

William J. Crawford – Enikő Csomay
Doing corpus linguistics
 [Gyakorlati korpusznyelvészet]

New York – London: Routledge, 2016. 163 p.
 ISBN: 978-1-138-02461-8

A szerkesztés során elkészült ennek a szövegnek a helyesírási szabályainkat szigorúan érvényesítő változata, ami lényegében kötőjeles alakok egybeírását jelentette. Az eredmény gondolkodásra késztetett. A google adataival megerősített tapasztalataink szerint például a korpuszalapú írásmód mellett a korpusz-alapú forma nagyon is él; máskor, például a textualitás-bázisú vs textualitás-bázisú pár esetében a kötőjelezés jelentősen megkönnyíti a szöveg olvasását; a regiszteranalízist viszont inkább egybeíránánk. Végül úgy döntöttünk, hogy egyes terminusok külön-külön mérlegelése helyett a recenzió kemény szaktudományos szövegében megőrizzük a terület kiváló művelőjeként ismert szerző írásmódját. – A szerk.

Egyre-másra jelennek meg napjainkban a korpusznyelvészettel foglalkozó, bevezető szintű tankönyvek nemzetközileg jelentős kiadók gondozásában. Az itt ismertetendő gyakorlati kézikönyv azonban tartalmát, profilját és metodikáját illetően markánsan eltér piacra dobott társaitól: mindhárom tekintetben kiváló. Nem csupán ismereteket nyújt a tárgyat képező korpusznyelvészetről mint nyelvtudományi diszciplínáról, hanem nagyon átgondoltan, fokozatosan bővíti a rá vonatkozó tudást, leszűrve róla a napjainkra túlzott méretűre dagadt szaktudományos argumentációt és mellőzve a gyakorlati útmutatók számára szükségtelen tudománytörténeti és elméleti jellegű fejtegetéseket. Arra szorítkozik, hogy praktikus ismereteket nyújtson a nyelvi korpuszoknak nevezett elektronikus lexikai adatbázisokról, ezek fejlesztéséről, valamint nyelvtudományi és egyéb célú alkalmazásairól. Ezért lehet a címe magyar fordításban *Gyakorlati korpusznyelvészet*. S mint ilyen funkciót ellátó szakkönyv,

szerény terjedelmű, ami rendkívül előnyös abból a szempontból, hogy a diszciplína bázistankönyveként a hazai felsőoktatásban könnyen alkalmazható a képzési programok féléves moduljaihoz, intervallumaihoz igazodva BA, MA vagy akár doktori szinten is. Nem igényel pallérozott előzetes nyelvtudományi ismereteket, ennek következtében az egyetemi nyelvészeti képzésen kívül a téma iránt érdeklődő, másféle képzettségű szakemberek is eredményesen forgathatják. Külön örömet jelent nekünk, magyaroknak, hogy a könyv egyik szerzője, Csomay Enikő magyar származású, ugyan doktori fokozatát már az Egyesült Államokban szerezte, a korpusznyelvészet amerikai fellegrárában, a Northern Arizona Universityn a diszciplína egyik kiemelkedő személyiségének, Douglas Bibernek a vezetése alatt. A korpusznyelvészeti vizsgálatok gyakorlatába bevezető szakkönyvet ízig-vérig áthatja az amerikai társadalmat jellemzően uraló szemléletmód: a hasznosságra, a közvetlen használhatóságra és ezáltal a sikerességre való törekvés, melynek értelmében a gyakorlati célokat szolgáló produktumok készítésekor a hosszas elméleti beágyazottság helyett a konkrétságnak van prioritása. Ezt a praktikus szemléletet testesíti meg a jelen kézikönyv is, magán viselve a kreatív eredetiség tipikus amerikai jegyeit. Olvasója már a kezdetektől fogva megtapasztalhatja, hogyan valósul meg mindez az ismeretanyag bemutatásában, a gyors és alapos elsajátítás metodikája révén, az ismeretek jól átgondolt adagolásával. A sikerhez vezető varázslatos módszertani fogás abban nyilvánul meg, hogy a diszciplináris szakfogalmak, tudományos tételek megismertetése itt a szokásossal ellentétes sorrendben történik: először az éppen tárgyalt jelenségek adat-orientált, értelmező bemutatását kapjuk, azaz a szerzők teljes körű leírást nyújtanak a vizsgált nyelvi tényanyagról, s csak ezt követően kerül sor a jelenségek szakterminológiai megnevezésére, értelmezésére és definiálására. Nem pedig fordítva, ahogy – főként Európában – általában megszoktuk.

Ebben a metodikai fogásban látom a praktikus tankönyvben nyújtott ismeretek sikeres elsajátíttatásának igazi kulcsát, hallatlan előnyét hasonló célú más korpusznyelvészeti bevezetőkkel szemben.

Most pedig tekintsük át részletesen a fejezetek tartalmát. Az előszóban (xi–xii) a szerzők kiemelik, hogy olyan útmutatót készítettek, amely gyakorlati technikai segítséget nyújt a korpusznyelvészet iránt érdeklődő olvasónak – tanulónak és tanárnak egyaránt – abban, hogy gyűjteni és egyúttal értelmezni tudja az elektronikus lexikai adatbázisokból, korpuszokból (ki)nyert eredményeket. Széleskörűen használt elektronikus korpuszokról közölnek ismereteket, kiegészítve őket *regiszter-analízis* alapján nyert nyelvi tényekkel. Az utóbbi aspektus meghatározó újítás a korpusznyelvészettel bevezető szinten foglalkozó kézikönyvek palettáján. Az általános ismereteket követően a saját adatgyűjtésre alapozott korpusz fejlesztésének módszereivel és technikájával foglalkoznak, külön hangsúlyt fektetve ezek nyelvpedagógiai alkalmazására.

A könyv három egységből, összesen kilenc rövid fejezetből áll. Az első egység (1–31) két fejezete általános és gyakorlatias ismereteket nyújt a nyelv rendszeréről, a korpusznyelvészetről, a nyelvi változatok sokféleségéről, valamint a nyelvi regiszterek fogalmáról és elemzésük megközelítésének módjairól. A nyelvszemléletet tekintve hangsúlyozottan a leíró célt támogatja, mely a nyelvi tényeket – a nyelv használatára támaszkodva – *megállapítja*, nem pedig – a használatot alakító módon – *előírja*. A nyelvhasználat tényeit a regisztervizsgálat faktorainak alapján szemléli és értelmezi. A bemutatott nyelvi példák gyakorisági mutatóit is regiszterbeli előfordulásuk tükrében tárgyalja, Biber és munkatársai korpusz-alapú monumentális grammatikájának (1999) megközelítését, a nyelvreírás adekvátságának követelményrendszerét idézve fel. Az ún. korpusz-alapú (corpus-based) nyelvelemzéssel szemben inkább az ún. korpusz-vezérelt (corpus-driven)

szemléletet részesíti előnyben. Előbbi a már előzetesen feltárt nyelvi adatok feldolgozásához, elemzéséhez keres korpusznyelvészeti eszközöket, míg utóbbi megközelítés alkalmazása során a nyelvész, nyelvelemző maga teremti meg a korpuszát, hozzá létre az azt elemezni, értelmezni képes technikai eszközöket, amelyek lehetővé teszik számára, hogy elérje nyelvreírása megbízhatóságának, validitásának kívánt fokát (Tognini-Bonelli 2001), rámutatva egyúttal a különféle regisztereket jellemző pragmatikai alapú eltérésekre. Az első egység 2. fejezetében a szerzők szemléletes példák elemzésén keresztül mutatják be a regiszter-elemzés és a korpusznyelvészeti vizsgálatok szoros összefüggését, a két diskurzus-bázisú elemzési terület kommunikációs alapú egymásra utaltságát a kontextus ismeretének függvényében. Biber és Conrad új korszakot nyitó nézeteit (2009) alkalmazva hangsúlyozottan tárgyalják a *regiszter*, a *stílus* és a *zsáner* összefüggéseit és eltéréseit. A különféle típusú szövegeket tartalmazó gazdag anyagból vett minták bemutatásán keresztül szemléletesen rajzolódik ki, hogy a nyelvi közlés szociolingvisztikai- és textualitas-bázisú produkciós körülményei miként hatnak magára a nyelvi anyagra és annak vizsgálhatóságára. Korpusznyelvészeti szempontból különösen nagy gyakorlati haszna van a több jelentéssel egységet tartalmazó szerkezetek, az ún. lexikai csomók (lexical bundle) kohéziós kötöttsége vizsgálatának.

Ezeknek a szövegszervező tényezőknek a vizsgálatát feltétlenül célszerű lenne kiegészíteni a szövegek szerkezetének magasabb szintű, már a koherencia körébe tartozó elemzéseivel, főként a fogalmi sémák körébe tartozó konceptuális *keretek*nek és az ezek szövegszervező szerepét felfedő ún. *kulcsszavak*nak a tanulmányozásával. Az ilyen irányú vizsgálatok ismertetése jóval meghaladja ugyan az ilyen bevezető jellegű, útmutató funkciójú szakkönyv céljait, azonban a diskurzus alapú korpusznyelvészeti kutatások jövőbeli irányvonalaira gondolva már itt

hasznos lett volna – figyelemfelkeltés céljából – megemlíteni őket. Alkalmat adott volna rá a 27. oldalon tárgyalt lexikális sűrűség (denzitás) fogalomkörének és szövegtani funkcionális státusának bemutatása. A szövegszervező erők, különösképpen a kohéziót és a koherenciát biztosító faktorok, valamint a kontextuális bázisú ún. világismereti tényezők összefüggéseire a szerzők – vizsgálatuk szűkebb, inkább leíró, nyelvi tények megállapítására összpontosító szemlélete miatt – sajnos nem hívták fel a figyelmet, annak ellenére, hogy ez a kitérő szervesen illeszkedett volna nézeteikhez. Különösképpen már csak azért is, mert a szövegkohéziót és koherenciát illető produktív és receptív, világtudás-alapú elvárások más és más nyomot hagynak az egyes szövegtípusok felszíni szerveződésében.

Az első egység végén a fejezetekben tárgyalt ismeretek elsajátítását ellenőrző feladatsort, valamint releváns irodalomjegyzéket adtak meg a szerzők. Igen hasznos, hogy ezt a további fejezetek végén, illetve egyes fejezeteknek a belsejében is megtették.

A könyv második egysége (33–72) szintén két rövid terjedelmű fejezetet foglal magában. Az első, tehát a könyv 3. fejezete (35–57) a nyelvi korpuszokban való keresés módjaiba és technikáiba vezet be az olvasót. Elsőként a kontextusba ágyazott kulcsszavak alapján történő adatkinyerést (Keyword in Context – KWIC) tárgyalja, a neves amerikai korpuszfejlesztő nyelvésznek, Mark Davisnek a *Mai amerikai angol nyelvezet korpusza (COCA)* elnevezésű, 450 millió lexikai tételt tartalmazó hatalmas adatbázisa alapján. A korpusz három nyelvhasználati regiszterből (beszélt nyelvi, média/újságnyelvi, valamint akadémiai írásbeli szövegek) nyújt adatokat, megadva egyúttal a disztribúciójukat is. Gyakorlatias stílusban, bőséges példaanyag tárgyalásával vezetik végig az olvasót a korpuszban való keresés folyamatán, közvetlenül megadva, megfogalmazva a releváns lekérdezési módokat. Ennek a fejezetnek nem a végén, hanem az egyes részekbe integrálva szerepelnek a gyakorlatok. Kár, hogy a nyomtatás techni-

kája nem kedvezett a COCA korpuszból nyert adatközlő táblák kellően éles, olvasóbarát megjelenítésének. Az már speciális szakmai kérdés, hogy a kulcsszóvá minősítés kapcsán meg lehetett volna világítani az alapvető eltérést a fogalom számítógépes nyelvészeti és szövegtani funkcionális értelmezése között: a számítógépes nyelvészek a statisztikailag kimutatható gyakoriság alapján azonosítják a KWIC adatközlés során a kulcsszavakat (lásd Scott–Tribble 2006), míg a szövegtan a szövegszervező státussal – a sémaszervezetben mentálisan tárolt, szövegreprezentációs potenciállal – rendelkező lexikális egységeket tekinti kulcsszavaknak (Andor 1989: 32–35).

Ezután igen rövid fejezetrész foglalkozik a kollokációk kérdéskörével, egy feladatot csatolva a bemutatáshoz. Ez a fejezetrész – a korpusznyelvészetben elfoglalt státusánál fogva – lehetett volna jóval részletesebb. Szerencsés lett volna kiegészíteni az ún. kolligáció, azaz a grammatikai bázisú illeszkedés tárgyalásával, valamint a kollokációs potenciál típusaiban tartozó polaritási-kérdések, s hozzá kötődően az ún. szemantikai prozódia, azaz a kollokáció által vizsgálható pozitív vagy negatív lexikális potenciál, illetőleg az azt érintő irányváltások kurrens kérdéskörének bemutatásával, felhasználva ehhez Stewart 2010-ben publikált útmutató monográfiáját.

A fejezet következő szakasza a korpusznyelvészet körében jártas kutató vagy oktató számára jól ismert fogalmat, az ún. *n-gramok* kérdéskörét tárgyalja. Ezek a szövegegységek több tagból álló, nagyobb gyakorisággal előforduló kollokatív lexikai szerkezetek, s attól függően, hogy hány tagból állnak, beszélhetünk *unigramról*, *bigramról*, *trigramról* és így tovább, egészen a Biber által definiált és körülírt lexikai konstrukciókig, az ún. lexikai csomókig. Vizsgálatuk a ténypuszta konstatálásos túl igényelné a további, főként a konstrukciós grammatika és az ún. keret-szemantika terén végzett összehangolt szemléletű, elméleti és alkalmazott nyelvészeti integrált kutatómunkát. Az ilyen vizsgálatok természetesen messze meghaladnák

a jelen tankönyv céljait. A 3. fejezet utolsó szakasza a korpusz-szövegek szófaj alapú tagolásának (POS tags), kódolásának módszerében nyújt rövid betekintést, egy feladattal kiegészítve a leírt ismereteket.

A könyv második tematikai egységében a 4. fejezet (58–72) a nyilvánosan hozzáférhető korpuszokra alapozott projekteket ismerteti, szintén röviden, két részbe tömörítve a tárgyalatokat. Az első a lexikai alapú, a második a grammatikai alapú projektekkel foglalkozik. A fejezet az angol nyelv változatait felvonultató hat korpuszra alapozott 12 projektet mutat be, ami nem kis teljesítmény. Az olvasó számára adott feladatok itt is a fejezet anyagába ágyazódnak be. A lexikai bázisú bemutatás diakrón és szinkrón szinten egyaránt meg-alapozott. Megjegyezzük, hogy a 4.2 és 4.5 számú projektben tárgyalt adverbiumok kezelése alapvetően lexikai szinten történik, holott ezek funkcionális státusa legalább annyira grammatikai jellegű, mint lexikális. Főként akkor, ha mondatrész szintű és propozicionális státusukat (sentential vs. non-sentential adverbs) megkülönböztetve, egymástól szétválasztva elemeznék a szerzők. Ugyanakkor impozáns az, ahogyan a különféle típusú – főként regionális alapon besorolt – korpuszokat használó bemutatások ilyen rövid bevezető tankönyvben is érdekes adatokat képesek szolgáltatni az egyes régiók eltérő nyelvváltozatairól mind lexikai, mind grammatikai szinten. A grammatikai kérdések korpusz-alapú bemutatásából kiviláglik az olvasó számára, hogy a lexikalitás a nyelvtaniaknak tartott jelenségkörökre is befolyással van, tehát éppen korpusznyelvészeti eszközökkel mutatható ki pontosan a két funkcionális sík összefonódottsága a nyelvi rendszer működésében. Ennek a komplex kérdéskörnek a nyelvhasználatban való megnyilvánulásait elméleti szinten a napjainkban egyre nagyobb teret nyerő kognitív és konstrukcionalista, alapvetően funkcionalista szemléletű nyelvtudomány képes modellálni.

A harmadik tematikai egység (73–155) a korpuszfejlesztést, a korpuszadatok kvantitatív elemzését, az észszerű adatfeldolgozás

lépéseit és tényezőit tárgyalja négy rövid fejezetben. Ezek közül az első, a könyv 5. fejezete (75–93) a saját korpusz fejlesztésével foglalkozik. Sorra veszi a korpuszfejlesztés alapvető céljait az általánosaktól a speciálisakig, beleértve a szocio-pragmatikai kérdéseket is. Típusaik szerint áttekinti a korpuszfejlesztést elősegítő szoftver-forrásokat, majd az AntConc és az AntWordProfiler programok felhasználásával különféle adatelemző technikákat mutat be. A korpuszanyagok gyűjtésével kapcsolatos jogi és technikai kérdésekre is felhívja a figyelmet. Értékes része a fejezetnek a korpusz reprezentativitásának és validitásának tárgyalása, összefüggésben az adatbázis terjedelmének jellemzőivel.

A 6. fejezet (94–116) a statisztikai alpműveletekbe nyújt rendkívül didaktikus bevezetést. Minden egyes alapfogalmat példák-kal illusztrálva körülír és definiál, különös tekintettel a variancia-analízisre, a nyelvileg releváns variabilitás és megoszlási adatok tárgyalására, a statisztikai vizsgálatok előkészítésére, szempontrendszerére, lényegében az SPSS [Statistical Package for the Social Sciences] programcsomag alapján. Kár, hogy akár csak röviden, de nem veti össze ennek eredményeit más eszközök, például a korpusz-alapú elemzésre egyre szélesebb körben alkalmazott R program eredményeivel.

A 7. fejezet (117–141) a variabilitás foka, a statisztikai változók releváns mutatói felbecsülésének során alkalmazott ANOVA (ANalysis Of VAriance, varianciaanalízis) chi-négyzet próba, valamint a Pearson-féle korrelációs differenciálesztetek nyelvészeti alkalmazásával foglalkozik, részletes és szemléletes elemző bemutatások keretében, konkretizáltan ismertette az egyes tesztek használatának céljait. Ebben a fejezetben sajnos már nem olyan módszeres a szakfogalmak azonnali értelmezése és definiálása, ahogyan azt a könyv korábbi fejezeteiben megszokhattuk.

A 8. fejezet (142–155) a saját projektet képező korpuszfejlesztésre vonatkozó egység utolsó része, a korpuszelemző munka ösz-

szefoglalása. A korpusztól a felhasználásig, lépcsőzetes formában áttekinti a komplex folyamatot, a szituatív tényezőktől a releváns nyelvészeti tény-alapú analízisen át a funkcionális szemléletű értelmezésig. A fejezetet az olvasó által készítendő projekt kívánalmainak ismertetése fejezi be. A 9. fejezet (156–159) már csak egy rövid, négyoldalas zárszó arról, hogy a korpusznyelvészet jövőjének útja a komplex, regiszterre alapozott lexikális-grammatikai bázisú szövegelemzés felé vezet.

A könyv végén hasznos tárgymutató található.

Megismételve a recenzió elején kifejezett álláspontomat, William J. Crawford és Csomay Enikő gyakorlati korpusznyelvészeti bevezető könyvét kitűnő minőségű, didaktikusan elkészített munkának tartom, alkalmazását tankönyvként javaslom a hazai felsőoktatási nyelvészképzés valamennyi szintjén: a BA, MA és a PhD programokban egyaránt, de ajánlom egyúttal a nyelvi lexikális adatbázisokat, korpuszokat használni kívánó, bármely tudományos diszciplínához tartozó szakemberek számára is.

Andor József

IRODALOM

- Andor, J. (1989): Strategies, tactics and realistic methods of text analysis. In: Heydrich, W. – Neubauer, F. – Petőfi, S. J. – Sözer, E. (eds.): *Connexity and coherence: Analysis of text and discourse*. Berlin – New York: Walter de Gruyter (= *Research in text theory* 12), 28–36.
- Biber, D. – Johansson, S. – Leech, G. – Conrad, S. – Finegan, E. (1999): *Longman grammar of spoken and written English*. New York: Longman.
- Biber, D. – Conrad, S. (2009): *Register, genre, and style*. Cambridge: Cambridge University Press.
- Scott, M. – Tribble, C. (2006): *Textual patterns: Key words and corpus analysis in language education*. Amsterdam – Philadelphia: John Benjamins.
- Stewart, D. (2010): *Semantic prosody*. New York – London: Routledge.
- Tognini-Bonelli, E. (2001): *Corpus linguistics at work*. Amsterdam: John Benjamins (= *Studies in corpus linguistics* 6).
- Yves Gambier – Annamaria Caimi – Cristina Mariotti (eds.)
Subtitles and language learning. Principles, strategies and practical experiences
 [Feliratok és nyelvtanulás. Alapelvek, stratégiák és gyakorlati tapasztalatok]
 Bern: Peter Lang, 2015. 351 p.
 ISBN: 978-3-0343-1529-6
- Közkeletű vélekedés, hogy az olyan országokban, ahol a televízióban és a mozikban eredeti nyelven, feliratozva vetítik a külföldi filmeket, a polgárok jobban ismernek idegen nyelveket, mint ahol a szinkronizálást részesítik előnyben. Bár ez az általánosítás nem minden esetben állja meg a helyét (gondoljunk csak az egyaránt feliratozó finnek és portugálok nyelvtudására!), mára bizonyított tény, hogy az eredeti nyelven, felirattal nézett audiovizuális anyagok lehetővé teszik a járulékos nyelvtanulást. A feliratokban rejlő nyelvoktatási potenciál azonban messze túlmutat ezen az összefüggésen: az alábbiakban bemutatandó kötetből megtudhatjuk, hogy mind a feliratok típusait, mind ezek felhasználását tekintve számtalan lehetőség áll a nyelvtanárok és a tanulni vágyók rendelkezésére.
- A tanulmánykötet előzménye a *Subtitles and language learning* címmel 2012-ben Paviában megrendezett konferencia, amely az *audiovizuális fordítás* (azaz a fordítástudománynak a feliratozást és szinkronizálást vizsgáló részterülete) és az *idegennyelv-oktatás* közös berkeiben kibontakozó új szakterület talán eddigi legnagyobb tudományos