

# Korpusznyelvészet és fordításkutatás

*Károly Krisztina*

## 1. Bevezető

Az elmúlt tíz évben a korpuszalapú kutatások mind nagyobb tért hódítottak a nyelvészet és az alkalmazott nyelvészet különböző területein. A számítógépes adattárolás és -elemzés olyan mennyiségű és minőségű információt bocsátott rendelkezésre, amelynek köszönhetően egyrészt újszerű megfigyelések láthatnak napvilágot a nyelvhasználattal kapcsolatban, másrészt a korábbi feltevések újragondolására készítették a kutatókat. A korpuszok alkalmazása pontos és megbízható megfigyelést tesz lehetővé a nyelvtanuló, a nyelvtanár, a fordító és a kutató számára is (Sinclair 1999). A korpusznyelvészet eszközei és módszerei segítségével ugyanis legyőzhetőek azok a nehézségek (pl. az időigényes elemzés miatti kis méretű adatbázisok, vagy gépi elemzés híján a nem következetes, megbízhatatlan, szubjektív elemzések), amelyek miatt a korábbi kutatási eredmények egy része nem szolgálhatott általánosítható következtetésekkel.

A dolgozat célja (1) a korpuszalapú kutatások – és ezeken belül különösen a korpuszalapú fordításkutatás – néhány elvi és módszertani kérdésének bemutatása, (2) a korpuszalapú vizsgálatok előnyeinek ismertetése, valamint (3) a korpuszalapú vizsgálatok néhány fontosabb eredményének bemutatása annak érdekében, hogy rávilágíthassunk arra, hogy a korpusznyelvészet eszközei és módszerei milyen újszerű eredményekkel szolgálhatnak a fordításkutatásban.

## 2. A korpusz meghatározása és fajtái

A **korpusz** olyan számítógépen tárolt szöveggyűjtemény, amelyet explicit módon megfogalmazott kritériumok szerint és meghatározott céllal hoztak létre (Atkins és Clear 1992: 5). A korpuszoknak számos fajtája létezik. A fordításkutatás szempontjából az egyik jelentős típus az **egynyelvű angol korpusz** (pl. a Brown Corpus, amelyet 1962-ben kezdtek el építeni; az International Corpus of English (ICE), amely brit, amerikai, ausztrál és kanadai angolt is tartalmaz; a Bank of English; a London-Lund Corpus; a British National Corpus (BNC); és az írott brit és amerikai angolt tartalmazó Longman-Lancaster Corpus). Léteznek ún. **nyelvtanulói korpuszok** (angolul „learner corpora”) is, amelyek olyan számítógépen tárolt szöveges adatbázisok, amelyek a nyelvtanulók által produkált nyelvet tartalmazzák (Leech 1998: xiv.o.). Ilyen például a – többek között – francia, kínai és lengyel anyanyelvű angol nyelvtanulók szövegeit tartalmazó International Corpus of Learner English (ICLE), to-

vábbá a Longman Learners' Corpus és a Hong Kong University of Science and Technology (HKUST) Learner Corpus.

A fordításkutatásban, ahol mind a forrásszöveg, mind a célnyelvi (fordított) szöveg, mind pedig az ún. nem fordított, eredeti cél- és forrásnyelvi szövegek vizsgálata folyik, speciális korpuszokat használnak. Az egyik ilyen korpuszfajta a **többnyelvű korpusz** (angolul "multilingual corpus"), amely hasonló szövegtípusokat tartalmaz több nyelven. A **kétnyelvű korpuszokon** ("bilingual corpora") belül két fő típus van. A **párhuzamos korpusz** ("parallel corpus") a forrásszövege(ke)t és annak fordításait tartalmazza (pl. az English-Norwegian Parallel Corpus az oslói egyetemen). Az **összehasonlítható korpusz** ("comparable corpus") nem fordításokat, hanem hasonló körülmények között, spontán alkotott, hasonló típusú szövegeket tartalmaz két különböző nyelven. A kétnyelvű korpusz ellentettje az **egynyelvű (összehasonlítható) korpusz** ("monolingual comparable corpus"), amely azonos nyelven született eredeti/autentikus és fordított szövegeket tartalmaz (pl. az English Comparable Corpus (ECC) a manchesteri UMIST egyetemen és a Finnish Comparable Corpus a joensuu egyetemen). Az utolsó típus, az ún. **fordítási korpusz** ("translation corpus") kizárólag fordításokból áll (pl. az English-Italian Translational Corpus (CEXI) a bolognai Forli Egyetemen).

### **3. A korpuszalapú kutatások főbb jellemzői és témakörei a nyelvészet és az alkalmazott nyelvészet területén**

Annak érdekében, hogy rámutathassunk a korpuszalapú kutatások hasznára, helyére és főbb eredményeire a fordításkutatás területén, az alábbiakban röviden bemutatjuk – általában véve – a korpuszkutatások célját, főbb területeit és az ilyen vizsgálatokat ért legfontosabb kritikákat. Rámutatunk továbbá arra, hogy a bírálat ellenére mégis miben és hogyan segítheti a korpusznyelvészet eszköztára mindennapi tanári, nyelvtanulói, fordítói és/vagy kutatói munkánkat.

#### **3.1. A korpuszalapú kutatások jellege**

A korpuszalapú kutatások a beszélt vagy írott **nyelvhasználatot** vizsgálják. Ez a hatalmas előnye azon kutatásokkal szemben, amelyek a nyelvet csupán izolált példákon keresztül látatják. A nyelvhasználat elemzése során a kutatásoknak két fő célja van: (1) a nyelvhasználatban fellelhető szabályszerűségek vizsgálata, illetve (2) a variabilitást befolyásoló, a szöveggörnyezetben fellelhető tényezők elemzése (Biber, Conrad és Reppen: 1998). E kettős cél megvalósításához egyrészt nagy mennyiségű adatra van szükség, hogy minimalizálható legyen a vizsgálat során az intuíció szerepe, valamint fontos, hogy e nagy adatbázis nagy számú beszélőtől származzon, hogy az egyéni sajátosságok befolyása jelentéktelen legyen.

### 3.2. A korpuszalapú kutatások főbb területei

A **leíró nyelvészet**en belül a korpuszok és a számítógépes elemzőprogramok (pl. konkordancia programok) segítségével leírhatók a nyelvhasználatot jellemző asszociációs minták (pl. milyen lexikai/grammatikai, vagy akár nem nyelvi asszociációk jellemeznék bizonyos nyelvi jelenségeket). A **lexikográfia** is nagymértékben épít rá, hiszen segítségével könnyedén megvizsgálható bizonyos szavak különféle jelentése, gyakorisága egy adott regiszterben, kollokációs mintája, stb. Bár a gyakran emberi értelmezést megkövetelő elemzések miatt ugyan korlátozott mértékben, de a **szövegnyelvészet** területén is hasznos információkkal szolgál. Segíthet a szövegszintű nyelvi eszközök (pl. referencia, ismétlés) használatának leírásában adott regiszterek vagy szövegtípusok esetében ugyanúgy, mint például a beszélt és az írott nyelv szövegszintű sajátosságainak elkülönítésében.

Az **alkalmazott nyelvészet** területén különösen fontos szerepet töltenek be a korpuszalapú kutatások. Csak hogy néhány példát említsünk: alkalmazható a szaknyelv-kutatás és -oktatás területén, az anyanyelv és az idegen nyelv elsajátításának vizsgálatában, a nyelvtanításban, a nyelvi mérés validálási folyamatában és a stilisztikában ugyanúgy, mint a leíró fordításkutatásban és a szakfordító-képzésben.

A **fordításkutatásban** a korpuszalapú vizsgálódás két leggyakoribb témája (1) a fordított és eredeti (nem fordított) szövegek összevetése és (2) a forrásnyelvi és célnyelvi szövegek nyelvi jellemzőinek leírása. Az előbbi a lexikai, szintaktikai és szövegszintű nyelvi jelenségek vizsgálatán keresztül rámutat a fordított szövegek olyan egyedi sajátosságaira, amelyek nem a forrás- vagy célnyelv hatására alakulnak ki. Az ilyen sajátosságok kimutatása több összehasonlítható korpuszban lehetővé teszi ún. fordítási univerzálék megfogalmazását (Baker 1998: 52). A másik terület, a forrásnyelvi és a célnyelvi szöveg összehasonlítása a nyelvpár-specifikus fordítói magatartás vizsgálatát teszi lehetővé. Bemutatja, hogy gyakorlott fordítók hogyan oldanak meg bizonyos fordítási problémákat és ezáltal az előíró jellegtől a leíró jelleg felé tolja a fordítástudományt (Kenny, idézi Baker 1998: 51).

### 3.3. A korpuszalapú kutatások általános bírálata

A korpuszalapú kutatásokkal kapcsolatban megfogalmazott kritikák leggyakrabban azok erősen kvantitatív, elsősorban az **előfordulási gyakoriságot** vizsgáló jellegét bírálják (Newmark 2002: 255). Bírálják azonban azért is, mert a folyamat helyett a nyelvi termékre és a kompetencia helyett a **performanciára** helyezi a hangsúlyt. Hátrányának róják fel továbbá azt is, hogy az automata gépi keresések csupán **felszíni nyelvi jelenségek** elemzésére korlátozódnak, valamint azt, hogy az összehasonlítás alapja általában az **anyanyelvi beszélő** performanciája, amely nehezen definiálható és nem is mindig kívánatos mérce (Leech 1998). Larsen-Freeman és Long (1991) a **természetes nyelvhasználat** nem minden esetben ideális voltát emelik ki: (a) bizonyos nyelvi jelenségek ugyanis egyáltalán nem vagy csak nagyon ritkán fordulnak elő, (b) nem lehet

minden, a nyelvhasználatot befolyásoló tényezőt kontrollálni, (c) nem hordozza a teljes nyelvi repertoárt, hiszen a beszélők a problémás jelenségeket inkább elkerülik (idézi Granger 1998: 4-5). Widdowson (2000) arra hívja fel a figyelmet, hogy mivel a korpuszalapú kutatások eredményei nem ritkán **ellentmondanak az anyanyelvi beszélő megérzéseinek**, valószínűleg ez a kutatási módszer sem oldja meg az összes kutatómódszertani és/vagy elméleti problémát. Ezek mellett az sem elhanyagolható deficitje, hogy jellegéből fakadóan (sok szöveg sok beszélőtől/írótól) figyelmen kívül hagyja az **egyéni sajátosságokat, egyedi előfordulásokat**, amelyek ugyanakkor fontos összetevői lehetnek bizonyos jelenségeknek.

### 3.4. A korpuszalapú kutatások előnyei

A fenti kritikai megjegyzések ellenére számos érv szól a korpuszalapú tudományos vagy éppen gyakorlati vizsgálódás mellett. Biber és Conrad (2001) szerint jóval többről van szó, mint egyszerű gyakorisági mutatók számításáról, hiszen a korpuszelemző szoftverek segítségével kimutatható a különféle **regiszterek jellegzetes grammatikája** és igazolhatóak vagy kijavíthatók az anyanyelvi beszélők – esetlegesen téves (!) – **intuíciói** a nyelvhasználattal kapcsolatban. Ehhez azonban elengedhetetlen a kvantitatív eredmények **kvalitatív, funkcionális értelmezése** (Biber, Conrad és Reppen 1998: 5). Ellis (2002) szerint az **input-gyakoriság** – és következésképpen annak elemzése is – fontos tényezője a nyelvvelsajátítási elméleteknek, mivel a nyelvtanulás és a nyelvi folyékonyság elsősorban nem elvont szabályok vagy szerkezetek ismeretén múlik, hanem azon, hogy a nyelvtanuló mennyire képes az emlékezetéből felidézni korábban hallott vagy olvasott megnyilatkozásokat (i.m. 166).

A korpuszalapú vizsgálatok nemcsak azt mutatják meg, amikor a nyelvtanulók hibáznak, hanem azt is, ha valamit sikeresen fogalmazzak meg. Ezért nagy előnyük az, hogy nemcsak bírálatot, hanem **elismerést** is indukálhatnak. A különféle nyelvi jelenségek általános vizsgálata mellett lehetőséget nyújtanak annak kimutatására is, hogy mely nyelvi jelenségeket használnak a nyelvtanulók jelentősen **ritkábban** vagy éppen sokkal **gyakrabban**, mint az anyanyelvi beszélők (Leech 1998), aminek kimagasló a jelentősége a nyelvtanulásban. Stubbs (2001) véleménye szerint pedig az, hogy a terméket vizsgálja és nem a folyamatot, nem kisebbíti az ilyen jellegű vizsgálatok érdemét, hiszen a nyelvi produktum az, ami ténylegesen **megfigyelhető**, elemezhető, úgymond „tapintható”.

Nagyon lényeges a korpuszalapú vizsgálatoknál a megfelelő módszertani hozzáállás: a gépi programok segítségével maximális objektivitással kell eljárni a nyelvi jelenségek leírásában, de a kutató megérzései, meglátásai és reakciói elengedhetetlenek a jelenségek kielégítő értelmezéséhez (Sinclair 1999: 21–22). A korpusznyelvészeti különlegessége és előnye éppen abban rejlik, hogy nem azzal foglalkozik, ami véletlenszerűen (tehát legalább egyszer) előfordul. Módszereit úgy alkották meg, hogy kizárják az egyedi eseteket, amelyek statisztikailag nem szignifikánsak, és olyan jelenségekre összpontosulnak, amelyek gyakran és jellemzően fordulnak elő (Stubbs 2001: 151).

#### 4. Korpuszalapú vizsgálatok a fordításkutatásban

A fordításkutatásban két terület válik markánsan szét: a hagyományos értelemben vett fordításkutatás (a fordítói magatartás vizsgálata) és a pedagógiai fordítások vizsgálata (pl. a fordítás mint nyelvoktatási vagy nyelvi mérési feladat). Ez utóbbi terület szakirodalmi jóval szegényesebb, de a fordítástudomány számos megállapítása fontos implikációkat hordoz a pedagógiai célú fordításkutatás számára is. A korpusznyelvészet mindkét terület számára hasznos információkkal szolgál. Az alábbiakban először röviden áttekintjük a korpuszalapú fordításvizsgálatok főbb témaköreit, majd a legfrissebb kutatások közül mutatunk be néhányat, amelyek szemléltetik a különféle témakörökben végezhető vizsgálatok lehetséges módszereit és eredményeit.

A korábbiakban részletezett **korpuszfajták** különböző jellegű elemzésekre adnak módot. A párhuzamos korpuszok segítségével bemutatatható, hogy a gyakorlott fordítók miként oldanak meg bizonyos fordítási problémákat. A többnyelvű korpuszok megmutatják a célnyelv jellegzetes (eredeti) nyelvezetét (szemben a fordítások nyelvezetével). Az egynyelvű összehasonlítható korpuszok segítségével összehasonlíthatók – egy adott nyelven – az eredeti (nem fordított) és a fordított szövegek, annak érdekében, hogy kimutathassuk a fordításokat jellemző eltéréseket. A nagyméretű referencia-korpuszok pedig a kulturális vonatkozású kérdések tisztázásában is segítenek.

A korpuszok számítógépes elemzésére szolgáló **konkordancia-programok** is nagyban segítik a fordítók és a kutatók munkáját. Lehetővé teszik az adat-alapú tanulást (angolul ún. „data-driven learning”) a kezdő fordítók vagy nyelvtanulók számára, akik e programok segítségével rákereshetnek bizonyos szabályszerűségekre és megpróbálhatják értelmezni azokat (pl. jellegzetes ekvivalenciák vagy megoldások). A konkordancia-programok a forrás- és a célnyelv aszimmetrikus jellegét is megmutatják (pl. a sok esetben hiányzó formai és/vagy strukturális ekvivalenciát). Összegzésül tehát elmondható, hogy a korpusznyelvészet eszköztára olyan információkkal szolgálhat, amelyekhez nem juthatunk hozzá hagyományos eszközökkel vagy csak nagyon sok idő és munka árán. Az alábbiakban a fenti lehetőségeket illusztráljuk néhány konkrét példán keresztül, kiemelve a fő fordítási problematikát vagy témát, amivel a tanulmány foglalkozik.

Stewart (2000) azt vizsgálja, hogy a korpuszok mennyire alkalmazhatók a **célnyelvi produkció elősegítésére** és a **kreatív íráskészség fejlesztésére**. Munkáját az a feltevés motiválta, miszerint a fordított szövegek jóval konvencionálisabbak, mint az eredeti szövegek. Különösen igaz ez az idegen nyelvre történő fordítás esetén, amikor a fordítók – a nyelvi korlátok miatt – kevésbé képesek „manőverezni” vagy kreatívan bánni a nyelvvel. A vizsgált szövegek egyik részét az egynyelvű Brit Nemzeti Korpusz, másik részét pedig forrásnyelvi szövegek (olasz turistabrosúrák) alkotják. Az elemzés adott szavak kollokációit vizsgálja (pl. olyanokét, amelyek esetében a szótár csak egy kollokációt ad meg vagy azt, hogy bizonyos tulajdonneveket le kell-e fordítani a célnyelvre). Számol kollokáció-gyakoriságot és kollokációs arányt (ami a kollokációs gyakoriság és a megoldások aránya). Az eredmények azt mutatják, hogy a nagy refe-

rencia-korpuszok két jelentős szempontból is hasznosak a fordításban: (1) idegen nyelvre történő fordítás során jóval több információt bocsátanak a fordító rendelkezésére, amelyek segítségével a nyelvi vagy kulturális korlátok/hiányosságok könnyebben leküzdhetők, mint például a hagyományos szótárakkal, (2) konvencionálisabb nyelvhasználatot mutatnak, mint ami általában az anyanyelvi beszélőket jellemzi.

Fekete (2002) a pedagógiai fordításkutatás területén a nyelvtudás mérésének szolgálatába állítja a korpusznyelvészet eszközeit: a **fordítás**, mint nyelvvizsga-feladat **validálásához** alkalmazza. Kutatásának célja kettős: egyrészt meghatározza a fordítói performancia azon aspektusait, amelyeket empirikus módon csak a konkordancia programok segítségével lehet megfelelően tanulmányozni, másrészt pedig számba veszi a fordítások értékelésének azon elemeit, amelyeket a vizsgáztatók bár nem tudnak empirikus módon validálni, mégis figyelembe vesznek a fordítások minőségének értékelésekor. A vizsgálathoz az ITK Korpusz Projekt adatbázisából egy angol újságcikk 113 magyar nyelvű fordítását használta fel. Az elemzés négy fő szempontra korlátozódik: azt vizsgálja, hogy konkordancia-programmal kimutatható-e kapcsolat (1) a nehézségi szint és a lexikai változatosság (angolul „lexical variation”: „type/token ratio”), valamint a mondatok hosszúsága (szó/mondat) között, (2) a tipikus vizsgázói fordítás (helyes megoldás), hiba (szám és típus) és a megfelelő szóválasztás között (ez utóbbi a tételanalízis alapjául szolgált), (3) a különféle elfogadható megoldások és az értékelés között, valamint (4) az értékelési kritériumok módosítása és a pontszámok alakulása között. E tényezők kapcsolatát gyakoriság és kontextus-elemzéssel, valamint más egyéb statisztikai számításokkal (pl. diszkriminációs index számítása) vizsgálja. Érdekes módon az eredmények a 2., 3. és 4. szempontban szereplő jelenségek között igazolt kapcsolatot, az elsőben – a nehézségi szint és a lexikai változatosság, valamint a mondatok hosszúsága (szó/mondat) között – nem. Összegzésül Fekete megállapította, hogy a konkordancia-programok alkalmazhatóak a fordításnak mint vizsgafeladatnak a validálásához (pl. tételanalízis vagy a sikeres vizsgázók fordítói stratégiáinak feltérképezése).

A fordításoktatáshoz Kohn (1999, 2000) párhuzamos korpuszokat és konkordancia-programokat hívott segítségül annak igazolására, hogy a számítógép felhasználása hozzájárul a fordítói tudatosság növekedéséhez és a **fordítási készség** kifejlődéséhez. Kohn munkáját néhány példán keresztül szemléltetjük. Az egyik elemzéshez a résztvevők a *Microconcord* nevű konkordancia-programot és német nyelvű eredeti szövegek magyar fordításait használták. A konkordanciák alapján feltárták a szövegek feltűnő lexikai és grammatikai sajátosságait, az esetleges fordítási törvényszerűségeket, valamint a hiányzó kategóriák kompenzációs lehetőségeit.

Más jellegű vizsgálatot tett lehetővé a *Multiconcord* nevű többnyelvű konkordancia-program, amelynek köszönhetően az összegyűjtött megfeleltetések egy időben mindkét nyelven és adott szöveggörnyezetben jelennek meg a képernyőn. E program segítségével három német szöveg különböző nyelvű (angol, francia, olasz, görög, dán) fordításait vizsgálták. Az egymással szembeállított szövegrészletek alapján strukturális és funkcionális eltéréseket állapítottak meg,

és tudatosították saját anyanyelvük néhány olyan sajátosságát, amire addig nem figyeltek. Az izolált megfigyelésekből törvényszerűségek rendszerét lehetett megfogalmazni. Kohn meglátása szerint a korpuszalapú fordításoktatás fő előnye az, hogy „optimálisan kapcsolja össze egymással a fordítói rutint és a problémamegoldási stratégiákat” (2000: 13).

Kenny (2000) is párhuzamos korpuszalapú dolgozott, amikor a **lexikai kreativitást** vizsgálta. Tanulmánya a német-angol, irodalmi szövegeket tartalmazó párhuzamos korpusz a GEPCOLT elemzése mellett kitér olyan módszertani kérdésekre is, amelyek a szoftver adta lehetőségek és a kutatási célok esetleges összeférhetetlenségeiből fakadnak. Nemcsak elemzési módszereket mutat be, hanem értékeli és „finomítja” is azokat. Kutatásának fő célja tisztázni azt a több kutató által is sugallt kérdést, hogy vajon a normalizáció (vagyis a fordított szövegek nyelvileg konzervatívabb jellege a célnyelv eredeti/autentikus szövegeihez vagy a forrásnyelvi szöveghez képest) jellemző-e a fordításokra. A forrásszövegekben rákeres a lexikai kreativitás különféle megnyilvánulásaira és az elemzőprogramok adta lehetőségek segítségével beazonosítja a kreatívknak minősíthető megfogalmazásokat, az íróra jellemző, speciális kifejezéseket és a szokásostól eltérő kollokációkat.

Mauranen (2000) kutatásának középpontjában a fordított és az eredeti/autentikus szövegek összevetése áll. Témája az olyan több szóból álló szóláncok vizsgálata, amelyek ún. „szöveg-reflexív” jelentést hordoznak. Ezen szóláncok elválnak a szöveg propozicionális tartalmától, feladatuk elsősorban a **nyelvi szövegszerkesztés**, az olvasói értelmezés megkönnyítése, segítése. Különösen gyakoriak ezek a tudományos szövegekben, és jóval gyakoribbak az angolban, mint például a finnben, a kutató anyanyelvében. Korpuszalapú vizsgálata megerősítette azt a feltevést, miszerint a tudományos szövegekben több a reflexivitás, mint a népszerű szövegekben, és a fordításokban is több van, függetlenül a forrásnyelvtől, amiről fordították. Elemzése arra is rámutattak, hogy a fordított szövegekben a többszavas szerkezetek kevésbé világosak és állandóak, mint az eredeti szövegekben, és el is térnek azoktól. Ez arra enged következtetni, hogy a fordított szövegek nem ugyanolyan ko-szelekciós mintákat mutatnak, mint a hozzájuk hasonlítható eredeti szövegek. Az adatok azt is tanúsították, hogy erősen célnyelv-specifikus kifejezések már sokkal kevésbé szerepelnek a fordításokban. Érdekes eredménye Mauranen kutatásának az, hogy a népszerű, nem irodalmi szövegek jobban eltérnek a célnyelvi normától, mint a tudományos szövegek, ami arra enged következtetni, hogy a kulturális presztizs kevésbé lényeges, mint a fordítás egyéb meghatározó tényezői.

Összehasonlítható korpuszokkal dolgozott Olohan és Baker (2000) is. Kutatásuk célja a fordításra jellemző **explicitáció** jelenségének igazolása volt. Elemzésük egy konkrét szerkezetre korlátozódott az angolban, mégpedig az opcionális *that* jelenlétének vagy hiányának vizsgálatára a függő beszédben használt *say* ill. *tell* igék után. Az elemzéshez az adatbázist a TEC (Translational English Corpus, magyarul Fordított Angol Nyelvi Korpusz), illetve a BNC (British National Corpus, magyarul Brit Nemzeti Korpusz) szolgáltatta. A konkordancia vizsgálat egyrészt kimutatta, hogy a *that* kötőszó sokkal gyakoribb a fordítási korpuszban (TEC), mint az eredeti szövegeket tartalmazó BNC kor-

puszban, másrészt, hogyha megvizsgáljuk mindkét ige összes alakját, akkor azt látjuk, hogy az eredeti szövegek (BNC) sokkal gyakrabban hagyják el, mint a fordított szövegek (TEC).

Jantunen (2001) a fordított szövegek egy másik – sokak által univerzálénak tartott – jellegzetességével foglalkozik, mégpedig az **egyszerűsődéssel** vagy más szóval szimplifikációval. A jelenséget először forrásnyelvi és célnyelvi szövegeken keresztül vizsgálták, majd pedig fordított és nem fordított (eredeti) szövegek összevetésén keresztül. Jantunen tanulmánya ez utóbbi módszert követi. A szimplifikáció gyakran a lexikai elemekben nyilvánul meg. A korpusz-alapú tanulmányok azonban azt sugallják, hogy a fordított szövegekben nem szegényesebb a lexikon. Ezt a szinonimák inherens természetével magyaráznak. A szinonimák használatának korlátai miatt ugyanis a fordítók sok esetben úgy használják őket, mint az eredeti szövegek írói. Jantunen kvantitatív elemzése ezt a problémát igyekszik megvilágítani. Vizsgálatát egy 2,2 millió szavas fordított és nem fordított szövegekből álló finn korpuszra építi. A fordítások angol nyelvű forrásszövegekből készültek, a szövegek műfaja pedig tudományos szaknyelv és szépirodalom. Kutatásai azt mutatják, hogy mindkét nyelvi variáns hasonlóan széles körű szinonimakészlettel rendelkezik.

## 5. Összegzés

E tanulmány legfőbb célja az, hogy felhívja a fordítók, nyelvtanulók, nyelv-tanárok és kutatók figyelmét arra, hogy a korpusznyelvészet módszerei és eszközei miként és milyen területen segíthetik munkájukat. A számítógépes adatbázisok alkalmazása és az elemzőprogramok által nyújtott eredmények megfelelő értelmezése a hagyományos eszközöknél (pl. szótárak) jóval korszerűbb és bővebb információkkal vértetik fel a felhasználót, melynek köszönhetően ma már elengedhetetlen részei a fordításkutatásnak, az alkalmazott és leíró nyelvészetnek egyaránt.

## Irodalom

- Atkins, S., Clear, J. 1992. Corpus Design Criteria. *Literary and Linguistic Computing*. Vol 7. N. 1. 1–16.
- Baker, M. (ed.) 1998. *Routledge Encyclopedia of Translation Studies*. London and New York: Routledge.
- Biber, D., Conrad, S., Reppen, R. 1998. *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: CUP.
- Biber, D., Conrad, S. 2001. Quantitative Corpus-Based Research: Much More Than Bean Counting. *TESOL Quarterly*, Vol. 35. N. 2. 331–336.
- Fekete, H. 2002. An Experimental Translation Corpus Project (Small Learner Corpus). Unpublished Ph.D. assignment, Language Pedagogy Ph.D. Program. Budapest: Eötvös Loránd University.

- Granger, S. 1998. The Computer Learner Corpus: a Versatile New Source of Data for SLA Research. In: S. Granger (ed.) *Learner English on Computer*. London: Longman. 3–18.
- Jantunen, J. H. 2001. Synonymity and Lexical Simplification in Translations: a Corpus-Based Approach. *Across Languages and Cultures* Vol. 2. N.1. 97–112.
- Kenny, D. 2000. Lexical Hide-and-Seek: Looking for Creativity in a Parallel Corpus. In: M. Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*. Manchester: St. Jerome Publishing. 93–104.
- Kohn J. 1999. Párhuzamos szövegek számítógéppel segített elemzése a fordításoktatásban (első rész). *Fordítástudomány* 1. évf.,1.szám. 67–78.
- Kohn J. 2000. Párhuzamos szövegek számítógéppel segített elemzése a fordításoktatásban (második rész). *Fordítástudomány* 2. évf.,1.szám. 5–16.
- Leech, G. 1993. Corpus Annotation Schemes. *Literary and Linguistic Computing* Vol. 8. N.4. 275–81.
- Leech, G. 1998. Preface. In: S. Granger (ed.) *Learner English on Computer*. London: Longman. (pp. xiv–xx).
- Mauranen, A. 2000. Strange Strings in Translated Language. A Study on Corpora. In: Olohan, M. (ed.) *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*. Manchester: St. Jerome Publishing. 119–142.
- Newmark, P. 2002. What makes me tick? Across-Interview by Anikó Sohár. *Across Languages and Cultures* Vol. 3. N.2. 251–255.
- Olohan, M., Baker, M. 2000. Reporting that in Translated English. Evidence for Subconscious Processes of Explication? *Across Languages and Culture* Vol. 3. N.2. 141–158.
- Sinclair, J. 1999. The Computer, the Corpus and the Theory of Language. *LMS Lingua*, 1, 21–32.
- 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 I. Textual and Cognitive Aspects*. Manchester: St. Jerome Publishing. 73–92.
- Stubbs, M. 2001. Texts, Corpora, and Problems of Interpretation: a Response to Widdowson. *Applied Linguistics* Vol. 22. N.2. 149–172.
- Widdowson, H. G. 2000. On the limitations of linguistics applied. *Applied Linguistics*, Vol. 21. N.1. 3–25.