kérdésekre csak a fordítási folyamat eredményeként létrejött fordított szövegeket vizsgáló eredményközpontú módszer tud pontos válaszokat adni, mivel csak a célnyelvi szöveg elemzése alapján kaphatunk pontos képet a fordító végső döntéseiről (vö. Hatim és Mason 1990:3).

A választott kutatási téma, nevezetesen a magyar nyelv vonatkozásában a fordított és nem fordított (autentikus) szövegek szövegépítő sajátosságainak egybevetése, két párhuzamos szövegtudományi összehasonlító elemzésén alapul. Az összehasonlító vizsgálatok ezen típusa lehetővé teszi a fordított és nem fordított 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 szövegtudományon végzett vizsgálatot, valamint kvantitatív mutatók kinyerését teszi lehetővé.

A vizsgálatot az alkalmazott nyelvészet számára az ezredfordulót követően új perspektívát kínáló nyelvtechnológiai alkalmazások (*Human Language Technologies*) segítették, amelyek „a nyelvleírás és a szoftvertudományi eszközök találkozásának” (Prószték 2005) köszönhetően lehetővé teszik a nyelvi tartalom szerinti szövegfeldolgozást. A szövegek intelligens kezelésére (fogalmazástámogatás, kereséstámogatás, megértéstámogatás, fordítástámogatás) hivatott számtalan új nyelvtechnológiai alkalmazás megjelenése még korántsem biztosítja az ilyen, vagy hasonló alkalmazások használatának elterjedését a nyelvészeti kutatásokban. A jelen kutatás újszerűsége éppen a kutatott témának, a lexikai kohézió vizsgálatának nyelvtechnológiai alkalmazások segítségével történő feltárásán alapszik. A kutatás során a következő lépések automatizálását lehetővé tevő 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 jelen vizsgálat számos területen kínál a nemzetközi gyakorlat szempontjából is újszerű megközelítést, amelyeket a következő pontokban összegezhünk. A kutatás újszerű elemei:

1. a nyelvtechnológiai eszközök alkalmazása az autentikus és a fordított magyar szövegek lexikai kohéziós relációinak feltárásában,
2. a nemzetközi kohéziókutatás gyakorlatától eltérő, nyelvtechnológiai eszközök segítségével megvalósított szövegkohéziós vizsgálat 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 nagyobb szövegtudomány vizsgálatára lényegesen kisebb időráfordítás mellett,
5. két párhuzamos, a keletkezésük körülményeiben eltérő, korpuszon végzett összehasonlító lexikai kohéziós vizsgálat,
6. az autentikus és a fordított magyar szövegek kohéziós mintázatának leírása a HuWordNet ontológiai kategóriáinak segítségével,
7. a kohéziós kapcsolatok bejelölésének következetessége a viszonylag nagyméretű szövegtudományok ellenére a HuWordNet kategóriák alkalmazása miatt.

A VII. fejezetben leírt kutatás igen fontos eredménye, hogy autentikus és fordított magyar szövegek esetében feltárja szövegtípus és nyelvpár függvényében a célnyelvi magyar szövegekre jellemző lexikai kohéziós jellemzőket. Az eredmények között egyaránt megtalálható a vizsgált szövegfajták (közéleti, szépirodalmi, tudományos és vallásos), illetve nyelvpárok (angol-magyar, francia-magyar, görög-magyar, latin-magyar, német-magyar, olasz-magyar) esetében az autentikus magyar szövegprodukcióna és a fordított magyar szövegekre jellemző lexikai kohéziós mintázatok különbsége.

A szövegfajták és a nyelvpárok ilyen széles skálájának vizsgálata a kutatásban alkalmazott nyelvtchnológiai módszereknek köszönhető. A magyar nyelv esetében a nyelvtchnológiai eszközök alkalmazásán alapuló lexikai kohéziós szövegjellemzők feltárását megvalósító vizsgálat úttörő jellegű, hozzájárul a nyelvléírás szempontrendszerének bővítéséhez. A mondathatárokon túlmutató kohéziós kapcsolatok automatikus feltárásának módját bemutató kutatás a gépi fordítás számára kínálja a WordNet-formalizmus alkalmazhatóságának bizonyítását. A Magyar Lexikai Kohéziós Projekt eredményei esetében ugyanakkor arról is kell szólni, hogy a WordNet-formalizmus kizárólag a főnévi csoportok kohéziós kapcsolatainak feltárását teszi lehetővé. Ez bizonyos szempontból jóval szűkebb megközelítés, mint az elsősorban manuális elemzés szempontjából népszerű Hoey (1991) vagy Károly (2002) féle, elméleti alapú lexikai ismétlésmodellek által lefedett lexikai relációk köre. Más szempontból a WordNet-formalizmus tartalmaz olyan relációkat is, amelyek az előző két modellnek nem részei (pl. *category domain*, *be in state* és *region domain* relációk), de ezen relációk alacsony száma az egyes szövegeken belül elhanyagolhatóvá teszi azok bevonását a kutatásba. A manuális elemzési taxonómiák (Hoey 1991, Károly 2002b) és a WordNet-formalizmus közötti különbségekből fakadó eltérések érintik ugyan a kohézió kutatás lényeges elméleti vonatkozásait, ugyanakkor a WordNet szűkített körű elemzési kerete nem gátolja meg a lexikai ismétlődési mintázat egybevetését, hiszen a kutatás ezt az elemzési keretet alkalmazta az összes szövegtípus és nyelvpár esetében.

A kutatás eredményei között figyelemre méltó az összehasonlíthatóság feltételeit kielégítő összehasonlító korpusz megalkotása, amelynek autentikus és fordított magyar szövegeket tartalmazó alkorpuszai az egyes szövegfajták és nyelvpárok tekintetében kiegyensúlyozottak. Az autentikus és fordított szövegeket tartalmazó alkorpuszok szétválasztása már önmagában is külön eredménynek tekinthető, mivel a hazai korpuszépítési gyakorlatra kevés kivételtől eltekintve – pl. Pannónia Korpusz (Robin és mtsai. 2016) – nem 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*¹ sem tesz különbséget a magyar szövegek között aszerint, hogy azok az elsődleges vagy a másodlagos szövegalkotás eredményeképpen keletkeztek-e. Ennek következtében a *Magyar Nemzeti Szövegtár* nem alkalmas arra, hogy referenciakorpuszként szolgáljon a fordított szövegek szövegépítő sajátosságainak feltárásához.

¹ <http://corpus.nytud.hu/mnsz>

II. A fordított szöveg

1. Bevezetés

Napjaink globálisan összekapcsolt világában a kultúraközi kommunikáció egyre fontosabb szerepet játszik a mindennapi életben és a kommunikációs rendszerek működtetésében. A különböző kultúrák közötti hatékony kommunikáció alapvető igénnyé vált a gazdasági, politikai és társadalmi kapcsolatok fenntartása szempontjából. Az Európai Unió is felismerte ennek a jelentőségét, amely jelenségre az Interkulturális Párbeszéd Európai Éve (2008) is felhívta a figyelmet. Az unió célul tűzte ki a különböző kulturális csoportok közötti párbeszéd erősítését, a kommunikációs partnerek közötti megértést, a tolerancia és az együttműködés elősegítését.

A kultúraközi kommunikáció erősödése közvetlenül befolyásolja a fordított szövegek mennyiségi növekedését. Ahogy a különböző kultúrák közötti kommunikáció szükségessége előtérbe kerül, úgy nő a fordítások iránti igény is. A fordítók kulcsszerepet játszanak abban, hogy az információk, hivatalos dokumentumok, tudományos kutatások eredményei – akárcsak az irodalmi művek – elérhetővé váljanak a más nyelveken beszélő közönség számára. A fordítások iránti növekvő igény azonban nem csupán a mennyiségi növekedést vonja maga után, hanem egyre nagyobb kihívást jelent a minőségi fordítások biztosítása terén.

A fordított szövegek számának növekedésével az autentikus, nem fordított szövegek és az anyanyelvhasználat is változásokon megy keresztül. A fordítás révén az egyes országokba áramló fordított szövegek idegen nyelvi strukturái és kifejezései hatással vannak az adott ország anyanyelvi szövegprodukcójára, amely sokszor átveszi a célnyelvi szövegek sajátosságait. Ez a jelenség különösen nyilvánvaló az olyan nyelvek esetében, ahol a fordított szövegek domináns szerepet töltenek be bizonyos területeken (pl. angolról fordított uniós dokumentumok).

A fordítástudományban a kontrasztív szövegnyelvészeti elemzések fontos szerepet játszanak az autentikus és a fordított szövegek közötti különbségek feltárásában. Ezek az elemzések segítenek megérteni, hogy a fordítási folyamat hogyan befolyásolja az autentikus a célnyelvi szövegprodukción, és milyen hatással van a célnyelv nyelvi normáira és a stílusra. Az explicitáció során például a fordító a forrásnyelvi szöveg implicit információit explicit módon fejez ki a célnyelvi szövegben. Ez a jelenség azt eredményezheti, hogy a fordított szövegek világosabbak, de terjedősebbek lehetnek az autentikus szövegekhez képest. A fordított szövegeket vizsgáló kutatások nemcsak a fordítók munkáját segítik, hanem hozzájárulnak ahhoz is, hogy a nyelvhasználók tudatosabban és kritikusabban viszonyuljanak a fordított szövegekhez és azok hatásaihoz, illetve hogy tudatosabban kezeljék a nyelvi és stilisztikai döntéseiket.

2. A fordított szövegek vizsgálata

A fordításról való gondolkodás több ezer éves múltjára (Cicero, Szt. Jeromos, Horatius, Quintilianus, Boéthius, Leonardo Bruni, Luther, Rotterdami Erasmus, Herder, Goethe, Schleiermacher, Arany János, Brassai Sámuel, Rajnis József, Batsányi János, Szász Károly stb.) a XX. század közepéig elsősorban az irodalmi megközelítés és a preskriptív normatív szemlélet volt jellemző, bár az évszádok során megfigyelhető a tartalom pontos visszaadása (pl. bibliafordítás) és a formális, -stiláris hűség (pl. versfordítás) előnyben részesítése közötti ingadozás. Ciceró (1903) az i.e. I. században a fordítót még szabad embernek tekinti, aki megválogathatja, hogy kinek, mit és hogyan fordít, aki nem nyelvközvetítőként, hanem alkotóként fordít, ezért nem szó szerinti hűséget, hanem a közvetített gondolat lényegét kívánja visszaadni, amihez a saját véleményét is hozzá engedi tenni. Csak a XX. század derekán terjed el a fordítások minőségével kapcsolatban ma is elvárt szempontok betartásának fontossága, így a szószerintiség, az eredeti mű gondolatainak közvetítése, az eredeti mű integritásának megtartása (a fordító semmit adhat hozzá az eredetihez és semmit sem vehet el az eredetiből), az eredeti mű stílusának tükrözése miközben a fordított szöveg a fordító saját korának elvárásainak is megfelel. A múlt század elején például Radó Antal 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). Radó felfogásában a „műfordítást, ép úgy mint a költészet bármely ágát, nem lehet és nem szabad gyárilag űzni; hangulat kell a költő tolmácsolásához is, türelem és kitartás a kész munka csiszolásához” (Radó 1909:135).

A XX. század közepétől az irodalmi fordítások mellett egyre nagyobb arányban megjelenő szakfordítások következtében kialakul a nyelvészeti fordítástudomány. Az önálló fogalom- és terminuskészlet, illetve az önálló tudományos diszciplína megeremtésében úttörő szerepet játszott Fjodorov (1953), Jakobson (1959), Mounin (1963), Revzin és Rozencveg (1964), Nida (1964) és 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 fordítónak 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 közvetítse. A szakirodalom a nyelvi közvetítés folyamat-, funkció- és eredmény orientált tudományos szintű vizsgálatakor (Holmes 1972) számba veszi a forrásnyelvi szöveg célnyelvre történő átültetését befolyásoló összes nyelvi és nem nyelvi tényezőt, és törekszik a fordítás eredményeként létrejött produktum lehetőleg objektív szempontrendszer szerinti elemzésére. Az addigi (a forrásnyelvi szöveg elsődleges-ségét hangsúlyozó) gyakorlat preskriptív megközelítését a (fordítási folyamat jellemzőit

leíró, az átvitel funkcionalitásait elemző és a célnyelvi szöveg jellegzetességeit feltáró deskriptív (*Descriptive Translation Studies*) megközelítés váltja fel (Holmes 1972).

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 interdiszciplináris és empirikus jellegének kialakulásához (Holmes 1972). 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)

2.1. A leíró fordításkutatás megjelenése

A fordítások számának ugrásszerű növekedése maga után vonta a fordítás mibenlétével, elveivel, követelményeivel foglalkozó tudományos gondolkodás iránti egyre nagyobb igényt. Megtörtént a fordítástudomány holmesi belső felosztása (Holmes 1972) elméleti, leíró és alkalmazott területekre, mely felosztás megteremtette a tudományág részterületeinek elkülönülését a gyakorlatban, és az igényt az egyes területek kutatási kérdéseinek és feladatkörének meghatározására. A történelmi jelentőségű *The Name and Nature of Translation Studies* című cikkében Holmes (1972:71) a leíró fordítástudomány számára a jelenségek leírása mellett az általános elvek kialakítását jelölte meg alapcélkitűzésként. A leíró fordítástudomány két fontos feladatköre szerinte: (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. Cikkben Holmes a leíró fordítástudományt további három nagy területre osztja, és rögzíti az eredmény-, a folyamat- és a funkcióközpontú kutatások mibenlétét.

A VII. fejezetben bemutatott kutatás szempontjából a szövegnyelvészeti megközelítésű, eredményközpontú megközelítés bír nagy jelentőséggel, amit Holmes a következőképpen definiált:

Az eredményközpontú leíró fordítástudomány – a létező fordításokat leíró fordításkutatási részterület – mindig is fontos területe volt a tudományos igényű fordításkutatásnak. Az ilyen típusú tanulmányok kiindulópontja az egyes fordítások leírása, vagyis a szövegközpontú fordításleírás. A második megközelítés az összehasonlító fordításleírás, amikor ugyanazon szöveg egyazon nyelven belüli, vagy különböző nyelvű fordításainak összehasonlító elemzését végzik el. (Holmes 1972:71, a szerző fordítása)²

Az eredmény-, azaz szövegközpontú kutatásoknak köszönhető többek között a célnyelvi szövegalkotás jellemzőinek leírása. A célnyelvi szövegek sajátosságainak feltárása automatikusan maga után vonta 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)

2.2. Interkulturális tevékenység a fordítási szituációban

A fordítási szituációban alapvetően interkulturális tevékenység zajlik, amelyben a fordító közvetítő szerepet tölt be két különböző kultúra között, mely közvetítő szerep nemcsak nyelvi, hanem kulturális dimenziókat is magában foglal. Szili (2006) erről a kérdésről a következőket fogalmazza meg a nyelvtudomány számára:

Az a vélemény, hogy a nyelv és kultúra elválaszthatatlanul egybefonódnak, a kultúra látható és láthatatlan szálakkal át- meg átszövi a nyelvet annak számos elemében, szinte minden megjelenésmódjában, úgy kétszáz éves (nem töretlen) fejlődés eredményeként a XX. század 80-as éveire jegecesedett ki. Az új értelmezés persze jóval nehezebb feladat elé állította a kutatókat, hiszen a nyelvet már mint társadalmi

² 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)

produktumot, a közösségi élet egyik meghatározó tevékenységformáját kellett vizsgálniuk, ebből következően kapcsolatuk leírása sem csupán a nyelvi formákban – főként a szókincsben – közvetlenül kimutatható kulturális jegyek feltárását követelte meg, hanem a nyelv használatában tetten érhető sajátosságokét is. (Szili 2006:34)

A fordító feladata ezek szerint, hogy a forrásnyelvi szöveg kulturális kontextusát átültesse a célnyelvi szövegbe, a fordítás tehát nem egyszerűen nyelvi átváltás, hanem kulturális transzfer is. A fordítási szituációban a fordítónak figyelembe kell vennie a forrás- és a célnyelvi közösség kulturális sajátosságait, ami különösen fontos a kulturálisan kötött szövegek esetében, mint például az irodalmi művek, a jogi dokumentumok, vagy a reklámszövegek.

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. „[A] beszélő és hallgató társadalomban elfoglalt helyének, a köztük levő távolságnak, a saját közösségükhöz való viszonyuknak a megítélése erősen kultúráktól függő tényező” (Szili 2006:39).

A legtöbb kutató a fordító legfontosabb feladatai közé sorolja a különböző kultúrák és nyelvközösségek közötti szellemi hídépítést (Hönig 1997:18), illetve közvetítést. Hönig magának a fordítónak legfontosabb készségei között említi a forrásnyelvi szöveg alapján a célnyelvre és a célnyelvi kultúrába „belesimuló” szöveg előállításának képességét (Hönig 1997:26). Időszerű tehát minden, az interkulturális kommunikáció jellemzőit – így a célnyelvi szövegek sajátosságait – feltáró kutatás.

A kultúrák hasonlóságai és különbözőségei felőli közelítés jellemzi Vermeer (1994) translációelméletét is, aki minden emberi cselekvést, így a szöveget is a kultúrába ágyazottan szemlél. 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).

Budin (2002:81) egészen konkrétan veti fel a fordított szövegek és a fordítási szituáció interkulturális beágyazottságát. Az interkulturális szakmai kommunikáció egyik formájának tekinti a nemzetközi szakkommunikációt és a többnyelvű információmenedzsmentet 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. Földes (2007:25) felfogása

szerint „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”. Ez a „kommunikációs hozzáállás” kétségkívül jellemzi a fordító tevékenységét.

3. A fordított szövegek hatása az anyanyelvhasználatra

Az élet minden területére egyre inkább jellemző fordítási tevékenység, valamint a dokumentumok terjesztését felgyorsító internethasználat következtében a társadalom egyre szélesebb rétegeihez egyre nagyobb számban jutnak el fordított dokumentumok. Napjainkra a hétköznapi életet fokozott mértékben hatja át az interkulturális kommunikáció, s e jelenség közvetett és közvetlen hatása az anyanyelvhasználatra egyre egyértelműbben kimutatható. Az anyanyelvhasználat terén tapasztalt eltulodások következtében egyre több kutató vizsgálja a célnyelvi szövegalkotás anyanyelvre gyakorolt hatását.

Klaudy már 2001-ben jelezte, hogy a feltáró és leíró célú kontrasztív fordításelemzések új perspektívát nyithatnak az anyanyelvhasználat tudatosításában, és ennek szükségyszerűsége egyre sürgetőbb. A fordítandó szövegmenyiség exponenciális növekedése láttán előre lehetett látni, hogy „ezt hivatásos fordítókkal elvégeztetni lehetetlen. A nyelvek közötti megfélemtetés a jövőben nem a hivatásos nyelvi közvetítők monopóliuma, hanem a művelt magyar állampolgár napi rutinja lesz” (Klaudy 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 hozzáegíthetnek „a magyar nyelvleírás egy korszerűbb, funkcionális megközelítéséhez” (Klaudy 2001:55).

Heltai a fordító nyelvi normáival foglalkozó tanulmányorozatának bevezetőjében tárgyalja a fordításnyelv problematikáját:

A fordításnyelv kifejezéshez erős negatív konnotációk járulnak: 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, 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 szöveg-szintű sajátosságainak tárgyalása során is” (Károly 2007:30). A célnyelvi szövegek szövegépítő sajátosságainak leírásánál azonban figyelembe kell venni, hogy e szövegek a kultúrák közötti közvetítés termékei, s így „a szöveget nem önmagában vizsgáljuk, hanem felhasználója világról és társadalomról alkotott tudása függvényében” (Károly 2007:18).

3.1. 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 információáramlás felerősödésének következtében a hétköznapi élet számos területén megtöbbszöröződött a fordított szövegek mennyisége (vásárlói tájékoztatók, hírek, uniós dokumentumok stb.), melyeknél azonban gyakran az autentikus 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 (vö. Károly 2007:47), és pedig két különböző kultúrához kapcsolódva. A fordítási szövegalkotás tehát a kultúrák közötti 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 alkotja meg Klaudy (2007) a „nyelvek fordítási viselkedése” (*translational behaviour of languages*) terminust. A „nyelvek fordítási viselkedése” ebben az esetben azt a jelenséget jelöli, amikor az adott forrásnyelvű szöveg másképpen viselkedik, ha különböző célnyelvekre kell lefordítani (Klaudy 2007:196).

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 és Mason 1990:223–224). Megfogalmazásukban a fordítási folyamat egyik jellemzője abból a kettősségből fakad, 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 esetlegessége ellenére is szerepeltetni kell a fordító olvasását. Ez a gondolat jelenik meg Bassnett cikkében is a fordítások értékelésével kapcsolatban: „A fordítási tevékenység alapvetően nemcsak az írást, hanem az olvasást is magában foglalja. A fordítások olvasásakor a forrásnyelvi szövegnek az egyes fordítók egyéni értelmezése szerinti olvasatát látjuk” (Bassnett 2003:42, a szerző fordítása)³.

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 az elsődleges és másodlagos szövegalkotásra ebből a szempontból jellemző különbséget. Az elsődleges szövegalkotás során a szövegépítő az egymással valamilyen logikai kapcsolatban álló gondolatai mentén haladva, a szövegfolyam kisebb egységeiből építi fel 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 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 (Péché 2010:76).

A fordított szövegek relativitása már a fordítói olvasásélmény egyedisége kapcsán is felvetődött. Ez a relativitás-gondolat azonban Neubertnél még hangsúlyosabban megjelenik, aki a fordított szövegek relativitására a forrásnyelvi szövegre gyakran jellemző eltérő értelmezési lehetőségek, illetve az egyfajta értelmezés esetén sem ritka különböző célnyelvi megfeleltetések kapcsán utal (Neubert 1994:97). A fordított szövegek relativitását Paepcke a forrásnyelvi szöveg kulturális és történelmi beágyazottságából fakadó, szövegszinten nem feltétlenül megfogható értelmezésbeli különbségeknek tulajdonítja

³ 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)

(Paepcke 1994:106–108). A fordított szövegek relativitását értelmezésében tovább árnyalja a fordítók szociokulturális kompetenciája, amely már az elsődleges szövegalkotást is nagymértékben befolyásolja. A szociokulturális kompetenciához köthető relativitás okozza a fordított szövegek feldolgozásakor a szükséges erőfeszítés nagyságának eltéréseit, amelyet 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)

A forrásnyelvi és célnyelvi szöveg kognitív környezete közötti különbség könnyen belátható, hiszen a „fordító kognitív környezete nem azonos a szerzőével, kivéve, ha saját művét fordítja” (Heltai 2009:20).

3.2. A célnyelvi szövegalkotásra jellemző sajátosságok

A fordított szövegek kontrasztív elemzésének központi témája az adott nyelv grammatikai és mondatalkotási szabályai szerint helyes, ám az anyanyelvi beszélő „szövegről alkotott intuitív elképzeléseinek” mégsem megfelelő célnyelvi szövegtulajdonságok feltárása (vö. Papp 1972/2006:122–123). 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 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 ered-

ményezik a különböző nyelveknél az egymástól eltérő kohéziós eszközhasználatot (Halliday és 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 alaposságára. 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 az eltolódást tovább árnyalják a különböző nyelvek stilisztikai preferenciái bizonyos kohéziós eszközök iránt (Blum-Kulka 1986).

A célnyelvi normától való eltérések okai között kell megemlíteni a fordító forrásnyelven és célnyelven történő gondolatainak szükségszerű interferálását, hiszen „interkulturális fordítást hajt végre, amely egy adott kultúrában létező fogalmaknak egy másik kultúrába való átültetését jelenti” (Simigné 2008:126). A fordító a két kultúrában egy időben cselekszik, és egyszerre befolyásolja a keletkező célnyelvi szöveget a két kultúrában gyökerező, eltérő gondolkodásmód. 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 ugyanakkor hangsúlyozza „a világ nyelvi képének rekonstrukciójával” való foglalkozást, amelyet a kognitív nyelvészet maga „a nyelvben rögződött objektív valóság interpretációjaként értelmez” (Bańczerowski 1999:82). Számos modern nyelvészeti kutatás kiindulópontja az a tény, hogy a különböző nyelvek eltérő módon ragadják meg a valóságot (pl. metaforakutatás). Ez az eltérő interpretáció hatással lehet a fordított szövegek keletkezésére, mivel ugyanazon gondolatok manifesztációja az eltérő nyelvek nyelvi jelei segítségével eltérhet egymástól.

Fontos a fordított szövegek tanulmányozása 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 kontrasztív elemzéseknek köszönhetően felszínre kerülhetnek az egyes nyelvek vonatkozásában a denotatív világ elemei 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)

3.3. Fordításnyelv – kvázi-helyesség

A fordított szövegek szövegépítő sajátosságaira már korábban is történt utalás. Az autentikus szövegprodukciónál az anyanyelvi olvasó számára feltűnően különböző célnyelvi szövegekkel a hazai és a nemzetközi kutatások is egyre gyakrabban foglalkoznak. „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 kódznak (Frawley 1984) nevezi. A fordított szöveg nyelvi mintázatainak vizsgálatáról írt átfogó cikkében Balaskó a következőképpen vélekedik: „a szöveg nem véletlenszerűen kiválasztott egymást követő szavakból keletkezik, hanem számos előre vagy félig előre gyártott szerkezetből áll össze, amelyek adott helyzetben az egyetlen választási lehetőséget jelentik akkor is, ha látszólag ezek a szerkezetek szétválaszthatók” (Balaskó 2020:58). Egyazon szöveg különböző nyelvű változatai esetében a nyelvi szerkezetek nem feltétlenül fedik egymást teljes mértékben, hiszen az egyes nyelvek eltérően strukturálják a valóságot. A fordított szövegek esetében a forrásnyelvi szöveg szövegépítő sajátosságainak hatása miatt előfordulhat, hogy a célnyelvben bizonyos nyelvi megformálások idegenül hatnak.

House (1981, 1997) a fordítások minőségének értékelésére kidolgozott modellje a forrásnyelvi és célnyelvi szöveg szemantikai, pragmatikai és textuális egybevetésével vizsgálja az ekvivalencia szinteket. A hibák azonosításánál egyértelmű hibákról (*overtly erroneous errors*) és burkolt hibákról (*covertly erroneous errors*) beszél:

Ha a megfelelően fordított szövegre érvényes a terjedelmi és ennek következtében a funkcionális egyezés követelménye, akkor ennek értelmében minden terjedelmi eltérés hibának minősül. Az ilyen terjedelmi eltéréseket burkolt hibáknak neveztük. Ezeket megkülönböztettük azoktól az egyértelmű hibáktól, amelyek vagy a forrás- és célnyelvi szövegelemek denotatív jelentéseinek eltéréseiből, vagy a célnyelv nyelvi rendszerének megsértéséből erednek. (House 1997:45, a szerző fordítása)⁴

Ide kapcsolódik az a kérdés, hogy különbséget kell tenni hiba és kvázi-helyesség között. A különbség elsősorban úgy írható le, hogy míg a kvázi-helyes fordított szöveg híven visszaadja a forrásnyelvi szöveg jelentését és megfelel a célnyelv nyelvi rendszerének, addig a hibás célnyelvi szöveg egyes részleteinek denotatív jelentése eltér az azonos

⁴ If a translation text, in order to be adequate, is to fulfill the requirement of a dimensional, and as a result of this, a functional match, then any mismatch along the dimensions is an error. Such dimensional errors were referred to as covertly erroneous errors. These were differentiated from those overtly erroneous errors which resulted either from a mismatch of the denotative meanings of source and translation text elements or from a breach of the target language system. (House 1997:45)

forrásnyelvi szövegrészlet denotatív jelentésétől és/vagy a fordított szöveg megsérti a célnyelv nyelvi rendszerét. Az előbbi logika alapján a fordító által nem az anyanyelvére fordított szövegben megjelenő eltérések is hibás célnyelvi szöveget eredményeznek.

A szövegszint elsődlegességét hangsúlyozó kutatások az 1970-es évektől kezdve egyre gyakrabban foglalkoznak az autentikus- (primer) és a fordítási (másodlagos) szövegprodukcóra jellemző szövegépítési sajátosságokkal. A fordított szövegek esetében gyakran megfigyelhetők az elsődleges szövegalkotástól eltérő, vagy részben eltérő sajátosságok, amelyet a szakirodalom kvázi-helyességnek (*quasi-correctness*) nevez. Ezeket a szövegszinten jelentkező különbségeket sokszor nem is lehet a mondatok szintjén észlelni és/vagy kimutatni. A kvázi-helyesség ezen értelmezése az adott célnyelv grammatikai szabályainak tökéletesen megfelelő, a szövegszinten érzékelhető eltérések (pl. egyes szerkezetek, lexikai egységek szokatlan disztribúciója vagy szokatlan retorikai szerkezetek) miatt mégis idegenül ható fordítások minőségére utal. A fordításnyelv (*translationese, translated language*) tehát a természetes anyanyelvi szövegalkotástól eltérő célnyelvi szövegprodukciónak jellemzője, az autentikus nyelvi környezetbe nem tökéletesen belesimuló fordított szöveg sajátossága (vö. Péch 2011).

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 a formailag helyes, de bizonyos minőségi követelményeket nem kielégítő fordított szövegek leírására és jellemzésére. (Schmidt 1995:20, a szerző fordítása)⁵

A fordított szövegek kvázi-helyes volta kapcsán Papp a mondathatárokon túlmutató összefüggések figyelembevételének szükségességét hangsúlyozza (Papp 1972/2006:133). Lindström a finnből fordított turisztikai brosúrák szavainak szófaji vizsgálata kapcsán állapítja meg, hogy a fordított szövegek szófaji eloszlása eltér az autentikus angol brosúrákra jellemző szófaji statisztikai jegyeitől és ezért a finnből fordított szövegek kvázi-helyesek (Lindström 1996). A kvázi helyesség kapcsán Klaudy mindezek mellett a kommunikatív szempontokra is felhívja a figyelmet.

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 eredménye-

⁵ 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)

képpen létrejött célnyelvi szöveg egésze olyan legyen, mint amilyenek az eredetileg célnyelven fogalmazott szövegek szoktak lenni. (Klaudy 1987:36)

A Klaudy-féle fordítási szövegalkotási gondolatmenetet folytatva Balaskó 2007-ben az eredeti angol, fordított angol és fordított magyar szövegeket tartalmazó korpuszt vizsgálja a fordított szövegekre jellemző, valamint az autentikus szövegalkotási sajátosságoktól eltérő harmadik kód jellegzetességei alapján. A fordításnyelv újradefiniálása során Balaskó különbséget tesz a fordítási univerzálék és a célnyelvi jegyek célnyelvi normától eltérő használata között:

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).

3.4. A szövegszintű eltolódások okai

Az eltolódások okainak feltáráásával foglalkozik Catford (1965), aki a formális megfelelést megkülönbözteti a szövegekvalenciától. Míg a formális megfelelés szerinte a nyelvi rendszerek különbözősége miatt nagyon ritka, addig a mondat szintjén megvalósuló ekvivalencia az egymásnak megfelelő forrásnyelvi és célnyelvi egységek 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. Catford 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*) esetében különbséget tesz a nyelvtani szerkezet különbözősége vagy a szórend változása miatt bekövetkező strukturális eltolódások (*structure shifts*), a morféimák, szavak, tagmondatok szintjén végbemenő egységbeli eltolódások (*unit shifts*), az adott elem szófajától eltérő szófajú elemmel történő fordításból 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 érvényű 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 esetében három eltérő okot különböztethetünk meg: (i) a fordítói hibákat, (ii) azokat az eseteket, amikor a fordító meg kívánja könnyíteni a célnyelvi befogadók szövegfeldolgozását (iii) és a célnyelv és forrásnyelv rendszeréből adódó

különbségeket. A fordítói megértésből, transzferből, szövegalkotásból vagy figyelmetlenségből adódó hibák jellegüknél fogva a fordításoktatás és a fordítói kompetenciafejlesztés területéhez tartoznak. Ezeket a fordítók szövegértési, szövegalkotási, idegennyelvi és anyanyelvi kompetenciájának fejlesztésével, továbbá a fordítás közben aktivált szöveg-szintű műveletek (grammatikai, lexikai és mondatszerkezeti 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 olvasó szövegfeldolgozását. Ezeket az eseteket a fordítástudományi szakirodalom több kutatási terület kapcsán is behatóan tárgyalja. Itt kell említést tenni többek között a műfajkutatásról, az ekvivalenciakutatásról, a fordítási univerzálék kutatásáról (explicitáció, implicitáció, egyszerűsítés, normalizáció), a reáliakutatásról és a relevanciaelméletről.

A fordítói hibák és a célnyelvi befogadók szövegfeldolgozásának megkönnyítése miatti eltérések a fordító személyével és a fordítás körülményeivel kapcsolatba hozható, előre pontosan nem meghatározható esetekben következnek be, elsősorban a fordítás adott pillanatára jellemző szövegértelmezés és az átvitelrel kapcsolatos egyedi döntések függvényében. A célnyelv és forrásnyelv nyelvi rendszerének különbözőségeire visszavezethető eltérések ezzel szemben bizonyos mértékig előre jósolhatók, hiszen ezekben a rendszernyelvészeti magyarázattal bíró esetekben számolni kell az egyes nyelvpárok, illetve a különböző fordítók által előnyben részesített átváltási műveletekkel. Ide kell sorolni a szövegfelszín szemantikai kapcsolatainak nyelvspecifikus voltára visszavezethető, a lexikai kohéziós elemek eloszlásában bekövetkező változásokat.

4. A fordítói kompetencia

A fordítói kompetencia leírásakor szakmai alapvetés, hogy a „nyelvi kompetencia szükséges, de nem elégséges feltétele a fordításnak” (Heltai 2002:112). A barcelonai PACTE kutatócsoport 2003-as fordítókompetencia-modellje hat alkompetenciát különböztet meg: a két nyelv ismeretének alkompetenciája mellett megtalálható a nyelvektől független enciklopédikus tudás alkompetenciája, a fordítással kapcsolatos ismeretek alkompetenciája, az eszközhasználatra vonatkozó instrumentális alkompetencia, a stratégiai kompetencia és a pszicho-fiziológiai kompetencia (PACTE 2005).

Az autentikus és a fordítói szövegalkotás folyamat- és produktumorientált megközelítései a két művelet különbségeire alapozva tárgyalják a két eltérő szövegalkotási folyamathoz szükséges kompetenciák mibenlétét. „Szükség van egy forrásnyelvi szövegre, amelyet a fordító a maga módján fordít le. A fordítás tehát igen bonyolult tevékenység, olvasási és (újra)írási szakmai jártasságot is feltételez” (Bassnett 2003:42, a szerző fordítása)⁶. Napjainkra ugyanakkor egyre inkább jellemző az az álláspont, amely a fordított

⁶ 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)

szöveget a forrásnyelvi szöveggel kommunikatív és funkcionális aspektusból is egyenértékű szöveggént kezeli (Klaudy 2001). E felfogás szerint a célnyelvi szöveg önmagában is képes helyállni a célnyelvi kontextusban, azaz az adott információt a célnyelvi normáknak megfelelő nyelvi, pszichológiai és társadalmi kontextusba helyezni. A célnyelvre vonatkozó nyelvi és szövegalkotási kompetenciák mellett a fordítónak szüksége van forrásnyelvi kompetenciára, 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 szamontartja az adott forrásnyelvre és célnyelvre vonatkoztatva a nyelvi, szociolingvisztikai és 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), akárcsak 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 szókinccs gazdagságára és használatának pontosságára, a nyelvtani pontosságra és a helyesírás pontosságára (Európai Tanács Nyelvpolitikai Osztály 2001/2005:57–58). 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)” (Európai Tanács Nyelvpolitikai Osztály 2001/2005:31). 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 (Európai Tanács Nyelvpolitikai Osztály 2001/2005:31). A KER foglalja továbbá a stratégiai kompetenciák meglétével, amelyek segítenek a végszó felismerésben és a következtetések levonásában, valamint a tervezés, a kompenzálás, a monitorozás és a javítás folyamatainak végrehajtásában (Európai Tanács Nyelvpolitikai Osztály 2001/2005:58).

A fordítási kompetencia értékelése során a KER a következő kompetenciák aktivizálását ellenőrzi a kommunikatív nyelvi tevékenységekre vonatkozó skálák alapján: (i) forrásnyelvi olvasási szövegértési kompetencia, (ii) a szöveg információtartalmának részletes megértéséhez szükséges forrásnyelvi kompetencia, (iii) a forrás- és célnyelv rendszerére vonatkozó szociolingvisztikai ismeretek, (iv) a célnyelv nyelvi sémáinak aktivizálása és (v) a célnyelvi rendszerben a szövegalkotáshoz szükséges nyelvi kompetencia.

A fordítói kompetenciák leírására és modellálására született számtalan tudományos igényű munka ismertetése helyett csupán annak hangsúlyozása fontos a jelen kontextusban, hogy a fordítónak a nyelvi kompetencián belül külön alkompetenciákra van szüksége ahhoz, hogy a célnyelvi szöveg esetében megteremtse a célnyelvi normáknak megfelelő tulajdonságokat. Erre vonatkozó iránymutatást tartalmaz a KER előzőekben ismertetett kompetenciaskálája, vagy például Károly *Szövegtan és fordítás* (2007) című munkája:

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 fordításoktatás és a fordítóképzés feladata tehát a hatékony kommunikáció megteremtéséhez szükséges nyelvi viselkedésforma tudatosítása és fejlesztése, ahogy ez az anyanyelvi kommunikáció fejlesztése esetében is fontos kritérium.

5. Norma és ekvivalencia a fordításkutatásban

Az előzőekben a fordított szövegek keletkezését esetlegesen befolyásoló tényezőkről esett szó, most a hivatásos fordítók által készített minőségi fordításokhoz kapcsolódó fordítási norma, illetve a fordítási ekvivalencia kerül középpontba.

5.1. A fordítási norma

A szakma által alkalmazandó, széles körben elfogadott módszerek és stratégiák a tekintéllyel, befolyással bíró kompetens fordítók tevékenységéhez köthetők. Ezek persze korántsem abszolút érvényű elvárások, hanem mindig az adott kor, adott szociokulturális környezet igényeinek felelnek meg. Másként fogalmazva a normatív fordítási magatartás olyan normavezérelt stratégiát jelent, amelyet a kompetens hivatásos fordítók adott köre az adott szociokulturális környezetben nagy valószínűséggel alkalmaz. A fordítási norma tehát az adott korban, az adott szociokulturális környezetben rendszeresen választott, tipikus fordítói magatartás.

Tourynál (1995) a kompetencia az összes rendelkezésre álló megoldási lehetőséget, a performancia az ezek közül az adott pillanatban éppen aktuálisan megvalósuló megoldást

dást jelenti. 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 felosztás szerint napjainkban előtérbe kerül a célnyelv követelményeinek való megfelelés elve; sőt egyes területeken – mint például a multinacionális cégek kommunikatív szempontból meghatározó dokumentumai (pl. termékleírások, tenderdokumentációk) – kifejezetten hangsúlyossá válik a célnyelvi olvasó elvárásainak való megfelelés.

Chesterman (1993) normaelmélete 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 megoldását – a forrásnyelvi szöveg, a megbízás körülményei, a célnyelv és a célnyelvi közönség határozza meg. A kompetens fordítók által rendszeresen alkalmazott stratégiák idővel normatív törvényszerűséggé válhatnak.

Heltai (2004) 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” (Heltai 2004:414). Míg egy nyelv grammatikai struktúráját kötelező érvényű szabályok határozzák meg, addig Heltai a lexika, a mondat és a szöveg-szintű 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 mérvadó szempont a célnyelvi közösség igényeinek elsődlegessége:

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 kifejezésvariáns 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, amelyet a kommunikációs cél érdekében a kommunikációs célcsoport igényeihez kell igazítani.

Wilssnél (1996) ez a kérdés a fordítási nehézséggel kapcsolódik össze, azaz a fordító számára kevésbé nehéz feladat a szöveg fordításközpontú elemzése, mint a forrásnyelvi szövegből a funkció-azonos célnyelvi szöveg létrehozása. Wilss a nyelvhasználat következő területein vizsgálja a fordítói tevékenységet meghatározó szabályokat: 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, másrészt az olyan állandósult kifejezéseket, amelyek használatuk gyakorisága miatt kvázi szabályértékűvé emelkedtek. Ezeket a szituációfüggő, ismétlődő kifejezéseket tekinti Wilss kommunikációs kompetenciánk meghatározó elemeinek. Az egyéni stílus esetében pedig éppen a példaértékű kifejezésmódokkal szemben megjelenő nyelvfórmáló erők szabad játékát hangsúlyozza.

5.2. Az ekvivalencia

A minőségi fordításokkal szemben támasztott elvárások alappillérei a forrásnyelvi szöveg (1) tartalmának, (2) hangulatának és stílusának lehető legteljesebb visszaadása, (3) a célnyelvi szöveg természetes és közérthető megfogalmazása és (4) a forrásnyelvi szöveg hatásához hasonló célnyelvi hatás kiváltása. E négy kritérium közül bármelyik elsődlegessége szükségessé teszi a különböző ekvivalenciafelfogások áttekintését. Hause megfogalmazásában „[a] fordított szöveg alapvető jellemzője a kettős kötődés: egyrészt a forrásnyelvi szöveghez, másrészt a célnyelvi befogadó kommunikációs állapothoz. Ez a kettős kötődés az alapja annak, amit számos szövegnyelvészeti megközelítés ekvivalencia-relációnak nevez” (Hause 1997:24, a szerző fordítása)⁷.

A forrásnyelv és célnyelv különböző nyelvi egységei közötti ekvivalencia viszony szintjei Baker (1992) szerint: (1) a szavak szintje, (2) a szavak szintje fölötti, de még nem mondatszintű lexikai egységek (kollokációk, idiómák, állandósult szókapcsolatok) közötti szint, (3) a lexikai kategóriába nem sorolható (például számok, nemek miatti) nyelvtan szintje, (4) a szövegszint, ahol ekvivalenciaviszony jöhet létre (i) a tematikus- és információstruktúra terén, (ii) a szövegkohézió terén és (iii) a szövegek különböző kommunikációs szituációkban történő használatát tükröző pragmatika terén.

Catford a *Linguistic Theory of Translation* című művében (1965) megkülönbözteti a formális megfelelést és a szövegek ekvivalenciáját. A formális megfelelést (*formal correspondence*) azon célnyelvi kategóriák (lexikai egységek, grammatikai struktúrák, szerkezetek, szerkezeti elemek), esetében alkalmazza, amelyek a célnyelv rendszerében ugyanazt a helyet foglalják el, mint a forrásnyelvi elem a forrásnyelv rendszerében. Catford szerint a formális megfelelés a nyelvi rendszerek különbözősége folytán nagyon ritka és sokkal inkább

⁷ The fundamental characteristic of a translation is that it is a text that is doubly bound: on the one hand to its source text and on the other hand to the recipient's communicative conditions. This double-binding nature is the basis of what has been called in many linguistic-textual approaches the equivalence relation. (Hause 1997:24)

a mondat szintjén hozható létre ekvivalenciaviszony. Ekkor viszont már szövegekvivalenciáról (*textual equivalence*) beszél, amely bármely a forrásnyelvi szöveggel vagy szövegrésszel egyenértékű célnyelvi szöveg vagy szövegrész esetében megvalósul, amikor a forrásnyelvi és célnyelvi ekvivalensek esetében a szituáció szempontjából releváns jegyek azonosak.

A forrásnyelvi üzenet formájának és tartalmának lehető leghűbb visszaadására törekvő, a forrásnyelvi szövegre és kontextusra összpontosító formális ekvivalenciára épülő fordítás Nida (1964) ekvivalenciafelfogásának is részét képezi. Nida ezzel szembeállítja a célnyelvi befogadó kulturális közegének szövegalkotási elvárásait szem előtt tartó, a forrásnyelvi befogadó forrásnyelvi szöveghez való viszonyával azonos célnyelvi hatás kiváltására törekvő fordítói magatartást. A dinamikus ekvivalencia (*dynamic equivalence*) meglétekor a fordítások a legtermészetesebb módon illeszkednek a célnyelvbe és a célnyelvi kultúrába, így a célnyelvi közönség számára természetesen hatnak. A természetes hatás érdekében a fordítók a grammatikai és lexikai elemek átváltásánál a megfelelő stilisztikai elemek kiválasztására törekszenek.

Toury (1980, 1991, 1993) az ekvivalencia viszony kritériumainak feltárása helyett az adekvátságra (*adequacy*), a célnyelvi követelményeknek való megfelelésre (*appropriacy*) és az elfogadhatóságra (*acceptability*) 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 szövegek elfogadhatóságát. Ennek értelmében, amikor az adott célnyelvi befogadó közösség különbséget tesz a megfelelő és a nem megfelelő fordítások között, voltaképpen az ekvivalenciának tulajdonít egyfajta funkcionális jelentőséget.

Klaudy (1987:37) 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 a szövegeket a kommunikatív ekvivalencia hiánya miatt. Klaudy a kommunikatív ekvivalencia előfeltételének tartja a tartalmi, a kontextuális és a funkcionális ekvivalenciát.

Wilss (1996) a fordítási szabályokkal kapcsolatban két esetet különböztet meg a fordítói gyakorlatban. A fordító bizonyos esetekben olyan szöveghelyekkel találja magát szemben, amelyeket könnyen megért és így a célnyelvi megfelelőket is könnyedén előállítja. Ebben az esetben Wilss „prestabilizált ekvivalencia viszonyról” beszél. Ugyanakkor azoknál a szövegrészeknél, amelyek fordítása jó problémamegoldó készséget igényel, a fordító csak viszonylagos ekvivalencia megteremtésére törekedhet.

Neubert (1994) a translációs folyamat viszonylagosságára hívja fel a figyelmet, mivel még a legtapasztaltabb és legrutinósabb fordítók munkáinak minősége is némi ingadozást mutat. Ezt szerinte sokan – tévesen – a fordítói folyamat szubjektív jellegével magyarázzák, holott a viszonylagosság a fordítások forrásnyelvi szöveghez való viszonyából fakadó objektív adottság. Mivel a szöveg maga határozza meg a grammatikai és lexikai választások lehetőségének körét, ezért a szöveg lesz felelős az alsóbb szintekre jellemző relatív ekvivalenciáért.

Reiss (1984) szerint az ekvivalencia szintjét a célnyelvi szöveg (gyakran a forrásnyelvi szöveg céljától 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 fordítás a szószint mellett a mondszerkezetek helyességét is előtérbe helyezi. A továbbra is forrásnyelviszöveg-orientált filológiai fordítás az előbbieken túl a stílus tekintetében is törekszik az ekvivalenciára, míg a kommunikatív fordítás elsődlegesen a célnyelvi befogadó igényeihez igazodik. E kategóriák mentén Reiss különbséget tesz a két nyelv nyelvi jelei közötti megfeleltetésekre vonatkozó, kontrasztív nyelvészeti (langue típusú) és a fordítástudomány szövegközpontú (parole típusú) ekvivalenciája között. Értelmezésében a szövegközpontú ekvivalencia az adott szöveg nyelvi jeleinek a két nyelvközösség adott szociokulturális kontextusába ágyazott megfeleltetése. Egy az egyhez megfeleltetés szerinte csak a gépi fordítás esetében lehetséges, mivel a humán fordítás a nyelvközösségek szituatív nyelvhasználati különbségeiből adódó eltérések következtében kizárja azt (Reiss 1984).

A példa alapján három következtetést lehet levonni: 1. A forrás- és célnyelvi szöveg közötti ekvivalencia megteremtését a nyelvi szerkezetek különbözősége befolyásolja és irányítja, és ez a különböző nyelvpárok esetében különböző megoldásokhoz vezethet. 2. Az ekvivalencia megítélése egyben értékelési szempont is [...] 3. Az ekvivalencia nem jelent 1:1 megfelelést, amelynek nincs alternatívája. Több fordítás is lehet egymással ekvivalens. (Reiss 1984:166, a szerző fordítása)⁸

A fordítók döntéseit egyre kevésbé a szavakhoz és a szűkebb értelemben vett szövegűséghez való ragaszkodás jellemzi, bár a szöveg információközvetítő funkciójának tiszteltben tartása továbbra is alapkövetelmény. Egyre inkább előtérbe kerül a kommunikatív hatékonyság igénye, ami a célnyelvi befogadó szövegfeldolgozásának megkönnyítését tartja szem előtt. Heltai (2009) a fordítás és a 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 azonosságként 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 koncentrálnak (Heltai 2005b). Így például ma már az interkulturális gazdasági kommunikációval, a szakszövegek fordításával foglalkozó tudományos meg-

⁸ 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. (Reiss 1984:166)

közelítés nem a kultúraspecifikus normák és hagyományok leírására helyezi a hangsúlyt. Ugyanakkor jellemző az egységes tárgyalási kontextus, amely a különböző kultúrkörökből származó emberek interakciója és közös munkája során megjelenő eredményeket helyezi előtérbe (Kelz 2002).

Sager (1994) szövegalkotó- és szövegbefogadó stratégiákat különböztet meg (*sender-orientierte und empfängerorientierte Strategie*), aszerint, hogy a fordítás megrendelője magának a forrásnyelvi szövegnek tulajdonít nagyobb jelentőséget, vagy inkább a célnyelvi befogadó feldolgozását tartja szem előtt.

A különböző ekvivalenciafelfogások alapján le lehet szögezni, hogy egyrészt a teljes ekvivalencia illúziójával szakítva a fordított szövegekben az ekvivalencia relativitásáról kell beszélni, másrészt viszont a célnyelvi szövegek esetében léteznek már feltárt univerzális szövegalkotási jellemzők, amelyeket a következő fejezet tárgyal.

6. A fordítási univerzálé

A nyelvészet és nyelvelírás hagyományosan foglalkozik az emberi nyelv univerzális tulajdonságainak feltárásával. Az univerzálék csoportjába a nyelv azon tulajdonságait szokás sorolni, amelyek nemcsak egy nyelvre vagy nyelvcsoporthoz, hanem minden nyelvre, vagy legalábbis a nyelvek többségére jellemzőek. A kutatók az univerzálékat fajtájuk szerint három nagyobb csoportba sorolják, amelyek (i) a nyelvi elemzési kategóriákat jelentő szubsztantív (főnév, igealak stb.), (ii) az elemkombinációk absztrakt szabályait tömörítő formai (nyelvtan) és (iii) a nyelv két tulajdonsága közötti állandó összefüggéseket jelölő implikációs (ha – akkor) univerzálék. A nyelv univerzális tulajdonságait kutató megközelítés előbb vagy utóbb szembesül a reprezentativitás kérdésével, és a leírni kívánt univerzális tulajdonság relatív vagy abszolút jellegével (vö. Crystal 1998:112).

Az univerzálékutatás hagyományaihoz igazodva fordítási univerzálénak azokat a nyelvi jegyeket tekintjük, amelyek a fordítási folyamatban aktuálisan szereplő nyelvpároktól függetlenül jellemzik a fordított szövegeket. A fordítástudományi szakirodalom a célnyelvi szövegek vonatkozásában a *Routledge Encyclopedia of Translation Studies*ban leírt ismérvek (Laviosa 2001:288–291) alapján öt alapvető univerzális tulajdonságot különböztet meg, melyek (i) az egyszerűsítés és a forrásnyelvi szöveg ismétléseinek kerülése, (ii) az explicitáció, (iii) az úgynevezett „normahűség” vagy konzervativizmus, (iv) a diskurzustranszfer és az interferenciatörvény, valamint (v) a célnyelvi lexikai elemek eltérő előfordulási gyakorisága.

Az egyszerűsítés lehet lexikai (tágabb jelentéskörű kifejezések alkalmazása – pl. *kajfli/péksütemény*), szintaktikai (bonyolult mondatstruktúrák egyszerűsítése) vagy stilisztikai (hosszú mondatok felbontása, körülírások lerövidítése, bonyolult kifejezésmód helyetti rövidebb szókapcsolatok). A forrásnyelvi szövegben előforduló szó szerinti ismétlések kerülése talán a célnyelvi szövegek egyik legjellegzetesebb tulajdonsága.

Már korábban is szó esett Blum-Kulka explicitációs hipotéziséről, amely szerint a fordítási folyamat – mint kommunikációs aktus – szükségképpen kohéziót és koherenciát érintő eltolódásokkal jár. Kohézió alatt a szövegelemek közötti nyílt/látható (grammatikai), a nyelvspecifikus eszközök (markerek) által jelölt összefüggéseket érti, míg a koherencia a szövegelemek közötti belső tartalmi (szemantikai, pragmatikai), az olvasóban zajló értelmezési folyamatok révén megnyilvánuló kapcsolat. Az explicitációs hipotézis szerint a szöveg szerveződését meghatározó kohéziós-jelölők a célnyelvi szövegekben gyakran eltolódnak. Az esetleges fordítói betoldások következtében a célnyelvi szöveg terjedelme általában meghaladja a forrásnyelvi szöveg terjedelmét. A forrásnyelvi szöveg értelmezését segítő nyíltabb, világosabb kifejezésmódnak köszönhetően – azaz a nyelvi közvetítésre jellemző általános stratégia következtében – a célnyelvi szöveg a forrásnyelvi szöveghez képest explicitebb. Akkor beszélünk explicitációról, ha a forrásnyelvi szövegben található implicit információkat tartalmazó szemantikai elemeket a befogadó nyelvben explicit módon kell kifejezni. A leggyakrabban alkalmazott explicitációs technikák: a megjegyzések közbevetése; a kötőszók, a módosítószók, a minőségjelzők, a kiegészítések, a magyarázatok, a kulturális háttérinformációk betoldása; a kétértelmű információk pontosítása (vö. Blum-Kulka 1986).

A fordításokra jellemző normahűségről 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, az írásjelek módosítása, a befejezetlen mondatok befejezése, a logikusabb szövegtagolás megvalósítása, az elbeszélésmódban a hagyományosabb formákhoz való igazodás, az idejétmúlt kifejezések helyett modern kifejezések használata stb. (vö. Vanderauwera 1985). A fordítás során lejátszódó mentális folyamatok hatására a fordítók hajlamosak a forrásnyelvi szöveg sajátosságait átvinni a célnyelvi szövegbe. Minél inkább hatnak a forrásnyelvi szöveg sajátosságai a célnyelvi szövegre, annál inkább tetten lehet érni a célnyelvi szövegben az interferencia jeleit. A forrásnyelvi szöveg sajátosságainak helytelen átvitele a negatív transzfer (vö. Toury 1995). A célnyelvi szöveg lexikai elemeinek eltérő előfordulási gyakoriságát Baker a nyelvi közvetítési folyamat alapvető tulajdonságai közé sorolja, mely folyamat bizonyos esetekben a fordított szövegek egyes lexikai elemeinek szokatlan szövegbeli eloszlását is eredményezheti (vö. Baker 1993, 1996, Blum-Kulka 1986).

III. Szövegnyelvészeti megközelítés

1. Bevezetés

„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) (vö. 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 és Hasan 1976). A mondatszintű 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: „Tekintettel arra, hogy a nyelvi kommunikáció mindig szövegek segítségével történik, az olyan területeket, mint az anyanyelv- és idegennyelv-oktatás, a fordítás stb. egyre inkább szövegelméleti alapokra kell helyezni. A nyelvészeti szövegelmélet megalkotása tehát nemcsak a nyelvészet, hanem a társadalom egésze számára is sürgető” (Isenberg 1976:47)⁹.

Kifejezetten a szöveg primátusát hirdette a német nyelvterületen megjelenő, a nyelvészetet a szövegnyelvészettel azonosító irányzat: „*Linguistik ist Textlinguistik*” (Weinrich 1967:109). A szöveget elsődleges nyelvi jelnek tekintő irányzat szerint az emberi írott és szóbeli kommunikáció nem mondat-, hanem szövegszinten zajlik, és a nyelv alapegysége

⁹ 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)

(vö. Schiewe 2009:95), illetve a fordítás egyetlen lehetséges egysége is a szöveg (vö. Dressler 1973:12), bár a kölcsönösen releváns szövegek maguk is összekapcsolódnak tartalmilag, formailag és funkcionálisan is a diskurzus keretében. 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 kommunikációelmélet, a pszicholingvisztika, a beszédaktuselmélet, a tartalomelemzés, a diskurzuselemzés, a modern retorika és a kognitív nyelvészet is.

A szöveg tanulmányozásával foglalkozó területek 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*). A jelen munka a szöveggel foglalkozó tudományterület megjelölésére a magyar szakirodalomban elterjedt szövegkutatás kifejezést (vö. Károly 2007:16) használja.

2. A szövegkutatás feladata

A nyelvi szerkezeteken túllépő és már a teljes szövegegységre kiterjedő, a szövegek felépítésének kritériumait, illetve a szövegek használatának módját és célját kutató vizsgálatok elterjedésével egy időben megkezdődött a diszciplína szerepének és feladatkörének, a szöveg nyelvészeti alapú tudományának meghatározása. Dressler (1973:1) minden olyan vizsgálatot a szövegkutatás tárgykörébe utal, amely a szöveget tekinti az írásbeli vagy szóbeli emberi kommunikáció alapegységének, vagy amelyik legalább a mondatgrammatika határain túllépve mondatfolyamok vagy nagyobb szövegegységek vizsgálatával foglalkozik. A Kugler és Tolcsvai szerzőpáros (2000) a szövegkutatás 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, forгатókönyvekből származó) ismeretekkel, és azt, hogy e tényezők hogyan alakítanak ki a szövegben összefüggő jelentéshálózatot [...] (Kugler és 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.

A szövegten vizsgálati tárgya a szöveg, amely komplex nyelvi, kognitív és társadalmi jelenség: egyszerre nyilvánul meg produktumként (mint szerkezet) és folyamatként (mint műveletek sora). 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 az emberi interakciók során, így többek között a fordítási szituációban használt szövegek „működésének” leírásával foglalkozik. A fordított szövegek szövegközpontú vizsgálatánál az utóbbi évtizedekben három vizsgálati témakör került előtérbe: (i) a Halliday és Hasan (1976, 1980) óta népszerű kohéziós vizsgálatok, (ii) a swalesi (1990) hagyományt folytató műfajelemzések és (iii) a Dijk és Kintsch (1983) nyomdokait követő makrostrukturális tanulmányok (vö. Károly 2007:33). A jelen mű a lexikai kohéziós vizsgálatok sorát kívánja gazdagítani, ezért a szövegkutatás elméleti kereteit meghatározó megközelítések közül azokat a szakirodalmi hivatkozásokat tárgyalja, 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 szövegkutatás legfőbb feladatának. Károly Sándor (1979:55) ugyanakkor a szövegszerkesztés általános szabályainak, a „szövegtenilag jó mondat”, „jó mondatsor, jó szöveg fogalmának” tisztázását és a lineáris kohézióval való foglalkozást tekinti a szövegkutatás központi témáinak. Szikszainé Nagynál (1999:18) a szövegekkel foglalkozó tudományág kutatási céljai között szerepel „azoknak a feltételeknek, lehetőségeknek, általános szabályszerűségeknek, eszközöknek a feltárása, amelyek a szövegeképítéshez kellenek, illetve magának a szövegalkotás folyamatának kiderítése” valamint „a gyakorlati teendők szolgálata (a kommunikációs zavarok csökkentése vagy kiszűrése, a fordítás könnyítése stb.)”. Fehérnél 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. (2004) szövegdefiníciója a „legtágabb értelemben vett nyelvhasználatot” veszi alapul, és a világ éppen aktuális szegmentumára vonatkozó tudásunkat is a szöveg része-

ként 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 szövegkutatás feladata e felfogás szerint „valamennyi olyan nyelvészeti módszerekkel kimutatható reláció elemzése és leírása, amely ’átlép’ a »szövegmondatok«-nak tekinthető kommunikátum-összetevők közötti határokon” (Petőfi S. 2004:74). A napjainkban egyre hangsúlyosabbá váló szövegkutatási irányzat önmaga szűkebb keretein túllépve, „széleskörű interdiszciplináris megalapozásra” törekszik, a „szövegeket komplex jelekként kezeli, és vizsgálja azokat mind szintaktikai (vagy pontosabban: formai), mind szemantikai, mind pragmatikai nézőpontból” (Petőfi S. 2004:86). Petőfi S. a szövegkutatás területéhez sorol minden olyan kutatást, „amelynek célja, hogy a szövegek formai és szemantikai architektonikája nyelvészeti és hermeneutikai elemzéséhez és leírásához adekvát elméleti keretet biztosítson” (Petőfi S. 2004:73). Több évtizedes szövegkutatási tapasztalataira támaszkodva hangot ad továbbá annak a véleményének is, hogy a szövegkutatással 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, szövegszerkezet, szövegstilisztika, szövegtipológia, szövegfonetika, szövegformatan, amelyeket a szövegalkotás és szöveg-befogadás vizsgálata egészít ki (Szikszainé Nagy 1999:23).

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 szövegtermékek 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 szövegkutatás vizsgálati kérdései jelentőségükhöz mérten egyre több interdiszciplináris területet is érintenek. A fordítástudományi kutatások közül például az eredményközpontú megközelítés 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 ez a megközelítés „mind elméleti, mind módszertani szempontból jelentős elvekkkel és eredményekkel gazdagíthatja a fordítási szöveg 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 Tourytól (1995) származik, aki azt az irodalmi fordítások 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 (vö. Nord 2005). A német szakirodalom ugyanakkor szívesen és behatóan foglalkozik a fordító számára mindenképpen megoldandó feladatot jelentő, a forrásnyelvi szövegben tetten érhető defektusokkal is (vö. Schmitt 1998:147–151).

A fordítás kommunikatív és interaktív szerepének előtérbe kerülésével a figyelem egyre inkább a fordítás funkciójára, a befogadói elvárásokra, a célnyelvi kultúrára és az egyes szövegfajták követelményeit szem előtt tartó fordítói szándéokra irányul. Ezzel a megközelítéssel a hangsúly áthelyeződik a forrásnyelvi szövegről és a szöveg reprodukcióról a célnyelvi szövegprodukcióna. 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 szövegproduktum (Neubert és Shreve 1992). A fordított szöveg tehát egyedülálló intertextuális kapcsolatban áll a forrásnyelvi szöveggel és egyre

hangsúlyosabb annak feltárása, 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 ekvivalenciaközpontú megközelítés helyett egyre gyakrabban kerül előtérbe a célnyelvi befogadó számára könnyebben feldolgozható szöveget előállító funkcionális fordítás.

A fordítástudományon belül hangsúlyossá váló, a célnyelvi szövegre fókuszáló megközelítés (*target text oriented translation studies*) (vö. Toury 1995) a fordítási folyamat eredményeként 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 (vö. Komisszarov 2002).

4. A szöveg meghatározása

Az 1960-as évektől egyre nagyobb számban megjelenő szövegnyelvészeti tanulmányok eleinte a szöveget egyszerűen csak a mondatnál nagyobb egységként kezelték, majd az „egymást megszakítás nélküli követő névmási láncokból álló nyelvi egységek” (Harweg 1968:148)¹⁰, az egymást követő jól formált mondatok sorozataként definiálták. Csak azután tulajdonítottak a szöveg meghatározások a szöveg pragmatikai és funkcionális vetületének is kiemelkedően fontos szerepet, miután bizonyítást nyert, hogy a szövegszerűségnek nem elégséges kritériuma az egymással szomszédos mondatok szintaktikai-szemantikai összekapcsolódása. E felfogás szerint a nyelvi jelsor „egy tényleges vagy feltételezett kommunikáció-szituációban egy tényleges vagy feltételezett kommunikatív funkció betöltésére alkalmas” (Petőfi S. 1988:43).

A szövegvértékűség 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:

¹⁰ ein durch ununterbrochene pronominale Verkettung konstituiertes Nacheinander sprachlicher Einheiten. (Harweg 1968:148)

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, helyetszerűség, intertextualitás) eleget tesznek. A hét ismérv 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, amelyek 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) „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 szövegmeghatározások sorát a szövegszerveződés ismérveinek feltárására irányuló vizsgálatok és egy univerzálisan érvényes szöveggrammatika leírásának igénye előzte meg. A számos meghatározás közül Szikszainé Nagy (1999) *Leíró magyar szöveg-tan* című munkájából áll itt a szakirodalom legjelentősebb szövegdefinícióit összegző szövegmeghatározás:

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)

5. A szövegszerűség kritériumai

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

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 szövegelemeket a közös kontextus és a közös szituáció tartja össze. A szövegelemek közötti kapcsolatok rendszere ugyanakkor már a szövegszerűség szintaktikai aspektusának része.

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

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 jellemző a szövegszerűség hét ismérvének realizációja (De Beaugrande és Dressler 2000:23), amelyek segítik a szövegalkotót és a befogadót, 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 ismérve (szándékolttság, elfogadhatóság, hírérték, helyzetszerűség, intertextualitás) felhasználó-központú.

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

A kohézió a szövegfelszín alkotóelemei, azaz az egyes lexikai elemek 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 (részben) eltérő értelmet tulajdoníthat ugyanazon szövegnek, illetve „különböző mennyiségű és minőségű viszonyok hozzátételére van szüksége ugyanazon szöveg megértéséhez” (Károly 2007:31).

A szándékoltás 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 szövegbefogadó feldolgozási 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 anyanyelvi szövegalkotáshoz képest komplexebb fordítási szövegprodukciónál a nyelvi- és szövegalkotási kompetencia mellett kiemelt szerepet kap a kommunikációs aktus kultúra- és helyzetspecifikus vonatkozása. A fordítási szöveg mint kommunikatív esemény, akkor képes elérni a célját, ha funkcionálisan és a szövegszerűség kritériumai szerint is adekvát, azaz ha releváns a kommunikáció és a fordítás célja szempontjából 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 produktív és reprodukív 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 reprodukációjáért 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, a koherenciát illetően ezzel szemben csak az átvitelért felelős, mivel a fogalmak közötti viszonyokat nem ő, hanem a forrásnyelvi szöveg szerzője teremti meg. Előfordulhat

viszont, hogy a fogalmak közötti koherenciaviszonyok egy az egyben történő reprodukciója olyan értelemszerkezetű célnyelvi szöveget teremt, amelyet a célnyelvi befogadó nem tud feldolgozni és amely számára nem lesz koherens, mivel a forrásnyelvi szöveg szerzőjének és/vagy befogadójának háttértudásától eltérő háttértudással rendelkezik. Amennyiben a fordító ezt a veszélyt felismeri és képes a forrásnyelvi szöveg fogalmi és relációs hálóját a célnyelvi közönség számára értelmezhető módon átültetni a célnyelvre, akkor valójában aktívan részt vesz a célnyelvi szöveg koherenciájának megteremtésében, vagyis magában a szövegprodukciónban.

Klaudy (2006) szerint a fordító feladatai közé tartozik a forrásnyelvi szöveg intertextuális viszonyainak helyes dekódolása is, aminek alapján eldönti, hogy a fordítás milyen típusú célnyelvi szövegek közé fog majd illeszkedni. A fordító szövegtipológiai ismeretei meghatározóak a megfelelő szövegtípus megteremtéséhez szükséges célnyelvi eszközök használatához. A fordítás célja és funkciója, valamint a szöveg típusa és 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 töltse be kommunikatív funkcióját.

IV. A lexikai kohézió vizsgálata

1. Bevezetés

A fordítóknak, és a gépi fordítóprogramok készítőinek legfontosabb célkitűzése a célnyelvi kultúrától és nyelvezettől nem idegen célnyelvi szövegek előállítására (vö. Hönic 1997:26). A legtöbb fordítás esetében mégis éppen ezen a területen érhetjük tetten az idegenszerűséget. Az anyanyelvi olvasó számára a szövegszint jegyei a legárukodóbbak, azaz a szövegszerűség meglétéért felelős jegyek valamelyikének hiánya, vagy azoknak az adott nyelven, az autentikus szövegprodukciónak 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 vonatkozásait.

2. A kohézió kutatás legismertebb képviselői

A lexikai kohézió vizsgálata a szövegtudomány egyik vitatott ám meghatározó irányzata, melynek tárgya elsősorban írott szövegek leírása, funkcionális elemzése. A szövegtudomány 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 szövegnyelvészet megközelítés a szöveg szempontjából releváns lexikai elemek ismétlődését vizsgálja. Halliday és Hasan (1976) 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 kohézió 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 (De Beaugrande és Dressler 2000). Azonban De Beaugrande és Dressler mellett számos kutató felismerte a kohézió meglétének jelentőségét a szövegalkotás 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, kohéziót is eltérő minőségű és mennyiségű eszköz segítségével fejezik ki. (Károly 2007:62)

A lexikai kohézió vizsgálata iránti érdeklődés növekedése miatt napjainkra egyre népszerűbbek lettek a témakörben folytatott korpuszalapú kutatások. A lexikai kohéziókutatás korpusznyelvészeti feltárásának növekvő igényét bizonyítja többek között a Benjamins kiadó gondozásában megjelent, kifejezetten a témakörben megvalósított kutatások bemutatására szentelt gyűjteményes kötet. A *Lexical Cohesion and Corpus Linguistics*-ben (Flowerdew és Mahlberg 2009) megjelentetett hat tanulmány közül egy a kohézió és a retorikai struktúra kapcsolatát (Morley 2009), három a lexikai kohézió hangzó szövegekben történő vizsgálatát (Nesi és Basturken 2006, Warren 2006, Cheng 2006), kettő pedig a kohézió tanulói korpuszokban történő realizációját kutatja (Flowerdew 2006, Mahlberg 2006).

Morley (2009) kutatása egyrészt a klasszikus megközelítést követve az azonos szemantikai mezőbe tartozó szavak közötti logikai kapcsolatokat vizsgálja, másrészt az újságcikkek főcímei és szövegtartalma közötti összefüggéseket kutatja. A vizsgálat alátámasztja, hogy a címek kohéziós szerepe a sajtószövegekben igen fontos, mivel a címek alapján a cikkek lexikai tartalma előrevetíthető. 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 jelenthet a szöveg értelmezésénél nagy segítséget.

Nesi és Basturken (2006) a *British Academic Spoken English* (BASE)¹¹ és a *Michigan Corpus of Academic Spoken English* (MICASE)¹² tudományos előadásokat tartalmazó korpuszokat vizsgálták, melyekben a lexikai kohéziós elemeket, többszavas kifejezéseket és a gyakori együttes szóelőfordulásokat kutatták. A korpuszokban a WordSmith Tools cluster funkciójának segítségével azonosították a minimum tízszer ismétlődő 4 szavas együttes előfordulásokat, amelyek százalékos előfordulási gyakoriságát a *Longman Grammar of Spoken and Written English* (Biber és mtsai. 1999) társalgásra és tudományos cikkekre jellemző adataival vetették egybe. Az eredmények ismeretében a szerzőpáros rávilágít a négyszavas ismétlődő előfordulások szövegépítő jelentőségére és a kohézió megteremtésében játszott szerepére.

Warren (2006) az ötszavas hongkongi hotelekben készített, a pályázók szakmai karrierje szempontjából nagy jelentőségű állásinterjúkat elemezte, ahol a felek által az aktuális szövegtörzsetben az informativitás és a szöveg kohéziója szempontjából

¹¹ <http://www2.warwick.ac.uk/fac/soc/al/research/collect/base>

¹² <http://quod.lib.umich.edu/m/micase>

előnyben részesített lexikaválasztást vizsgálta. A vizsgált anyag a kétmillió szövegszavas *Hong Kong Corpus of Spoken English* (HKCSE)¹³ anyanyelvi angol beszélők és hongkongi kínaiak párbeszédeit tartalmazó korpusz része, amelyet Warren prozódiailag is annotált. A szógyakorisági listák, kulcsszólisták, konkordanciavizsgálatok és prozódiai elemzések eredményeként a kutatás végkövetkeztetése, hogy a beszélők a lexikai kohéziós elemeket részesítik előnyben a nem lexikaiakkal szemben. Megállapításuk összefüggésbe hozható azzal a feltételezésükkel, hogy a beszédhelyzetben aktiválódó 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*) egészségügyi szöveganyagot elemezte. A ConcGam©¹⁴ programmal azonosított hat leggyakoribb lexikai elem kollokációs kapcsolatainak feltárása kapcsán megállapította, hogy az együttes előfordulások identifikációja árnyaltabb képet ad a szövegek koherenciájáról annál, mintha csak a lexikai kohéziót vizsgálnánk a szövegekben.

Flowerdew (2006) és Mahlberg (2006) kutatásai az angol nyelv tanítása kapcsán foglalkoznak a kohézió kérdésével. 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. A kohézió témakörének az idegennyelv-oktatásban való integrálására a korpuszalapú módszert tartja a legmegfelelőbbnek.

3. 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 a befogadó is szöveggként értelmezze. A szövegen belül mind a makroszerkezeti, mind a mikroszerkezeti kapcsolódások felelősek a szövegegység és szövegösszefüggés megteremtéséért. A kapcsolódásokat Szikszainé Nagy (1999) a megvalósulásuk szerinti szövegszint alapján és a kapcsolatot jellemző utalás iránya alapján csoportosítja, ahogy ezeket a következőkben bemutatom. A szövegen belüli kapcsolatok fajtái szerint a pragmatikai jellegűek a koherencia, a szemantikai jellegűek a kohézió, a szintaktikai jellegűek a konnexitás, a szerkezeti szintűek a kompozicionális konnexitás, a stilisztikai jellegűek a stíluskohézió, a fonetikai-prozódiai jellegűek pedig az akusztikai kohézió. A szövegen belüli kapcsolatokat megteremtő eszközök eltérőek a különböző szövegszinteken, valamint változik az összekapcsoltság mértéke a különböző szövegtípusok esetében. A szövegen belüli utalás iránya alapján megkülönböztetünk anaforikus (visszafelé utaló), kataforikus (a szövegben előre utaló) és deiktikus (rámutató funkciójú) kapcsolódásokat (vö. Szikszainé Nagy 1999:57–72).

¹³ <http://rcpce.engl.polyu.edu.hk/HKCSE>

¹⁴ A ConcGram© Chris Greaves (Department of English, The Hong Kong Polytechnic University) által fejlesztett keresőmotor

A szövegegység és szövegösszefüggés megteremtését többek között a szövegelemek szövegen belüli változatlan és koreferenciális (ugyanarra a valóságszegmensre vonatkozó) ismétlődése biztosítja. Ezen ismétlődések részben a 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 szöveggörnyezetre, vagy akár a teljes szövegre jellemző globális kohézió között. A VI. és VII. fejezetben bemutatott kutatás a szöveg kisebb egységein belüli – az automatikus elemzések számára szövegszerkesztési és központoszási tulajdonságai révén könnyen megragadható, (az egyik beljebb kezdéstől a következő beljebb kezdésig tartó) bekezdés szintű – szemantikai jellegű összefüggés-kapcsolatok révén megvalósuló kohéziós viszonyokat írja le.

3.1. A kohézió meghatározása és a kohéziós eszközök

A lineáris kohézió a szövegek mikroszerkezeti szintjén, az egymást követő mondatokban, szövegegységekben megvalósuló grammatikai-szemantikai összefüggésláncolat. 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 következőkben ezek bemutatása kerül előtérbe. A kohéziós taxonómiák egymásra épülnek, mivel az újabb rendszerek iránti igény mindig akkor merült fel, amikor az azt megelőző nem bizonyult kielégítőnek, részben a taxonómia által használt kategóriák elmosódottsága, részben a hiányzó kategóriák miatt.

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 és 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 megléte.

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, amely a szöveg egy másik elemére való utalás által 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 előre utaló (katafora) vagy hátrafelé mutató (anafora).

A referencia (i) utalhat személyes és birtokos névmások használatával személyre (*personal reference*) (ii) lehet a mutatószó és a határozott névelő használata miatt mutató (*demonstrative reference*) vagy (iii) az azonosságot, hasonlóságot, különbözőséget, mennyiségi vagy minőségi hasonlítást kifejező jelzők és határozók használata 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 ellipszis esetén a szöveg egyik eleme kimarad, azaz a magától értetődő szövegrész esetében zéró helyettesítés történik. Az angolban a helyettesítést végző elem névszói, igei, vagy mellékmondati funkciót tölthet be és így az ellipszisnek is három fajtája létezik: névszói-, igei- és mondatellipszis. 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 biztosító kötőszavak. 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 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, ezért 1984-ben átdolgozta Halliday és Hasan taxonómiáját és Gutwinski (1976) közeliismétlés-kategóriái alapján újraértelmezte a szöveg lexikai egységei közötti különféle kohéziós relációkat. Az addigi kutatások cáfolták azt az eredetileg megfogalmazott hipotézist, miszerint a kohéziós elemek láncban belüli arányának növekedésével a szöveg koherensebb lesz. Hasan a lexikai kohéziós kapcsolatokat ekkor 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*), a konkrét – általános összefüggést jelölő hiponímiát (*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 egyazon 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-e a láncalkotásban. A szövegszerveződések szempontjából meghatározó lexikai elemek (*relevant tokens*) részt vesznek a láncalkotásban, míg a szöveg témájához nem kapcsolódóak (*peripheral tokens*) nem alkotnak láncokat.

Halliday (1985) felfogásában fontos szerepet játszik a szövegalkotásban a tematikus struktúra és a fókusz. Az elbeszélő és érvelő szövegekben a téma többnyire a szövegen

belül fókusz pozícióba kerül, míg más regiszterek esetében általában az előző tagmondat rémája lesz az aktuális tagmondat témája. A fókusz a legfontosabb információt tartalmazza és az információs fókusz a mondat végére kerül, míg a téma rendszerint a mondat elején dominál.

Halliday a szövegalkotás másik fontos elemének a kohéziót tekinti, amely a jelentés-tani kapcsolatok tartományában az úgynevezett referenciális láncok révén realizálódik. A szövegszerkezet maga ugyan nem grammatikai, hanem szemantikai természetű, de a különböző referenciális láncok alkotóelemei más láncok elemeivel grammatikai viszonyban is állhatnak, ami a szövegkohéziót tovább erősíti. A kohézióra Halliday a szöveghez hasonlóan folyamatként tekint és az angol nyelv esetében a kohézió megvalósulásának négy fajtáját különbözteti meg, 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 fajtáit Halliday (1) a lexikai elem pontosan egyező, inflexiós vagy derivációs változatainak ismétlésében (*repetition*) (2) a szinonímia viszonyban és (3) a kollokációban (*collocation*) határozza meg. 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 az 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 a szövegszerűség hét ismérvének egyikeként szerepel, grammatikai függőségeken alapul és azt tükrözi, hogy „a hallott vagy látott szavak között milyen kölcsönös összefüggések vannak egy adott szószorozaton belül” (De Beaugrande 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 ismétlését (pl. *forog – forgó*), (2) a teljes szerkezetet új elemek beiktatásával ismétlődő párhuzamot, (3) az adott szövegegység tartalmát más kifejezésekkel körülíró parafrázist, (4) a tartalmas elemeket tömörítési céllal helyettesítő szómásokat (pl. névmások, igemások), (5) a szövegfelszín bizonyos elem-sorozatának részleges ismétlésén alapuló ellipszist, (6) és az igeidő, az igeszemlélet és a kötőszavak révén történő beillesztéseket.

A kohéziókutatás további eredményeire támaszkodva dolgozta ki taxiómáját Hoey (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. Az ilyen jellegű kapcsolódás az egymástól távol eső mondatok esetében is megvalósulhat. A kapcsolódások megléte alapján elkülönülnek egymástól a szöveg tartalma 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 ismétléskapcsolatok között.

Szikszainé Nagy (1999) a szöveg jelentésbeli egységét a téma változatlanóságához és a szerkezeti egységek jelentésének egymásutánosságához köti, a szövegjelentés összetartó elemeként pedig a kohéziót jelöli meg. A szöveg lineáris kohéziójának megteremtődését a fogalmi kategóriák egymásba épülésével és a lexikai egységek jelentésbeli viszonyainak meglétével magyarázza. A kohéziós eszközök között megkülönbözteti a koreferenciát, a mezőösszefüggést, az asszociációs mezőt, a mellérendelő kötőszavakat és az ellipszist. Az azonos valóságalemeire utaló, koreferenciális szövegelemek esetében szemantikai ekvivalenciáról beszél, amely 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, 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, így megkülönböztetik a morfológiai (kötőszavas, névszói kapcsolat és egyezőségen alapuló viszony), a szintaktikai (ellipszisen alapuló kapcsolat, mondaton belüli szórenden alapuló kapcsolat), a lexikai (deriváción és összetételen alapuló viszony, szinonima- és hiperonimaviszony), a stilisztikai (anaforikus kapcsolat és ismétlés) és a prozódiai (intonáción alapuló viszony) kapcsolódásokat.

A modern szövegnyelvészet a kohéziót gyakran a koherenciával együtt tárgyalja. Többek megítélése szerint „azonos jelenségre vonatkoznak a konnexió, kohézió és a koherencia fogalmak, amelyek (a szemiotika háromelemű jelfelfogása értelmében) a szövegkapcsolódás 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 Csátár és Haase 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 és Haase 2011:60)

3.2. A lexikai kohézió

A kohéziós eszközök közül szövegalkotó, szövegszervező szerepe miatt a lexikai kohézió különös figyelmet érdemel. Ezt az előzőekben bemutatott, a mondathatárokon túlmutató szemantikai-szintaktikai kapcsolatokat vizsgáló ismétlésmodellek mindegyike hangsúlyozza (Halliday és Hasan 1976, 1989, Hasan 1984, Hoey 1991). A magyar szövegnyelvészeti szakirodalom is a kezdetektől tárgyalja a szövegek lexikai mintázatának leírását, illetve a lexikai elemek ismétlődéseinek és szemantikai kapcsolatainak feltárására irányuló kutatásokat.

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

Békési (1979) 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” (Békési 1979:56–57). E lényegileg nem grammatikai kapcsolatok esetében fontos tudnunk, milyen a „kapcsolás típusa, formája, alakja és módja”, illetve „terjedelme, ereje, szorossága”, mivel Békési a konstrukciós forma felépítésének szempontjából megkülönböztet fő- és alkapsolásokat.

Károly Krisztina (2002) felhívja a figyelmet arra, hogy a kohéziós eszközök segítségével nemcsak a mondatokat és tagmondatokat összekapcsoló nyelvtani struktúrák azonosítása lehetséges, hanem a szövegelemek közötti mondathatárokon túlmutató kapcsolatok megértése is (Károly 2002b:19). Hoey (1991) taxonómiájából kiindulva alkotta meg Károly 2002b-ben a lexikai ismétlődés szövegszervező funkcióját vizsgáló rendszert. Károly a szövegfelszín kohézióért felelős elemeknek elsődleges funkcióját a kapcsolatok jelzésében és ezáltal a szöveg folyamatosságának biztosításában látja (Hoey 1991:65). Taxonómiájában a szöveg kohézióját az olyan lexikai egységek biztosítják, amelyeknek jelentése nem vezethető vissza alkotóelemeik 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 valamilyen formában (szinonima, ellentét, hiponima, meronima) (Hoey 1991:103–106).

3.3. 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”. A Petőfi S. által említett információ-hordozó elemek szemantikai-szintaktikai ismétléseit kívánják a különböző lexikai kohéziós taxonómiák megragadni.

Károly 2002-es ismétlésmodellje a kohéziós elemzés megkezdése előtt szükségesnek tartja a szöveg kohézióját biztosító lexikai egység meghatározását, mivel a szövegek kohéziós kapcsolatait nem feltétlenül a helyesírási szavak szemantikai kapcsolataiban látja. Károly maga lexikai egységként az olyan egységet definiálja, amelynek jelentése nem vezethető vissza alkotóelemeinek jelentésére, így elemzéseiben lexikai egységként jelennek meg a különálló szavak mellett a több szóból felépülő nyelvi egységek, a kollokációk és az állandósult szókapcsolatok is. Károlynál 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 2002b:95).

4. A lexikai kohézió vizsgálata fordított szövegekben

Az átvitel következtében a célnyelvi szövegek lexikai kohéziós mintázatában bekövetkező változások meghatározó jellemzői a fordításnyelvnek. A VI. és VII. fejezetekben bemutatott kutatások központi gondolata, hogy a fordítás során a célnyelvi szövegek kohéziós eszközhasználatában bekövetkező változások az eltolódásjelenségek sorába tartoznak. Ennek következtében a fordított szöveg kohéziós mintázata nem egyezik sem a kiindulásként szolgáló forrásnyelvi szöveg, sem a célnyelven írt autentikus szövegek kohéziós mintázatával. A következőkben bemutatott négy szerző argumentumai is alátámasztják, hogy kiemelten érdekes a lexikai kohézió vizsgálata a fordításkutatás számára. (i) Hoey (1991) véleménye szerint a kohézió inkább lexikai, mint grammatikai relációk eredménye, így a szövegszerveződés szempontjából a lexikai kohézió kitüntetett jelentőségű. (ii) Hatim és Mason (1990) megállapítják, hogy a lexikai kohézió a szövegalkotó szándékos lexikaválasztásán múlik, szemben a többnyire az egyes nyelvre jellemző jóformáltsági előírásait kielégítő grammatikai kohézióval. (iii) Halliday és Hasan (1976) szerint a kohézió nyelvi jelenség és az egyes nyelvekre jellemző eltérő nyelvi sajátosságok következtében a különböző nyelvek kohéziós eszközhasználatára sem fedheti száz százalékosan egymást. (iv) Blum-Kulka (1986) pedig megállapítja, hogy ha a különböző nyelvek eltérő módon tolerálják a kohéziós elemek ismétlését, akkor a fordításra eltolódás lesz jellemző ebben a tekintetben is.

[...] a kohézió, vagy a felszíni elemek egymás utáni kapcsolódása sokkal inkább nyelv- vagy szövegspecifikus jelenség. Számos lehetséges kohéziós eszköz létezik, amely képes a mondatok közötti adott kapcsolatot visszaadni. Egy adott nyelv esetén némelyik választása valószínűleg előnyt élvez. (Hatim és Mason 1990:195, a szerző fordítása)¹⁵

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 fordított szövegek tulajdonságai eltérőek az autentikus szövegek tulajdonságaitól, mivel „ha a fordításnyelv önálló nyelvi rendszer, akkor megkülönböztethető más nyelvi rendszerektől, azaz felismerhető” (Balaskó 2005:113).

Az eddigi tudományos eredmények ismeretében megállapítható, hogy az egy nyelven belüli szövegprodukciónál esetében nagy jelentőségű a kohéziós eszközhasználat megléte és a használata milyensége. Ahogy Klaudy fogalmaz „anyanyelvünkön nem tudatosítjuk azt, hogy mondat szerkesztési műveleteinknek milyen szövegszintű következményei vannak” (Klaudy 2006:142), viszont 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 fordítástudomány deskriptív elemzéseiben, mind a fordításoktatás preskriptív követelményrendszerének kialakításánál. Ma már a szakirodalom is hangsúlyozza, 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 figyelmet. Mondat és szövegszinten a célnyelvi normák részét képezi tehát a megfelelő kohéziós eszközhasználat is (Heltai 2005b:169).

Klaudy (1987) a fordított szövegek esetében a kohézió megteremtésének funkcionális jelentőségére utal a kvázi-helyesség fogalom bevezetésével, mely szerint „a fordított magyar szövegek kohéziója csökken, az olvasó késve azonosítja a mondatok

¹⁵ [...] 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 rephrasing, say, a given relationship between propositions. And in a given language, some are likely to be preferred options. (Hatim és Mason 1990:195)

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 (2003) 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 elem a „retorikai cél; a műfaj; a grammatikai és lexikai regiszter – a szaknyelv; a tematikus rend” mellett (Szabó Mihály 2003:58).

A fordításoktatás értékelésére kidolgozott Dróth-féle hibatipológia (2002) formatív szempontrendszerének öt kritériuma között is szerepel a lexikai kohézió. A hibák súlyozásának rendszere előtérbe helyezi azt az igényt, melynek célja „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 fordított szövegek értékelésekor figyelembe vett kritériumok elsődlegesen azt értékelik, hogy a fordított 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” (Dróth 2002:98).

A feltáró kontrasztív fordításelemzés új perspektívát nyithat az anyanyelv tudatosításában, a kutatási eredmények többek között hozzásegíthetnek „a magyar nyelvleírás egy korszerűbb, funkcionális megközelítéséhez” (Klaudy 2007:55). A fordítandó szövegmennyiség exponenciális növekedése következtében a teljes fordítási volumen célnyelvi megfeleltetését „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 2007:55). E folyamat miatt válik a fordítás útján keletkezett szövegek kontrasztív vizsgálata, illetve a fordításokra jellemző fordításnyelvi sajátosságok feltárása az alkalmazott nyelvészet kiemelkedően fontos kutatási területévé.

A fordítási szövegalkotással foglalkozó munkák közül Heltai *A fordító és a nyelvi normák* című tanulmány sorozata a „kohéziós eszközök használatát” a „szövegalkotási normák közé” sorolja (Heltai 2004:417), ugyanakkor a szerző azt is megállapítja, „hogy a szövegalkotási normák nem fogalmazhatók meg olyan egyértelműen és explicit módon, mint a morfológiai és szintaktikai normák” (Heltai 2004:417).

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 nyilvánvalók. (Heltai 2004:417)

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 is.

Szikszainé Nagy (1999:58–72) a szövegösszetartó erő meglétében látja a szöveggé szerveződés kritériumát. Szerinte a szövegösszetartó erő különböző fajtáinak működése eredményezi a textualitást. Szikszainé 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 a szöveg jelentéshordozó elemeinek továbbra is dominánsan lexikális jellege, az írásos vagy hangzó manifesztáció, és az autonóm szöveg teljes, vagy részszoveg jellege.

V. A korpusznyelvészeti módszer

1. Bevezetés

Az utóbbi évtizedek robbanásszerű számítástechnikai fejlődésének köszönhetően számos új eszköz, számítógépes alkalmazás segíti az alkalmazott nyelvészeti kutatások megvalósítását. A fordítástudomány számára is új paradigmát jelentenek a korpusznyelvészeti kutatások (Seidl-Péché 2018, 2020), mert 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 hangsúlyozza, hogy a korpuszban összegyűjtött szövegek a ténylegesen használt nyelv olyan előfordulásai, amelyek a nyelvész beavatkozása nélkül jöttek létre. A korpuszalapú megközelítés legfőbb előnye 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, a forrás- és célnyelvi, valamint az autentikus és fordított szövegek közötti különbségek feltárására (Baker 1995).

2. A korpusznyelvészet mint módszer

A *Corpus Linguistics* (Lüdeling és Kytö 2008) antológia a következőképpen határozza meg a korpusznyelvészeti módszer szerepét az alkalmazott nyelvészeti kutatások között:

A korpusznyelvészetet gyakran tekintik ma a nyelvészet egy viszonylag új ágának, amely a „valós” nyelvhasználat empirikus tanulmányozásával foglalkozik számítógépek és elektronikus korpuszok segítségével. Elsődlegesen a „korpusz” nem más, mint írott vagy beszélt nyelvi szövegek egyszerű gyűjteménye. Amikor azonban ezt a kifejezést a modern nyelvészettel összefüggésben használjuk, akkor az számos további ismérvet is magában foglal, többek között a gépileg olvasható formát, a mintavételt és reprezentativitást, a véges méretet, valamint azt az elképzelést, hogy a korpusz az általa reprezentált nyelvváltozat standard referenciája. (Lüdeling és Kytö 2008:5, a szerző fordítása)¹⁶

¹⁶ 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

A korpusznyelvészeti módszer számos nyelvészeti terület fontos kutatási paradigmája. Így például a leíró nyelvészet az adott nyelv nyelvhasználatára jellemző lexikai/grammatikai és nem nyelvi asszociációs mintákat kutatja korpuszalapú módszerrel, mely vizsgálatok azt tárják fel, hogy milyen nyelvi jelenségek milyen lexikai/grammatikai asszociációkkal társulnak. A lexikográfia korpuszok segítségével írja le az egyes szavak különböző jelentéseinek reprezentációját, az adott regiszterre jellemző gyakoriságát és kollokációs mintázatát. A szövegnyelvészet az adott regiszterre és szövegtípusra jellemző szöveg-szintű nyelvi eszközök (referencia, ismétlés) leírásához támaszkodik a korpuszalapú módszerre. Az alkalmazott nyelvészetben belül a korpuszalapú kutatások a szaknyelvkutatásban és a szaknyelvoktatásban, az anyanyelv-elsajátítás és az idegennyelv-tanulás vizsgálatánál, a stilisztikában, a fordításkutatásban és a fordítóképzésben jellemzőek.

A múlt század közepétől a számítógépes tárolókapacitás növekedésének köszönhetően a szövegek véletlenszerű gyűjteményeit a szövegek mérete, fajtája, forrása vagy más meghatározó kritériuma alapján tervszerűen gyűjtött szövegbankok váltották fel. A modern korpuszok tehát gépileg olvasható, explicit módon megfogalmazott szempontok szerint összeállított szöveggyűjtemények, amelyek a nyelv empirikus vizsgálatát valódi nyelvhasználati minták alapján teszik lehetővé (vö. McEnery és Wilson 1996:1). Mivel a korpuszban tárolt szövegek a nyelv természetes megvalósulásai és a nyelv adott állapotát vagy változatát testesítik meg, alkalmasak nyelvészeti kutatások számára.

3. Korpusznyelvészet a fordításkutatásban

A korpuszokban található szövegek jellemzésénél Sinclair kiemeli, hogy azok valódi nyelvhasználati minták és természetes nyelvi szövegek, amelyek jellemzőek a nyelv adott állapotára vagy változatára: „[a korpusz] természetesen előforduló nyelvi szövegek gyűjteménye, amelyet egy nyelv adott állapotának, vagy változatának jellemzésére választanak ki” (Sinclair 1991:171, a szerző fordítása)¹⁷. Sinclair máshol a szöveggyűjtés szempontjai kapcsán azt emeli ki, hogy a szövegek lehetnek teljes szövegek, vagy szövegrészek, illetve hogy a korpusz összeállítója explicit módon megfogalmazott nyelvészeti szempontok szerinti és nem véletlenszerű adatgyűjtés eredményeként hozza létre az adott nyelv vagy nyelvváltozat mintájául szolgáló korpuszt. „A korpusz olyan nyelvi anyagok gyűjteménye, amelyeket meghatározott nyelvészeti kritériumok alapján választanak ki és rendeznek, azzal a céllal, hogy az adott nyelv mintájaként használhassák” (Sinclair 1996:4)¹⁸.

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 és Kytö, 2008:5.)

¹⁷ [corpus is] a collection of naturally-occurring language text, chosen to characterize a state or a variety of a language. (Sinclair 1991:171)

¹⁸ A corpus is 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 1996:4)

A korpuszalapú fordításkutatás ismérvei között szerepel, hogy empirikusan közelíti meg a nyelv leírását, ragaszkodik a hiteles használati előfordulások elsődlegeséhez, azaz a beszélt vagy írott nyelvhasználatot nem izolált példákkal szemlélteti. A korpuszalapú fordításkutatás vizsgálatának fókuszában egyrészt az autentikus (nem fordított) és a fordított szövegek összevetése, másrészt a forrásnyelvi és célnyelvi szövegek nyelvi jellemzőinek kontrasztív leírása áll. Az első esetében az egynyelvű korpuszokra jellemző lexikai, szintaktikai és szövegszintű nyelvi jelenségek vizsgálata a szövegek olyan egyedi sajátosságaira mutat rá, amelyek kialakulása sem a forrásnyelv, sem a célnyelv hatásával nem magyarázható és amelyek nagy valószínűséggel a fordítások univerzális tulajdonságai közé sorolhatók. A forrásnyelvi és célnyelvi szövegek nyelvi jellemzőinek kontrasztív leírása esetén a kutató kétnyelvű párhuzamos korpuszok segítségével a nyelvpárspecifikus fordítói magatartást, a gyakorlott fordítók fordítási problémákra adott megoldási javaslatait tárja fel. A valós nyelvi előfordulások vizsgálatának növekvő népszerűségét mutatja a korpuszalapú fordítástudományi kutatások magas száma a múlt század végétől, így például Baker (1993, 1995, 1996), Laviosa (1997, 1998, 2002), Kennedy (1998), Tymoczko (1998) Stubbs (2001), Kenny (2001), Olohan (2004), Teubert és Cermáková (2007) kutatásai.

A kulturális vonatkozású és nyelvhelyességi kérdések tisztázását segítő nagy referenciakorpuszok kapcsán Stewart (2000) megemlíti, hogy a fordító sokkal jobban tudja használni a nagy autentikus nyelvi mintákat tartalmazó korpuszokat a hagyományos szótáraknál a nyelvi és kulturális korlátok leküzdésére.

A korpuszalapú fordítástudomány további térhódítását vetíti előre Tymoczko (1998) a kutatási irány távlati lehetőségeinek számbavételével. A fordítástudomány igényeinek megfelelő technológiák alkalmazása és adaptálása mellett Tymoczko kiemeli a nyelvészeti és kulturális megközelítés integrálását és az ideológiák fordításokra gyakorolt befolyásának vizsgálatát. Tymoczko továbbá annak a reményének is hangot ad, hogy a korpuszalapú fordítástudománynak köszönhetően megnyílik a lehetőség a fordítás elméleti és gyakorlati területének egymáshoz közelítése, sőt akár újraegyesítése előtt is.

4. 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 korpuszokkal kapcsolatos nyelvészeti tevékenységet Kennedy (1998) négy nagyobb területre osztja, annak megfelelően, hogy a nyelvészeti tevékenység középpontjában (i) a korpuszépítés (számítógépes tárolásra és feldolgozásra alkalmas szövegek

gyűjtése, és ezekből adatbázis építése), (ii) a nyelvreírás (lexikon és/vagy nyelvtan leírása adott nyelv esetében), (iii) a statisztikán alapuló elemzés (az adathalmaznak a számítógépes nyelvészeti statisztikai módszereivel történő elemzése), vagy (iv) valamely a hétköznapi életben is elterjedt alkalmazás (lexikográfia, nyelv- és/vagy fordításoktatás és tanulás stb.) áll.

4.1. Az elektronikusan elérhető korpuszok

A jelen fejezet a nyelvreírás és a fordítástudományi vizsgálatok számára meghatározó online és offline korpuszok közül terjedelmi keretek miatt a kutatási témához illeszkedő írott nyelvi korpuszokra összpontosít, és nem tér ki a beszélt nyelvi korpuszok (*Spoken Language Corpora*) bemutatására (vö. Seidl-Pécs 2018). A beszélt nyelvi korpuszok (pl. *EUR-ACCOR* korpusz¹⁹) esetében a feldolgozhatóság megkönnyítése érdekében ma már általában nyelvtechnológiai szoftvereket alkalmaznak a lejegyzés automatizálásához.

A vizsgálat tárgya lehet írott vagy beszélt nyelvi korpusz, mely utóbbinál mindenképpen szükséges a beszélt nyelvi anyag átírása az élőbeszéd jellemzőinek (szünet, hangereőség, 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 elemzett korpuszok is. (i) A szegmentálás és tokenizálás azonosítja a bekezdés-, mondat- és szóhatárokat, (ii) a lemmatizáló program az azonos szótöbbszámú alakokat szűri ki, (iii) a lexikai elemek meghatározásához nélkülözhetetlen a morfológiai elemzés, a jelentéségyértelműsítés (*word sense disambiguation*) és a tulajdonnévfelismerés, (iv) míg a mondat szerkezeti egységeinek azonosítását, és a közöttük lévő összefüggések feltárását a szintaktikai elemzés (*parsing*) valósítja meg.

Az írott nyelvi online és offline 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*²⁰ (Francis és Kucera 1964) volt. A *Brown Corpus* számos egynyelvű korpusz (*monolingual corpora*) követte, melyek közül – a teljesség igénye nélkül – a fordítástudomány számára legfontosabbak bemutatása következik.

A legnagyobb méretű, angol nyelvű szövegeket tartalmazó korpuszok közül a *British National Corpus*²¹ (BNC) 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 a késő XX. századi modern brit nyelvet reprezentálja, ezért a szö-

¹⁹ <http://www.cstr.ed.ac.uk/research/projects/artic/accor.html>

²⁰ <http://icame.uib.no/brown/bcm.html>

²¹ www.natcorp.ox.ac.uk

vegek különböző témaköröket, stílusokat, műfajokat és regisztereket képviselnek, illetve a kiegyensúlyozottság biztosításának érdekében a szerzők teljes szövegeket és rövidebb szövegrészeket is rögzítettek. A BNC első verzióját követte 2001-ben a *BNC World*, 2007-ben a *TEI szabvány (The Text Encoding Initiative)* standardizált annotációját integráló *BNC XML Edition*, majd 2014-ben a beszéltnyelvi szövegeket tartalmazó *Spoken British National Corpus 2014*.

Mark Davies 2008-ban indította el a *Brigham Young University*n a legnagyobb szabadon hozzáférhető angol nyelvű korpusz, a *Corpus of Contemporary American English*²² (COCA) építését, melyet azóta is évente egyszer vagy kétszer frissítenek évi további 20 millió szövegszóval. Az utolsó frissítés a honlap tanúsága szerint 2019-ben történt. A mára már több mint egy billió szövegszót tartalmazó *Corpus of Contemporary American English* részkorpuszai között található a *British National Corpus*, a 100 millió szövegszót tartalmazó *TIME Corpus* (1920-2000) és a 400 millió szövegszót tartalmazó *Corpus of Historical American English* (COHA; 1810-2009). A korpuszban egyenlő eloszlásban 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 német nyelvű szövegeket tartalmazó *TIGER Korpus*²³ célkitűzése egy nyelvészetileg elemzett német sajtószöveg-gyűjtemény létrehozása volt. A projekt első szakasza 700 000 szövegszót és 40 000 mondatot tartalmazott, míg a második szakasz során további 200 000 szövegszót és 10 000 mondatot annotáltak félautomatikus gépi elemzéssel. A jelenleg publikált *Version 2.1* és *2.2* a Frankfurter Rundschau szövegeiből létrehozott (900 000 token, 50 000 mondat), morfológiaiilag elemzett és lemmatizált német sajtókorpusz.

A magyar egynyelvű korpuszok közül a *Magyar Nemzeti Szövegtár*²⁴ (MNSz) építésére 1998 és 2005 között került sor az *MTA Nyelvtudományi Intézetének Korpusznyelvészeti Osztályán*. A projekt célja egy 100 millió szavas szöveggörpusz létrehozása volt, amely a mai magyar nyelv jellegzetes megnyilvánulásait tartalmazza. A korpusz 187,6 millió szövegszót tartalmaz, amelyek az öt kárpátmedencei regionális nyelvváltozatot (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.

A *Szeged Treebank 2.0*²⁵ létrehozását 1999-ben kezdték meg. A korpusz legfontosabb ismérve, hogy morfoszintaktikailag elemzett és kézzel egyértelműsített természetes nyelvi magyar szövegeket tartalmaz, mérete 1 millió szövegszó. A kézzel egyértelműsített

²² <https://www.english-corpora.org/coca>

²³ <https://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger>

²⁴ <http://corpus.nytud.hu/mnsz>

²⁵ <https://rgai.inf.u-szeged.hu/node/113>

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).

Baker (1995) iskolateremtő írása a fordításkutatás és fordításpedagógia számára a párhuzamos (*Parallel Corpora*), a többnyelvű (*Multilingual Corpora*) és az összehasonlítható korpuszok (*Comparable Corpora*) vizsgálatát tartja célravezetőnek. A fordítástudományban széleskörűen elterjedt felosztás alapján többnyelvű korpusz például a *MLCC Multilingual and Parallel Corpora*²⁶.

Az összehasonlító korpusz (pl. *International Corpus of English*) egy adott nyelven tartalmaz autentikus szövegeket és egy vagy több másik nyelvről az adott nyelvre fordított szövegeket. Az összehasonlító korpusz tehát célnyelvi szövegminták gyűjteménye, és nem tartalmazza a párhuzamos korpuszokhoz hasonlóan a fordított szövegek forrásnyelvi megfelelőit.

A párhuzamos korpuszok ezzel szemben az autentikus szövegek mellett ezek fordítását is tartalmazzák. A napjainkban már több nyelvpár esetében szabadon hozzáférhetőek párhuzamos korpuszok, amelyek lehetnek egy- vagy kétirányúak, mondat vagy szó szinten párhuzamosítottak. Az egyirányú *Parallel Corpora in Uppsala*²⁷ például csak eredeti svéd szövegekből, és azok fordításaiból áll, míg a kétirányú *English-Norwegian Parallel Corpus*²⁸ angol és norvég eredeti szövegeket, és azok fordítását is tartalmazza a másik nyelven. A fordítástudomány számára különösen érdekesek a mondatillesztéssel, illetve szó szinten is párhuzamosított korpuszok, mint az *Aligned Hansards of the 36th Parliament of Canada*²⁹, és a *European Parliament Proceedings 1996-2001*³⁰. Az Európai Unió joganyagát és más dokumentumokat tartalmazó EUR-Lex³¹ a 24 hivatalos nyelv ingyenesen elérhető párhuzamos korpusza.

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*³². Egy adott nyelv szövegeinek több különböző nyelvre történő fordítását tartalmazza továbbá például a *Pedant Corpus* (Danielsson, Ridings 1996), az *Oslo Multilingual Corpus* (Johansson 1997) és a *Triptic Corpus* (Paulussen 1999) is.

A többnyelvű korpuszok tárgyalásánál meg kell említeni a fordítási korpuszokat (*translational corpus*) is, amelyeket a szakirodalom speciális funkcionalitásuk és felhasználóik meghatározott köre miatt szokott hol az összehasonlító, hol a párhuzamos korpuszon belül elkülöníteni (vö. Lüdeling és Kytö 2008:272). A fordítási korpusz elnevezést

²⁶ <https://catalogue.elra.info/en-us/repository/browse/ELRA-W0023>

²⁷ <http://xml.coverpages.org/etap-over.html>

²⁸ <https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc>

²⁹ <https://www.isi.edu/division3/natural-language/download/hansard>

³⁰ <http://www.statmt.org/europarl>

³¹ <https://eur-lex.europa.eu/homepage.html>

³² <https://www.ice-corpora.uzh.ch/en.html>

a szakirodalom jelenleg 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 elnevezést alkalmazta a párhuzamos korpusz megnevezés helyett, azaz gyakorlatilag a forrásnyelvi szövegeket és azok fordításait egyaránt tartalmazó szöveggyűjteményekre (vö. Lüdeling és Kytö 2008:277). A bakeri felfogást tükrözi a különböző nyelvekről csak angol nyelvre fordított szövegek gyűjteményét tartalmazó, mintegy 10 millió szövegszavas *The Translational English Corpus*³³ (TEC), amelyet a Manchester-i Egyetem Fordító és Tolmács Központjában állítottak össze.

4.2. Korpuszok és fordításkutatás

A fordítási korpusz különösen alkalmas a fordított szövegek sajátosságainak vizsgálatára, szövegek, lexikai elemek vagy kifejezések különböző fordításainak tanulmányozására. A fordított szövegeket tartalmazó korpusz egynyelvű korpuszsal történő egybevetése magyarázatot adhat a fordításüzűség szövegszinten kimutatható okaira, illetve feltárhatja az adott nyelvpárra és fordítási irányra jellemző fordítási mintázatokat.

Ha a forrás- és a célnyelvi szöveg is rendelkezésre áll, akkor a korpuszban szegmenspárok, párhuzamos fordítási egységek hozhatók létre. A szöveg párhuzamosítása történhet manuálisan vagy automatikusan, attól függően, hogy a szinkronizált szövegpár mondat/tagmondat szintű szegmenseit egymásnak hogyan feleltették meg. A szöveg párhuzamosítása során a mondat/tagmondat az a legkisebb nyelvi egység, amelynek szintjén egy szövegpár még pontosan szinkronizálható, de létezik bekezdésszintű szinkronizálás is. A szegmentálás automatizálásának nehézségei között említendő, hogy a szegmenshatárok (mondat/pont) definiálása a számítógépes feldolgozás számára egyáltalán nem egyértelmű feladat, hiszen a szövegben az elemző rendszerek nemcsak a mondathatárokon azonosítanak pontokat (pl. dátum, rövidítés stb.). Ezért a szövegpárhuzamosítás számára fontos támpontot jelentenek a számok, időpontok, mennyiségek, tulajdonnevek is (vö. Prószycki és Kis 1999:263–266). A szinkronizálással létrehozott, az autentikus és fordított szövegeket tartalmazó párhuzamos korpuszok speciális használati lehetőségük miatt fontosak a fordítástudomány számára, mivel alkalmasak a fordítási szövegminták tanulmányozására (vö. Johansson 1997), és a fordítási univerzálék feltárására (vö. Baker 1995). Ezek a korpuszok továbbá kulcsfontosságúak a nem angol korpusznyelvészet (ázsiai nyelvek, kis nyelvek) számára.

A McEnery és Xiao (2007:18) szerzőpáros a párhuzamos és összehasonlító korpuszok előnyeként említi, 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 forrás- és célnyelvi szövegek, illetve az autentikus és fordított szövegek közötti különbségeket. E korpuszok alkalmazása a fordítástudományon kívül számos más területen is hasznos, így például a lexikográfiai kutatások során és a nyelvoktatásban is.

³³ <https://research.manchester.ac.uk/en/projects/translational-english-corpus-tec>

Klaudy (2005) arra hívja fel a fordításkutatók figyelmét, hogy az 1990-es évek közepén a szövegkorpuszok felfedezése elősegítette, hogy a fordítástudomány mint diszciplína át tudjon alakulni normatív (előíró) tudományból, deskriptív (leíró) tudományá. E szemléletváltásnak köszönhetően a kutatók a fordított szöveget már nem rontott szöveggént, hanem kutatásra alkalmas és kutatási szempontból érdekes szöveggént kezdték kezelni. A korpuszalapú fordításkutatás térnyerésének tehát mindenképpen pozitív hozadéka a forrásnyelvi szöveg primátusának megszűnése, és a célnyelvi szöveg kutathatóságának előtérbe kerülése.

Klaudy hangsúlyozza továbbá azt is, 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 (1995) is a fordított szövegek lexikai jellegzetességeinek feltárását emeli ki a korpuszalapú vizsgálatokkal kapcsolatban. A korpuszok 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 nyelven keletkezett autentikus szövegekben. Vizsgálható továbbá 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 munkáiban (pl. 1993, 1996) többször is felvetődik az a kérdés, hogy a fordított szövegek esetében azonosíthatóak-e olyan általánosan jellemző tulajdonságok, amelyek a fordításokat megkülönböztetik az autentikus szövegproduktiótól, azaz beszélhetünk-e ebben az értelemben fordítási univerzálék-ról. Baker szerint éppen az intuíció és a manuális elemzések alapján feltételezhető fordítási univerzálék létezését kell és lehet a korpuszok segítségével bizonyítani vagy cáfolni.

Károly (2002) a korpusznyelvészet és a fordításkutatás összekapcsolásának előnyét az írott vagy beszélt nyelv szabályszerűségeinek vizsgálatában, illetve a variabilitást befolyásoló szöveggörnyezeti tényezők és a nyelvhasználat szabályszerűségeinek feltárásában látja. Az univerzálékutatást és a nyelvpárspecifikus fordítói magatartás feltárását segítheti szerinte a forrás- és célnyelvi szövegjellemzők összevetése. Károly a korpuszalapú kutatások előnyei között tartja számon, hogy a gyakran és jellemzően előforduló esetekkel foglalkoznak, kizárva a statisztikailag nem szignifikáns eseteket. Segítségükkel kimutatható az egyes regiszterekre jellemző grammatika és igazolható az anyanyelvi beszélők intuíciója, hogy mely nyelvi jelenséget használnak gyakrabban vagy ritkábban a nyelvtanulók, mint az anyanyelvi beszélők. A korpuszalapú kutatások megmutatják tehát a forrás- és célnyelv aszimmetrikus jellegét. Károly ugyanakkor bírálja is a korpuszalapú kutatásokat, melyek szerinte csupán a szöveg felszíni jelenségeit (pl. gyakoriság) vizsgálják és az összehasonlítások alapjául az anyanyelvi beszélők performanciáját veszik alapul. E kutatások további hátrányát abban látja, hogy figyelmen kívül hagyják az egyéni sajátosságokat és az egyedi előfordulásokat.

A fordítástudományban a korpuszalapú kutatások esetében szerteágazóan széleskörű a vizsgálatok spektruma. Kenny (1998, 2001) például a GEPOLT korpusz

(*German-English Parallel Corpus of Literary Texts*) segítségével a fordításokra jellemző lexikai kreativitást vizsgálta, míg Holmes (1972) és Laviosa (2002) a *The Translation English Corpus* használta fel a fordított szövegek általános tulajdonságainak feltárásához: vizsgálatuk középpontjában a magyarázó betoldások miatti explicitégségi szint növekedése, valamint a bonyolult mondatok feloldásának köszönhető egyszerűsítés; normalizáció áll.

A fordítástudományi kutatások gyakran próbálják meg számszerű adatokkal bizonyítani, hogy bizonyos nyelvi jellegzetességek eloszlása más a fordított szövegekben, mint az autentikus szövegekben. Blum-Kulka (1986) ilyen szempontból vizsgálta a kohéziós eszközök eloszlását a fordított szövegekben, Johansson (1995) pedig az explicitációt az angol és norvég párhuzamos szövegekben. Hansen-Schirra és munkatársai (2006) német–angol, illetve angol–német relációban vizsgálták a fordítások explicitégségi szintjének növekedése és a kohéziós eszközök eloszlásának változása közötti kapcsolatot. Baker (2000) és Olohan (2004) a fordító stílusát kutatta, Mauranen (1996) a fordítások interkulturális jellemzőit, Doherty (1999) német–angol nyelvpár esetében vizsgálta az információs struktúrát, míg Hasselgård és Oksefjell (1999) a tematikus szerkezetet.

4.3. Saját korpusz építése

A fordításkutatók körében egyre népszerűbb az egyetemi kutatóhelyeken létrehozott, és elektronikusan elérhető korpuszok mellett a saját kutatási célokat jobban kiszolgáló, egyénileg épített korpuszok vizsgálata. Ezt az állítást a jelen kutatást követő, az ELTE BTK Fordítástudományi Doktori Programban publikált számos korpuszalapú kutatás is alátámasztja (pl. Polcz 2012, Lengyel 2013, Mohácsi-Gorove 2014, Robin 2015, Sato 2014, Somodi 2014, Bozsik 2015, Kovács 2015, Makkos 2015, Nagy 2015, Szijj 2015, Götz 2019, Mány 2021, Szegh 2022, Stern 2022, Dankó 2022, Ugrin 2023, Horváth 2023, Juhász-Koch 2023), melyek közül több kutatás vizsgálati szövegei a későbbiekben a Pannónia Korpusz (Robin és mtsai. 2016) szövegállományát is gazdagították. A korpuszépítés maga több egymásra épülő részfeladatrészre bontható, kezdve a tervezés és szöveggyűjtés fáziséval, amelyet a szöveg(részlet)ek tárolása követ, majd végül az annotálás, azaz a nyelvészeti elemzés eredményeit tartalmazó többletinformáció feltüntetése.

A korpusztervezés kapcsán alapvető szempontként merül fel a reprezentativitás igénye, a korpuszok lekérdezésével nyert eredmények összevethetősége érdekében. A korpusz tervezőjének elsődleges célja, hogy garantálja kutatási eredményeinek érvényességét és megbízhatóságát (vö. Kennedy 1998, Károly 2002a), ezért az eredmények általánosíthatósága érdekében ügyel a korpusz reprezentativitására. A kiegyensúlyozott (*well-balanced*) reprezentatív korpusz a mindennapi nyelvhasználatnak megfelelő arányban tartalmazza a szövegtípus, a műfaj, a szövegfunkció, a szövegalkotó neve, a térbeli és időbeli lefedettség stb. szempontjából különböző jellegű szövegeket. A korpusz reprezentativitása összefügg a korpusz méretével és a mintavétel megbízhatóságával is.

A korpusz méretét a szövegszavak számával szokás jelölni. Bár a számítástechnika tárolási kapacitásának ugrásszerű növekedése lehetővé teszi a több százmillió szavas kor-

puszok létrehozását is, a kutatási céltól és a lekérdezéshez rendelkezésre álló eszközöktől függően sok esetben elegendő a hipotézisek vizsgálatát kisebb korpuszokon megkezdeni.

4.3.1. 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 (vö. Kennedy 1998). A teljes szövegekből felépített korpusz esetében nehezebb megszerezni a szerzők jogi hozzájárulását (Láncos 2022), ezért gyakori a részszövegek gyűjtése, ami azonos korpuszméret mellett több szerző szövegrészletének tárolását teszi lehetővé. Szerzői jogi hozzájárulás szükséges mind az elektronikus formában történő tároláshoz, mind a szövegek későbbi kutatásban való felhasználásához (vö. Láncos 2022). A szövegek letöltését, ma már megkönnyíti számos elektronikusan elérhető adatbázis (pl. *Digitális Irodalmi Akadémia*³⁴, *Littérature de langue française en ligne*³⁵, *Literatur im Netz*³⁶, *Magyar Elektronikus Könyvtár*³⁷ stb.)

4.3.2. Szövegválasztás

A statikus korpusz a létrehozása után változtatás nélkül tartalmazza a meghatározott szempontok alapján válogatott szövegeket, míg a dinamikus korpusz a belső arányok megtartása mellett folyamatosan bővül. Az eltérő időszakokban gyűjtött szövegeket tartalmazó dinamikus korpusz 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, illetve a vizsgált szövegekkel milyen időperiódust szeretne felölelni. A kiválasztott szövegeknek illeszkedniük kell a vizsgálat tárgyához, amely irányulhat a forrás- vagy a célnyelvi szöveg vizsgálatára, illetve lehet összehasonlító elemzés. A vizsgálandó szövegek gyűjtését befolyásolja továbbá a kutatási eredmények tervezett felhasználása aszerint, hogy a lekérdezési eredményeket a leíró fordítástudomány, az oktatás, vagy valamely alkalmazott nyelvészeti terület (pl. lexikográfia) kívánja hasznosítani. További szövegválasztási szempontot jelentenek a szöveg csak elemzéssel meghatározható belső tulajdonságai (pl. tulajdonnevek aránya), illetve a szövegalkotás jellemzőiként számontartott külső tulajdonságok. A külső tulajdonságok közé szokás sorolni például a szövegtípust és a szövegek beszélt vagy írott nyelvi voltát, vagy az olyan tematikus szempontokat, mint a szerző, a terület, a közzététel helye és a befogadó.

³⁴ <https://dia.hu>

³⁵ <http://www.lehman.cuny.edu/depts/langlit/french/lit.html>

³⁶ <https://www.dla-marbach.de/bibliothek/literatur-im-netz>

³⁷ <https://www.mek.oszk.hu>

4.4. A korpuszok annotálása

Az egyre nagyobb méretű szövegdatabázisok, korpuszok létrejöttével párhuzamosan fejlődött a számítógépes nyelvészet területére tartozó nyelvi tartalom szerinti feldolgozás és a szövegek intelligens kezelése, illetve az automatikus gépi tanuló eljárások fejlesztése. Az ember és a számítógép közötti természetes nyelven történő kommunikációt megvalósítani kívánó és az ehhez szükséges számítógépes programok megalkotásával foglalkozó nyelvtechnológiának köszönhetően a nyelvi-elemzőrendszerek széles skálája vált a kutatók számára elérhetővé. Minden nyelvi-elemzőrendszer esetében két együttes cél érvényesül, amelyek a minél több nyelvi elem azonosítását, illetve kódolását biztosító ún. lefedettség (*recall*) és az azonosítás vagy kódolás hibamentességét mutató ún. pontosság (*precision*) maximalizálását jelentik. 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. Minden nyelvtechnológia-elemző szoftver legoptimálisabb találati listájának beállítása tehát e két ellentétes művelet aprólékos hangolásának eredményéből fakad.

Mivel a számítógépes elemzőrendszerek számára a természetes nyelvi megnyilatkozások csak karakterek sorozataiként jelennek meg, ezért szükség van a homogén karaktersorozatokat 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 rögzítésével megvalósítani, hanem a morfémák és jellemző tulajdonságaik felsorolását a szótárban a szóalakok felépítéséhez szükséges szabályok megadásával kell kiegészíteni. A toldalékkolt szavak gépi reprezentációja így a tő és toldalékok által hordozott információk összességéből tevődik össze. A magyar morfológiai elemző, a HUMor (*High-speed Unification MORpology*) (Prószéky 1995) 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, amely során a rendszer az adott szóalak adott kontextusba illő elemzését választja ki relatív gyakoriságon és szóhármás-vizsgálatokon alapuló lekérdezések segítségével. A lemmatizálás során az elemző rendszer az azonos szótókból származó szóalakokat felelteti meg az adott nyelv szavait tartalmazó szótár szavainak. A magasabb szintű gépi műveletek (pl. fordítástámogatás) feltételezik már a statisztikai és grammatikai modellek alapján végrehajtott szintaktikai elemzést (*parsing*) is.

4.5. Egyszerű korpuszalapú elemzések

A korpuszok elemzéséhez használt számítógépes programokra az információlekérdezés pontossága, gyorsasága és jó minősége jellemző (Szirmai 2005). A ma rendelkezésre álló eszközökkel (pl. Sketch Engine³⁸) könnyen lekérdezhető statisztikák, kulcsszó-keresések, konkordanciák segítenek a korpusz lexikális jellemzőinek feltérképezésében (vö. Seidl-Pécs 2024, Szlávik 2023b). Az így kapott eredmények további kutatási hipotézisekre inspirálhatnak.

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 változatosságot mutatja meg. 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. További lekérdezési lehetőség a kis- és nagybetűs betűgyakorisági lista, az alfabetikus elrendezésű vagy gyakoriság szerinti szólista, vagy az átlagos szó- és mondatosság vizsgálat. Ez utóbbi a mondatathár felismerés miatt nehezen automatizálható, és mindenképpen 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 töreváló visszavezetésének, lemmatizálásának problematikáját (pl. *alma, almát, almás*), a többszavas kifejezések egy szótári egységként való kezelésének nehézségét (pl. *bázi feladat*), az azonos szó különböző írásmódjainak egységes kezelését (pl. *szőlő – szőlő*), és a kategóriakülönbségek felismerését, mint például a kis- és nagy betűs írásmód hatására módosuló jelentést (pl. *attila* = alkalmi öltöny, *Attila* = férfinév).

A kulcsszókereső-program a vizsgált szöveg szólistáját hasonlítja össze egy hosszabb szöveg szólistájával, 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 levest eszik.*” mondatot 3-gram esetében a következő egységekre bontható: (i) *Pisti minden délben,* (ii) *minden délben levest* és (iii) *délben levest eszik.* A konkordancia (*concordance*) az „adott szó vagy kifejezés szövegben szereplő összes előfordulását szöveggörnyezetében bemutató lista” (Szirmai 2005). Az eredményeket a csomóponttól jobbra vagy balra található első/második/stb. szó alapján listáztathatjuk ki a programmal. A konkordanciára Sinclair (1991) mint a korpusznyelvészet egyik legfontosabb vizsgálati eszközére tekint, mivel az a szöveg mintázataira hívja fel a kutatók figyelmét.

³⁸ <https://www.sketchengine.eu>

5. A WordNet-formalizmus

A természetes nyelvi szövegek gépi feldolgozását segítő eszközök között jelentek meg a nyelvek lexikális elemeit tartalmazó fogalomtárak, ontológiák. A WordNet (WN) formalizmus eredetileg a mentális lexikon számítógépes modelljeként jött létre, a tartalmas szakosztályok (főnevek, igék stb.) lexikai elemeit és a köztük levő szemantikai kapcsolatokat 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 kapcsolatai a *synset*-relációk. A WordNet-formalizmus alapvető jellemzője, hogy a synset alapján nyilvántartott szóhalmozok 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.

5.1. Princeton WordNet

A Princeton WordNet³⁹ 1990-ben indult a Princeton Egyetem Kognitív Tudományok Laboratóriumának professzora, George A. Miller kezdeményezésére, és az angol nyelv teljes szókészletének gépileg olvasható adatbázisát tartalmazza. Az azóta is folyamatosan bővülő ontológia a szavak és fogalmak felsorolásából és a közöttük levő viszonyok leírásából áll. A lexikai kapcsolatok esetében a Princeton WordNet a szinonímia, antonímia, hiponímia, meronímia, troponímia viszonyokat különbözteti meg.

5.2. EuroWordNet és BalkaNet

A Princeton WordNet ontológiai célkitűzései alapján kezdték meg 1996 és 1999 között az Európai Közösség által támogatott EuroWordNet⁴⁰ projekt fejlesztési munkálatait, ami nyolc nemzeti nyelv (az angol, holland, spanyol, olasz, francia, német, cseh és észt) esetében jelentette egy ontológikus adatbázis kiépítését. 2001 és 2004 között a BalkaNet⁴¹ keretén belül a WordNet projekthez tartozó ontológiaépítés további nyelvek (görög, román, szerb, török, bolgár, koreai, kínai, hindi és tamil) adatbázisaival bővült.

5.3. Hungarian WordNet

A 2005 és 2007 között fejlesztett magyar Hungarian WordNet (HuWN)⁴², a magyar nyelv lexikai-szemantikai kapcsolatait tartalmazó gépileg olvasható adatbázis, a szavak je-

³⁹ <http://wordnet.princeton.edu>

⁴⁰ <https://archive.illc.uva.nl/EuroWordNet/database.html>

⁴¹ <http://www.dblab.upatras.gr/balkanet/index.htm>

⁴² <https://rgai.inf.u-szeged.hu/node/159>

lentései mentén azonosítható lexikai relációk (pl. szinonima, antonima, hiponima, meronima) hálójá. Mivel a VII. fejezetben bemutatott Magyar Lexikai Kohéziós Projekt során kizárólag a főnevek kohéziós kapcsolatainak feltárása valósult meg, a leírás 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 feldolgozás egy elektronikus magyar-angol kétnyelvű szótár főnév állománya és az egynyelvű magyar Értelmező Kéziszótár számítógéppel feldolgozható strukturális és szemantikai információi alapján történt. Az eredményeket az adatbázis egy részkorpuszának kézi ellenőrzésével validálták (vö. Miháltz 2003).

A HuWN projekt a főnévi synsetek esetében öt fő relációtípust használ, a szinonímiát, az antonímiát, a hiponímiát, a meronímiát és a doménrelációt. A szinonímia esetében az egy synsetbe tartozó főnevek jelentései közel azonosak (pl. *autó* – *kocs*) és maguk a szinonimahalmazok rendkívül fontos szerepet játszanak a wordnetek felépítésében. Az antonimareláció a lexikális főnévi párok közötti fogalmi ellentétet jelöli (pl. *férj* – *feleség*). A hiponimaviszony a generikus fogalmak és az azonos fogalmi körhöz tartozó specifikus egyedek közötti kapcsolatra utal (pl. *virág* – *rózsa*). A meronimaviszony a részegész relációkat jelöli (pl. *fej* – *száj*). A domén relációk három fajtáját különbözteti meg WordNet a *category*, a *region* és az *usage* relációkat. 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* synset földrajzi területek térbeli egységeinek és a hozzájuk kapcsolódó fogalmi, kulturális, történelmi, szervezeti stb. elemeknek a jelölésére szolgál (pl. *kilenc lyukú híd* – *Hortobágy*). Az *usage domain* synset nyelvhasználati kategóriát jelöl (pl. költői, szleng), míg a *be in state* reláció egy melléknévi synsetet kapcsol egy főnévhez, a főnév által leírt állapotban (pl. *éber* – *éberség*).

VI. A pilot kutatás

1. Bevezetés

A pilotprojektben a kézíleg annotált, kisméretű német–magyar párhuzamos korpusz gépi lekérdezésével mutattam ki a német–magyar nyelvpár esetében az autentikus és a fordított szövegekre jellemző lexikai kohéziós mintázatot. A pilot kutatás a fordítástudomány és kohéziókutatás területének összekapcsolásával újszerű megközelítést kínált, melynek lényegét az összehasonlító elemzés, illetve a fordítás során a kohéziós elemek használatában bekövetkezett változások feltárása jelentette. A fordítások lexikai kohéziós elemeinek egybevetése az autentikus szövegprodukcóra jellemző megvalósulásokkal alapvetően az új korpusznyelvészeti technológia elterjedésének és hozzáférhetőségének köszönhető, mivel a korpusznyelvészeti módszer segítségével és a felhasználói szintű számítógépes ismeretekkel vált reális célkitűzéssé a szövegek, illetve szövegrészletek párhuzamosítása és szegmenspárokénti vizsgálata.

2. A pilot kutatás szövegtörzsének

A pilot kutatáshoz hivatalos internetes honlapokról töltöttem le a szövegeket, amelyeket Microsoft Word formátumban archiváltam. A leíró szerepet betöltő, informatív szövegek egy-egy intézményt, tájegységet, szervezetet stb. mutatnak be. Annak ellenére, hogy internetes felületekről töltöttem le a szövegeket, fontosnak tartottam, hogy ezek ne legyenek multimodálisak és multidimenzionálisak, mivel ebben az esetben a szövegek kohéziója csak részben támaszkodna nyelvi eszközökre, másik részük a vizualitás eszközeire épülne (vö. Háhn 2011). A szövegválasztásban három további fő szempont érvényesült:

(i) A kiválasztott szövegek stílusa hasonló a leíró jellegből fakadóan, ami kizárja az eltérő stílusú szövegekre jellemző, pragmatikai és stílári okokra visszavezethető különbségeket a lexikai kohéziós elemek használatánál (vö. 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). A szövegtípus figyelembevételét a gyakorisági vizsgálatok, és azok eredményeinek összehasonlíthatósága indokolja.

(ii) A szövegek német fordítását feltehetően profi fordítók készítették, hiszen ezek a weboldalak és információk marketing érdekeket szolgálnak. A fordítások jó minősége garancia arra, hogy a vizsgált célnyelvi szövegek esetében a lexikai kohéziós

elemek használatában tapasztalható eltérések nem a suta fordítói kompetenciával, hanem a nyelvek közötti, a kohéziós eszközök alkalmazásakor megfigyelhető eltérésekkel magyarázhatók.

(iii) A harmadik, ám nem elhanyagolható szempont, hogy a nyilvánosan hozzáférhető szövegek letöltésével időt lehet megtakarítani.

Az elemzett szövegeket 2006-ban gyűjtöttem, adataikat az 1. táblázat tartalmazza, míg a pontos bibliográfiai adatokat a Források.

| | cím | nyelv | szószám | mondatszám |
|-----------|-------------------------------|-------|---------|------------|
| 1. szöveg | Magyar Fordítóház | HU | 202 | 15 |
| | Das Ungarische Übersetzerhaus | DE | 283 | 17 |
| 2. szöveg | Balatoni hajózás | HU | 223 | 11 |
| | Balaton Schiffahrts AG. | DE | 273 | 11 |
| 3. szöveg | Művészetek Palotája | HU | 292 | 17 |
| | Palast der Künste | DE | 360 | 17 |
| ÖSSZESEN | | | 1633 | 88 |

1. táblázat: A pilotprojekt szövegei

3. A Károly-féle taxonómia

A pilot kutatás elméleti háttérét Károly 2002-ben kidolgozott rendszere adja, amely a lexikai ismétlődés szövegszervező funkcióját vizsgálja. Károly 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 szövegalapú 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 szövegalapú 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 2002b:95-107).

A Károly-féle taxonómia a szövegfelszín szintaktikai-szemantikai hálóját tehát kilenc lexikai kohéziós kategória segítségével írja le, melyek az egyszerű és derivációs ismétlés, az egyszerű és derivációs szinonima, az egyszerű és derivációs antonima, a meronima, a hiponima és a szövegalapú relációk. Ezen kategóriák segítségével tartam fel a pilot kutatás során a szövegszerveződés szempontjából meghatározó szemantikai kapcsolatokat.

4. A gépi feldolgozás előkészítése

A kiválasztott szövegek feldolgozása, azaz a kohéziós eszközök bejelölése a CLaRK System (Simov 2004) elemzést segítő nyelvtchnológiai eszköz segítségével történt. A CLaRK System alkalmazással végzett nyelvészeti elemzés lehetővé teszi a számítógépes feldolgozásra szánt szövegek strukturált kódolását. A strukturált kódolás során az elemzett szöveg(részlet) integritása megmarad, ugyanakkor a többletinformáció, azaz a nyelvészeti elemzés eredménye annotációként kerül a szövegbe (Prószéky és Kis 1999). Ilyen szövegbe ágyazott többletinformáció a jelen elemzés esetében a mondathatárok, illetve az ismétlődő lexikai elemek bejelölése, valamint a kohéziós kapcsolat típusának jelölése (1., 2., 3. ábra).

A CLaRK Systemben lehetőség van arra, hogy az eredeti szövegrészek és az annotált információk a képernyőn különböző színekkel jelenjenek meg. Az 1. ábrán az annotált információkat tartalmazó kék és zöld tagek közé fekete színnel ékelődik az elemzett szöveg.

A kutatás során vizsgálni kívánt kategóriákat a kutatás kezdetekor a CLaRK Systemben meg kell határozni. A kutatás céljának megfelelően tetszőleges számú és elnevezésű kategóriát lehet meghatározni, illetve rendelkezhetünk arról is, ha sorszámossal akarjuk ellátni őket, és/vagy valamilyen jellemző tulajdonságo(ka)t kötelező vagy választható jelleggel kívánunk hozzájuk rendelni. Így a konkrét elemzés során a kurzor mozgásával már csak a legördülő menüből kell a kívánt értékeket kiválasztani.

5. Annotálás és elemzés

A manuális elemzés első lépése a CLaRK Systembe beolvasott folyamatos szövegekben a mondathatárok bejelölése volt, illetve a mondatok sorszámozása (1. ábra). A mondathatárookra és mondatsorszámra vonatkozó többletinformáció értelmezhető adat a korpusznyelvészeti program számára, ugyanakkor elengedhetetlen feltétele az eredeti és fordított szöveg mondatszintű párhuzamosításának. Az ábrán látható mondatsorszám (pl. Sentence No="No6", Sentence No="No7") a magyar szöveg hatodik, hetedik stb. mondatát jelzi (1. ábra), míg a német szöveg megfelelő fordított mondatainak sorszámai mellett egy „d” betű is található (pl. Sentence No="No6d", Sentence No="No7d").

```
<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>
```

1. ábra: A mondathatárok jelölése és sorszámozása

A kutatás következő lépése az első mondat elemzése volt, amelyben jelöltem azokat a lexikai egységeket, amelyek a Károly-féle ismétlődési kapcsolatok szempontjából meghatározóak (2. ábra). A későbbiekben ismétlődő lexikai egységeket (RLU = *repeated lexical unit*) hivatkozási számmal láttam el. A hivatkozási adatok (H.1.1, H.1.2, H.1.3 stb.) első tagja (H) arra utal, hogy a magyar szöveg valamely mondatáról van szó, míg a H betűt követő két számjegy közül az első a mondat sorszámára, míg a második a jelölt elem sorszámára utal. Hasonló módon jelöltem a többi mondat ismétlődő lexikai elemeit is. A hivatkozási számok segítségével lehetséges a későbbi ismétlődéseknél az előző mondatok lexikai elemeire visszaautalni.

```
<Sentence No="No1">A
<RepeatedLexUnit RLU="H.1.1.">MagyarFordítóbász</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>
```

2. ábra: Az első szöveg első mondatának ismétlődő lexikai elemei

Ezt követte az első mondatban korábban jelölt lexikai egységek (*lexical unit*) ismétlődéseinek meghatározása az összes többi mondatban. A kohéziós kapcsolódás jellegét a Károly-féle taxonómia kilenc kategóriája szerinti jelöltem (3. ábra). A kohéziós kapcsolat meghatározását egy olyan referenciális szám követi (pl. LexicalUnit "InstantialRelation" reference="H.1.1."), amely az első mondatban hivatkozási számmal ellátott ismétlődő egységre utal vissza (jelen esetben ez a H.1.1.). Ennek a referenciális számnak a kapcsolódások azonosítása során van jelentősége.

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

3. ábra: Az első szöveg első mondatának ismétléskapcsolatai a 15. mondatban

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. Így a teljes szöveg elemzése során minden mondatot mindegyik mondattal összevettem.

A pilotprojekt során három teljes magyar szöveget és ezek német fordításait elemeztem. A magyar szövegek mondatszámra 11 és 17 között váltakozott. Minden szöveg elemzésénél az első mondat lexikai elemeinek ismétléskapcsolatait jelöltem a második, harmadik, negyedik stb. mondatban, majd a második mondat ismétléskapcsolatait a harmadik, negyedik, ötödik stb. mondatban és így tovább a szöveg utolsó előtti mondatáig. Az elemzés során mindhárom forrásnyelvi és célnyelvi szöveg minden egyes mondatának lexikai elemei esetében megvizsgáltam az utána következő mondatokban realizálódó ismétlés kapcsolatait.

6. A pilot kutatás eredményei

A 2. táblázat tartalmazza a három magyar (HU) forrásnyelvi és három német (DE) célnyelvi szöveg esetében az összes olyan lexikai egységet, amelyre a szöveg további részében visszautalás történik (*RepeatedLexUnit*), illetve ezen lexikai egységek ismétlésének konkrét megvalósulását (*LexUnit*). A lekérdezési eredmények szövegenként és nyelvenként csoportosítva jelennek meg a 2. táblázatban, ahol a közel azonos magyar és német szövegenkénti mondatszám ellenére eltérések körvonalazódnak a kapott eredmények tanúsága szerint a szövegek lexikai kohéziós mintázatában.

| | | 1. szöveg | 2. szöveg | 3. szöveg |
|----|-----------------|-----------|-----------|-----------|
| HU | Mondatszám | 15 | 11 | 17 |
| | RepeatedLexUnit | 51 | 49 | 62 |
| | LexUnit | 177 | 268 | 412 |
| DE | Mondatszám | 17 | 11 | 17 |
| | RepeatedLexUnit | 59 | 58 | 66 |
| | LexUnit | 213 | 308 | 462 |

2. táblázat: A pilotprojekt kohézióért felelős elemei

A 2. táblázatban megjelenített értékek alapján látható, hogy a pilotprojekt során vizsgált három szöveg esetében az annotált szövegeken futtatott lekérdezések tapasztalata alapján a fordított német szövegekben nagyobb a lexikai elemek ismétlésének sűrűsége, mint amilyen az eredeti magyar szövegekben. A lekérdezésekkel sikerült a kirajzolódó tendenciákat szemléltetni és azt a kiindulási hipotézist alátámasztani, miszerint a fordítási tevékenység befolyásolja a szövegfelszínen megjelenő kohéziós kapcsolatokat.

A részletes eredményeket, amelyek a kutatás automatizálására és nagyobb szövegmennyiség vizsgálatára ösztönöztek, több nyilvános előadáson ismertettem, valamint magyar és idegen nyelven írt tanulmányokban publikáltam (Péché 2007a, 2007b, 2007c, 2007d, 2009a, 2009b, 2011). A pilotprojekt méretét, az elemzett szövegek számát a kutatás szempontjából ígéretes eredmények ellenére sem volt célszerű tovább bővíteni a kézi annotálás időigénye miatt. A fordítás során a kohéziós mintázatban bekövetkező változások további vizsgálatát nagyobb korpuszon, több nyelvpár és szövegtípus automatikus elemzése útján kívántam folytatni.

VII. A Magyar Lexikai Kohéziós Projekt

1. Bevezetés

Az autentikus és fordított szövegek lexikai kohéziós mintázatának egybevetésére létrehoztam a Magyar Lexikai Kohéziós Projektet (MLKP). Kutatásom interdiszciplináris eredményei három tudományterület, a fordítástudomány, a szövegnyelvészet és a korpusznyelvészet számára relevánsak. A leíró fordítástudomány számára kiemelkedően fontos a célnyelvi szövegek jellemzőinek és szövegépítő sajátosságainak, ezen belül a fordítási szövegprodukción eltolódásjelenségeinek feltárása. Kutatásom központi feladata a kohéziós eszközök terén tapasztalható eltolódások feltárása, amelyet az autentikus szövegek kohéziós jellemzőinek és a fordított szövegek kohéziós sajátosságainak egybevetésével vizsgálok.

A vizsgálat több új szempontot kínál a fordításkutatás, azon belül a célnyelvi szövegalkotás leírásához. Leírja egyrészt a célnyelvi szövegalkotásra jellemző kohéziós eszközhasználatot, másrészt hozzájárul a fordított szövegek nyelvi normáinak leírásához és a fordításízűség egyik okának feltárásához, továbbá felhívja a figyelmet a mondathatáron túlmutató szemantikai kapcsolatok fontosságára. A kutatás eredményei relevánsak a fordításoktatás számára, mivel segíthetik a megfelelő kohéziós eszközhasználat tudatosítását a fordításoktatásban, illetve adatokkal támasztják alá az értékelési kritériumok (hibatipológia) szempontrendszerében megjelenő kohéziós eszközhasználat jogosultságát. Az eredmények hasznosak lehetnek továbbá a gépi fordítás számára, mivel az elemzések párhuzamos szövegek nyelvi szerkezetére vonatkoznak. A kutatásban bemutatott WordNet-formalizmus alkalmazása a mondathatárokon túlmutató kohéziós kapcsolatok azonosítására hatékony megoldást kínálhat a gépi fordítás számára a szöveggörnyezetnek megfelelő jelentés kiválasztásához.

Kutatásom szövegnyelvészeti jelentősége abban áll, hogy korábban nem történt meg nagyméretű magyar nyelvű szöveggörnyezet lexikai kohéziós elemzése és a magyar autentikus és célnyelvi szövegprodukción jellemző kohéziós mintázat feltárása és leírása. Kutatásom újszerűsége, hogy a magyar szövegek lexikai kohéziós mintázatának vizsgálatához a számítógépes nyelvészet legmodernebb eszközkészletét használta és az időigényes kézi annotálás helyett automatikus kategóriafelismerést alkalmazott.

A korpusznyelvészeti módszer alkalmazása tette lehetővé a nagy mennyiségű szöveg hatékony elemzését. A kutatás keretében létrehozott 3 050 000 szövegszóból álló fordított magyar korpuszt (FMK) szövegtípus és forrásnyelv szerinti elrendezésben tartalmazza a magyar célnyelvi szövegeket. A kutatás hasznai között kell említeni a WordNet-formalizmus felhasználását a mondathatárokon túlmutató szövegsemantikai relációk feltárásában. Ez egyrészt alkalmat adott a HuWordNet ontológia működésének nagyméretű szöveggörnyezeten való tesztelésére, másrészt a kutatás a WordNet adatbázis relá-

cióinak felhasználásával a mondathatárokon túlmutató lexikai-szemantikai kapcsolatok meghatározására kínál automatikusan megvalósítható gépi eljárást.

A Magyar Lexikai Kohéziós Projekt számos kritérium tekintetében különbözik a pilot kutatástól. A végrehajtott változtatások hátterében alapvetően a vizsgált korpusz méretének megsokszorozása áll. Míg a pilotprojekt korpusza 1633 szövegszó, addig a Magyar Lexikai Kohéziós Projekt közel 4 000 000 szövegszóból áll. Kutatásmódszertani szempontból a nagyobb szövegtípus vizsgálatát indokolt, hiszen így valósítható meg eltérő nyelvpárok és szövegtípusok bevonása a kutatás menetébe, mely eljárás lehetővé teszi a kapott adatok alapján megbízhatóbb következtetések levonását. A nagyobb szövegtípus indokolja az elemzés automatizálását, illetve ennek következtében a Károly-féle (2002) taxonómia helyett a WordNet-formalizmus alkalmazását a lexikai kohéziós relációk azonosítására. (Ezekről bővebben a következő fejezetekben: 4. A Magyar Lexikai Kohéziós Projekt annotálása, 5. A Magyar Lexikai Kohéziós Projekt elemzése, 6. A Magyar Lexikai Kohéziós Projekt eredményei.)

2. A Magyar Lexikai Kohéziós Projekt kutatási kérdései és hipotézisei

A kutatás előzményeként bemutatott pilotprojekt (lásd VI. fejezet) lexikai kohéziós vizsgálata kisméretű, kézi címkézéssel annotált korpuszon történt. A lekérdezési eredmények ismeretében megfogalmazott hipotézisek a fordításnyelvre jellemző, az autentikus szövegalkotási szokásoktól eltérő kohéziós mintázatra vonatkoznak.

A pilot projekt eredményeinek ismeretében további szövegjellemzők mentén tartottam a kutatás folytatását indokoltnak. Kutatásra érdemesnek látszott például annak vizsgálata, hogy bekövetkezik-e a némettől különböző forrásnyelvekről magyarra fordított szövegek esetében valamilyen kohéziós eltolódás, illetve hogy kimutatható-e a különböző forrásnyelvekről magyarra fordított szövegek esetében a fordítás következtében megváltozó lexikai kohéziós eszközhasználat. További kutatandó kérdésként merült fel, hogy milyen mértékben egyezik a magyar nyelvű autentikus szövegek lexikai kohéziós mintázata a magyar célnyelvi szövegek lexikai kohéziós mintázatával. A magyar nyelv vonatkozásában ezért fontos megvizsgálni, hogy azonosíthatók-e nyelvpárspecifikus fordításnyelvi kohéziós eltolódások.

A Magyar Lexikai Kohéziós Projekt során a felmerülő lehetséges kutatási kérdések alapján összehasonlító elemzést végeztem, melyben feltártam a fordítási tevékenység és a forrásnyelvi szöveg hatására bekövetkező eltolódásokat a magyar szövegek lexikai kohéziós elemeinek disztribúciójában. A szakirodalom tanulmányozását követően, illetve a pilot projekt elemzési eredményeinek tükrében a következő eredményekre számítottam a magyar nyelv vonatkozásában a kutatás végén:

- A fordított szövegekre jellemző kohéziós eszközhasználat nem egyezik az autentikus szövegprodukciónak jellemző kohéziós eszközhasználattal. Ezt a hipotézist

tézist az autentikus és fordított magyar szövegek kohéziós mintázatának egybevetésével kívánom vizsgálni.

- A fordított szövegek kohéziós eszközhasználatát befolyásolja a forrásnyelvi szöveg kohéziós eszközhasználat. Ezt a hipotézist az autentikus és a különböző forrásnyelvekről magyarra fordított szövegek kohéziós mintázatának egybevetésével kívánom vizsgálni.
- A fordított szövegek kohéziós eszközhasználatának eltolódása azonos szövegtípus esetén nyelvpárfüggő. Ezt a hipotézist az azonos szövegtípusú autentikus és különböző forrásnyelvekről magyarra fordított szövegek kohéziós mintázatának egybevetésével kívánom vizsgálni.
- A fordított szövegek kohéziós eszközhasználatának eltolódása fordítási tendencia. Ezt a hipotézist az előző hipotézisek beigazolódása után lehet a magyar nyelv és a vizsgált korpusz esetében feltételezni.

Az elméleti fejezetek szakirodalmi áttekintése alapján kijelenthető, hogy a fordítások minőségének lényegi eleme a kohézió megléte, illetve a célnyelvi rendszernek (mind fajtáját, mind mennyiségét tekintve) megfelelő kohéziós eszköztár alkalmazása. A most bemutatott kutatás összehasonlító elemzés keretében fókuszál a fordítás során a kohéziós elemek használatában bekövetkező változások feltárására. A kutatás az autentikus magyar szövegekre jellemző kohéziós eszközhasználatot veti egybe a fordított magyar szövegekre jellemző kohéziós eszközhasználattal.

A Magyar Lexikai Kohéziós Projekt 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 vizsgálandó 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 esetében realizálódik, a kutatás az elemzési idő és az elemzési eredmények optimalizálása érdekében a névszói csoportok kohéziós kapcsolatainak identifikációjára korlátozódik.

A kutatás alapcélkitűzése az autentikus és fordított magyar szövegek szövegépítő sajátosságainak összehasonlítása révén a fordítás következtében a célnyelvi magyar szövegek kohéziós elemeinek disztribúciójában bekövetkezett változások jellemzőinek feltárása. A kézi annotáció számára elérhető kisméretű szöveggörpusz vizsgálata helyett az automatikus eszközökkel segített elemzés megvalósítását több tényező is ösztönözte. Tekintettel arra, hogy a gépi fordítás számára az egyik legkomolyabb kihívást az egybefüggő gondolatmenetek egyes lexikai elemeinek a szöveggörnyezetbe illeszkedő, adekvát fordítása jelenti, ezért a szöveg bekezdésein belüli lexikai kohéziós kapcsolatok automatikus felismerése hozzájárulhat ennek a problémának a kezeléséhez. A Magyar Lexikai Kohéziós Projekt alkorpuszaiba gyűjtött szövegek elemzése a korpusznyelvészeti módszernek köszönhetően következetes, gyors és megbízható eredményeket jelent.

3. A Magyar Lexikai Kohéziós Projekt szövegkorpusza

A jelen kutatáshoz hasonló összehasonlító elemzésre természetesen nem véletlenül nem került sor a szövegkutatás korábbi időszakában, hiszen a hagyományos eszközzel végzett lexikai kohéziókutatás keretein belül egy ilyen típusú összehasonlító elemzés igen időigényes lett volna, mivel a hagyományos eszközzel végzett kutatás egy-egy szó kapcsolatait a szóból kiindulól, s az adott lexikai egység ismétléses kapcsolataihoz érkező vonalak segítségével jelöli (vö. Károly 2002b:80). Egyszerűen belátható, hogy a jelen vizsgálat alapjául szolgáló nagyobb korpuszhoz hasonló szövegkorpuszok esetén a hagyományos vizsgálat embert próbáló feladatot jelentene.

Az összehasonlító elemzés eredményeinek általánosíthatósága érdekében mind az autentikus magyar szövegprodukción, mind a célnyelvi magyar szövegprodukción reprezentáló korpusz összeállításánál fontos szempont volt a különböző szövegtípusú szövegek vizsgálata. A célnyelvi magyar szövegeket tartalmazó korpusz összeállításánál további fontos szempont volt, hogy a fordított magyar szövegek forrásnyelvi eredetijei különböző forrásnyelveket reprezentáljanak. A korpusz összeállítása ezen paraméterek mentén kizárta, hogy az elemzés során nyert eredmények, illetve az eredmények alapján levont következtetések csak egy szövegtípus és egy nyelvpár esetén legyenek validak. A vizsgált négy szövegtípus és hat forrásnyelv vonatkozásában kapott lexikai kohéziós adatok tágabb általánosíthatóságot tesznek lehetővé.

3.1. Korpuszépítés

Az elemzett szövegek a szövegprodukciónban részt vevők száma és egymáshoz való viszonya alapján mind monologikus 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 (vö. 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

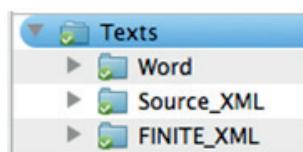
megszerkesztett volta teszi lehetővé a kohéziós elemek vizsgálatát, hiszen az írásbeli kommunikáció során a szövegalkotónak lehetősége nyílik a gondolatok körültekintő kifejtésére és a gondolatokat legpontosabban kifejező lexikai elemek és szerkezetek választékos, a kommunikációs szándékuk és a szövegtípusnak legmegfelelőbb alkalmazására.

A szövegeket részben nyilvános internetes oldalakról, adatbázisokból töltöttem le, részben ismerősök bocsátották fordításaikat a nyelvészeti elemzés rendelkezésére. E szövegek esetében is fontosnak tartottam, hogy ne multimodálisak és multidimenzionálisak legyenek, mivel ebben az esetben a szövegek kohézióját csak részben biztosítanák nyelvi eszközök, másik részük vizuális eszközökre épülne (vö. Háhn 2011). A mindenki számára legálisan elérhető internetes honlapokról letöltött szövegek esetében nem okozott gondot a szerzői jogi kérdések tisztázása és a begépelés. Az elemzés számára a szövegeket Microsoft Word formátumban archiváltam.

3.2. A korpusz szerkezete

A Word-dokumentumként lementett és az elemzés különböző szakaszaiban keletkezett szövegek archiválására a számítógépes dokumentumok tárolásánál bevált folderes struktúrát választottam. Az egyes foldereken belül a kiválasztott paraméterek alapján további, egymással azonos szinten elhelyezkedő, illetve egymásba ágyazott alfolderek definiálására nyílik lehetőség. Ez a folderstruktúra az emberi keresés számára is könnyen átlátható elrendezést tesz lehetővé.

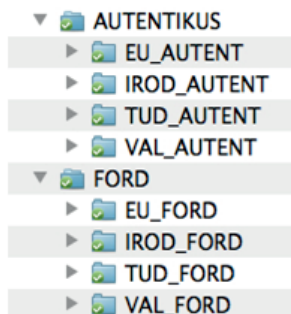
A legfelső szinten a korpuszban tárolt szövegeket a dokumentum formátuma szerint rendeztem három különböző folderbe. Így elkülönítettem az eredetileg archivált Word-dokumentumokat (WORD) és az OpenOffice program segítségével átalakított, az elemzés bemeneti fájljaként használt XML fájlokat (Source_XML), illetve az elemzés végeredményeként kapott, a lexikai kohéziós kapcsolatokat tartalmazó kimeneti XML dokumentumokat (FINITE_XML). A szövegfájlok fájlformátum szerinti elkülönítése (4. ábra) csupán az emberi keresést és munkavégzést könnyíti meg, mivel az egyes fájl-típusok explicit tulajdonságok alapján is könnyen megkülönböztethetőek egymástól (pl. kiterjesztés, belső struktúra).



4. ábra: Az MLKP korpusz legfelső folderszintje

A mindhárom fájl-típus alapján kialakított alforderben a szövegfájlokat további alfolderekbe csoportosítottam autentikus magyar (AUTENTIKUS) és célnyelvi magyar (FORDÍTOTT) mivoltuk alapján. A folderstruktúra következő szintjén – továbbra is egészen az ezen a szinten található összes folderben – alakítottam ki a kommuni-

kációs szintér típusa szerinti négy kategóriának megfelelően, a közéleti (EU), az irodalmi (IROD), a tudományos (TUD) és a vallásos (VAL) szövegeket tartalmazó alfordereket (5. ábra).



5. ábra: Az MLKP korpusz folderstruktúrája

Az autentikus szövegprodukciónál az autentikus szintér típusa szerinti folderek (EU_AUTENT, IROD_AUTENT, TUD_AUTENT, VAL_AUTENT) tartalmazzák az adott kategóriába tartozó szövegfájlokat. A fordított szövegeket tartalmazó, kommunikációs szintér szerinti alfolderek (EU_FORD, IROD_FORD, TUD_FORD, VAL_FORD) esetében viszont a forrásnyelvek alapján kialakítottam további alfoldereket. Ezen a szinten már nem törekedtem az egységes struktúra továbbvitelére, hanem az alfordereket 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 rendelkezésre: angol, francia, német, olasz, latin, görög (lásd 3. táblázat). A szövegtípusokon belül a forrásnyelvek szerinti folderek tartalmazzák az egyes szövegfájlokat.

| Fordított alkorpusz | Forrásnyelvek |
|---------------------------|-------------------------------------|
| közéleti (EU) szövegek | angol, francia |
| irodalmi (IROD) szövegek | angol, francia, német, olasz |
| tudományos (TUD) szövegek | angol, francia, német, olasz |
| vallásos (VAL) szövegek | angol, francia, német, latin, görög |

3. táblázat: A fordított alkorpusz szövegeinek forrásnyelvei

3.3. A korpusz összetétele

A korpuszépítés legfontosabb kritériuma az autentikus és célnyelvi magyar szövegprodukciónak a reprezentálása volt. A szövegek gyűjtésekor elsődleges szempontként szerepelt, hogy a szövegek elektronikus fájlok formájában álljanak rendelkezésre. A do-

küldetések gyűjtésének második meghatározó kritériuma a nyilvános elérhetőség, vagy a kutatási célokra való felhasználás engedélyezése 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” (Prószték, Olasz és Váradi 2006:1054). A szövegek többségének internetes elérhetősége ugyan kétségkívül hozzájárult például az összehasonlításban szereplő szövegek forrásnyelvek szerinti diverzitásához, ugyanakkor éppen a szövegek nyilvános elérésének korlátai határozták meg legfőképpen a korpusz összetételének lehetőségét.

3.3.1. Az alkorpuszok tervezése

A projekt tervezése során egyértelműen kirajzolódott, hogy a kutatás lefolytatásához mely nyelvtechnológiai eszközök alkalmazása vált reálisan elérhetővé. A rendelkezésre álló nyelvtechnológiai eszközök köre magyar nyelvű szövegek automatikus elemzését tette lehetővé. Az eszközök szabta kereteknek megfelelően a vizsgálat fókuszába az autentikus és fordítás során keletkezett szövegproduktó szövegépítő sajátosságai kerültek. A gyűjtött szövegek esetében mindezek következtében igen fontos szempontot jelentett, hogy a szövegek keletkezésével kapcsolatos információk egyértelművé tegyék, hogy a szöveg eredetileg magyar nyelven keletkezett-e vagy fordítás eredményeként jött létre, illetve hogy pontosan be lehessen azonosítani a szöveg keletkezésének eredeti nyelvét.

Az egybevetettség érdekében az autentikus és a fordított magyar szövegeket tartalmazó alkorpuszok a szövegek szövegépítő tulajdonságai és a szövegek keletkezésének körülményei szempontjából hasonló jellemzőket mutatnak. A közéleti alkorpuszokba az Európai Unióval kapcsolatos szövegek kerültek, a szépirodalmi alkorpuszba alapvetően regények és novellák. A vallási szövegek alkorpusza a hitbeli kérdésekkel foglalkozó egyházi szövegeket tartalmazza. Ez a szövegtípus azért releváns a vizsgálat szempontjából, 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 alkorpusza történelmi, filozófiai, nyelvészeti és pszichológiai témájú tudományos szövegeket tartalmaz.

3.3.2. Autentikus magyar szövegek

A korpuszban található autentikus közéleti magyar szövegek részben „A 2011. évi magyar EU-elnökség” honlapjáról⁴³, részben az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár⁴⁴ archívumából származnak. A dokumentumok kiválasztásakor meghatározó szempont volt, hogy a szövegek információtartalmuk szerint ne jogszálló

⁴³ www.mfa.gov.hu/kum/hu/bal/eu/2011_eu_elnokseg

⁴⁴ <http://mek.oszk.hu>

felsorolások legyenek, hanem egybefüggő gondolatokat tartalmazó szöveges állományok. Az autentikus magyar szépirodalmi alkorpusz szövegei az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából származnak. A magyar nyelven írt vallásos témájú szövegek forrásai egyrészt magyar papok nyilvános weboldalaikon közzétett írásai, másrészt az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumának és a Magyar Katolikus Egyház hivatalos honlapjának⁴⁵ anyagai. A letöltött dokumentumok között megtalálhatóak elektronikus formában elérhető könyvek, továbbá a Magyar Katolikus Püspöki Konferencia körlevelei és iránymutató írásai. A tudományos szövegtípus autentikus magyar szövegei közé különféle tudományos honlapokról letöltött történelmi, filozófiai, nyelvészeti és pszichológiai témájú tudományos szövegek kerültek.

3.3.3. Fordított magyar szövegek

A fordított magyar korpusz szövegállományának összegyűjtése hosszabb kutatómunka eredménye volt, mivel igen nehéz volt az autentikus magyar korpusz kritériumaihoz hasonló célnyelvi magyar korpuszt építeni. A különböző forrásnyelvekről fordított célnyelvi magyar szövegeket tartalmazó alkorpusz összeállítása a kritériumoknak megfelelő szövegek elérhetősége miatt vált nehezen megoldhatóvá, ezért az alkorpusz építésénél a már felkutatott weboldalokról származó, vagy az ismerősöktől kapott dokumentumok archiválása volt megvalósítható. Az adatok összehasonlíthatósága érdekében a szövegek forrásnyelvenkénti és szövegtípusonkénti jellemzői alapján kellett az autentikus magyar szövegeket tartalmazó alkorpuszok méretéhez igazodó fordított szövegeket tartalmazó alkorpuszokat létrehozni.

A fordított magyar korpusz esetében az angolból fordított közéleti szövegek nagy részét az Európai Unió hivatalos weboldaláról⁴⁶, másik részét pedig az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából töltöttem le. Más weboldalak nem tartalmaztak a kutatás időszakában hasonló témában és ellenőrizhetően fordítás eredményeként létrejött dokumentumállományokat. A fordított magyar korpusz szépirodalmi kategóriájába tartozó angol, francia, német, olasz szövegek az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából származó dokumentumok. A vallásos témájú latinból fordított szövegek a Vatikán honlapjáról⁴⁷ származó, nyilvánosan elérhető pápai enciklikák. A német nyelvről magyarra fordított Anselm Grün könyveket a fordító bocsátotta a kutatás rendelkezésére. Az angol nyelvű vallásos témájú műveket az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívumából töltöttem le. A görögökből fordított Újszövetséget a Bibliatársulat nyilvánosan elérhető honlapjáról⁴⁸ archiváltam. A tudományos tárgyú szövegek esetében a szövegek egy részét az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár archívu-

⁴⁵ <http://www.katolikus.hu>

⁴⁶ http://europa.eu/index_hu.htm

⁴⁷ <http://www.vatican.va>

⁴⁸ <http://www.biblia-tarsulat.hu/biblia/parhuzam.htm>

mából vagy az Educatio Társadalmi Szolgáltató Nonprofit Kft.⁴⁹ által üzemeltetett és működtetett oldalról töltöttem le, másik részük ismerősök által készített, a kutatás rendelkezésére bocsátott fordítások.

3.4. A korpusz mérete

A korpusz végső méretét a kutatás lefolytatásakor digitális formában elérhető dokumentumok száma határozta meg, ami nem zárja ki, hogy bizonyos nyelvek és bizonyos szövegtípusok esetében ne lett volna könnyű tovább bővíteni a korpuszban található dokumentumok számát. Ugyanakkor a vizsgált forrásnyelvi irányok zöménél erre nemigen kínálkozott lehetőség, ezért az adatok összevethetősége érdekében nem volt célravezető tovább folytatni a szöveggyűjtést. A korpuszban archivált anyagok összetétele és mennyisége alapján a megfogalmazott kutatási hipotézisek (lásd VII./2 fejezet) bizonyítására a jelen kutatás alapjául szolgáló korpuszállomány megfelelő méretű és összetételű és a korpusz szövegeinek elemzésével kapott eredmények lehetővé teszik a hipotézisek bizonyítását vagy cáfolatát.

A korpuszok méretét a szakmai gyakorlat a tokenszám megadásával határozza meg. A Magyar Lexikai Kohéziós Projekt teljes szövegállományának méretét az autentikus és a célnyelvi magyar alkorpuszokra vonatkozó szövegtípusonkénti adatok megadása szemlélteti.

| AUTENTIKUS | TOKENSZÁM | FORDÍTOTT | TOKENSZÁM |
|---------------|------------------|---------------|------------------|
| EU | 249 315 | EU | 249 420 |
| IROD | 255 699 | IROD | 1 070 126 |
| TUD | 257 442 | TUD | 853 869 |
| VAL | 245 009 | VAL | 873 597 |
| ÖSSZES | 1 007 465 | ÖSSZES | 3 047 012 |

4. táblázat: Az autentikus és a célnyelvi magyar alkorpusz

A 4. táblázat adataiból látható, hogy az autentikus magyar szövegeket tartalmazó alkorpusz 1 007 465 tokenszámú, míg a célnyelvi magyar szövegeket tartalmazó alkorpusz 3 047 012 tokenszámú. Összességében megállapítható, hogy az autentikus és a célnyelvi magyar szövegprodukciónak lexikai kohéziós szövegépítő sajátosságait a Magyar Lexikai Kohéziós Projekt közel 4 000 000 tokenszámú szöveg elemzéseinek eredményein keresztül vizsgálja. Az elemzett szövegállomány a szövegtípusok és nyelvpárok diverzitásának tekintetében, illetve a korpusz mérete alapján is úttörő jellegű vállalkozás a lexikai kohéziós fordítástudományi kutatások között.

⁴⁹ <http://www.tankonyvtar.hu>

4. A Magyar Lexikai Kohéziós Projekt annotálása

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 (*Hyper Text Markup Language*)⁵⁰, az SGML (*Standard Generalized Markup Language*)⁵¹ vagy az XML (*Extensible Markup Language*)⁵². A humán tudományok esetében általában a számítógépes feldolgozásra szánt digitális dokumentumokat SGML- vagy XML-alapú, platformfüggetlen kódolással látják el. A jelölőelemek (*tag*) szabadon bővíthetőek, de használatuknak szigorú szintaxisa van, amit a TEI (*The Text Encoding Initiative*)⁵³ ajánlásai szabályoznak.

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 többletinformáció, például a nyelvészeti elemzés eredményei kerülnek a szövegbe (Prószéky és 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 mondatathárok, a főnévi csoportok és az ismétlődő lexikai elemek bejelölése, illetve a kohézió típusának megjelenítése. A jelen kutatás során az elemzendő szövegek feldolgozását, azaz a bekezdések, mondatathárok és kohéziós eszközök bejelölését egy erre a feladatra készített korpusz nyelvészeti program segítette. Az elemzés eredményeként kapott annotált dokumentumban való eligazodást segíti, ha az annotált információkat különböző színek jelenítik meg (6. ábra).

```

5073 <sentence><origin>Az ő ismeretek ő közvetítése ő külföldiek ő számára ő sokféle ő formában ő és ő szinten ő történhet ő
• országon ő belüli ő és ő külföldi ő helyszíneken ő egyaránt ő. </origin><token ő stem="az" ő tag="[REJT]">Az </
• token><token ő stem="ismeret" ő tag="[EN][PL][NOM]">ismeretek </token><token ő stem="közvetítés" ő tag="[FN]
• [PSz3][NOM]">közvetítése </token><token ő stem="külföldi" ő tag="[FN][PL][NOM]">külföldiek </token><token ő
• stem="számára" ő tag="[NY][PSz3]">számára </token><token ő stem="sokféle" ő tag="[MN][NOM]">sokféle </
• token><token ő stem="forma" ő tag="[FN][INE]">formában </token><token ő stem="és" ő tag="[KOJT]">és </
• token><token ő stem="szint" ő tag="[EN][SU.P]">szinten </token><token ő stem="történik" ő tag="[IGE][HAT]"
• [e3]">történhet </token><token ő stem="," ő tag="[PUNCT]">,</token><token ő stem="ország" ő tag="[FN]
• [SU.P]">országon </token><token ő stem="belüli" ő tag="[MN][NOM]">belüli </token><token ő stem="és" ő
• tag="[KOJT]">és </token><token ő stem="külföldi" ő tag="[MN][NOM]">külföldi </token><token ő stem="helyszín" ő
• tag="[EN][PL][SU.P]">helyszíneken </token><token ő stem="egyaránt" ő tag="[HA]">egyaránt </token><token ő
• stem="." ő tag="[PUNCT]">.</token></sentence>

```

6. ábra: Az MLKP korpusz annotált szövegrészlete

⁵⁰ <https://www.w3schools.com/html>

⁵¹ <https://www.w3.org/TR/WD-html40-970708/intro/sgmltut.html>

⁵² https://www.w3schools.com/xml/xml_what_is.asp

⁵³ <http://www.tei-c.org/index.xml>

5. A Magyar Lexikai Kohéziós Projekt elemzése

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 jelölése (Péché 2007). A nyelvtechnológiai eszköz az azonos kategóriaelemeket mindig azonos metaadatot tartalmazó címkével látja el, ami különösen a nagy szövegtörzsek lefuttatott elemzéseik esetében meghatározó jelentőségű. A számítógéppel segített elemzés tehát nagymértékben hozzájárul a kutatások megbízhatóságához (*reliability*) (vö. Károly 2002a, Seidl-Péché 2024), hiszen az automatikus elemzést megismételve mindig ugyanazt az eredményt kapjuk.

A kohéziós elemek identifikációját a Magyar Lexikai Kohéziós Projekt céljainak megfelelően összekapcsolt nyelvtechnológiai eszközök segítették. A kiindulási korpuszban archivált Word-dokumentumokból első lépésben az OpenOffice program segítségével a bekezdéseket és mondathatárokat megtartó XML dokumentumokat készítettem. A Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar PhD hallgatói fejlesztettek a Magyar Lexikai Kohéziós Projekt célkitűzéseinek megfelelő elemzőt (Laki 2011, Novák, Orosz és Indig 2011). Az XML szövegfájlok szolgálták az elemző bemenetét. Az elemző először a mondathatárokat és a tokeneket azonosítja, majd a HuMor program (Prószéky 1995) segítségével végzi el a morfológiai annotációt és egyértelműsítést és végül az egyes bekezdéseken belül megkeresi a főnevek más mondatokban történő ismétléses előfordulásait, azaz a magyar WordNet Program (vö. Prószéky és Mihaltz 2008, Mihaltz 2005) élei mentén azonosítja a mondathatárokon túlmutató – szinonímia, antonímia, meronímia, hiponímia vagy doménreláción alapuló – szemantikai kapcsolatokat.

A HuWordNet Program bevonása a lexikai kohéziós kapcsolatok feltárásába eltér a lexikai kohézió kutatásának eddigi gyakorlatától. Mivel a WordNet-formalizmus kizárólag a főnévi csoportok kohéziós kapcsolatainak feltárását teszi lehetővé és a kapcsolatok azonosítása a jelen kutatás során csak a bekezdéseken belül valósul meg, a Magyar Lexikai Kohéziós Projekt jóval szűkebb megközelítést kínál a lexikai relációk azonosítására, mint a manuális elemzésre adaptált Hoey (1991)- vagy Károly-féle (2002) lexikai ismétlésmodellek. Más szempontból a WordNet-formalizmus tartalmaz olyan relációkat is, amelyek az előző két modellnek nem részei (például *be in state* relációk, *region domain* relációk), de ezen relációk alacsony számú előfordulása az egyes szövegeken belül lehetővé teszi ezen relációk figyelmen kívül hagyását a kutatás szempontjából.

A manuális elemzési taxonómiák (Hoey 1991, Károly 2002b) és a WordNet-formalizmus közötti különbségekből fakadó eltérések érintik a kohéziókutatás lényeges elméleti vonatkozásait, és az elemzés figyelmen kívül hagyja például azokat a lexikai relációkat, amelyek a bekezdések között alkotnak köteléket vagy nem egy bekezdésen belül vannak. Ugyan Hoey elemzési módszerének újszerűsége éppen abban rejlik, hogy lehetőséget teremt a szöveg közelebb és távolabb eső, kisebb és nagyobb részei közötti relációk azonosítására, az itt kidolgozott módszer ezt mégsem teszi lehetővé. A kutatás

során alkalmazott elemző (Laki 2011, Novák, Orosz és Indig 2011) továbbfejlesztése kínálhat arra megoldást, hogy a lekérdezések a szöveg nem egy bekezdésen belüli, távolabb eső részei közötti lexikai kohéziós kapcsolatokat is azonosítsák. A jelen kutatás esetében ugyanakkor a WordNet szűkített körű elemzési kerete és a lexikai relációk bekezdésen belüli azonosítása ellenére megvalósul az autentikus és fordított szövegek lexikai ismétlési mintázatainak egybevetése, hiszen a kutatás elemzési kerete az összes szövegtípus és nyelvpár esetében változatlan. Az elemzések alapján levont következtetések hatóköre egyelőre a bekezdésen belüli főnévi lexikai kohéziós kapcsolatokra terjed ki.

5.1. Szegmentálás és morfológiai elemzés

A lexikai kohéziós elemek identifikációja céljából annotált korpuszban a szegmentálás első lépése a téma tagolódásának vizuális jelölésére szolgáló bekezdések gépiles olvasható metainformációval (tag) való bejelölése. A vizuálisan az egyik bentebb kezdéstől (azaz behúzástól) a másik bentebb kezdésig tartó 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 szintén szükséges lépés, mert a lexikai kohéziós relációk a mondathatárokon túlmutató szövegsemantikai kapcsolatok. Az automatikus gépi eszköz tehát az egyes bekezdéseken belül található különböző mondatok lexikai elemei közötti kapcsolatok identifikációját tette lehetővé.

A szegmentálás következő lépéseként 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*), amely információ magában foglalja 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. A kutatás számára készített elemző (Laki 2011, Novák, Orosz és Indig 2011) feladata a következő lépésben az egy bekezdéseken belül található főnevek felkínálása a magyar WordNet adatbázis számára és annak segítségével az ismétlésrelációk megtalálása.

5.2. A lexikai kohéziós kapcsolatok 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 alkalmazásának szerepe az annotáció során tehát annak vizsgálata, hogy az adott bekezdés bármely főneve esetében az adott főnév WordNet adatbázisban nyilvántartott szemantikai relációi mentén található lexikai elemek valamelyike előfordul-e a vizsgált főnévvel azonos bekezdés többi mondatában. Az elemző egyrészt az egyes paragrafusok végén

összegezi a kapcsolat típusa szerint a bekezdésben talált lexikai-szemantikai kapcsolatok számát (7. ábra), másrészt a szöveg végén az egyes bekezdések után található végeredményeket összegezve adja meg a főveleknek az adott dokumentum különböző bekezdéseiben azonosított kohéziós ismétléses kapcsolatait. Ez az összesítés hasonlít a 7. ábra által bemutatott összesítéshez, csak a teljes dokumentumban előforduló ismétléskapcsolatokra vonatkoznak az adatok.

```
<statistic>
<numberOf edge="hyponym">4</numberOf>
<numberOf edge="synonym">1</numberOf>
<numberOf edge="antonym">1</numberOf>
<numberOf edge="repetition">10</numberOf>
<tokens>81</tokens>
</statistic>
```

7. ábra: Példa az ismétléskapcsolatok összegzésére

A lexikai kohéziós kapcsolatok azonosítása során az elemző bejelöli az elemzés főkuzába tartozó ismétléseket, szinonimákat, antonimákat, meronimákat és hiponimákat. A következőkben néhány szövegminta érzékelteti azt a folyamatot, ahogy az internetről vagy egyéb forrásokból letöltött szövegeket az elemző feldolgozza és ahogy a kohéziós kapcsolatokat azonosítja.

5.2.1. A lexikai kohéziós kapcsolatok realizációja

Az első példa (8. ábra) a közéleti szövegtípusba tartozó *Az unió kapujában*⁵⁴ című autentikus magyar Word-dokumentum szövegrészlete:

– 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 közgazdaság pedig jól felfogott érdekei szerint védekezik. Tehát ez is örök háború. Egy maffiának szüksége van arra, ha külföldi, hogy útlevelet, betelepedési, bontási, 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 tehesen szert, hozzájusson közpénzekhez is. Az alvilág részéről rendszeresen érzékelhető, kitalálható ez a támadás. Hogy hol állunk ebben a küzdelemben, ez az igaz kérdés. Vannak olyan pozíciók, például a közgazdaságban, amit aktívan és nem eredménytelenül támadnak a maffiák. Nagyon sokszor előfordult a közpénzek felhasználásakor, az állami megrendelésekkor, hogy valószerűsítendő volt: a maffiák vitték el ezeket a pénzeket.

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

A 9. ábra a szöveg szegmentálását és morfológiai elemzését szemlélteti a második mondat bemutatásával.

⁵⁴ Molnár P. (szerk.) 2000. *Az unió kapujában*. Budapest: Mundus Magyar Egyetemi Kiadó. Elérhető: mek.oszk.hu/01900/01989

```

<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][PSz3][DAT]">felhasználásának</token>
<token stem="eldöntés" tag="[FN][PSz3][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][PSz3][ACC]">lábát</token>
<token stem="." tag="[PUNCT]">.</token></sentence>

```

9. ábra: Példa a 8. ábra második mondatának morfológiai annotálására

A 9. ábra fekete színnel jeleníti meg a példamondatot, illetve utána soronként a mondat tokenizált egységeit, beleértve a lexikai elemeket és a központozást. Kék színnel jelennek meg a nyelvészeti elemzés során kapott metainformációk, amelyek minden esetben ún. kacsacsőrök között elhelyezkedő nyitó- és zárótagokból állnak.

| | | | |
|---------------|--------------|---|----------------|
| "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": | "avilág" | ➤ | "avilág" |
| "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": | "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

A kutatás szempontjából az elemzés következő lépése az igazán érdekes, amikor a bekezdésen belüli lexikai kohéziós kapcsolatokat azonosítja a program. Az elemző kimeneti dokumentumaként kapott XML fájl tartalmazza a nyelvészeti annotációkat. Az eredeti XML fájl struktúráját szemlélteti leegyszerűsítve (terjedelmi keretek miatt az annotáció jelölésére szolgáló tageket nem a teljesség igényével megjelenítve) a 10. ábra. A tagekben található információk közül az ábra csak a lexikai kohéziós kapcsolatok fajtájára utaló kifejezéseket jeleníti meg kék színnel.

A 10. ábra a szövegben lineárisan haladva szemlélteti a példaként hozott közéleti szövegtípushoz tartozó szövegrészletben realizálódó kohéziós kapcsolatokat. Az elemző 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 előtti tagmondat utolsó lexikai eleméhez. A megtalált kohéziós kapcsolatokat minden esetben a kapcsolat típusa szerint is azonosítja az elemző. A 8. ábra szövegrésze azt is jól szemlélteti, hogy a gépi módszer nem alkalmas a szövegalapú relációk azonosítására, azaz azoknak a logikai kapcsolatoknak a feltárására, amelyek az adott szövegkörnyezetre jellemzőek. A 8. ábra szövegrészeletében erre példa a fegyver – alvilág/fegyver – embercsempész/fegyver – maffia logikai kapcsolat.

– 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, kacsá mindig legyen a menüben, mert a kacsát egyetlen konyha sem tudja olyan porhanyósra sütni, mint a kínai.**

11. ábra: Szövegrészlet A kitagadott című regényből

Szintén nyelvi példaként szolgál a szépirodalmi szövegtípusba tartozó Ballai László *A kitagadott*⁵⁵ című regényéből (11. ábra) származó szövegrészlet. Az ábra szövegrészletében a kohéziós kapcsolatok száma sokkal kevesebb, mint a 8. ábra szövegrészletében. A 11. ábra szövegrészletében a szemléltetés miatt jelennek meg félkövér betűtípussal kiemelve a kohéziós kapcsolatokat megteremtő lexikai egységek. Itt kell még egyszer felhívni a figyelmet arra, hogy a jelen kutatás csak a névszói elemek lexikai kohéziós kapcsolatainak beazonosítására szorítkozik.

⁵⁵ Ballai L. 2000. A kitagadott. Elérhető: www.vasaros.com/art/ballai

| | | | |
|---------------|-------------|---|--------------|
| "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 12. ábra esetében jól látszik, hogy az *étterembe* lexikai elemnek csak olyan ismétléskapcsolata van, amelyet változatlan formájú lexikai elem biztosít (*repetition*). Az adott szövegrészletben hasonlóan viselkednek a *konyháról*, a *vacsorára*, a *menüt* és a *kacsa* elemek is, ezek esetében is változatlan formájú lexikai elem (*repetition*) biztosítja a kohéziós kapcsolatot. A *szószokat* lexikai elem esetében több meronimakapcsolatot azonosított az elemző, melyek: az *ételek*, az *ételeiket*, a *fogásoktól* és az *ételekig*. Az *ételek* lexikai egység esetében találunk szó szerinti ismétléseket (*ételeiket*, *ételekig*) és szinonima-kapcsolatokat (*fogásoktól*) is.

A tudományos szövegtípusra mutat példát *A magyar történelem tizedei*⁵⁶ című Bonfini-mű magyar fordítása Kulcsár Péter fordításában (13. ábra). A 13. ábra szövegrészletében is félkövér betűtípussal kiemelve jelennek meg a lexikai kohéziós kapcsolatrendszer elemei, melyeket a 14. ábra mutat be.

Theodosius parancsára Ephesusban zsinatot tartanak, amelyen **Dioscorus** elnökeletével 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 gyűlekezet** megerősítette, hogy **Krisztusban** kettős természet van, és ugyanaz az egyetlen **Krisztus** kell istennek és embernek hinni. Aztán valamennyi **atya** egyetértésével nyomban kiátkozták a manicheus fertőzés szerzőit, és **könyveiket** nyilvánosan elégették; **Dioscorus** dühjét megaposták, és az eretnokséget eltörölték.

13. ábra: Szövegrészlet A magyar történelem tizedei műből

⁵⁶ 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>

| | | | |
|------------------------|------------------------|---|------------------------|
| " <i>repetition</i> ": | " <i>Theodosius</i> " | ➤ | " <i>Theodosius</i> " |
| " <i>synonym</i> ": | " <i>parancsára</i> " | ➤ | " <i>utasítására</i> " |
| " <i>hyponym</i> ": | " <i>zsinatot</i> " | ➤ | " <i>gyűlést</i> " |
| " <i>repetition</i> ": | " <i>Dioscorus</i> " | ➤ | " <i>Dioscorus</i> " |
| " <i>hyponym</i> ": | " <i>gyűlést</i> " | ➤ | " <i>gyülekezet</i> " |
| " <i>repetition</i> ": | " <i>atya</i> " | ➤ | " <i>atya</i> " |
| " <i>hyponym</i> ": | " <i>atya</i> " | ➤ | " <i>szerzőt</i> " |
| " <i>repetition</i> ": | " <i>Krisztusban</i> " | ➤ | " <i>Krisztust</i> " |

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

A 13. ábra és 14. ábra szemlélteti azt az esetet, amikor egy lexikai elem egy korábban előforduló lexikai elem ismétléses kapcsolata (*zsinatot*–*gyűlést*), majd később a szövegben az adott elemnek is megtalálható egy másik lexikai elem által biztosított ismétléses kapcsolata (*gyűlést*–*gyülekezet*). A *zsinatot* elem hiponimakapcsolatban van a *gyűlést* lexikai elemmel, majd a *gyűlést* elem hiponimakapcsolataként jelenik meg a *gyülekezet* kifejezés.

Az utolsó példa (15. ábra és 16. ábra) a *Caritas in veritate*⁵⁷ 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 15. ábra szövegrészletében ebben az esetben is félkövér betűtípus emeli ki a kohéziós kapcsolatokat biztosító lexikai elemeket.

VI. Pál a *Populorum progressióban* olyan piactudományi 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 *Respublicae* 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 önigazgatásához **szüksége** van az állam osztó célzati 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

⁵⁷ 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

| | | | |
|---------------|-------------------|---|------------------|
| "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

5.2.2. Példák a kohéziós kapcsolatok típus szerinti realizációjára

A már említett terjedelmi keretek miatt a Magyar Lexikai Kohéziós Projektben szereplő antonima-, meronima- és hiponimakapcsolatok bemutatása néhány nyelvi példára szorítkozik és a kohéziós kapcsolatok jelen korpuszban megtalálható sokszínű megvalósulását hivatott alátámasztani. A példák a Függelék 1-ben szerepelnek.

5.3. Az automatikus elemzés ellenőrzése

A Magyar Lexikai Kohéziós Projekt (MLKP) során a kohéziós elemek identifikációja nyelvtechnológiai eszközök összekapcsolásával történik, ami nagyobb szövegtípus elemzését teszi lehetővé. Az MLKP korpusz mérete ugyanakkor nem teszi lehetővé a lexikai kohéziós kapcsolatok automatikus bejelölésének ellenőrzését minden szövegfájl esetében. Ezért az elemzési lánc eredményeit a korpuszban található szövegek egy részén végzett elemzések tesztadatai alapján lehet validálni.

A validálásnál abból a már elfogadott téziszból indultam ki, 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 (vö. Károly 1979:55). Az elemzések eredményeit ezért először az autentikus magyar szövegek kisebb korpuszán ellenőriztem. Az eredmények a különböző szövegtípusba sorolható autentikus magyar szövegek egymással összehasonlítható nagyságú szövegtípus szerinti alkorpuszai esetében, az egyes szövegtípusok szövegépítő tulajdonságainak különbözőségét tükröző kohéziós mintázatbeli eltéréseket mutatnak.

Az ellenőrzéshez szükséges adatok lekérdezéséhez összeállított kísérleti korpusz tartalmazza az MLKP korpusz összes szövegtípus kategóriájának megfelelő 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 jelölése az V./5 fejezetben bemutatott magyar WordNet (Prószték és Mihaltz 2008, Mihaltz 2005) formalizmus segítségével történik, megkülönböztetve az ismétlés (R), szinonima (S), az ellentét (A), hiponima (H) és meronima (M) kategóriákat. Mivel a relációk közül a *category domain*, a *be in state* és a *domén (category/region/usage domain)* relációk számszerű előfordulása elenyésző (az utóbbi kettő még a teljes korpusz esetében sem éri el, vagy csak éppen a tokenszámhoz viszonyított 0,01%-os előfordulási gyakoriságot), ezért vizsgálatukat nem integrálta az MLKP, illetve az előfordulásukra vonatkozó adatok közlésétől is teljes mértékben eltekintett.

| ELLENŐRZŐ KORPUSZ | | TOKENEK | R | S | A | H | M | ÖSSZES |
|-------------------|-----------------------------|---------|--------|-----|-----|-------|-----|--------|
| EU | DB | 71 427 | 5 235 | 455 | 80 | 1 751 | 143 | 7 664 |
| | 1000 tokenre vetített érték | | 73 | 6 | 1 | 25 | 2 | 107 |
| IROD | DB | 71 244 | 6 342 | 348 | 159 | 1 141 | 97 | 8 087 |
| | 1000 tokenre vetített érték | | 89 | 5 | 2 | 16 | 1 | 114 |
| TUD | DB | 75 930 | 12 613 | 490 | 126 | 1 886 | 144 | 15 259 |
| | 1000 tokenre vetített érték | | 166 | 6 | 2 | 25 | 2 | 201 |
| VAL | DB | 71 038 | 12 894 | 563 | 479 | 1 580 | 648 | 16 164 |
| | 1000 tokenre vetített érték | | 182 | 8 | 7 | 22 | 9 | 228 |

5. táblázat: Az autentikus magyar szövegek egy részén mért adatok

Az automatikus ellenőrzés eredményeit összegző táblázat (5. táblázat) soraiban először az egyes szövegtípusokra jellemző értékek jelennek meg (DB) majd az összehasonlíthatóság érdekében ezen értékek 1000 tokenre vetített mennyisége (1000 tokenre vetített érték), míg az oszlopok a különböző ismétléskategóriák (R: ismétlés, S: szinonima, A: ellentét, H: hiponima, M: meronima) esetében tüntetik fel a tokenszámban kifejezett eredményeket, illetve az alkorpuszra jellemző, 1000 tokenre vetített mennyiségeket. Az egyes szövegtípusok esetében megfigyelhető a lexikai kohéziós elemek 1000 tokenre vetített mennyisége közötti eltérés, ami arra utal, hogy az ellenőrzésre összeállított korpusz igazolja az automatikus gépi eszközök összekapcsolásával kivitelezett kutatás mérési módszertanát.

6. A Magyar Lexikai Kohéziós Projekt eredményei

A közel 4 000 000 szövegszavas korpusz szövegeinek gépi elemzése és a korpuszban a lexikai kohéziós kapcsolatok automatikus azonosítása a vizsgált autentikus és fordított magyar közéleti, szépirodalmi, tudományos és vallási szövegek esetében eddig példa nélküli. A Magyar Lexikai Kohéziós Projekt a kohézió kutatás eddigi mennyiségi kereteit kitérítve teszi lehetővé az egyes szövegtípusokra jellemző lexikai kohéziós kapcsolatok konkrét nyelvi realizációjának vizsgálatát. A lexikai kohéziós kötések típusa szerinti nyelvi példák lekérdezése az annotált szövegegységekben hozzájárulhat a szövegek szövegszemantikai szerveződéséről alkotott elképzelés és a fordítás során e téren meglévő ismeretek pontosításához.

| AUTENT | TOKE- NEK | ÖSSZES LEX. KOH. KAPCS. | 1000 TOKENRE VETÍTETT LEX. KOH. KAPCS. | FORD | TOKE- NEK | ÖSSZES LEX. KOH. KAPCS. | 1000 TOKENRE VETÍTETT LEX. KOH. KAPCS. |
|--------|--------------|----------------------------------|---|--------|--------------|----------------------------------|--|
| EU | 249 315 | 27 142 | 109 | EU | 249 420 | 26 107 | 105 |
| IROD | 255 699 | 30 998 | 121 | IROD | 1 070 126 | 111 305 | 104 |
| TUD | 257 442 | 58 414 | 227 | TUD | 853 869 | 181 762 | 213 |
| VAL | 245 009 | 53 348 | 218 | VAL | 873 597 | 172 197 | 197 |
| ÖSSZES | 1 007 465 | 169 902 | 169 | ÖSSZES | 3 047 012 | 491 371 | 161 |

6. táblázat: A lexikai kohéziós kapcsolatok száma az MLKP-ben

A 6. táblázat az autentikus (AUTENT) és célnyelvi (FORD) magyar alkorpuszokat jellemző szövegtípusonkénti (közéleti: EU, szépirodalmi: IROD, tudományos: TUD és vallásos: VAL) tokenszámot (TOKE-NEK) és a szövegtípusban előforduló összes lexikai kohéziós kapcsolat számát (ÖSSZES LEX. KOH. KAPCS.) tartalmazza. A táblázatban a fordított szövegeket tartalmazó korpuszra jellemző adatok összesítve jelennek meg. Fontos megjegyezni, hogy a fordított szövegeket tartalmazó korpusz mérete közel háromszorosa az autentikus szövegeket tartalmazó korpusznak, mivel a fordított szövegek esetében a korpusz tartalmazza az összes nyelvi irányból magyarra fordított szövegeket. A táblázat utolsó oszlopa (1000 TOKENRE VETÍTETT LEX. KOH. KAPCS.) az összevethetőség érdekében az ezer tokenre standardizált adatokat tartalmazza. Az ezer tokenre vetített adatok esetében jól kirajzolódó tendencia, hogy a magyarra fordított szövegeket tartalmazó korpusz minden szövegtípus esetében kevesebb lexikai kohéziós elemet tartalmaz, mint az autentikus szövegek korpusza.

Az elemző (Laki 2011, Novák, Orosz és Indig 2011) által dokumentumonként ki-mutatott lexikai kohéziós kapcsolatszámot az autentikus és a célnyelvi szövegprodukcóra vonatkozóan külön Excel-táblázatban összegeztem. A táblázatokban a szövegekre vo-

natkozó lexikai kohéziós adatokat az egyes szövegtípusok és forrásnyelvek bontásában külön lehet vizsgálni. A táblázatok sorai tartalmazzák az egyes szövegekre, míg oszlopai az egyes lexikai kohéziós kategóriákra jellemző eredményeket. A 7. táblázat szemlélteti az egyes elemzett szövegekre vonatkozó adatokat a lexikai kohéziós kapcsolat típusa szerint.

| XX_sz. | SZÖVEG | TOKENEK | R | S | A | H | M |
|--------|-----------------------------|----------------|---------------|--------------|------------|--------------|------------|
| 1 | Ballai | 48 187 | 8 649 | 312 | 615 | 1 569 | 511 |
| 2 | Fekete | 83 292 | 5 588 | 656 | 166 | 2 069 | 199 |
| 3 | Sánta | 71 244 | 6 342 | 348 | 159 | 1 141 | 97 |
| | DB | 202 723 | 20 579 | 1 316 | 940 | 4 779 | 807 |
| | 1000 tokenre vetítve | | 102 | 6 | 5 | 24 | 4 |

7. táblázat: Részlet az eredményeket archiváló táblázatból

A táblázat oszlopaiban található adatok a tokenszám (TOKENEK), az ismétlések száma (R), a szinonimák száma (S), az antonimák száma (A), a meronimák száma (M) és a hiponimák száma (H). Az egy szövegtípuson belül az azonos forrásnyelvekről fordított szövegekre jellemző adatokat lexikai kohéziós kategóriánként összegzi az utolsó előtti sor (DB), majd a következő sor az adott érték ezer tokenre vetített előfordulási gyakoriságát adja meg egészekre kerekítve. Az így kapott értékek megmutatják, hogy melyik szövegtípusra melyik lexikai kohéziós eszköz használata a jellemzőbb.

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. A jelen fejezet mutatja be a forrásnyelvek szerint összesített eredményeket, amelyeket a részkorpuszok részben eltérő méretei miatt az összehasonlíthatóság érdekében minden esetben 1000 tokenre vetítve is tartalmaz a táblázat.

Az 8. táblázat az autentikus magyar szövegtípusok, míg a 9. táblázat az angol nyelvből (FORD ANGOL) fordított célnyelvi magyar szövegtípusok lexikai kohéziós elemeinek számát tartalmazza szövegtípusonkénti bontásban (közéleti [EU], szépirodalmi [IROD], tudományos [TUD] és vallásos [VAL]). A táblázat sorai a szövegtípusokra, oszlopai a tokenszámra és a lexikai kohéziós kapcsolatok típusára jellemző adatokat (tokenszám [TOKENEK], ismétlések [R], szinonimák [S], antonimák [A], meronimák [M], hiponimák [H] száma) mutatják. A szövegtípusra jellemző adatoknál a felső sorban (DB) az adott szövegtípusban talált összes lexikai kohéziós kapcsolat száma jelenik meg a kapcsolat típusa (R, S, A, H, M) szerinti bontásban. Az utolsó oszlop a lexikai kohéziós kapcsolatok számát tartalmazza összesítve (ÖSSZES). A szövegtípusra jellemző adatoknál az alsó sor (1000 tokenre vetítve) az adott szövegtípusra jellemző lexikai kohéziós kapcsolatszám 1000 tokenre vetített értékét mutatja. A 8. táblázat jól szem-

lélteti, hogy az autentikus magyar szövegek jelen korpuszában az egyes szövegtípusok esetében eltérő a lexikai kohéziós eszközök használatának gyakorisága. Ez az eltérés az egyes szövegtípusok eltérő szövegépítő sajátosságaival magyarázható.

| AUTENTIKUS | | TOKENEK | R | S | A | H | M | ÖSSZES |
|------------|----------------------|---------|--------|-------|-----|-------|-------|--------|
| EU | DB | 249 315 | 19 018 | 1 766 | 249 | 5 471 | 459 | 26 963 |
| | 1000 tokenre vetítve | | 76 | 7 | 1 | 22 | 2 | 108 |
| IROD | DB | 255 699 | 22 094 | 1 427 | 956 | 5 223 | 941 | 30 641 |
| | 1000 tokenre vetítve | | 86 | 6 | 4 | 20 | 4 | 120 |
| TUD | DB | 257 442 | 45 317 | 1 717 | 241 | 8 632 | 1 691 | 57 598 |
| | 1000 tokenre vetítve | | 176 | 7 | 1 | 34 | 7 | 224 |
| VAL | DB | 245 009 | 41 262 | 2 886 | 789 | 6 760 | 1 269 | 52 966 |
| | 1000 tokenre vetítve | | 168 | 12 | 3 | 28 | 5 | 216 |

8. táblázat: Az autentikus magyar szövegek kohéziós elemeinek száma

A 9. táblázat jól szemlélteti, hogy az angol nyelvből (FORD ANGOL) fordított jelen célnyelvi magyar szövegkorpusz esetében az egyes szövegtípusokra eltérő lexikai kohéziós eszközhasználat jellemző. A 8. táblázat és 9. táblázat egybevetése azt is megmutatja, hogy az autentikus és az angoltól fordított magyar szövegek esetében a lexikai kohéziós eszközhasználatban bekövetkező változás iránya szövegtípusonként eltérő (pl. a tudományos szövegtípus esetén az angoltól fordított szövegeknél erőteljesen nő a lexikai kohéziós eszközhasználat, míg a vallási szövegek esetében jelentősen csökken a számuk).

| FORD ANGOL | | TOKENEK | R | S | A | H | M | ÖSSZES |
|------------|----------------------|---------|--------|-------|-----|-------|-------|--------|
| EU | DB | 249 420 | 21 370 | 735 | 240 | 2 603 | 198 | 25 146 |
| | 1000 tokenre vetítve | | 86 | 3 | 1 | 10 | 1 | 101 |
| IROD | DB | 278 909 | 23 443 | 1 538 | 575 | 4 169 | 1 077 | 30 802 |
| | 1000 tokenre vetítve | | 84 | 6 | 2 | 15 | 4 | 110 |
| TUD | DB | 187 334 | 49 107 | 818 | 140 | 3 325 | 213 | 53 603 |
| | 1000 tokenre vetítve | | 262 | 4 | 1 | 18 | 1 | 286 |
| VAL | DB | 169 363 | 13 705 | 1 226 | 250 | 2 555 | 282 | 18 018 |
| | 1000 tokenre vetítve | | 81 | 7 | 1 | 15 | 2 | 106 |

9. táblázat: Az angoltól fordított magyar szövegek kohéziós elemeinek száma

A 10. táblázat a franciából fordított (FORD FRANCIA) célnyelvi magyar szövegtípusok, a 11. táblázat a németből fordított (FORD NÉMET) célnyelvi magyar szövegtípusok, míg a 12. táblázat az olaszról fordított (FORD OLASZ) célnyelvi magyar szövegtípusok és a 13. táblázat a latinból és görögből (FORD LAT, FORD GÖR) fordított célnyelvi magyar szövegtípusok lexikai kohéziós elemeinek számát tartalmazza szövegtípusonkénti bontásban (közéleti [EU], szépirodalmi [IROD], tudományos [TUD] és valóságos [VAL]). A táblázat sorai a szövegtípusokra, míg oszlopai a tokenzámra és a lexikai kohéziós kapcsolatok típusára jellemző adatokat (tokenzám [TOKENEK], ismétlések [R], szinonimák [S], antonimák [A], meronimák [M], hiponimák [H] száma) mutatják.

| FORD FRANCIA | | TOKENEK | R | S | A | H | M | ÖSSZES |
|--------------|----------------------|---------|--------|-------|-----|-------|-------|--------|
| IROD | DB | 257 731 | 12 392 | 897 | 396 | 2 617 | 1 460 | 17 762 |
| | 1000 tokenre vetítve | | 48 | 3 | 2 | 10 | 6 | 69 |
| TUD | DB | 280 259 | 33 988 | 1 807 | 464 | 6 006 | 768 | 43 033 |
| | 1000 tokenre vetítve | | 121 | 6 | 2 | 21 | 3 | 154 |

10. táblázat: A franciából fordított magyar szövegek kohéziós elemeinek száma

A szövegtípusra jellemző adatoknál a felső sorban (DB) az adott szövegtípusban talált összes lexikai kohéziós kapcsolat száma jelenik meg a kapcsolat típusa (R, S, A, H, M) szerinti bontásban. Az utolsó oszlop a lexikai kohéziós kapcsolatok számát tartalmazza összesítve (ÖSSZES). A szövegtípusra jellemző adatoknál az alsó sor (1000 tokenre vetítve) az adott szövegtípusra jellemző lexikai kohéziós kapcsolatszám 1000 tokenre vetített értékét mutatja.

A franciából fordított (FORD FRANCIA) célnyelvi magyar szövegtípusok (10. táblázat) csak irodalmi és tudományos alkorpuszokat tartalmaz. Mindkét szövegtípus esetében megállapítható, hogy a jelen korpusz franciából magyarra fordított szövegei az 1000 tokenre vetített értékek alapján lényegesen kevesebb lexikai kohéziós kapcsolatot tartalmaznak, mint az azonos szövegtípus autentikus alkorpusza.

| FORD NÉMET | | TOKENEK | R | S | A | H | M | ÖSSZES |
|------------|----------------------|---------|--------|-------|-------|-------|-------|--------|
| IROD | DB | 260 803 | 14 916 | 909 | 514 | 2 493 | 708 | 19 540 |
| | 1000 tokenre vetítve | | 57 | 3 | 2 | 10 | 3 | 75 |
| TUD | DB | 257 079 | 31 373 | 2 002 | 618 | 8 452 | 3 995 | 46 440 |
| | 1000 tokenre vetítve | | 122 | 8 | 2 | 33 | 16 | 181 |
| VAL | DB | 259 514 | 52 878 | 2 318 | 1 147 | 6 781 | 642 | 63 766 |
| | 1000 tokenre vetítve | | 204 | 9 | 4 | 26 | 2 | 246 |

11. táblázat: A németből fordított magyar szövegek kohéziós elemeinek száma

A németből fordított (FORD NÉMET) célnyelvi magyar szövegtípus (11. táblázat) irodalmi, tudományos és vallási alkorpuszokat tartalmaz. A 11. táblázat jól szemlélteti, hogy a jelen korpuszban található németről fordított magyar szövegek esetében az autentikus szövegprodukciónak képest a lexikai kohéziós eszközhasználatban bekövetkező változás iránya szövegtípusonként eltérő (pl. az irodalmi és a tudományos szövegtípus esetén a németből fordított szövegeknél erőteljesen csökken a lexikai kohéziós eszközhasználat, míg a vallási szövegek esetében nő a számuk).

Az olaszról fordított (FORD OLASZ) célnyelvi magyar szövegtípus (12. táblázat) csak irodalmi és tudományos alkorpuszokat tartalmaz. Mindkét szövegtípus esetében megállapítható, hogy a jelen korpusz olaszról magyarra fordított szövegei az 1000 tokenre vetített értékek alapján több lexikai kohéziós kapcsolatot tartalmaznak, mint az azonos szövegtípus autentikus alkorpusza.

| FORD OLASZ | | TOKENEK | R | S | A | H | M | ÖSSZES |
|------------|----------------------|---------|--------|-------|-------|-------|-------|--------|
| IROD | DB | 272 683 | 31 918 | 2 172 | 1 586 | 6 010 | 798 | 42 484 |
| | 1000 tokenre vetítve | | 117 | 8 | 6 | 22 | 3 | 156 |
| TUD | DB | 129 197 | 25 488 | 2 411 | 494 | 6 174 | 1 495 | 36 062 |
| | 1000 tokenre vetítve | | 197 | 19 | 4 | 48 | 12 | 279 |

12. táblázat: Az olaszról fordított magyar szövegek kohéziós elemeinek száma

A latinból és görögből (FORD LAT, FORD GÖR) fordított célnyelvi magyar szövegtípus (13. táblázat) csak vallási szövegeket tartalmaz. Mindkét nyelv esetében azt mutatják az adatok, hogy a magyarra fordított szövegek az 1000 tokenre vetített értékek alapján kevesebb lexikai kohéziós kapcsolatot tartalmaznak, mint az azonos szövegtípus autentikus alkorpusza.

| FORD VAL | | TOKENEK | R | S | A | H | M | ÖSSZES |
|----------|----------------------|---------|--------|-------|-----|--------|-------|--------|
| FORD LAT | DB | 269 713 | 37 236 | 4 488 | 530 | 11 471 | 1 062 | 54 787 |
| | 1000 tokenre vetítve | | 138 | 17 | 2 | 43 | 4 | 203 |
| FORD GÖR | DB | 175 007 | 28 532 | 1 812 | 702 | 2 517 | 547 | 34 110 |
| | 1000 tokenre vetítve | | 163 | 10 | 4 | 14 | 3 | 195 |

13. táblázat: A latinból és görögből fordított magyar szövegek kohéziós elemeinek száma

Az itt bemutatott eredmények mindegyike alátámasztja és az ezer tokenre vetített értékeknek köszönhetően jól szemlélteti, hogy a különböző forrásnyelvekről fordított magyar szövegek esetében a kohéziós elemek gyakorisága és eloszlásának aránya a vizsgált szövegtípusoknál különböző mértékben tér el az autentikus szövegprodukciónak jellemző

értékektől. Ez az eltérés valószínűleg azzal magyarázható, hogy az egyes nyelvekre eltérő lexikai kohéziós eszközhasználat jellemző és az eltérő forrásnyelvek eltérő módon befolyásolják a célnyelvi szövegalkotást a lexikai kohéziós eszközhasználat tekintetében.

6.2. A lexikai kohéziós eltolódások

Az előző fejezet forrásnyelvek szerint összesített eredményeit az adatok könnyebb egybevetethetősége, illetve a lexikai kohéziós eltolódásjelenség megragadása érdekében most szövegtípusonkénti bontásban vizsgáljuk.

6.2.1. Lexikai kohéziós eltolódások a közéleti szövegekben

Közéleti szövegeket a Magyar Lexikai Kohéziós Projektben csak az autentikus magyar alkorpusz, valamint angoltól fordított magyar alkorpusz tartalmaz. Ezen alkorpuszok esetében mutatja be a lexikai kohéziós eszközhasználatot a 14. táblázat.

| EU | | TOKENEK | R | S | A | H | M | ÖSSZES |
|---------|----------------------|---------|--------|-------|-----|-------|-----|--------|
| AUTENT | DB | 249 315 | 19 018 | 1 766 | 249 | 5 471 | 459 | 26 963 |
| | 1000 tokenre vetítve | | 76 | 7 | 1 | 22 | 2 | 108 |
| FORD EN | DB | 249 420 | 21 370 | 735 | 240 | 2 603 | 198 | 25 146 |
| | 1000 tokenre vetítve | | 86 | 3 | 1 | 10 | 1 | 101 |

14. táblázat: A közéleti szövegtípusra jellemző kohéziós elemek eloszlása

A 14. táblázat sorai az autentikus (AUTENT) és az angoltól fordított (FORD EN) közéleti témájú magyar szövegek lexikai kohéziós elemeinek számát és eloszlásának arányait tartalmazza. A táblázat oszlopai a tokenszámmra és a lexikai kohéziós kapcsolatok típusára jellemző adatokat (tokenszám [TOKENEK], ismétlések [R], szinonimák [S], antonimák [A], meronimák [M], hiponimák [H] száma) mutatják. Az utolsó oszlop a lexikai kohéziós kapcsolatok számát tartalmazza összesítve (ÖSSZES). Az alkorpuszra jellemző adatoknál a felső sorban (DB) az adott alkorpuszban talált összes lexikai kohéziós kapcsolat száma jelenik meg a kapcsolat típusa (R, S, A, H, M) szerinti bontásban. Az alkorpuszra jellemző adatoknál az alsó sor (1000 tokenre vetítve) az adott alkorpuszra jellemző lexikai kohéziós kapcsolatszám 1000 tokenre vetített értékét mutatja.

A lekérdezések alapján elmondható, hogy a jelen közéleti szövegek tekintetében közel azonos tokenszámú részkorpuszok esetében az angoltól 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 hiponimakapcsolat a jellemző, addig az angoltól fordított magyar közéleti szövegek esetében magasabb az ismétlések és kisebb a szinonima- és hiponimakapcsolatok

száma. Ez az eltérés valószínűleg azzal magyarázható, hogy az angol nyelvre a magyar-tól eltérő lexikai kohéziós eszközhasználat jellemző és az angol forrásnyelv befolyásolja a magyar célnyelvi szövegalkotást a lexikai kohéziós eszközhasználat tekintetében.

6.2.2. Lexikai kohéziós eltolódások a szépirodalmi szövegekben

Szépirodalmi szövegeket a Magyar Lexikai Kohéziós Projektben az autentikus magyar alkorpusz, valamint az angolból, franciából, németből és olaszból fordított magyar alkorpusz tartalmaz. Ezen alkorpuszok esetében mutatja be a lexikai kohéziós eszközhasználatot a 15. és 16. táblázat.

| AUTENTIKUS | | TOKENEK | R | S | A | H | M | ÖSSZES |
|------------|----------------------|---------|--------|-------|-----|-------|-----|--------|
| IRODALOM | DB | 255 699 | 22 094 | 1 427 | 956 | 5 223 | 941 | 30 641 |
| | 1000 tokenre vetítve | | 86 | 6 | 4 | 20 | 4 | 120 |

15. táblázat: Az autentikus magyar szépirodalmi szövegtípusra jellemző kohéziós eszközhasználat

| FORDÍTOTT IRODALOM | | TOKEN-SZÁM | R | S | A | H | M | Összes |
|--------------------|----------------------|------------|--------|-------|-------|-------|-------|--------|
| ANGOL IRODALOM | DB | 278 909 | 23 443 | 1 538 | 575 | 4 169 | 1 077 | 30 802 |
| | 1000 tokenre vetítve | | 84 | 6 | 2 | 15 | 4 | 110 |
| FRANCIA IRODALOM | DB | 257 731 | 12 392 | 897 | 396 | 2 617 | 1 460 | 17 762 |
| | 1000 tokenre vetítve | | 48 | 3 | 2 | 10 | 6 | 69 |
| NÉMET IRODALOM | DB | 260 803 | 14 916 | 909 | 514 | 2 493 | 708 | 19 540 |
| | 1000 tokenre vetítve | | 57 | 3 | 2 | 10 | 3 | 75 |
| OLASZ IRODALOM | DB | 272 683 | 31 918 | 2 172 | 1 586 | 6 010 | 798 | 42 484 |
| | 1000 tokenre vetítve | | 117 | 8 | 6 | 22 | 3 | 156 |

16. táblázat: A fordított magyar szépirodalmi szövegtípusra jellemző kohéziós eszközhasználat

A 15. táblázat az autentikus (AUTENTIKUS), a 16. táblázat pedig az angolból (ANGOL IRODALOM), franciából (FRANCIA IRODALOM), németből (NÉMET IRODALOM) és olaszból (OLASZ IRODALOM) fordított szépirodalmi magyar szövegek lexikai kohéziós elemeinek számát és eloszlásának arányait tartalmazza. A táblázat oszlopai a tokenszámra és a lexikai kohéziós kapcsolatok típusára jellemző adatokat (tokenszám [TOKENEK], ismétlések [R], szinonimák [S], antonimák [A], meronimák [M], hiponimák [H] száma) mutatják. Az utolsó oszlop a lexikai kohéziós kapcsolatok számát tartalmazza összesítve (ÖSSZES). Az alkorpuszra jellemző adatoknál a felső sorban (DB) az adott alkorpuszban talált összes lexikai kohéziós kapcsolat száma jelenik meg a kapcsolat típusa (R, S, A, H, M) szerinti bontásban. Az alkorpuszra jellemző ada-

toknál az alsó sor (1000 tokenre vetítve) az adott alkorpuszra jellemző lexikai kohéziós kapcsolatszám 1000 tokenre vetített értékét mutatja.

A lekérdezések alapján elmondható, hogy a jelen szépirodalmi szövegek hasonló tokenszámú részkorpuszai esetében a célnyelvi magyar szövegek kohéziós elemeinek száma és eloszlásának aránya mindegyik forrásnyelv esetében eltér az autentikus magyar szépirodalmi szövegek lexikai kohéziós eszközhasználatától. Az adatok alátámasztják azt a feltételezést is, hogy az egyes forrásnyelvek különbözőképpen befolyásolják a célnyelvi magyar szépirodalmi szövegekben a lexikai kohéziós elemek disztribúcióját. Az autentikus magyar szépirodalmi szövegekre jellemző lexikai kohéziós kapcsolatok számához viszonyítva az angolból fordított szépirodalmi magyar szövegek esetében a hiponimaviszonyok számának csökkenése tapasztalható, míg a franciából és németből fordított szépirodalmi magyar szövegeknél az ismétlések, szinonima-, antonima-, és hiponimakapcsolatok 1000 tokenre vetített száma csökken. Ezzel szemben az olaszból fordított szépirodalmi magyar szövegekre az autentikus magyar szépirodalmi szövegekre jellemző lexikai kohéziós kapcsolatok számához viszonyítva 1000 tokenre vetítve több ismétlés, szinonima-, antonima- és hiponima- és kevesebb meronimakapcsolat a jellemző. Ezek az eltérések valószínűleg azzal magyarázhatók, hogy az egyes nyelvekre jellemző lexikai kohéziós eszközhasználat eltérő módon befolyásolja az irodalmi szövegek esetén a célnyelvi szövegalkotást a lexikai kohéziós eszközhasználat tekintetében.

6.2.3. Lexikai kohéziós eltolódások a tudományos szövegekben

Tudományos szövegeket a Magyar Lexikai Kohéziós Projektben az autentikus magyar alkorpusz, valamint az angolból, franciából, németből és olaszból fordított magyar alkorpusz tartalmaz. Ezen alkorpuszok esetében mutatja be a lexikai kohéziós eszközhasználatot a 17. és 18. táblázat. Az angolból és olaszból fordított magyar alkorpusz méretét tekintve kisebb, mint az autentikus magyar alkorpusz, ezért itt különösen nagy a jelentősége az 1000 tokenre vetített értékek egybevetésének.

A 17. táblázat az autentikus (AUTENTIKUS), a 18. táblázat pedig az angolból (ANGOL TUDOMÁNYOS), franciából (FRANCIA TUDOMÁNYOS), németből (NÉMET TUDOMÁNYOS) és olaszból (OLASZ TUDOMÁNYOS) fordított tudományos magyar szövegek kohéziós elemeinek számát és eloszlásának arányait tartalmazza. A táblázat oszlopai a tokenszámra és a lexikai kohéziós kapcsolatok típusára jellemző adatokat (tokenszám [TOKENEK], ismétlések [R], szinonimák [S], antonimák [A], meronimák [M], hiponimák [H] száma) mutatják. Az utolsó oszlop a lexikai kohéziós kapcsolatok számát tartalmazza összesítve (ÖSSZES). Az alkorpuszra jellemző adatoknál a felső sorban (DB) az adott alkorpuszban talált összes lexikai kohéziós kapcsolat száma jelenik meg a kapcsolat típusa (R, S, A, H, M) szerinti bontásban. Az alkorpuszra jellemző adatoknál az alsó sor (1000 tokenre vetítve) az adott alkorpuszra jellemző lexikai kohéziós kapcsolatszám 1000 tokenre vetített értékét mutatja.

| AUTENTIKUS | | TO- KENEK | R | S | A | H | M | Összes |
|------------|----------------------|--------------|--------|-------|-----|-------|-------|--------|
| TUDOMÁNYOS | DB | 257 442 | 45 317 | 1 717 | 241 | 8 632 | 1 691 | 57 598 |
| | 1000 tokenre vetítve | | 176 | 7 | 1 | 34 | 7 | 224 |

17. táblázat: Az autentikus magyar tudományos szövegtípusra jellemző kohéziós eszközhasználat

| FORDÍTOTT TUDOMÁNYOS | | TOKENEK | R | S | A | H | M | Összes |
|-------------------------|-------------------------|---------|--------|-------|-----|-------|-------|--------|
| ANGOL TUDOMÁNYOS | DB | 187 334 | 49 107 | 818 | 140 | 3 325 | 213 | 53 603 |
| | 1000 tokenre vetítve | | 262 | 4 | 1 | 18 | 1 | 286 |
| FRANCIA TUDOMÁNYOS | DB | 280 259 | 33 988 | 1 807 | 464 | 6 006 | 768 | 43 033 |
| | 1000 tokenre vetítve | | 121 | 6 | 2 | 21 | 3 | 154 |
| NÉMET TUDOMÁNYOS | DB | 257 079 | 31 373 | 2 002 | 618 | 8 452 | 3 995 | 46 440 |
| | 1000 tokenre vetítve | | 122 | 8 | 2 | 33 | 16 | 181 |
| OLASZ TUDOMÁNYOS | DB | 129 197 | 25 488 | 2 411 | 494 | 6 174 | 1 495 | 36 062 |
| | 1000 tokenre vetítve | | 197 | 19 | 4 | 48 | 12 | 279 |

18. táblázat: A fordított magyar tudományos szövegtípusra jellemző kohéziós eszközhasználat

A lekérdezések 1000 tokenre vetített értékeinek egybevetése szerint a tudományos szövegtípus esetében a célnyelvi magyar szövegek lexikai kohéziós elemeinek száma és eloszlásának aránya mindegyik forrásnyelv esetében eltér az autentikus magyar tudományos szövegek lexikai kohéziós eszközhasználatától. Az adatok alátámasztják azt a feltételezést is, hogy az egyes forrásnyelvek különbözőképpen befolyásolják a célnyelvi magyar tudományos szövegekben a lexikai kohéziós elemek disztribúcióját.

Az angoltól fordított célnyelvi tudományos magyar szövegek esetében nagyobb az ismétlések, viszont kisebb a szinonima-, antonima-, hiponima- és meronimakapcsolatok 1000 tokenre vetített száma az autentikus magyar tudományos szövegekre jellemző adatokhoz viszonyítva. A franciából fordított tudományos szövegek esetében az autentikus magyar tudományos szövegekre jellemző adatokhoz viszonyítva csökken az ismétlések, az antonima-, a hiponima- és a meronimakapcsolatok 1000 tokenre vetített száma. A németből fordított tudományos szövegek esetében az autentikus magyar tudományos szövegekre jellemző adatokhoz viszonyítva csökken az ismétlések és a hiponimakapcsolatok 1000 tokenre vetített száma, viszont növekszik a szinonima-, az antonima- és a meronimakapcsolatok száma. Az olaszról fordított tudományos magyar szövegekre viszont az autentikus magyar tudományos szövegekre jellemző adatokhoz

viszonyítva 1000 tokenre vetítve nagyobb számú ismétlés, szinonima-, antonima-, hiponima- és meronimakapcsolat a jellemző. Ezek az eltérések valószínűleg azzal magyarázhatók, hogy az egyes nyelvekre jellemző lexikai kohéziós eszközhasználat eltérő módon befolyásolja a tudományos szövegtípus esetén a célnyelvi szövegalkotást a lexikai kohéziós eszközhasználat tekintetében.

6.2.4. Lexikai kohéziós eltolódások a vallási szövegekben

Vallási szövegeket a Magyar Lexikai Kohéziós Projektben az autentikus magyar alkorpusz, valamint az angolból, németből, latinból és görögből fordított magyar alkorpusz tartalmaz. Ezen alkorpuszok esetében mutatja be a lexikai kohéziós eszközhasználatot a 19. és 20. táblázat. Az angolból és görögből fordított magyar alkorpusz méretét tekintve kisebb, mint az autentikus magyar alkorpusz, ezért itt különösen nagy a jelentősége az 1000 tokenre vetített értékek egybevetésének.

A 19. táblázat az autentikus (AUTENTIKUS), a 20. táblázat pedig az angolból (ANGOL VALLÁSI), németből (NÉMET VALLÁSI), latinból (LATIN VALLÁSI) és görögből (GÖRÖG VALLÁSI) fordított tudományos magyar szövegek kohéziós elemeinek számát és eloszlásának arányait tartalmazza. A táblázat oszlopai a tokenszámra és a lexikai kohéziós kapcsolatok típusára jellemző adatokat (tokenszám [TOKENEK], ismétlések [R], szinonimák [S], antonimák [A], meronimák [M], hiponimák [H] száma) mutatják. Az utolsó oszlop a lexikai kohéziós kapcsolatok számát tartalmazza összesítve (ÖSSZES). Az alkorpuszra jellemző adatoknál a felső sorban (DB) az adott alkorpuszban talált összes lexikai kohéziós kapcsolat száma jelenik meg a kapcsolat típusa (R, S, A, H, M) szerinti bontásban. Az alkorpuszra jellemző adatoknál az alsó sor (1000 tokenre vetítve) az adott alkorpuszra jellemző lexikai kohéziós kapcsolatszám 1000 tokenre vetített értékét mutatja.

| AUTENTIKUS | | TOKENEK | R | S | A | H | M | Összes |
|------------|----------------------|---------|--------|-------|-----|-------|-------|--------|
| VALLÁSI | DB | 245 009 | 41 262 | 2 886 | 789 | 6 760 | 1 269 | 52 966 |
| | 1000 tokenre vetítve | | 168 | 12 | 3 | 28 | 5 | 216 |

19. táblázat: Az autentikus magyar vallásos szövegtípusra jellemző kohéziós eszközhasználat

| FORDÍTOTT VALLÁSI | | TO-KENEK | R | S | A | H | M | Összes |
|-------------------|----------------------|----------|--------|-------|-------|--------|-------|--------|
| ANGOL VALLÁSI | DB | 169 363 | 13 705 | 1 226 | 250 | 2 555 | 282 | 18 018 |
| | 1000 tokenre vetítve | | 81 | 7 | 1 | 15 | 2 | 106 |
| NÉMET VALLÁSI | DB | 259 514 | 52 878 | 2 318 | 1 147 | 6 781 | 642 | 63 766 |
| | 1000 tokenre vetítve | | 204 | 9 | 4 | 26 | 2 | 246 |
| LATIN VALLÁSI | DB | 269 713 | 37 236 | 4 488 | 530 | 11 471 | 1 062 | 54 787 |
| | 1000 tokenre vetítve | | 138 | 17 | 2 | 43 | 4 | 203 |
| GÖRÖG VALLÁSI | DB | 175 007 | 28 532 | 1 812 | 702 | 2 517 | 547 | 34 110 |
| | 1000 tokenre vetítve | | 163 | 10 | 4 | 14 | 3 | 195 |

20. táblázat: A fordított magyar vallásos szövegtípusra jellemző kohéziós eszközhasználat

A lekérdezések 1000 tokenre vetített értékeinek egybevetése alapján a vallási szövegek esetében a célnyelvi magyar szövegek lexikai kohéziós elemeinek száma és eloszlásának aránya mindegyik forrásnyelv esetében eltér az autentikus magyar vallási szövegek lexikai kohéziós eszközhasználatától. Az adatok alátámasztják azt a feltételezést is, hogy az egyes forrásnyelvek különbözőképpen befolyásolják a célnyelvi magyar vallási szövegekben a lexikai kohéziós elemek disztribúcióját.

Az angolból fordított magyar vallási szövegek esetében csökken az 1000 tokenre vetített ismétlések, a szinonima-, antonima-, hiponima- és meronimakapcsolatok száma az autentikus magyar vallási szövegekre jellemző adatokhoz viszonyítva. A németből fordított vallási szövegek esetében az autentikus magyar vallási szövegekre jellemző adatokhoz viszonyítva növekszik az 1000 tokenre vetített ismétlések és antonimakapcsolatok száma, viszont csökken a szinonima-, a hiponima- és a meronimakapcsolatok száma. A latinból fordított vallási szövegek esetében az autentikus magyar vallási szövegekre jellemző adatokhoz viszonyítva csökken az 1000 tokenre vetített ismétlések és a meronimakapcsolatok száma, viszont növekszik az 1000 tokenre vetített szinonima-, az antonima- és a hiponimakapcsolatok száma. A görögből fordított vallási magyar szövegekre az autentikus magyar vallási szövegekre jellemző adatokhoz viszonyítva 1000 tokenre vetítve nagyobb számú antonimakapcsolat, viszont kevesebb ismétlés, szinonima-, hiponima- és meronimakapcsolat a jellemző. Ezek az eltérések valószínűleg azal magyarázhatók, hogy az egyes nyelvekre jellemző lexikai kohéziós eszközhasználat eltérő módon befolyásolja a vallási szövegek esetén a célnyelvi szövegalkotást a lexikai kohéziós eszközhasználat tekintetében.

VIII. Összegzés

A jelen mű alkalmazott nyelvészeti megközelítésből foglalkozik a fordításokkal, a magyar nyelvű célnyelvi szövegek szövegépítő sajátosságainak feltárásával. A szövegek komplexebb vizsgálatát tűzi célul maga elé, amit a bemutatott kutatási projekt során a szövegnyelvészet és korpusznyelvészet eszközeivel, kvantitatív módszerekkel kutattam. A korpuszt jellemző, számszerűsített lexikai kohéziós adatok igyekeztek a fordított szövegekre jellemző lexikai kohéziós eltolódásjelenséget megragadni és a forrásnyelvi szöveg hatására bekövetkező célnyelvi lexikai kohéziós mintázatot feltárni.

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 magyar szövegek lexikai kohéziós eszközhasználatának összevetése, a lexikai kohéziós eszközhasználat szempontjából a nyelvpárspecifikus célnyelvi szövegjellemzők feltárása, valamint a célnyelvi szövegproduktiót jellemző lexikai kohéziós eltolódások mentén a lexikai kohéziós fordítási univerzálé hipotézisének megfogalmazása. A bemutatott kutatás három tudományterület – a fordításkutatás, a szövegnyelvészet és a korpusznyelvészet – kutatási szférájának metszéspontjában helyezkedik el, eredményeivel ezek elméleti és gyakorlati tudásbázisát gazdagítja.

Az összehasonlító elemzés nagyobb mennyiségű nyelvi adat, mintegy négymillió szövegszót tartalmazó, többféle műfaji és szövegtipológiai csoportba sorolható magyar autentikus és magyarra fordított korpusz mondathatárokon túlmutató lexikai kohéziós szövegépítő sajátosságai alapján készült. Az elemzett elbeszélő, leíró és érvelő szövegek a közéleti (hivatalos), irodalmi, tudományos és vallási szövegtípusokba tartoznak. A Magyar Lexikai Kohéziós Projektben a lexikai kohéziós elemek identifikációját a projekt céljainak megfelelően és érdekében összekapcsolt nyelvtechnológiai eszközök (Laki 2011, Novák, Orosz és Indig 2011) automatizálták, melyek közül a magyar WordNet Program (Prószéky és Mihaltz 2008) segítette a szemantikai kapcsolatok feltárását. Az eredményeket az egyes szövegtípusok és nyelvpárok vonatkozásában külön-külön táblázat összesíti (lásd VII./6. fejezet).

A projekt tapasztalatainak összegzése előtt fontos kiemelni, hogy a magyar autentikus és célnyelvi szövegek lexikai kohéziós kapcsolatainak feltárására bemutatott módszertan sohasem kacérkodott azzal, hogy egy nyelv szövegeire jellemző kohéziós mintázat pusztán kvantitatív adatokkal pontosan megragadható. Egy adott nyelv lexikai kohéziós eszközhasználatát nem hasonlítható azon mérnöki létesítményekhez, amelyeknél egy elemet csak egy vele teljesen kompatibilis elemmel lehet kiváltani. A szövegalkotó az adott nyelvű, szövegtípusú szövegben az adott gondolatmenet kohezivitását több lehetséges elemmel is biztosíthatja nyelvi és kognitív kompetenciáknak függvényében.

A szövegalkotó választását befolyásolja többek között maga a szövegalkotó személyisége, a kommunikációs helyzet és cél, a szövegtípus, a szöveg keletkezésének körülménye, a szöveg keletkezésének nyelve stb., ahogy ezt a II./3.1. fejezet bővebben tárgyalja. A jelen kutatás a lexikai kohéziót befolyásoló összes tényező közül elsősorban a forrásnyelvi szöveg kohéziós mintázatának hatását törekedett megragadni.

A bemutatott eljárás nem törekedhetett a lexikai kohéziós elemek előfordulásának pontos adatolására. A jelen kutatás a vizsgált szöveganyag nagyságának köszönhetően arra bizonyosan alkalmas, hogy segítsen feltárni az adott szövegtípuson belüli, ám a keletkezésük körülményét tekintve különböző forrásnyelvek alapján készített fordítások lexikai kohéziós mintázatának eltéréseit. A jelen vizsgálat elsődleges célkitűzése tehát az, hogy megfelelő szöveg minta alapján jelezze azokat a forrásnyelvi hatásoknak tulajdonítható eltolódásjelenségeket, 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 több szövegtípusú és eltérő forrásnyelvű alkorpuszok vizsgálatát. A lexikai kohéziós kapcsolatok automatikus azonosítása és a négy szövegtípusra és a különböző nyelvpárokra jellemző disztribúciójának leírása tette lehetővé az autentikus és a célnyelvi magyar szövegekre jellemző lexikai kohéziós mintázatok egybevetését, valamint a szövegtípusok és a forrásnyelvek függvényében jellemző lexikai kohéziós mintázatok leírását.

A kutatás eredményeit összegezve lehetséges a vizsgált korpusz vonatkozásában néhány megállapítás, mely jó kiinduló pontnak bizonyul a célnyelvi szövegek lexikai kohéziós mintázatával kapcsolatos hipotézisek megfogalmazásához és a vizsgált kérdéskörben kvalitatív kutatások lefolytatásához.

1. A kutatási eredmények alapján megállapítható, hogy a vizsgált korpusz esetében a fordított szövegekre jellemző lexikai kohéziós eszközhasználat nem egyezik az autentikus szövegprodukcióna jellemző lexikai kohéziós eszközhasználattal. A VII./6.1. fejezetben bemutatott eredmények alátámasztják, hogy a vizsgált négy szövegtípus esetében az autentikus magyar szövegprodukcióna jellemző lexikai kohéziós eszközhasználattól eltérő lexikai kohéziós mintázat jellemző a célnyelvi magyar szövegekre minden forrásnyelv – az angol, francia, német, olasz, latin és görög – viszonylatában. A négy szövegtípus és a hat forrásnyelv viszonylatában tett megállapítás összesen tizennégyféle relációban bizonyult igaznak.
2. A kutatási eredmények alapján megállapítható, hogy a vizsgált korpusz esetében a célnyelvi szövegek lexikai kohéziós eszközhasználatát a forrásnyelvek lexikai kohéziós eszközhasználatára befolyásolja. A VII./6.2. fejezetben bemutatott eredmények alátámasztják, hogy a forrásnyelvek hatására történnek változások a lexikai kohéziós eszközök használatának gyakoriságában és eloszlásában. Az egyes forrásnyelvek által előnyben részesített lexikai kohéziós eszközhasználat

befolyásolja a célnyelvi szövegek lexikai kohéziós mintázatát. Az autentikus szövegprodukcióna jellemző lexikai kohéziós mintázattól való eltérés hat forrásnyelv viszonylatában bizonyult igaznak.

3. A kutatási eredmények alapján megállapítható, hogy a vizsgált korpusz esetében a célnyelvi szövegek lexikai kohéziós eszközhasználatának eltolódása azonos szövegtípus esetén nyelvpárfüggő. A VII./6.2. fejezetben bemutatott szövegtípusonkénti összevetések rámutattak arra, hogy a célnyelvi szövegek esetében a lexikai kohéziós eszközök használatának gyakoriságában és eloszlásában bekövetkező változások adott szövegtípuson belül a forrásnyelvek függvényében különböznek egymástól. Az egyik szövegtípus esetében sem született olyan eredmény, ahol a különböző forrásnyelvekről fordított szövegek esetében a lexikai kohéziós eszközök disztribúciójában bekövetkezett változások egymással megegyező tendenciákat mutattak volna. A három szövegtípus és a hat forrásnyelv viszonylatában tett megállapítás összesen kilencféle relációban bizonyult igaznak.
4. A kutatási eredmények alapján megállapítható, hogy a vizsgált korpusz esetében a célnyelvi szövegek lexikai kohéziós eszközhasználatának eltolódása fordítási tendencia. A kutatásban szereplő négy szövegtípus és hat forrásnyelv relációjában, tehát összesen tizennégy vizsgált kategóriában bizonyult igaznak, hogy a forrásnyelv hatására eltolódás következik be a célnyelvi szövegek lexikai kohéziós eszközhasználatának gyakoriságában és disztribúciójában. A kutatás ezzel szemben a négy szövegtípus és a hat forrásnyelv relációjában vizsgált tizennégy kategória egyikében sem talált a forrásnyelv lexikai kohéziós eszközhasználati preferenciájának hatásától mentes célnyelvi alkorpuszt, amely esetében a lexikai kohéziós eszközhasználat gyakorisága és disztribúciója megegyezne az autentikus szövegalkotásra jellemző lexikai kohéziós eszközhasználattal.
5. A kutatási eredmények alapján megállapítható, hogy a szövegszerveződés lexikai-szemantika szerkezetére jellemző lexikai kohéziós mintázat feltárása automatizálható a WordNet-adatbázis synset relációi alapján történő lexikai kohéziós kapcsolatok bejelölésével. A közel négymillió szövegszavas korpusz elemzése alapján igaznak bizonyult, hogy a WordNet adatbázis relációi alapján a szövegekben feltárt lexikai kohéziós kapcsolatok hozzájárulhatnak az adott szövegtípusra jellemző lexikai kohéziós mintázat leírásához.

Az összegzésben újra fel kell hívni a figyelmet arra, hogy a korábbi kézi annotálásra alkalmas elméleti alapú modellek (Hoey 1991, Károly 2000b) taxonómiájával ellentétben a jelen kutatás esetében az automatizálás révén a lexikai kohéziós elméletek néhány aspektusa hiányzik a vizsgálat tárgyköréből. A kutatás eredményei alapján levont következtetések ennek következtében kizárólag a WordNet-formalizmus segítségével feltárt főnévi csoportok esetében igazak. A kutatás során kapott eredmények összevetése az eddigi hazai és nemzetközi lexikai kohézió kutatási eredményekkel mindezek tükrében korai lenne, hiszen a korábbi kutatások fókusza kiterjed a lexikaikohézió-kutatás más aspektusaira is. Az összevethetőség hiánya mindenképpen a jelen kutatás esetleges korlátai közé tartozik.

A most bemutatott kutatás viszont jó kiindulási alapot szolgáltat további kutatásokhoz. A lehetséges további kutatási irányok között mindenképpen említésre érdemes a lexikai kohéziós kapcsolatok automatikus feltárásába bevont további szófajok (pl. melléknevek, igék) lexikai kohéziós kapcsolatainak feltárása. A jövőben érdemes lenne továbbá a mostani kvantitatív vizsgálatokat kvalitatív vizsgálatokkal kiegészíteni, amelyek magyarázatot adhatnak a nyelvpárfüggő célnyelvi kohéziós mintázatok kialakulására.

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. Helsinki. Doctoral Thesis.
- Aligned Hansards of the 36th Parliament of Canada*. Online elérés: <https://www.isi.edu/division3/natural-language/download/hansard>
- Baker, M. 1992. *In Other Words. A coursebook on translation*. London: Routledge. <https://doi.org/10.4324/9780203133590>
- 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. <https://doi.org/10.1075/z.64.15bak>
- Baker, M. 1995. Corpora in Translation Studies. An Overview and Some Suggestions for Future Research. *Target*. 7/2. 223–243. <https://doi.org/10.1075/target.7.2.03bak>
- 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. <https://doi.org/10.1075/bl.18.17bak>
- Baker, M. 2000. Towards a methodology for investigating the style of a literary translator. *Target*. 12/2. 241–266. <https://doi.org/10.1075/target.12.2.04bak>
- 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 szöveg 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.
- Balaskó M. 2020. Korpuszalapú fordítástudomány, és ami mögötte van. In: Robin E., Seidl-Pécs O. *Fókuszban a fordított és a tolmácsolt szöveg: Korpuszalapú fordításkutatás Magyarországon*. 55–77. <https://doi.org/10.36252/Nyelvikozvszegedkonyv1.3>
- BalkaNet*. Online elérhető: <http://www.dblab.upatras.gr/balkanet/index.htm>
- 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. <https://doi.org/10.1080/02690050308589869>
- Békési I. 1979. A kapcsolás problematikája. In: Kornya L., Czellerné Farkas M. (szerk.) 1996. *Magyar nyelvészeti cikkek- 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., House, J. (eds.) 1986. *Interlingual and Intercultural Communication*. Tübingen: Narr. 17–35.
- Bozsik Gy. 2015. *A lexikai kohézió és az énekelhetőség vizsgálata operaszövegek fordításában*. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2015.074>
- British National Corpus*. Online elérés: www.natcorp.ox.ac.uk
- Brown Corpus*. Online elérés: <http://icame.uib.no/brown/bcm.html>
- 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 Tankönyvkiadó.
- 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. (eds.) 2009. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics*. Vol. 17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 65–83. <https://doi.org/10.1075/bct.17.05che>
- Chesterman, A. 1993. From ‘Is’ to ‘Ought’. Laws, Norms and Strategies in Translation Studies. *Target. International Journal of Translation Studies*. Vol. 5, Issue 1. 1–20. <https://doi.org/10.1075/target.5.1.02che>
- Chou, I., Li, W., Liu, K. 2023. Representation of interactional metadiscourse in translated and native English: A corpus-assisted study. *PLoS ONE 18(7): e0284849*. <https://doi.org/10.1371/journal.pone.0284849>
- Cicero, M. T. 1903. *De optimo genere oratorum*. Oxford: Oxford University Press. (ed. Wilkins, A. S.)
- Corpus of Contemporary American English*. Online elérés: <https://www.english-corpora.org/coca>
- 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: Språkdata.
- Dankó Sz. 2022. *Alul- és felülreprezentált elemek a fordított szövegekben. Korpuszalapú vizsgálat*. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.156476/ELTE.2021.030>

- de Beaugrand, R., Dressler, W. 2000. *Bevezetés a szövegnyelvészetbe*. Budapest: Corvina.
Digitális Irodalmi Akadémia. Online elérés <https://dia.hu>
- Doherty, M. 1999. Clefts in Translations between English and German. *Target* 11/2. Amsterdam & Philadelphia: John Benjamins Publishing Company. 289–315. <https://doi.org/10.1075/target.11.2.06doh>
- Dressler, W. U. 1973. *Einführung in die Textlinguistik*. Tübingen: Max Niemeyer. <https://doi.org/10.1515/9783111358932>
- Dróth J. 2002. Az értékelés mint rendszeres visszajelzés a fordítás oktatásában. In: Feketé né S. M. (szerk.) *Porta Lingua. 2002. Szaknyelvoktatásunk az EU kapujában*. Debrecen: Debreceni Egyetem. 94–100.
- English-Norwegian Parallel Corpus*. Online elérés: <https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc>
- EUR-ACCOR. Online elérés: <http://www.cstr.ed.ac.uk/research/projects/artic/accor.html>
- EuroWordNet*. Online elérés: <https://archive.illc.uva.nl/EuroWordNet/database.html>
- EUR-Lex. Online elérés: <https://eur-lex.europa.eu/homepage.html>
- European Parliament Proceedings 1996-2001*. Online elérés: <http://www.statmt.org/europarl>
- 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. literaturi na inosztrannih jazikah.
- Flowerdew, J. 2006. Use of signalling nouns in a learner corpus. In: Flowerdew, J., Mahlberg, M. (eds.) 2009. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 85–102. <https://doi.org/10.1075/bct.17.06flo>
- Flowerdew, J., Mahlberg, M. (eds.) 2009. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/bct.17>
- Földes Cs. 2007. „Interkulturális kommunikáció”: koncepciók, módszerek, kérdőjelek. *Fordítástudomány*. IX. évf. 1. szám. 14–39.
- Francis, W. N., Kucera, H. 1964. *Manual of Information to accompany a Standard Corpus of Present-Day. Edited American English, for use with Digital Computers*. Brown University.
- 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.
- Götz A. 2018. Az explicitáció és az -e partikula fordított magyar szövegekben. *Fordítás-tudomány* 20. évf. 1. szám. 5–17.
- Götz A. 2019. Linguistic contrast and translation universals: A corpus-based investigation of the translational use of the Hungarian pragmatic marker vajon ('I wonder'). Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.156476/ELTE.2018.048>
- Gutwinski, W. 1976. *Cohesion in Literary Texts : A Study of Some Grammatical and Lexical Features of English Discourse*. The Hague: Mouton. <https://doi.org/10.1515/9783111352176>
- 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. <https://doi.org/10.3115/1621034.1621040>
- 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.
- Hasselgård, H., Oksefjell, S. (eds.) 1999. *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi. <https://doi.org/10.1163/9789004653689>
- Hatim, B., Mason, I. 1990. *Discourse and the Translator*. London: Longman.
- Háhn J. 2011. *Promóciós műfajok a virtuális térben*. Pécs: Pécsi Tudományegyetem Nyelvtudományi Doktori Iskola.
- 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. https://doi.org/10.1163/9789004486669_008
- Horváth Á. 2023. *Betegbiztonság és kommunikáció a fordításban: A magyarról angolra fordított belegegyező nyilatkozat fordításának korpuszalapú műfaji elemzése*. Doktori értekezés. Kéz-

- irat. Budapest: ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2022.208>
- House, J. 1981. *A Model for Translation Quality Assessment*. Tübingen: Narr.
- House, J. 1997. *Translation Quality Assessment. A Model Revisited*. Tübingen: Narr.
- Hönig, H. 1997. *Konstruktives Übersetzen*. Tübingen: Stauffenburg.
- HTML (*Hyper Text Markup Language*). Online elérés: <https://www.w3schools.com/html>
- Hungarian WordNet (*HuWTN*). Online elérés: <https://rgai.inf.u-szeged.hu/node/159>
- H. Varga Gy. (szerk.) 2007. *A verbális üzenet. Segédanyag a Nyelvi kommunikáció c. tárgyhoz*. Eger: Eszterházy Károly Főiskola.
- International Corpus of English*. Online elérés: <https://www.ucl.ac.uk/english-usage/projects/ice.htm>
- Isenberg, H. 1976. Einige Grundbegriffe für eine linguistische Texttheorie. In Danes, F., Viehweger, D. (eds.) 1976. *Probleme der Textgrammatik*. Berlin: Akademie-Verlag. 47–145. <https://doi.org/10.1515/9783112709573-006>
- 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.
- Juhász-Koch M. 2023. *A bangalámondás fordításának műfaji és nyelvi jellegzetességei angol-magyar nyelvi irányban*. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2022.211>
- Károly K. 2002a. Az alkalmazott nyelvészeti kutatások néhány alapvető módszertani kérdéséről. *Alkalmazott Nyelvtudomány* 2. évf. 1. szám. 77–87.
- Károly K. 2002b. 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 K. 2021. Láthatatlan fordítás? A tudományos szakfordítás empirikus kutatásának eredményeiről funkcionális megközelítésben. *Fordítástudomány* 23. évf. 1. szám. 21–48. <https://doi.org/10.35924/fordtud.23.1.2>
- Károly K., Csiborné Horváth A., Engel I., Van Waarden F. 2022a. A szövegkohézió újrateemtése műfordításban: esettanulmány Salinger *The Catcher in the Rye* című művének két magyar fordításáról. *Fordítástudomány* 24. évf. 2. szám. 5–40. <https://doi.org/10.35924/fordtud.24.2.1>
- Károly K., Csiborné Horváth A., Engel I., Van Waarden F. 2022b. A makrostruktúra újrateemtése műfordításban: esettanulmány Salinger *The Catcher in the Rye* című

- művének két magyar fordításáról. *Fordítástudomány* 24. évf. 2. szám. 41–81. <https://doi.org/10.35924/fordtud.24.2.2>
- Károly S. 1979. A szöveg és a jelentés szerepe a kommunikációs szemléletű nyelvészeti törekvésekben. In: Kornya L., Czellerné Farkas M. (szerk.) 1996. *Magyar nyelvészeti cikke- 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 cikke- 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: Lansztyák I., Vančóné Kremmer, I. 2005. *Nyelvészetéről - 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.
- Klenk M. 2019. Lexikai kohéziós eszközök változása egy sajtószöveg fordítása és lektorálása során. *Fordítástudomány* 21. évf. 2. szám. 32–45. <https://doi.org/10.35924/fordtud.21.2.3>

- Kocsány P. 2002. Szöveg, szövegtípus, jelentés: a mondás mint szövegtípus. Budapest: Akadémia Kiadó.
- Komisszarov, V. N. 2002. *Sovremennoe perevodovedenie: uchebnoe posobie*. Moszkva: ÉTS.
- Kovács M. 2015. *A frazeológiai univerzálék fordítási aspektusai és üzenetközvetítő szerepe európai uniós kontextusban*. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítás-tudományi Doktori Program. <https://doi.org/10.15476/ELTE.2015.089>
- Kugler N., Tolcsvai N.G. 2000. *Nyelvi fogalmak kiegészítője*. 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 Tan-székcsoport. 12–23.
- Laviosa, S. 1997. How Comparable Can 'Comparable Corpora' Be? *Target*. 9/2. 289–319. <https://doi.org/10.1075/target.9.2.05lav>
- 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. <https://doi.org/10.1163/9789004485907>
- Láncos P. L. 2022. A tudományos kutatás jogi szabályozása. Klaudy K., Robin E., Seidl-Pécs O. (szerk.) 2022. *Bevezetés a fordítás és a tolmácsolás kutatásmódszertanába I. Általános rész*. Budapest: ELTE FTT–MANYE Fordítástudományi Szakosztály. 123–140. <https://doi.org/10.21862/kutmodszertan1/8>
- Lengyel I. 2013. *A fordítási hiba fogalma funkcionális megközelítésben*. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program. Elérhető: <http://doktori.btk.elte.hu/lingv/lengyelivan/diss.pdf>
- Literatur im Netz*. Online elérés: <https://www.dla-marbach.de/bibliothek/literatur-im-netz>
- Littérature de langue française en ligne*. Online elérés: <http://www.lehman.cuny.edu/depts/langlit/french/lit.html>
- 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*. Vol. 1. Berlin-New York: Walter de Gruyter. <https://doi.org/10.1515/booksethsk29>
- 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. (eds.) 2009. *Lexical Cohesion*

- and Corpus Linguistics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 103–122. <https://doi.org/10.1075/bct.17.07mah>
- Makkos A. 2015. *Összehasonlítható kompetenciák anyanyelvi és fordított szövegekben. Az anyanyelvi fogalmazási kompetencia és a fordítási kompetencia összefüggései egyetemi hallgatók magyar nyelvű, összehasonlítható szövegei alapján. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2014.034>*
- 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. <https://doi.org/10.1075/pbns.41.13mau>
- Mány D. 2021. *Eufemizmusok az orvosi szakfordításban. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2020.107>*
- McEnergy, T., Wilson, A. 2001. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnergy, 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. <https://doi.org/10.21832/9781853599873-005>
- 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: <https://catalogue.elra.info/en-us/repository/browse/ELRA-W0023>
- Mohácsi-Gorove A. 2014. *A minőség fogalma a fordítástudományban és a lektorálás mint minőségbiztosítási garancia. Doktori értekezés. Kézirat. ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.156476/ELTE.2014.011>*
- Mounin, G. 1963. *Les problémes théoriques de la traduction*. Paris: Gallimard.
- Morley, J. Lexical cohesion and rhetorical structure. 2009. In: Flowerdew, J., Mahlberg, M. (eds.) 2009. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 23–44. <https://doi.org/10.1075/bct.17.02mor>
- Nagy J. 2015. *A kommunikatív dinamizmus (relatív) egyensúlya a fordításban. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2015.204>*
- Nesi, H., Basturkmen, H. 2006. Lexical bundles and discourse signalling in academic lectures. In: Flowerdew, J., Mahlberg, M. (eds.) 2009. *Lexical Cohesion and Corpus Ling*

- istics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 23–44. <https://doi.org/10.1075/bct.17>
- 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. <https://doi.org/10.1163/9789004495746>
- Nord, Ch. 2005. *Text Analysis in Translation: Theory, Methodology, and Didactic Application of a Model for Translation-oriented Text Analysis*. Amsterdam: Rodopi. <https://doi.org/10.1163/9789004500914>
- 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. <https://doi.org/10.4324/9780203640005>
- 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. <https://doi.org/10.7202/011004ar>
- Paepcke, F. 1994. Textverstehen – Textübersetzen - Übersetzungskritik. Snell-Hornby, M. (Hrsg.) *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 op 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. / Jünger, 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 – Nyelvhasználát*. 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 világyelv. 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 találkozásása a fordításban. Doktori kutatások Klauy 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., Czellerne Farkas M. (szerk.) 1996. *Magyar nyelvészeti cikke- és tanulmánygyűjtemény szakfordítóknak. II. Nyelvművelési és szakmai nyelvhasználát*. 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.
- Polcz K. 2012. *Konvencionálisan indirekt beszédaktusok az angol–magyar filmfordításban*. Doktori értekezés. Kézirat. ELTE BTK Fordítástudományi Doktori Program. Elérhető: <http://doktori.btk.elte.hu/lingy/polczkaroly/diss.pdf>
- Princeton WordNet*. Online elérhető: <http://wordnet.princeton.edu>
- 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. <https://doi.org/10.1556/mterm.1.2008.1.4>
- Radó A. 1909. *A fordítás művészete*. Budapest: Franklin Társulat.

- Räsänen, 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.
- Riley, P., Caswell, I., Freitag, M., Grangier, D. 2020. Translationese as a Language in „Multilingual” NMT. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, July 5 – 10. 2020 Association for Computational Linguistics. 7737–7746. <https://doi.org/10.18653/v1/2020.acl-main.691>
- Robin E., Dankó Sz., Götz A., Nagy A. L., Pataky É., Szegh H., Török G., Zolczer P. 2016. Fordítástudomány és korpuszkutatás: bemutatkozik a Pannónia korpusz. *Fordítástudomány* 18. évf. 2. szám. 5–26.
- Robin E. 2014. *Fordítási univerzálék a lektorált szövegekben*. Doktori értekezés. Kézirat. Budapest. ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2014.031>
- Robin E. 2018. *Fordítási univerzálék és lektorálás*. Budapest: ELTE Eötvös Kiadó, Szent Jeromos Alapítvány.
- Robin E. 2023. Fordítási univerzálék a lektorált szövegekben. In: Robin E., Malaczkov Sz., Kovalik-Deák Sz. (szerk.) *Így született a magyar fordítástudomány: Doktori kutatások 2003–2023*. Budapest: ELTE BTK Fordító- és Tolmácsképző Tanszék. 139–144.
- Sager, J. C. 1994. Die Übersetzung im Kommunikationsprozess: der Übersetzer in der Industrie. In: Snell-Hornby M. (ed.) *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 Translator Students Translating from Russian into Swedish. Stockholm: Almqvist and Wiksell International.
- Seidl-Péché O. 2018. Melyek a (szak)fordító és a fordításkutató munkáját segítő legfontosabb nyelvi korpuszok? In: Robin E., Zachar V.(szerk.) *Fordítástudomány ma és holnap*. Budapest: L'Harmattan Kiadó. 175–191. Online elérés: https://real.mtak.hu/120310/1/FT_maesholnap_2018.pdf
- Seidl-Péché O. 2020. Korpuszalapú fordításkutatás: lehetőségek és nehézségek. Fókuszban a korpuszépítés és a korpuszalapú elemzés. In: Robin E., Seidl-Péché O. (szerk.) *Fókuszban a fordított és a tolmácsolt szöveg: Korpuszalapú fordításkutatás Magyarországon*. Budapest: Eötvös Loránd Tudományegyetem, Bölcsészettudományi Kar, Magyar Alkalmazott Nyelvészek és Nyelvtanárok Egyesülete. 79–94. <https://doi.org/10.36252/Nyelvkozvszegedkonyv1>

- Seidl-Péché O. 2024. Korpusznyelvészeti módszerek a fordított szövegek kutatásában I. Általános rész. In: Klaudy K., Robin E., Seidl-Péché O. (szerk.) *Bevezetés a fordítás és a tolmácsolás kutatómódszertanába*. Budapest: ELTE BTK Fordító- és Tolmácsolás Kutatóközpontja, MANYE Fordítástudományi Szakosztály 235–251. <https://doi.org/10.21862/kutatomodszertan2/13>
- SGML (*Standard Generalized Markup Language*). Online elérés: <https://www.w3.org/TR/WD-html40-970708/intro/sgmltut.html>
- 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. 1996. *EAGLES (Expert Advisory Group on Language Engineering Standards). Preliminary Recommendations on Corpus Typology*. Online elérés: <https://www.ilc.cnr.it/EAGLES96/corpusstyp/corpusstyp.html>
- Sketch Engine*. Online elérés: <https://www.sketchengine.eu>
- Schmitt, P. A. 1998. *Handbuch Translation*. Tübingen: Stauffenburg.
- Somodi J. 2014. *Megszólítások pragmatikája japán–magyar összevetésben. A japán apellatív megszólítások fordításának vizsgálata magyar filmszövegekben. Doktori értekezés*. Kézirat. Budapest. ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2014.035>
- Stern A. L. 2022. *A terminológiai konzisztencia vizsgálata műszaki szövegekben. Doktori értekezés*. Kézirat. ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.156476/ELTE.2021.142>
- 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. <https://doi.org/10.4324/9781315759951-6>
- Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Szabómihály G. 2003. A Szlovákiai magyar szakfordítások minőségének javításáról és az objektív fordításkritika megeremtésének 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 cikkek- és tanulmánygyűjtemény szakfordítóknak. II. Nyelvművelés és szakmai nyelvhasználat*. Debrecen: Kossuth Lajos Tudományegyetem Idegennyelvi Központ. 12–16.

- Szegh H. 2022. *Harmadik kód a tolmácsolásban: Vajon létezik tolmácsolási szöveg?* Doktori értekezés. Kézirat. ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.156476/ELTE.2021.084>
- Szeged Treebank 2.0*. Online elérés: <https://rgai.inf.u-szeged.hu/node/113>
- Szlávik Sz. 2023a. Az utószerkesztés hatása a neurális gépi fordítás lexikai kohéziójára angol–magyar szakfordításokban. *Fordítástudomány* 25. évf. 1. szám. 68–80. <https://doi.org/10.35924/fordtud.25.1.4>
- Szlávik Sz. 2023b. Sketch Engine. *Modern Nyelvoktatás* 29. évf. 3–4. szám. 201–208.
- Szijj M. 2015. *Műfordítás nem anyanyelvre. Magyarok a magyar irodalom spanyol fordításában*. Doktori értekezés. Kézirat. ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2015.141>
- 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.
- TEI. (*The Text Encoding Initiative*). Online elérés: <http://www.tei-c.org/index.xml>
- The Translational English Corpus-t (TEC)*. Online elérés: <https://research.manchester.ac.uk/en/projects/translational-english-corpus-tec>
- Tirkkonen-Condit, S. 2002. Translationese — a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target* 14/2. Amsterdam & Philadelphia: John Benjamins Publishing Company. 207–220. <https://doi.org/10.1075/target.14.2.02tir>
- TIGER Korpus*. Online elérés: <https://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/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. <https://doi.org/10.1075/btl.4>
- 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. <https://doi.org/10.7202/004515ar>
- XML (*Extensible Markup Language*). Online elérés: https://www.w3schools.com/xml/xml_what.asp
- Ugrin Zs. 2023. *A számítógéppel támogatott fordítás szövegszintű jellemzői a fordításpedagógia szémszögéből*. Doktori értekezés. Kézirat. ELTE BTK Fordítástudományi Doktori Program. <https://doi.org/10.15476/ELTE.2022.061>

- Vanderauwera, R. 1985. Dutch Novels Translated into English: The Transformation of a 'Minority' Literature. Amsterdam: Rodolpi. <https://doi.org/10.1163/9789004490284>
- 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.) *Übersetzungswissenschaft. Eine Neuorientierung*. Tübingen, Basel: Francke Verlag. 30–53.
- Warren, M. 2006. Cohesive chains and speakers' choice of prominence. In: Flowerdew, J., Mahlberg, M. (eds.) 2009. *Lexical Cohesion and Corpus Linguistics. Benjamins Current Topics*. Vol.17. Amsterdam & Philadelphia: John Benjamins Publishing Company. 45–63. <https://doi.org/10.1075/bct.17.04war>
- Weinrich, H. 1967. *Syntax als Dialektik (Bochumer Diskussion)*. *Poetica* 1. 109–126. <https://doi.org/10.30965/25890530-00101010>
- Wilss, W. 1996. *Übersetzungsunterricht. Eine Einführung*. Tübingen: Gunter Narr Verlag.
- Wittgeinstein, L. 2004/1963. *Logikai–filozófiai értekezés*. Budapest: Atlantisz Kiadó.
- WordSmith Tools*. Online elérés: <http://www.lexically.net/LexicalAnalysisSoftware/index.html>
- Zhang, M., Toral, A. 2019. *The Effect of Translationese in Machine Translation Test Sets*. Online elérés: <https://arxiv.org/abs/1906.08069>, <https://doi.org/10.48550/arXiv.1906.08069>

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%C3%B6n>

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>

Függelékek

1. Függelék

Példák a Magyar Lexikai Kohéziós Projektben szereplő antonima-, meronima- és hiponimakapcsolatokra

Példák az antonimakapcsolat realizációjára

A következő példák a Magyar Lexikai Kohéziós Projektben realizálódó antonimakapcsolatokat szemléltetik az autentikus és fordítási alkorpuszban szövegtípusok szerint.

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 adattár⁵⁸ című szövegből

⁵⁸ Bíró A.-M. (szerk.) 2002. *Kisebbségi Adattár*. Budapest: Teleki László Alapítvány. Elérhető: mek.oszk.hu/05900/05981/05981.pdf

| | | | |
|------------|-----------------|---|---------------|
| "antonym": | "munkavállalók" | ➤ | "munkáltatók" |
| "antonym": | "alkalmazott" | ➤ | "munkáltató" |
| "antonym": | "jövedelem" | ➤ | "kiadásokat" |

19. ábra: Példák az IP-07-752 azonosítójú EU-s dokumentum⁵⁹ szövegéből

| | | | |
|------------|-----------|---|-----------|
| "antonym": | „férfiak” | ➤ | „nők” |
| "antonym": | „lányok” | ➤ | „fiúknál” |

20. ábra: Példák az IP-07-1115 azonosítójú EU-s dokumentum⁶⁰ 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" | ➤ | "boltakat" |
| "antonym": | "anyjánál" | ➤ | "apja" |
| "antonym": | "férfi" | ➤ | "asszonyok" |
| "antonym": | "szülei" | ➤ | "gyermekekből" |
| "antonym": | "szegényt" | ➤ | "gazdagot" |
| "antonym": | "múltra" | ➤ | "jövőt" |
| "antonym": | "gyűlölettel" | ➤ | "szeretettel" |
| "antonym": | "rendbe" | ➤ | "zűrzavarában" |

⁵⁹ 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

⁶⁰ 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>

| | | | |
|------------|--------------|---|----------------|
| "antonym": | "kérdésre" | ➤ | "válasz" |
| "antonym": | "örömet" | ➤ | "gyötrelmeket" |
| "antonym": | "örömet" | ➤ | "fájdalom" |
| "antonym": | "világosság" | ➤ | "sötétbe" |

22. ábra: Példák a *Tájfun*⁶¹ című antológiából

AUTENTIKUS ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

| | | | |
|------------|-----------------|---|------------------|
| "antonym": | "nő" | ➤ | "ember" |
| "antonym": | "lányaim" | ➤ | "fiú" |
| "antonym": | "melegben" | ➤ | "hidegben" |
| "antonym": | "vége" | ➤ | "elején" |
| "antonym": | "fogyásával" | ➤ | "növekedésével" |
| "antonym": | "kérdésre" | ➤ | "válasz" |
| "antonym": | "tökéletessége" | ➤ | "tökéletlensége" |
| "antonym": | "békét" | ➤ | "háborút" |
| "antonym": | "kapitalizmus" | ➤ | "szocializmus" |
| "antonym": | "ugyanaz" | ➤ | "más" |
| "antonym": | "anód" | ➤ | "katód" |
| "antonym": | "zűrzavarában" | ➤ | "rendet" |

23. ábra: Példák a *Négy könyv*⁶² című gyűjteményből

FORDÍTOTT ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

| | | | |
|------------|---------------|---|----------------------|
| "antonym": | "rendnél" | ➤ | "rendtelenségtől" |
| "antonym": | "háborút" | ➤ | "békével" |
| "antonym": | "erényeid" | ➤ | "tökéletlenségemmel" |
| "antonym": | "szeretetet" | ➤ | "gyűlöletet" |
| "antonym": | "élőkhoz" | ➤ | "halottak" |
| "antonym": | "utódának" | ➤ | "őseikéért" |
| "antonym": | "apját" | ➤ | "anyját" |
| "antonym": | "mások" | ➤ | "ugyanazon" |
| "antonym": | "asszonyokat" | ➤ | "férjükhöz" |
| "antonym": | "gyalázatban" | ➤ | "tisztetélettel" |

⁶¹ Osztoivits L. (szerk.) 1968. *Tájfun. Klasszikus angol kisregények*. Budapest: Európa Kiadó. Elérhető: <http://mek.niif.hu/03600/03689/html/02.htm>

⁶² Németh L. 1988. *Négy könyv*. Budapest: Szépirodalmi Kvk. Elérhető: mek.oszk.hu/01000/01014

| | | | |
|------------|---------------|---|--------------|
| "antonym": | "bátorságai" | ➤ | "félelemtől" |
| "antonym": | "szzenvedése" | ➤ | "örömeiben" |

24. ábra: Példák A magyar történelem tizedei című műből

AUTENTIKUS ALKORPUSZ – VALLÁSOS SZÖVEGTÍPUS

| | | | |
|------------|-------------------|---|------------------|
| "antonym": | "születések" | ➤ | "haláláról" |
| "antonym": | "termékenység" | ➤ | "terméketlenség" |
| "antonym": | "gyermek" | ➤ | "szülőknek" |
| "antonym": | "férfi" | ➤ | "nő" |
| "antonym": | "férj" | ➤ | "feleség" |
| "antonym": | "emberek" | ➤ | "nők" |
| "antonym": | "fiatalok" | ➤ | "felnöit" |
| "antonym": | "fiatalok" | ➤ | "öregék" |
| "antonym": | "anya" | ➤ | "apák" |
| "antonym": | "csökkenését" | ➤ | "növekedését" |
| "antonym": | "mérséklődése" | ➤ | "növekedése" |
| "antonym": | "munkanéklültség" | ➤ | "állásukat" |

25. ábra: Példák A boldogabb családokért⁶³ körlevélből

FORDÍTOTT ALKORPUSZ – VALLÁSOS SZÖVEGTÍPUS

| | | | |
|------------|--------------------|---|------------------|
| "antonym": | "férfiak" | ➤ | "nők" |
| "antonym": | "más" | ➤ | "ugyanaz" |
| "antonym": | "embert" | ➤ | "nőnek" |
| "antonym": | "gyermekek" | ➤ | "szülők" |
| "antonym": | "apa" | ➤ | "anya" |
| "antonym": | "élők" | ➤ | "holtak" |
| "antonym": | "világosság" | ➤ | "sötétség" |
| "antonym": | "vereséget" | ➤ | "győzelmet" |
| "antonym": | "egyenlőtlenségek" | ➤ | "egyenlőségével" |
| "antonym": | "kezdet" | ➤ | "vég" |

26. ábra: Példák a Gaudium et spes⁶⁴ kezdetű lelképásztori konstitúcióból

⁶³ A Magyar Katolikus Püspöki Konferencia körlevele. 1999. *A boldogabb családokért*. Budapest. Elérhető: <http://uj.katolikus.hu/nyomtat-konyvtar.php?h=157>

⁶⁴ II. Vatikáni Zsinat. 1965. *Gaudium et spes* kezdetű lelképásztori 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

Példák a meronimkapcsolat realizációjára

A következő példák a Magyar Lexikai Kohéziós Projektben realizálódó meronimkapcsolatokat szemléltetik az autentikus és fordítási alkorpuszban szövegtípusok szerint.

AUTENTIKUS ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

| | | | |
|------------|--------------|---|--------------------|
| "meronym": | "április" | ➤ | "tavaszt" |
| "meronym": | "éveken" | ➤ | "évtizednek" |
| "meronym": | "hetekben" | ➤ | "hónapon" |
| "meronym": | "bónaptól" | ➤ | "év" |
| "meronym": | "polgárok" | ➤ | "országok" |
| "meronym": | "polgárai" | ➤ | "nemzetekkel" |
| "meronym": | "Budapestre" | ➤ | "Magyarországra" |
| "meronym": | "Erdély" | ➤ | "Románia" |
| "meronym": | "Mallorcán" | ➤ | "Spanyolországban" |
| "meronym": | "színtig" | ➤ | "épületet" |
| "meronym": | "gyerekek" | ➤ | "családjában" |

27. ábra: Példák Az unió kapujában című szövegből

FORDÍTOTT ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

| | | | |
|------------|-----------------|---|-------------------|
| "meronym": | "szöveg" | ➤ | "kiadványainak" |
| "meronym": | "személyiséget" | ➤ | "személyek" |
| "meronym": | "rész" | ➤ | "egészek" |
| "meronym": | "cikkek" | ➤ | "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": | "polgárokka" | ➤ | "országokai" |
| "meronym": | "Coventry" | ➤ | "Angliában" |
| "meronym": | "légekör" | ➤ | "földön" |
| "meronym": | "padlásra" | ➤ | "bázakból" |

| | | | |
|------------|--------------|---|-------------|
| "meronym": | "keze" | ➤ | "ember" |
| "meronym": | "tenyeréről" | ➤ | "kezeiből" |
| "meronym": | "kezeiből" | ➤ | "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": | "bét" | ➤ | "bónap" |
| "meronym": | "órákor" | ➤ | "nappal" |
| "meronym": | "napig" | ➤ | "bét" |
| "meronym": | "birtokról" | ➤ | "föld" |
| "meronym": | "birtokról" | ➤ | "ország" |
| "meronym": | "felbőjét" | ➤ | "ég" |
| "meronym": | "fenekére" | ➤ | "tó" |
| "meronym": | "alakulatáboz" | ➤ | "hadseregébe" |
| "meronym": | "csata" | ➤ | "báborútkban" |
| "meronym": | "gyerekek" | ➤ | "családhoz" |
| "meronym": | "szemem" | ➤ | "arca" |
| "meronym": | "kezéire" | ➤ | "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": | "Nilus" | ➤ | "Egyiptom" |

31. ábra: Példák a Négy könyv című gyűjteményből

FORDÍTOTT ALKORPUSZ – TUDOMÁNYOS SZÖVEGTÍPUS

| | | | |
|------------|--------------|---|---------------|
| "meronym": | "arcol" | ➤ | "emberek" |
| "meronym": | "lábát" | ➤ | "testének" |
| "meronym": | "szeme" | ➤ | "arcol" |
| "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": | "polgárai" | ➤ | "nemzei" |
| "meronym": | "testvéreinket" | ➤ | "család" |
| "meronym": | "gyermekéinek" | ➤ | "család" |
| "meronym": | "kezéd" | ➤ | "ember" |
| "meronym": | "részük" | ➤ | "egészségben" |
| "meronym": | "jellegzetességei" | ➤ | "természetébe" |
| "meronym": | "személyiségének" | ➤ | "Lelke" |
| "meronym": | "keresztények" | ➤ | "Egyház" |

34. ábra: Példák a Gaudium et spes kezdetű lelképásztori konstitúcióból

Példák a hiponimakapcsolat realizációjára

A következő példák a Magyar Lexikai Kohéziós Projektben realizálódó hiponimakapcsolatokat szemléltetik az autentikus és fordítási alkorpuszban szövegtípusok szerint.

AUTENTIKUS ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

| | | | |
|------------|-----------------|---|----------------|
| "hyponym": | "Uniónak" | ➤ | "Amerika" |
| "hyponym": | "Európa" | ➤ | "földrész" |
| "hyponym": | "kontinens" | ➤ | "Európa" |
| "hyponym": | "integráció" | ➤ | "unió" |
| "hyponym": | "bazánk" | ➤ | "országainak" |
| "hyponym": | "tagállamokban" | ➤ | "országok" |
| "hyponym": | "kísérletei" | ➤ | "lehetőségét" |
| "hyponym": | "próbálkozás" | ➤ | "lehetőségét" |
| "hyponym": | "kiadások" | ➤ | "vesztését" |
| "hyponym": | "kísállalat" | ➤ | "cégek" |
| "hyponym": | "kísállalat" | ➤ | "társaságokat" |

35. ábra: Példák Az unió kapujában című szövegből

FORDÍTOTT ALKORPUSZ – KÖZÉLETI SZÖVEGTÍPUS

| | | | |
|---------------------|-------------------|---|-------------------|
| " <i>hyponym</i> ": | "munka" | ➤ | "termék" |
| " <i>hyponym</i> ": | "vállalkozások" | ➤ | "munka" |
| " <i>hyponym</i> ": | "munka" | ➤ | "munkahely" |
| " <i>hyponym</i> ": | "munkavállalónak" | ➤ | "dolgozók" |
| " <i>hyponym</i> ": | "költségek" | ➤ | "kiadásait" |
| " <i>hyponym</i> ": | "egészsége" | ➤ | "jólléte" |
| " <i>hyponym</i> ": | "nyereségességét" | ➤ | "állapot" |
| " <i>hyponym</i> ": | "gyakorlat" | ➤ | "tevékenységeket" |
| " <i>hyponym</i> ": | "tagállamra" | ➤ | "ország" |

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

AUTENTIKUS ALKORPUSZ – SZÉPIRODALMI SZÖVEGTÍPUS

| | | | |
|---------------------|---------------|---|---------------|
| " <i>hyponym</i> ": | "könyvtárban" | ➤ | "helyiségben" |
| " <i>hyponym</i> ": | "árubázásban" | ➤ | "üzletből" |
| " <i>hyponym</i> ": | "világon" | ➤ | "területére" |
| " <i>hyponym</i> ": | "emberek" | ➤ | "valaki" |
| " <i>hyponym</i> ": | "halott" | ➤ | "valaki" |
| " <i>hyponym</i> ": | "könyvei" | ➤ | "szakasza" |
| " <i>hyponym</i> ": | "szakasza" | ➤ | "zene" |
| " <i>hyponym</i> ": | "szag" | ➤ | "érzés" |
| " <i>hyponym</i> ": | "illata" | ➤ | "szag" |
| " <i>hyponym</i> ": | "ingeket" | ➤ | "ruhában" |
| " <i>hyponym</i> ": | "Duna" | ➤ | "folyónál" |
| " <i>hyponym</i> ": | "Volga" | ➤ | "folyónál" |

37. ábra: Példák A kitagadott című regényből

FORDÍTOTT ALKORPUSZ – SZÉPIRODALMI SZÖVEGTÍPUS

| | | | |
|---------------------|-------------------|---|---------------|
| " <i>hyponym</i> ": | "dolgról" | ➤ | "gondra" |
| " <i>hyponym</i> ": | "kötelességemnek" | ➤ | "dolgában" |
| " <i>hyponym</i> ": | "adósság" | ➤ | "tartozásait" |
| " <i>hyponym</i> ": | "arcképén" | ➤ | "képét" |
| " <i>hyponym</i> ": | "összevgy" | ➤ | "asszony" |
| " <i>hyponym</i> ": | "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" | ➤ | "fémekbezz" |
| "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" | ➤ | "lelkész" |
| "hyponym": | "képében" | ➤ | "arcol" |
| "hyponym": | "palástot" | ➤ | "ruházatával" |
| "hyponym": | "köpenye" | ➤ | "ruhája" |
| "hyponym": | "ütközetből" | ➤ | "nézeteltérésének" |
| "hyponym": | "problémáinak" | ➤ | "nehézségével" |
| "hyponym": | "halálból" | ➤ | "vége" |
| "hyponym": | "idő" | ➤ | "pillanatra" |

| | | | |
|------------|--------------|---|---------------|
| "hyponym": | "meghívásai" | ➤ | "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": | "fejlődési" | ➤ | "folyamatát" |
| "hyponym": | "segítség" | ➤ | "tevékenységnek" |
| "hyponym": | "összhangban" | ➤ | "rendben" |

42. ábra: Példák a Gaudium et spes kezdetű lekipásztori konstitúcióból

2. Függelék

Az ábrák jegyzéke

| | |
|--|-----|
| 1. ábra: A mondathatárok jelölése és sorszámozása | 73 |
| 2. ábra: Az első szöveg első mondatának ismétlődő lexikai elemei | 74 |
| 3. ábra: Az első szöveg első mondatának ismétléskapcsolatai a 15. mondatban..... | 74 |
| 4. ábra: Az MLKP korpusz legfelső folderszintje | 81 |
| 5. ábra: Az MLKP korpusz folderstuktúrája | 82 |
| 6. ábra: Az MLKP korpusz annotált szövegrészlete | 86 |
| 7. ábra: Példa az ismétléskapcsolatok összegzésére..... | 89 |
| 8. ábra: Részlet Az unió kapujában című dokumentumból..... | 89 |
| 9. ábra: Példa a 8. ábra második mondatának morfológiai annotálására | 90 |
| 10. ábra: A 8. ábra szövegrészletében található kohéziós kapcsolatok | 90 |
| 11. ábra: Szövegrészlet A kitagadott című regényből | 92 |
| 12. ábra: A 11. ábra szövegrészletében található kohéziós kapcsolatok | 93 |
| 13. ábra: Szövegrészlet A magyar történelem tizedei műből | 93 |
| 14. ábra: A 13. ábra szövegrészletében található kohéziós kapcsolatok | 94 |
| 15. ábra: Szövegrészlete a Caritas in veritate kezdetű enciklikából | 94 |
| 16. ábra: A 15. ábra szövegrészletében található kohéziós kapcsolatok | 95 |
| 17. ábra: Példák Az unió kapujában című szövegből | 129 |
| 18. ábra: Példák a Kisebbségi adattár című szövegből..... | 129 |
| 19. ábra: Példák az IP-07-752 azonosítójú EU-s dokumentum szövegéből | 130 |
| 20. ábra: Példák az IP-07-1115 azonosítójú EU-s dokumentum szövegéből | 130 |
| 21. ábra: Példák A kitagadott című regényből | 130 |
| 22. ábra: Példák a Tájfun című antológiából | 130 |
| 23. ábra: Példák a Négy könyv című gyűjteményből | 131 |
| 24. ábra: Példák A magyar történelem tizedei című műből | 131 |
| 25. ábra: Példák A boldogabb családokért körlevélből | 132 |
| 26. ábra: Példák a Gaudium et spes kezdetű lelkipásztori konstitúcióból | 132 |
| 27. ábra: Példák Az unió kapujában című szövegből | 133 |
| 28. ábra: Példák a Kisebbségi adattár című szövegből..... | 133 |
| 29. ábra: Példák A kitagadott című regényből | 133 |
| 30. ábra: Példák a Tájfun című antológiából | 134 |
| 31. ábra: Példák a Négy könyv című gyűjteményből | 134 |
| 32. ábra: Példák A magyar történelem tizedei című műből | 135 |
| 33. ábra: Példák A boldogabb családokért körlevélből..... | 135 |
| 34. ábra: Példák a Gaudium et spes kezdetű lelkipásztori konstitúcióból | 136 |
| 35. ábra: Példák Az unió kapujában című szövegből | 136 |
| 36. ábra: Példák az IP-07-752 azonosítójú EU-s dokumentum szövegéből | 137 |

| | |
|---|-----|
| 37. ábra: Példák A kitagadott című regényből | 137 |
| 38. ábra: Példák a Tájfun című antológiából | 137 |
| 39. ábra: Példák a Négy könyv című gyűjteményből | 138 |
| 40. ábra: Példák A magyar történelem tizedei című műből | 138 |
| 41. ábra: Példák A boldogabb családokért körlevélből | 139 |
| 42. ábra: Példák a Gaudium et spes kezdetű lelkipásztori konstitúcióból | 139 |

3. Függelék

A táblázatok jegyzéke

| | |
|---|-----|
| 1. táblázat: A pilotprojekt szövegei..... | 72 |
| 2. táblázat: A pilotprojekt kohézióért felelős elemei..... | 75 |
| 3. táblázat: A fordított alkorpusz szövegeinek forrásnyelvei | 82 |
| 4. táblázat: Az autentikus és a célnyelvi magyar alkorpusz | 85 |
| 5. táblázat: Az autentikus magyar szövegek egy részén mért adatok | 96 |
| 6. táblázat: A lexikai kohéziós kapcsolatok száma az MLKP-ben | 97 |
| 7. táblázat: Részlet az eredményeket archiváló táblázatból | 98 |
| 8. táblázat: Az autentikus magyar szövegek kohéziós elemeinek száma | 99 |
| 9. táblázat: Az angolból fordított magyar szövegek kohéziós elemeinek száma | 99 |
| 10. táblázat: A franciából fordított magyar szövegek kohéziós elemeinek száma | 100 |
| 11. táblázat: A németből fordított magyar szövegek kohéziós elemeinek száma | 100 |
| 12. táblázat: Az olaszból fordított magyar szövegek kohéziós elemeinek száma | 101 |
| 13. táblázat: A latinból és görögből fordított magyar szövegek kohéziós elemeinek száma | 101 |
| 14. táblázat: A közéleti szövegtípusra jellemző kohéziós elemek eloszlása | 102 |
| 15. táblázat: Az autentikus magyar szépirodalmi szövegtípusra jellemző kohéziós eszközhasználat | 103 |
| 16. táblázat: A fordított magyar szépirodalmi szövegtípusra jellemző kohéziós eszközhasználat | 103 |
| 17. táblázat: Az autentikus magyar tudományos szövegtípusra jellemző kohéziós eszközhasználat | 105 |
| 18. táblázat: A fordított magyar tudományos szövegtípusra jellemző kohéziós eszközhasználat | 105 |
| 19. táblázat: Az autentikus magyar vallásos szövegtípusra jellemző kohéziós eszközhasználat | 106 |
| 20. táblázat: A fordított magyar vallásos szövegtípusra jellemző kohéziós eszközhasználat | 107 |

4. Függelék

A pilotprojekt szövegei

Művészetek Palotája. Elérhető: http://www.mupa.hu/epulet_ahaz.jsp?language=hu&languageCode=0

Palast der Künste. Elérhető: http://www.mupa.hu/epulet_ahaz.jsp?language=ge&languageCode=2

Magyar Fordítóház. Elérhető: <http://www.c3.hu/~bfordhaz>

Das Ungarische Übersetzerhaus. Elérhető: <http://www.c3.hu/~bfordhaz>

Balatoni hajózás. Elérhető: <http://www.balatonihajozas.hu/hu/rolunk.htm>

Balaton Schifffahrts AG. Elérhető: <http://www.balatonihajozas.hu/de/rolunk.htm>

5. Függelék

A Magyar Lexikai Kohéziós Projektben elemzett szövegek

Autentikus magyar közéleti szövegek

- A. Gergely A. 2004. *Kisebbségek és az integrációs csatlakozás. Tematikus közelműhelyek Magyarországon európai integrációs csatlakozásának háttérreléséhez*. Budapest: MTA Politikatudományi Intézet. <http://mek.oszk.hu/01900/01999>
- Lelkes O. 2005. *Szegénység az Európai Unió országáiban*. Budapest: TÁRKI Társadalomkutatási Intézet Zrt. https://www.tarki.hu/sites/default/files/2020-10/08_Lelkes.pdf
- Magyar EU-elnökség 2011. 2011. *Az alapjogok tiszteletben tartása és előmozdítása*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Általános Ügyek (GAC)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Az elnökség városai*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Ázsia-Európa Találkozó (ASEM)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Bel- és Igazságügy (JHA)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Bővítés*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Az elnökség költségvetése*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Duna Régió Stratégia*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Sikeresen zárult az Európai Unió Energia Tanácsának ülése*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Európa 2020*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Foglalkoztatás, Szociálpolitika, Egészségügy és Fogyasztóvédelem (EPSCO)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Gazdasági kormányzás*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Gazdasági és Pénzügyek (ECOFIN)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *A gyermekszegénység elleni küzdelem*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *A gyógy-, termál- és ásványvizek hazája*. <http://www.eu2011.hu> [2011. augusztus 10.]

- Magyar EU-elnökség 2011. 2011. *Innovatív Unió - az Európa 2020 zászlóshajója*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Keleti Partnerség*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Környezetvédelem (ENVI)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Közlekedés, Távközlés és Energia (TTE)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *A kulturális sokszínűség és az Önkéntesség Európai Éve*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Külföldi (FAC)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Magas szintű uniós légiközlekedési konferencia kezdődött Budapesten*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Líbiai fejlemények: az EU válaszánaak áttekintése*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *A magyar elnökség prioritásai és programja*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Magyar EU elnökség (2011)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Mezőgazdaság és Halászat (AGRI)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *A NATO új stratégiai koncepciója – adalékok a magyar szempontból*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Oktatás, Ifjúság, Kultúra és Sport (EYCS)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Regionális politika*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Az európai romapolitika kialakítása*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *A soros elnökség*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Az elnökség struktúrája*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *A szociális párbeszéd Európában*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Magyar elnökség összefoglaló*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Orbán: Magyarország teljesítette elnökségi céljait*. http://www.grotius.hu/doc/pub/CCXIDT/2011_142_a_magyar_eu-elnökség_tevékenysége_összegzése_az_európai_parlamentben.pdf

- Magyar EU-elnökség 2011. 2011. *Magyarország átadta az elnökséget Lengyelországnak*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Medgyesi M. *A jövedelmek eloszlása az EU országokban*. http://www.tarsadalomkutatas.hu/kkk.php?TPUBL-A-811/publikaciok/tpubl_a_811.pdf
- Magyar EU-elnökség 2011. 2011. *Európa megerősödött – Orbán Viktor miniszterelnök cikke az eu2011.hu-nak*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Változó Európa*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Egészségfejlesztés a színtereken*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Magyar EU-elnökség 2011. 2011. *Versenyképesség (COMP)*. <http://www.eu2011.hu> [2011. augusztus 10.]
- Molnár P. (szerk.) 2000. *Az unió kapujában*. Budapest: Mundus Magyar Egyetemi Kiadó. <http://mek.oszk.hu/01900/01989>
- Scheiring G. (szerk.) 2003. *Beleéldülünk? Zöldek az EU-ról*. Budapest: Védegylet. <http://eu.zpok.hu>
- Schindler A. 2005. *Európa közepén*. Szigetvár: CyberBooks Kiadó. <https://mek.oszk.hu/03100/03198>

Fordított magyar közéleti szövegek

- Bárdi N. (szerk.) 2002. *Kisebbségi adattár. Kisebbségvédelem és a nemzetközi szervezetek*. Budapest: Teleki László Alapítvány. <http://mek.oszk.hu/05900/05981/05981.pdf> [2011. augusztus 10.]
- Európai Bizottság. 2005. *Az EU „Sokszínűségért. A diszkrimináció ellen” első díját francia újságírók nyerték el. IP/05/512*. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2005. 2006. *A dolgozók mobilitásának Európai Éve. IP/05/817*. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2005. *Az Európai Bizottság bemutatja a kohéziót, növekedést és munkahelyteremtést célzó új közösségi stratégiai iránymutatásokat. IP/05/850*. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *A Bizottság jelentése szerint a 2004-es bővítés óta a munkavállalók szabad mozgásának pozitív hatásai voltak. IP/06/130*. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Egymillió állás az interneten: megkezdődött a munkavállalók mobilitásának európai éve. IP/06/186*. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Az Európai Bizottság évi mintegy 500 millió eurós támogatást javasol a munkavállalók számára egy új, a globalizációhoz való alkalmazkodás elősegítését célzó európai alap keretében. IP/06/245*. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]

- Európai Bizottság. 2006. *A vállalkozások társadalmi felelősségével foglalkozó európai szövetség létrehozása.* IP/06/358. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *A Bizottság jobb munkalehetőségeket kínál a szolgáltatói ágazatban: gyorsabb és kevésbé bürokratikus kiküldetések.* IP/06/423. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Holland újságírók nyerték az ezévi „A sokszínűségért. A diszkrimináció ellen” elnevezésű EU- sajtódíjat.* IP/06/615. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Akcióban a munka és a növekedés: szeptemberben lesz az első európai állásbörze.* IP/06/1006. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Európai mobilitás busz turné az Európai Unión belüli földrajzi mobilitás elősegítése érdekében.* IP/06/1247. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *„Növekedés és Foglalkoztatás” a gyakorlatban: az első európai állásbörze segítséget kíván nyújtani az állampolgároknak a külföldön adódó lehetőségek kihasználásában.* IP/06/1275. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *A demográfiai időzített bomba hatástalanításának öt módja.* IP/06/1359. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Miniszterek és szakértők véleménycsereje az európai népesség elöregedéséből fakadó problémák megoldásáról.* IP/06/1490. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Javul az uniós foglalkoztatási helyzet, de a 2010-es célkitűzések eléréséhez a reformoknak új lendületre van szükségük.* IP/06/1506. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Az új uniós felmérés szerint az európaiak bátran néznek szembe a munkaerőpiac kihívásaival.* IP/06/1507. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *Nyilvánosságra hozták a munkavállalói mobilitás európai díjának 2006. évi nyertesét.* IP/06/1722. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2006. *2006: Fellendülőben a munkavállalói mobilitás, de vannak még megoldásra való kihívások.* IP/06/1723. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Az esélyegyenlőség európai éve lendületet adhat a sokféleségről szóló vitának.* IP/07/69. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Az Unió reformstratégiája munkahelyeket teremt – állítja egy új jelentés.* IP/07/202. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]

- Európai Bizottság. 2007. *A közös megközelítés javítja a társadalmi kohéziót Európában.* IP/07/203. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *2012-ig az üzemi balesetek 25 %-os csökkenését kívánja elérni az új uniós stratégia.* IP/07/204. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *A diszkrimináció titkos meséje nyerte el az EU sajtódíját.* IP/07/499. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Az esélyegyenlőség európai éve: a 2007. évi Eurovíziós Dalverseny hivatalos partnere.* IP/07/542. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Kohéziós politika 2007–2013: A Bizottság jóváhagyja Magyarország nemzeti stratégiáját és prioritásait.* IP/07/635. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *A Bizottság értékeli a régiók számára juttatott támogatások hatását, és vitát indít a kohéziós politika következő szakaszáról.* IP/07/721. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Legyen könnyebb a teher – a váz- és izomrendszeri zavarok kezelése.* IP/07/752. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *A Bizottság lépéseket tesz a bürokratikus akadályok csökkentése érdekében a szolgáltatási ágazatban, ugyanakkor biztosítja a kiküldött munkavállalók védelmét.* IP/07/817. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *A rugalmas biztonság több embert segít jó állásokhoz.* IP/07/919. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *A Bizottság fellép a férfiak és nők közötti fizetéskülönbség csökkentése érdekében.* IP/07/1115. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *A Bizottság célja, hogy jobban integrálja a fiatalokat a társadalomba.* IP/07/1281. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Az európai foglalkoztatási napokon több ezer munkáltató és álláskereső találkozik.* IP/07/1383. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Ünnepeljük meg az egyenlő esélyek érvényre juttatásának 1000 módját.* IP/07/1717. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Javult a foglalkoztatási teljesítmény az integrált munkaerő-piaci politikának köszönhetően.* IP/07/1761. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *A Bizottság jelentése szerint sikeres a növekedést és foglalkoztatást célzó megújult lisszaboni stratégia, de az Európai Uniónak további reformokat kell végrehajtania annak érdekében, hogy megfeleljen a globalizáció kihívásainak.* IP/07/1892. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]

- Európai Bizottság. 2007. *2010. a szegénység és a társadalmi kirekesztés elleni küzdelem európai éve* lezár. IP/07/1956. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2007. *Portugál újságíró nő nyerte az EU idei „A sokszínűségért – a diszkrimináció ellen” sajtódíját*. IP/07/1906. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *A Bizottság megszünteti a foglalkoztatási egyenlőségi szabályok hiányosságait*. IP/08/155. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *Foglalkoztatás az Unióban: a kilátások az egyenlőtlén fejlődés ellenére is reménykeltők*. IP/08/287. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *A reformok megtérülnek, de nem szabad megfeledkezni a leginkább kiszolgáltatottakról – uniós jelentés*. IP/08/296. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *Uniós jelentés: a nőket még mindig üvegplafon választja el a vezető pozícióktól*. IP/08/390. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *Az EU egymillió kiküldött munkavállaló foglalkoztatási körülményeinek javítását sürgeti*. IP/08/514. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *Az új uniós tanulmány szerint a szociális Európa sikeresen állhatja a globalizáció kibívásait*. IP/08/563. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *Az EU szerepe a szociális siker szempontjából kulcsfontosságú*. IP/08/678. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *A 2007–2013-as kohéziós politika: Hübnér és Špidla biztos üdvözlő a prioritásoknak a növekedés és foglalkoztatás irányába történő elmozdulását*. IP/08/744. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *A munkabehelyi balesetek és megbetegedések számának csökkentése jobb kockázattértékeléssel*. IP/08/931. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *A Bizottság egy megújított szociális menetrendet javasol a XXI. század Európájában élők támogatása és életkörülményeinek javítása érdekében*. IP/08/1070. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság. 2008. *Az Európai Unió a munkabehelyen kívüli diszkrimináció elleni védelemre tesz javaslatot*. IP/08/1071. Európai Bizottság. <https://commission.europa.eu> [2011. augusztus 10.]
- Európai Bizottság Foglalkoztatási és Szociális Ügyek Főigazgatósága. 2009. *Fogyatékos-ságtudományi tanulmányok III*. Budapest: Eötvös Loránd Tudományegyetem Bárczi Gusztáv Gyógypedagógiai Kar. <http://mek.oszk.hu/09400/09455/09455.pdf> [2011. augusztus 10.]

Autentikus magyar irodalmi szövegek

- Ballai L. 2000. *A kitagadott*. Kézirat. <http://www.vasaros.com/art/ballai> [2011. augusztus 10.]
- Fehér K. 1968. *A Földrengések Szigete*. Budapest: Móra kiadó. <https://mek.oszk.hu> [2011. augusztus 10.]
- Fekete J. J. 2002. *Imádságos kolostor*. Újvidék: Jugoszláviai Magyar Művelődési Társaság. <https://mek.oszk.hu> [2011. augusztus 10.]
- Sánta Cs. 2004. *Narkóma*. Miskolc: kézirat. <http://narkoma.uw.hu> [2011. augusztus 10.]

Fordított magyar irodalmi szövegek

- Apollinaire, G. 2002. *A meggyilkolt költő*. Magyar Elektronikus Könyvtár. <http://mek.niif.hu/00300/00312/html/index.htm> [2011. augusztus 10.]
- Boccaccio. 2010. *Dekameron*. Budapest: Kossuth kiadó. <https://mek.oszk.hu/00300/00334/00334.pdf> [2011. augusztus 10.]
- Fontane, Th. 1977. *Jenny Treibel Síne. Két kisregény*. Budapest: Európa Könyvkiadó. <https://mek.oszk.hu/03000/03091/03091.htm> [2011. augusztus 10.]
- Kästner, E. 2008. *A két Lotti*. Budapest: Móra kiadó. <https://www.literart.org.ro/files/A-ket-Lotti.pdf> [2011. augusztus 10.]
- Mann, Th. 1999. *Mario és a varázsló*. Budapest: Interpopulart kiadó. <https://mek.oszk.hu/00400/00438/00438.htm> [2011. augusztus 10.]
- Osztoivits L. (szerk.) 1968. *Tájjfun. Klasszikus angol kisregények*. Budapest: Európa Könyvkiadó.
- Raucat, Th. 1998. *Tiszteletreméltó kirándulás*. Budapest: Terebess Kiadó. <https://mek.oszk.hu/02900/02962/02962.htm> [2011. augusztus 10.]
- Rolland, R. 1964. *Pierre és Luce*. Budapest: Magyar Helikon. <https://mek.oszk.hu/00400/00476/00476.htm> [2011. augusztus 10.]
- Shaw, G. B. 1957. *Cashel Byron mestersége*. Budapest: Sport Lap- és Könyvkiadó.
- Schnabel, E., Frank, A. 1958. *Anne Frank naplója. (A hátsó traktus)*. Budapest: Európa Könyvkiadó. <https://mek.oszk.hu/00200/00261/00261.pdf> [2011. augusztus 10.]
- Schwartz-Barth, A. 1962. *Igazak ivadéka*. Budapest: Európa Könyvkiadó. <https://mek.oszk.hu/00500/00555/html/01.htm> [2011. augusztus 10.]
- Zweig, S. 1959. *Sakknovella*. Budapest: Európa Könyvkiadó. <https://mek.oszk.hu/00500/00538/00538.htm> [2011. augusztus 10.]

Autentikus magyar tudományos szövegek

- Karikó S. (szerk.) 1997. *Derrida Marx - szelleme*. Budapest – Szeged: Gondolat – Szegedi Lukács Kör.
- Lahdelma, T., Jankovics J., Nyerges J., Laihonon, P. (szerk.) 2002. *Hatalom és kultúra. Plenáris előadások & Kerekasztal vitaindítók. V. Nemzetközi Hungarológiai Kongresszus*

- (Jyvaskylä, 2001. augusztus 6-10.). Jyvaskylä. <http://real-eod.mtak.hu/id/eprint/1835> [2011. augusztus 10.]
- Németh L. 2007. *Négy könyv. Tudománytörténeti munkák.* Magyar Elektronikus Könyvtár. <http://mek.oszk.hu/01000/01014> [2011. augusztus 10.]
- Németh P. (szerk.) 2003. *A 130 éves Népszava. Töredékek a 130 éves Népszava történetéből.* <https://mek.oszk.hu/04000/04015/04015.pdf> [2011. augusztus 10.]

Fordított magyar tudományos szövegek

- Advocacy Institute. 1988. *A dohányipar praktikái. Hogyan szerezz a dohányipar batalmat és nyer profitot pénzzel és hazugságokkal.* <http://mek.oszk.hu/08000/08081> [2011. augusztus 10.]
- Bánfalvi A. (szerk.) 1966. *Pszichoanalízis, filozófia, metapszichológia (szöveggyűjtemény).* Debrecen: Kézirat. <http://mek.oszk.hu/02000/02012> [2011. augusztus 10.]
- Bonfini, A. 1995. *A magyar történelem tízedei.* Budapest: Balassi Kiadó. <http://www.tankonyvtar.hu/tortenelem/magyar-tortenelem-080903-50> [2011. augusztus 10.]
- Bölsche, W. 1919. *Az állatok családjája. Révai-kiadás.* <https://mek.oszk.hu/00500/00541/00541.pdf> [2011. augusztus 10.]
- Bölsche, W. 1919. *Az ember származása. Révai-kiadás.* <https://mek.oszk.hu/00500/00543/00543.pdf> [2011. augusztus 10.]
- Brehm, A. 1901. *Az állatok világa. Első kötet. Emlősök.* "Budapest: Légrády Testvérek."
- Burchard-Bélaváry, J.-C. 2004. *Családi történetek. Jules-Conrad Burchard-Bélaváry emlékei fia közreadásában. A törököknél. Magyar és lengyel emigránsok az 1848-49-es szabadságharc után.* Magyar Elektronikus Könyvtár. <http://mek.oszk.hu/02200/02212> [2011. augusztus 10.]
- Dewald, J. 2002. *Az európai nemesség 1400-1800.* Budapest: Pannonica kiadó. <http://www.tankonyvtar.hu/tortenelem/eurpai-nemesség-080904-9> [2011. augusztus 10.]
- Dumont L. 2002. *Bevezetés két szociálintropológiai elméletbe. Leszármazási csoportok és házassági szövetség.* Budapest: L'Harmattan Kiadó. <http://www.tankonyvtar.hu/szociologia/bevezetes-ket-080904-45> [2011. augusztus 10.]
- Haeberle, E.J. *Emberi reprodukció. Magyar Elektronikus Könyvtár.* <https://mek.oszk.hu/03600/03627/html/index.htm> [2011. augusztus 10.]
- Human Telex Consulting Kft 2010. *A cél. Tréningprogram.* Budapest: Human Telex Consulting Kft.
- Human Telex Consulting Kft 2010. *A teljesítmény számít: az építő kritika szerepe.* Témafüzet. Budapest: Human Telex Consulting Kft.
- Human Telex Consulting Kft 2010. *A vadon üzenete. Pávián az üzleti életben.* Témafüzet. Budapest: Human Telex Consulting Kft.
- Human Telex Consulting Kft 2010. *A vezetés kihívásai. Tréningprogram.* Budapest: Human Telex Consulting Kft.
- Human Telex Consulting Kft 2010. *Jól csinálom? Tréningprogram.* Budapest: Human Telex Consulting Kft.

Segesváry V. 2005. *Az iszlám és a reformáció. Tanulmány a zürichi reformátorok iszlámmal szembeni magatartásáról.* Hága: Mikes International. <http://mek.oszk.hu/02400/02498/02498.pdf> [2011. augusztus 10.]

Autentikus magyar vallási szövegek

- Bíró L. 2011. *Hívom a családokat 2011-ben, a Család évében – Bíró László püspök levele.* Magyar Kurír, 2011. 01. 01. <https://www.magyarKurir.hu> [2011. augusztus 10.]
- Erdő P. 2011. *Szűz Mária istenanyasága – Erdő Péter bíboros szentbeszéde.* Magyar Kurír, 2011. 01. 02. <https://www.magyarKurir.hu> [2011. augusztus 10.]
- Erdő P. 2011. *Erdő Péter szentbeszéde az ökumenikus imahét megnyitó istentiszteletén* Magyar Kurír, 2011. 01. 17. <https://www.magyarKurir.hu> [2011. augusztus 10.]
- Erdő P. 2011. *Erdő Péter bíboros szentbeszéde Szalézi Szent Ferenc ünnepén.* Magyar Kurír, 2011. 01. 25. <https://www.magyarKurir.hu> [2011. augusztus 10.]
- Erdő P. 2011. *Erdő Péter bíboros szentbeszéde Szent Margit ünnepén.* Magyar Kurír, 2011. 01. 17. <https://www.magyarKurir.hu> [2011. augusztus 10.]
- Gyulay E. 2002. Államegyház? *Tárlatok* 55. szám. 41–43.
- Komáromi Cs. 2006. *Magyar szentek és boldogok.* Pécs: Kézirat. https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://regi.katolikus.hu/csatolt/mszentekboldogok.rtf&ved=2ahUKEwi5utqp_cKHAxUPKRAIHWOsFSwQFnoECBkQAQ&usg=AOvVaw3l3V09CD19dzk6C9ZYFll- [2011. augusztus 10.]
- Magyar katolikus püspöki konferencia. 1999. *A boldogabb családokért. A Magyar Katolikus Püspöki Konferencia körlevele a hívekhez és minden jóakarátú emberhez a házasságról és a családról Magyarországon.* Budapest: Magyar katolikus püspöki konferencia. <http://uj.katolikus.hu/nyomtat-konyvtar.php?h=157> [2011. augusztus 10.]
- Magyar katolikus püspöki konferencia. 1996. *Igazságosabb és testvériesebb világot! A Magyar Katolikus Püspöki Kar körlevele.* Budapest: Magyar katolikus püspöki konferencia. <https://regi.katolikus.hu/konyvtar.php?h=184> [2011. augusztus 10.]
- Márfi Gy. 2011. *Márfi Gyula érsek vízkereszt-i körlevele.* Magyar Kurír, 2011. 01. 06. <https://www.magyarKurir.hu> [2011. augusztus 10.]
- Márton Á. 1973. *Válogatás Márton Aron írásaiból és beszédeiből.* <http://uj.katolikus.hu/nyomtat-konyvtar.php?h=172> [2011. augusztus 10.]
- Thordy A. 2004. A dicsőítés szolgálatában virrasztani. In: Puskás A. (szerk.) *A Szent Titok vonzásában: A hetven éves Fila Béla köszöntése.* 169–178. http://liturgika.htk.ppke.hu/attila/fest_fila_a_dicsoites_szolgalata.pdf [2011. augusztus 10.]
- Thordy A. 2002. Húsvét örömhíre szent Márk szerint. In: Fancsali A. (szerk.) *"Mindent az Evangéliumért": Gyürki László 70. születésnapjára.* 151–162 http://liturgika.htk.ppke.hu/attila/fest_gyurki_mark_16.pdf [2011. augusztus 10.]
- Thordy A. 2003. A 'megismerés' illetve az 'ismeret' fogalma a Corpus Paulinumban. In: Baán I. (szerk.) *A szent atyák nyomdokait követve : In memoriam Vanyó László.* Budapest:

- Bizantinológiai Intézeti Alapítvány. 395–404. http://liturgika.htk.ppke.hu/attila/megismeres_ismeret_fogalma_a_corp_paul.pdf [2011. augusztus 10.]
- Thordy A. 2005. A neaniskos szerepe a Márk megfogalmazása szerinti húsvéti misztériumban. In: Tarjányi B. (szerk.) *A 65 éves Rózsa Huba köszöntése*. Budapest: Szent István Társulat. http://liturgika.htk.ppke.hu/attila/neaniskos_szerepe_Mk_ban.pdf [2011. augusztus 10.]
- Thordy A. 2002. Ósegyházi hitvallás Jézus föltámadásáról. http://liturgika.htk.ppke.hu/attila/fest_bolberitz_az_osz_bolcsesseg.pdf [2011. augusztus 10.]
- Thordy A. 2002. Az Ószövetség bölcsesség-fogalma. "In: Rokay Z. (szerk.) *Egység a különbségben: A 60 éves Bolberitz Pál köszöntése*. 245–254." Budapest: Szent István Társulat.
- Thordy A. 2008. *Jézus programbeszéde: A Hegyi Beszéd*. Szeged: Gál Ferenc Főiskola.
- Thordy A. 2008. A zsoltárok az egyház imádságai. *Vigilia* 2008/4. 250–258. http://liturgika.htk.ppke.hu/attila/fest_jelenits_zsoltarok.pdf [2011. augusztus 10.]
- Victor J. 2008. "Boldog rabság. Vezérfonal Pál apostolnak a filippbeliekhez írott levele tanulmányozásához" Szövétnék-Kiadás. <http://mek.oszk.hu/06700/06747> [2011. augusztus 10.]

Fordított magyar vallási szövegek

- Burton, R. E. 2008. *Önemlékezés. Út az öröklétbe*. Magyar Elektronikus Könyvtár. <http://mek.oszk.hu/06400/06423> [2011. augusztus 10.]
- Collins, M. 1935. *Világosság az ösvényen és Karma*. Budapest: Magyar Teozófiai Társulat. <http://mek.oszk.hu/00100/00118> [2011. augusztus 10.]
- Grün, A. 2008. *Ünnepek és szokások, amelyek megépítik az életünket*. Budapest: Jel Könyvkiadó.
- Grün, A. 2011. *A mennyország benned kezdődik*. Budapest: Jel Könyvkiadó.
- Grün, A. 2011. *A válaszkönyv*. Budapest: Jel Könyvkiadó.
- Grün, A. 2010. *Mit tegyek?* Budapest: Jel Könyvkiadó.
- Grün, A. 2010. *Szentek a szükség idején*. Budapest: Jel Könyvkiadó.
- Grün, A. 2011. *Változtasd át félelmed!* Budapest: Jel Könyvkiadó.
- II. János Pál pápa. 1991. *Centesimus annus*. Magyar Katolikus Püspöki Konferencia. <https://regi.katolikus.hu/konyvtar.php?h=86> [2011. augusztus 10.]
- II. János Pál pápa. 2003. "Ecclesia De Eucharistia. II. János Pál pápa enciklikája az Eucharisztia és az Egyház kapcsolatáról." Libreria Editrice Vaticana. https://www.vatican.va/content/john-paul-ii/hu/encyclicals/documents/hf_jp-ii_enc_20030417_ecll-de-euch.html [2011. augusztus 10.]
- II. János Pál pápa. 1998. *II. János Pál pápa DIES DOMINI kezdetű apostoli levele a Püspököknek, Papoknak és Híveknek a Vasárnap megszenteléséről*. Magyar Katolikus Püspöki Konferencia. <http://uj.katolikus.hu/konyvtar.php?h=70> [2011. augusztus 10.]
- II. János Pál pápa. 1999. *II. János Pál pápa FIDES ET RATIO kezdetű enciklikája a Katolikus Egyház püspökeibez. A hit és az ész kapcsolatának természetéről*. Szent István Társulat.

- Az Apostoli Szentszék Könyvkiadója. <http://uj.katolikus.hu/konyvtar.php?h=69> [2011. augusztus 10.]
- II. János Pál pápa. 1996. "II. János Pál pápa *VITA CONSECRATA* kezdetű szinodus utáni apostoli buzdítása a püspökökhöz és papokhoz a szerzetesrendekhez és szerzetesi kongregációkhoz, az apostoli élet társaságaikhoz, a világi intézményekhez és minden hívőhöz az Istennek szentelt életről és annak küldetéséről az egyházban és a világban." Magyar Katolikus Püspöki Konferencia. <http://uj.katolikus.hu/konyvtar.php?h=73> [2011. augusztus 10.]
- Körner, R. 2011. *Akinek füle van, hallja meg!* Budapest: Jel Könyvkiadó.
- Pál püspök. Isten szolgáinak szolgája. A zsinati atyákkal együtt. 1963. *A II. Vatikáni Zsinat. Sacrosanctum Concilium kezdetű konstitúciója a szent liturgiáról.* Magyar Katolikus Püspöki Konferencia. <https://regi.katolikus.hu/konyvtar.php?h=1> [2011. augusztus 10.]
- Pál püspök. Isten szolgáinak szolgája. A zsinati atyákkal együtt. 1965. "A II. Vatikáni Zsinat *GAUDIUM ET SPES* kezdetű lelképásztori konstitúciója az Egyházzól a mai világban." Libreria Editrice Vaticana. https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19651207_gaudium-et-spes_hu.html [2011. augusztus 10.]
- Sampson, S. 2002. *Meghallbatod isten hangját.* Magyar Elektronikus Könyvtár. <http://mek.oszk.hu/00100/00186> [2011. augusztus 10.]
- Wagner, P. 1994. "Területi szellemek. Tizenkilenc keresztény vezető gondolatai a stratégiai szintű szellemi hadviselésről." Budapest: Amana. <http://mek.oszk.hu/00100/00196> [2011. augusztus 10.]
- XIII. Leó pápa. 1891. "Rerum Novarum. XIII. Leó pápa enciklikája a munkáskérdésről (az első pápai szociális körlevél)." Magyar Katolikus Püspöki Konferencia. <https://regi.katolikus.hu/konyvtar.php?h=125> [2011. augusztus 10.]
- XVI. Benedek pápa. 2011. XVI. Benedek pápa üzenete a betegek XIX. Világnapjára. Magyar Kurír, 2011. 02. 10. <https://www.magyarkurir.hu> [2011. augusztus 10.]
- XVI. Benedek pápa. 2005. XVI. Benedek *Deus caritas est* kezdetű enciklikája a püspököknek, a papoknak és diakónusoknak, az Istennek szentelt személyeknek és minden krisztushívőnek a keresztény szeretetről. Libreria Editrice Vaticana. http://www.vatican.va/holy_father/benedict_xvi/encyclicals/documents/hf_ben-xvi_enc_20051225_deus-caritas-est_hu.html [2011. augusztus 10.]
- XVI. Benedek pápa. 2009. "Caritas In Veritate. XVI. Benedek pápa enciklikája a püspököknek, papoknak és diakónusoknak az Istennek szentelt személyeknek, a krisztushívő laikusoknak és minden jóakarátú embernek. Az ember teljes értékű fejlődéséről a szeretetben és az igazságban." Libreria Editrice Vaticana. http://www.vatican.va/holy_father/benedict_xvi/encyclicals/documents/hf_ben-xvi_enc_20090629_caritas-in-veritate_ge.html [2011. augusztus 10.]
- XVI. Benedek pápa. 2007. "Sacramentum Caritatis. XVI. Benedek pápa szinodus utáni apostoli buzdítása a püspökökhöz és papokhoz, a szerzetesekhez és szerzetesnőkhöz, és a világi krisztushívőkhöz. Az eucharisziáról, az egyház életének és küldetésének forrásáról és csúcs-

pontjáról." Libreria Editrice Vaticana. http://www.vatican.va/holy_father/benedict_xvi/apost_exhortations/documents/hf_ben-xvi_exh_20070222_sacramentum-caritatis_hu.html [2011. augusztus 10.]

XVI. Benedek pápa. 2007. *XVI. Benedek pápa Spe salvi kezdetű enciklikája a püspököknek, papoknak és diakónusoknak, az istennek szentelt személyeknek és minden krisztushívőnek a keresztény reményről.* Libreria Editrice Vaticana. http://www.vatican.va/holy_father/benedict_xvi/encyclicals/documents/hf_ben-xvi_enc_20071130_spe-salvi_ge.html [2011. augusztus 10.]

Újszövetségi Szentírás <http://www.biblia-tarsulat.hu/biblia/parhuzam.htm> [2011. augusztus 10.]

A fordítási szövegalkotás leírása központi helyet foglal el a fordítástudományi kutatások között. A vizsgálatok eredményei arra utalnak, hogy a fordított szövegek sajátos jellegzetességei nem csupán a nyelvi struktúrák szintjén azonosíthatóak, hanem mélyebb, szemiotikai és pragmatikai szinteken is. A fordítások gyakran tartalmaznak olyan nyelvi elemeket és mintákat, amelyek az adott nyelvre nem, vagy csak másmilyen gyakorisággal és/vagy eloszlással jellemzőek. Az autentikus szövegproduktiótól eltérő fordítási mintázatok jelentős hatással lehetnek a szövegek befogadására és értelmezésére. A lexikai ismétlés szövegszervező funkciójának bemutatott vizsgálata korpusznyelvészeti eszközökkel valósult meg. Az integrált megközelítés nagymintájú szövegvizsgálatot tett lehetővé, s ezzel új perspektívákat nyitott a nyelvtechnológia számára a szövegelemzés területén.

A Fordítástudományi értekezések könyvsorozat negyedik köteteként megjelenő *A fordítások lexikai kohéziós mintázata* című tudományos munka az autentikus szövegproduktiótól eltérő fordítási mintázatok közül a lexikai ismétléskapcsolatokban bekövetkező változásokat vizsgálta. A téma és annak újszerű megközelítése számot tarthat gyakorló fordítók és fordításkutatók érdeklődésére is.