

**Hilde Hasselgård – Jarle Ebeling – Signe Oksefjell Ebeling**  
**(szerk.): Corpus Perspectives on Patterns of Lexis**

(Amsterdam/Philadelphia: John Benjamins Publishing Company. 2013. 307 p.)

A *Corpus Perspectives on Patterns of Lexis* egy tizenhárom tanulmányból álló konferenciakötet, amely a *Studies in Corpus Linguistics* sorozat részeként jelent meg 2013-ban. A kötetbe foglalt tanulmányok a 32. ICAME konferencia legszínvonalasabb előadásainak írott változatai. Ezek mindegyike a lexikon kisebb-nagyobb szegmenseivel foglalkozik, szavak és szó szerkezetek használati mintáit igyekeznek korpuszok segítségével feltárni és értelmezni. Eközben olyan módszertani kérdéseket vetnek fel, melyek általános érdeklődésre tarthatnak számot. Hogyan vizsgáljuk korábbi korok nyelvállapotát? Hogyan keressünk neologizmusokat? Hogyan vizsgáljunk szemantikai változásokat korpusz segítségével?

Tizenkét tanulmány az angol nyelv különféle változataival foglalkozik, egy pedig a norvégban megjelenő angol jövevényszavakkal. Tehát két olyan nyelvről olvashatunk, amely kiváló minőségű, nyílt hozzáférésű korpuszokkal bír. Itt érdemesnek tartom megjegyezni, hogy a magyar nyelvvel foglalkozó kutatásokhoz is számos korpusz áll rendelkezésre. Az általános célú MNSZ-en túl vannak diakrón kutatásra alkalmas történeti korpuszok (ÓMK, TMK, MTSZ), sőt beszélt nyelvek (pl. BUSZI) is. Ugyanakkor a magyar korpuszok esetében alapvető nehézségekkel szembesülünk, például az agglutináló jellegből adódó annotálási problémákkal. Ha az itt bemutatásra kerülő kutatásokhoz hasonlókat tervezünk magyar nyelvre, ennek a különbségnek tudatában kell lennünk.

A kötet elsősorban az angol nyelvvel foglalkozó lexikológusokat célozza meg, de úgy vélem, néhány tanulmány a nagyobb közönség számára is tanulságos. A korpusz használata mint kutatási módszer alapvető fontosságú számos nyelvészeti területen, ezért érdemes jobban megismerni a módszer részleteit.

Az **előszóban** a szerkesztők részletesen ismertetik a címben szereplő fogalmakat. A korpusz alapú perspektívákkal (*corpus perspectives*) olyan módszerekre utalnak, amelyeknél a kutató valamilyen korpuszt használ ahhoz, hogy a vizsgált szó vagy szerkezet kontextusai által kirajzolt mintákat feltárja és elemezze. Ezek nagyon sokszor olyan minták, amelyeket „szabad szemmel” vagy intuíció alapján nem lehetett volna észrevenni. A lexikon mintázatai (*patterns of lexis*) alatt pedig három dolgot értenek a szerkesztők: 1) a szavak kollokációs mintáit, 2) új szavak, szókapcsolatok keletkezését, 3) a minták lexiko-grammatikai tulajdonságait. Hangsúlyozzák, hogy a lexikon és a

grammatika szoros kapcsolatban állnak. Nem csak az figyelhető meg, hogy egyes szavak milyen más szavakat „vonzanak”, hanem az is, hogy milyen nyelvtani konstrukciókban mennyire tipikusan fordulnak elő. Fontos viszont látni, hogy ezek tendenciák, nincsenek kőbé vésve.

A kötet szerkezetét tekintve négy részre oszlik. Az első a **Sequence and order** címet viseli, amelyet magyarra talán így fordíthatnánk: sorrend és sorrendiség. A terminusok John Rupert Firth-től származnak. Ő – bár expliciten nem fogalmazta meg – *sequence* alatt a konkrét, korpuszból vett mintázatot értette, *order* alatt pedig azokat a törvényszerűségeket, amelyekre a korpusznyelvéssz az előbbiek alapján következtet. Az *order* tehát az egymáshoz hasonló *sequence*-ek halmaza feletti absztrakció.

Az első rész címéhez Michael Stubbs korpusznyelvéssz cikke adta az ötletet. Ő egyrészt újragondolja a Firth-től származtatott két fogalmat, másrészt megkérdőjelezi a korpuszalapú és korpuszvezérelt módszerek közti különbségtétel létjogosultságát. Cikkét szemléletformálónak tartom, és a későbbiekben részletesen ismertetem majd.

Doris Dant azt vizsgálja, hogy a két főnévi tagból álló frázisokban (pl. *mom and dad* ’anyu és apu’, *men and women* ’férfiak és nők’) a tagok sorrendje mennyire és milyen tényezők hatására cserélhető fel az amerikai angolban. Megerősíti korábbi tanulmányok megállapításait, miszerint a szótagszámnak befolyása van a sorrendre: a kevesebb szótagból álló, rövidebb főnév rendszerint előrébb kerül. A szerző megvizsgál egy szemantikai megszorítást is: igaz-e, hogy amikor férfira és nőre utaló főnevek állnak párban, az „erősebb referens” (általában a férfi) áll előbb? A korpuszadatok azt mutatják, hogy bár ez is adathozható, mégis meghatározóbbnak tűnik a szótagszám szerepe. Felmerül egy másik szemantikai faktor is: a családi viszonyt kifejező szavaké. Úgy tűnik, a családtagokra utaló főnév + főnév párok esetében a női referens áll az első helyen.

Stephen James Coffey a brit angolban előforduló, kvázi-szinonim jelzői mellékneveket vizsgálja (pl. *great big* ’hatalmas nagy’, *tiny little* ’iciri-piciri’). Azt találja, hogy ezek a melléknevek a beszélt nyelvben és a fiktív párbeszédekben a leggyakoribbak, tehát egyértelműen az informális regiszterhez köthetők. Ezek a szekvenciák változatosságot mutatnak a sorrendi kötöttség tekintetében. Például a már említett *great big* kifejezésben nem cserélhető fel az elemek sorrendje, míg a *tiny little* és *little tiny* egyaránt előfordulnak.

Ezt a három cikket a **Competing Constructions** ’Versengő konstrukciók’ rész tanulmányai követik. A közös bennük az, hogy mindegyik több, hasonló jelentésű, rivalizáló nyelvi szerkezetet vizsgál. A kérdés jellemzően az, hogy

milyen tényezők befolyásolják a versengés kimenetelét (tehát azt, hogy melyik konstrukciót produkálja a beszélő).

Zeltia Blanco-Suárez egy diakrón kutatást ismertet, amelynek célja kimutatni, hogy a melléknév előtti módosítói pozícióban a *dead* vagy *deadly* 'halálos, haláli' szóalakot használták-e inkább különböző nyelvtörténeti korokban (a késő 15. századtól napjainkig). A tendenciák azt mutatják, hogy kezdetben objektív, szó szerinti jelentésben szerepeltek ezek a szavak, és idővel egyre szubjektívebb, intenzifikáló jelentést nyertek. Ez egy grammatikalizációs folyamat, amely a *dead* esetében jóval előrehaladottnak tűnik, mint a *deadly*-nél. Ugyanakkor az is látszik, hogy a választás gyakran a rákövetkező melléknévtől függ. Bár a két szó közel szinonim, mégis eltérő kollokációs preferenciáik vannak.

Ingo Bachmann tanulmánya szintén diakrón jellegű. A *go* 'megy' + *ige* és a *go and* 'megy és' + *ige* konstrukciók gyakoriságát, eloszlási mintáit vizsgálja egy amerikai angol történeti korpuszban. Többek között azt mutatja ki, hogy a 19. században még az utóbbi szerkezet volt a gyakoribb, napjainkban viszont épp fordítottak az arányok. A *go* + *ige* minta legkorábban felszólító mondatokban jelent meg, majd infinitívuszi és jelen idejű formákban is elterjedt.

Günter Rohdenburg tanulmánya is igei frázisokkal foglalkozik: a *cannot help -ing* 'nem tehet mást, mint...' szerkezetet és ennek riválisait veszi górcső alá. Diakrón, de nemcsak különböző nyelvtörténeti korokat vet össze (kezdve a 17. századdal), hanem összehasonlítja a brit és az amerikai angolt is. Figyelembe veszi továbbá a szöveg regiszterét, mint a szerkezetek közti választás egyik befolyásoló tényezőjét.

David Lorenz kutatásának alapját egy beszélt nyelvi, amerikai angol korpusz, a Santa Barbara Corpus képezi. A kiinduló hipotézise a következő: azáltal, hogy a *gonna* 'fog valamit csinálni' (a *going to* redukált formája) gyakorisága nőtt, függetlenedett is az eredeti formától. A cikk alaposan számba veszi a relevánsnak tűnő faktorokat, amelyek meghatározhatják a formaválasztást. Ilyenek a beszélő kora, iskolázottsága, a régió, ahol él, a beszédtempója, a megelőző szó és a modalitás típusa. A legfontosabb faktornak a beszédtempó és a beszélő kora bizonyul.

Adam Smith a prepozíció + főnév + prepozíció szekvenciák (pl. *in [the] light of* 'valaminek a fényében') eloszlási mintáit vizsgálja, valamint azt, hogy a főnév határozottságának milyen hatása van a frázis jelentésére. Végül arra is kitér, hogy megfigyelhetők-e lényeges különbségek a brit és amerikai angol között ezeknek a frázisoknak a használatában.

Ezt a hosszú részt az **Emerging Patterns** 'Kiemelkedő mintázatok' tanulmányai követik. Ezek abból a fontos tényből indulnak ki, hogy bizonyos szavaknak, mondattani szerkezeteknek, morfológiai jelenségeknek életciklusai

vannak. Létrejönnek, egyre gyakoribbá válik a használatuk, majd elkezdenek kikopni a használatból, és végül (sok esetben) megszűnnek. A tanulmányok ennek az életciklusnak egészét vagy részeit vizsgálják.

Az előbbi esetre példa Antoinette Renouf tanulmánya. Egy diakrón, újságcikkeket tartalmazó korpuszon különböző esettanulmányokat végez, amelyek lexikai és szemantikai neologizmusok életciklusaiból mutatnak be egy-egy epizódot. Ezek a következők: létrejövés, a gyakoriság növekedése, ortográfiai szabályozás, produktivitás, kreatív használat, megszilárdulás, elavulás, eltűnés. A ciklikusság leginkább a szemantikai neologizmusoknál ragadható meg. Amikor egy régi szót új jelentéssel ruházunk fel, az életciklusa előről kezdődik, és jellemzően új kollokációs mintákat is felvesz.

Anne-Line Graedler a neologizmusoknak egy sajátos esetét vizsgálja. A kutatási kérdése: Milyen grammatikai integráción esnek át az angol főnevek, amelyek átkerülnek a norvégba – egy olyan nyelvbe, amely nyelvtani nemeket különböztet meg? A szerző egy norvég újságorpuszból származó, 950 szóból álló listát vizsgál meg aszerint, hogy milyen rendszer, analógia-sor alapján kapnak nyelvtani nemet az egyes kölcsönszavak.

Stefan Diemer tanulmánya az angol neologizmusok egy meglehetősen új formáját vizsgálja: az internetes, rövidsége törekvő írásokban megjelenő, prefixummal ellátott igéket (pl. *stand at* → *atstand* 'áll valahol', *go on* → *ongo* 'továbbmegy'). Fontos megjegyezni, hogy a vizsgált anyag tartalmaz olyan szövegeket is, amelyeket nem angol anyanyelvű emberek írtak. Tehát ennél a jelenségnél a nyelvi kreativitás és a rövidsége való törekvés mellett befolyásoló tényező lehet az idegen nyelvi hatás is.

A kötet utolsó része a **Correlating Patterns and Meaning** 'Mintázatok és jelentés összefüggésbe hozása' címet viseli. Az itt olvasható két tanulmány azt a kérdést járja körül, amelyet talán a legnehezebb megválaszolni a korpusznyelvészetben: hogyan vizsgálhatjuk a jelentést, jelentés és forma összefüggését korpuszok segítségével?

Benet Vincent tanulmánya az ige + *wh*-kérdőszó szekvenciák és a modalitás viszonyát vizsgálja a British National Corpus szövegein. A szerző négy fő modális jelentést talál a *wh*-kérdőszóval bevezetett tagmondatok esetében. Ezek a cél (*purpose*), a nehézség (*difficulty*), a deontikus jelentés (*deontic meaning*) és a szándék (*intention*). A felsorolt modális jelentésekhez tipikus frázis-sémák rendelhetők. Vincent munkája szemaziológiai kutatás abban az értelemben, hogy egy adott formához kapcsolható fogalmakat vizsgál.

A megközelítését tekintve ennek épp az ellentéte Diane Goossens tanulmánya, amely onomaziológiai kutatás, tehát egy fogalom különböző formáit, megnevezéseit keresi. A cikk a „mennyiségi becslés” (*quantity*

*approximation*) jelentésű kifejezések automatikus kinyerését tűzte ki célul. Ehhez két módszert használ: egy morfoszintaktikai, valamint egy szemantikai címkézésen alapulót. Ezeket ki is értékeli úgy, hogy egy manuálisan annotált korpusszal veti össze a saját automatikus megoldásainak a teljesítményét. Ezek lényegesen jobbnak bizonyulnak, mint az üzleti témájú korpuszokból egyszerű gyakorisági listaként kinyert frázisok.

Az alábbiakban részletesen bemutatom Michael Stubbs tanulmányát. Úgy vélem, a nagyközönség számára ez a munka lehet a legérdekesebb, de korpusznyelvészeknek is ajánlom, mert szemléletformáló cikkről van szó. A címe *Sequence and order: The neo-Firthian tradition of corpus semantics* 'Sorrend és sorrendiség: a korpuszalapú szemantika neo-Firth-i hagyománya'.

A szerző azzal kezdi az absztraktot, hogy felhívja a figyelmet a korpusznyelvészet egyik legfontosabb törekvésére: arra, hogy ez az irányzat alapvetően el akarja kerülni a „korpusz előtti időkre” jellemző feltételezéseket. A korpusznyelvészet módszere induktívnak tekinthető, mivel konkrét megfigyelésekből vonja le a következtetéseit, ezek alapján általánosít, absztrahál. Stubbs ugyanakkor azt is hangsúlyozza, hogy a tiszta indukciót már több mint négyszáz éve megkérdőjelezzük, az elmélettől mentes indukció lehetőségét pedig (legalábbis a társadalomtudományokban) általánosan elutasítják. Így Sinclair 1990-es években létrehozott frazeológiai modellje nem nevezhető nyers korpuszadatokra épülő, tisztán induktív általánosításnak, bármennyire is törekszik erre. Stubbs ezt a modellt vizsgálja felül egy kis esettanulmány segítségével. Ezzel párhuzamosan olyan alapvető korpusznyelvészeti fogalmakat gondol újra, mint a *sequence* (sorrend), *order* (sorrendiség), valamint a *corpus-driven* (korpuszvezérelt) és *corpus-based* (korpuszalapú) módszerek közti különbségtétel.

A bevezető részben megismerjük a címben is szereplő „neo-Firth-i hagyományt”. Ez a kifejezés John Rupert Firth nyelvészre, valamint tanítványaira és követőire utal. Az irányzat története az 1950-es évek közepéig nyúlik vissza, nagyjából arra az időszakra, amikor megjelentek az első digitális korpuszok. A kulcsfogalom a jelentés, a cél pedig a szavak, frázisok jelentésének feltárása korpusz segítségével. Stubbs itt rögtön felhívja a figyelmet arra, hogy az emberi nyelvekben a jelentés a társadalom által meghatározott dolog, nem pedig a szavak megmásíthatatlan, inherens tulajdonsága. Ezzel igyekszik az olvasó fejében elültetni a gondolatot, hogy sok nyelvészeti diszciplínában valóban nem lehet olyan cáfolhatatlanul eljárni, következtetni, mint a természettudományokban.



A kis kitérő után, a második fejezetben visszatér a neo-Firth-i hagyományra, és annak két kulcsfogalmára, a már sokat emlegetett *sequence* és *order* terminusokra. Firth az 1951-es *Modes of meaning* című munkájában a *sequence* alatt a megfigyelhető, szöveges adatot érti, *order* alatt pedig az elméleti szemantikai modelleket (és erről többet nem is igazán mond). Halliday (1961) és Palmer (1964) már alaposabban körüljárják a két fogalmat. A *sequence* a konkrét adat, amely lineáris – ha írott szöveg, akkor a szavak térben követik egymást, ha hangzó szöveg, akkor időben. Ez az, amely megfigyelhető, kimérhető. Az *order* absztrakt, elméleti konstrukció, amely nem figyelhető meg közvetlenül, több dimenziós, és a *sequence*-minták interpretációja által jön létre.

Stubbs szerint a korpusznyelvész két dolgot tud ténylegesen megfigyelni: a gyakoriságot és az eloszlást. Fontos, hogy a következtetésekhez minden esetben szükség van mindkét típusú adatra. Teljesen máshogy kezelendő adat az, ha valami nagyon gyakori és egyenletes az eloszlása a korpuszban (tipikusan ilyenek a funkciószavak: névelők, prepozíciók, stb.), mint az olyan esetek, amikor valamilyen szó, kifejezés gyakori, de csak a korpusz egy kis szeletében. Ha pedig a jelentés modellálása a cél, a korpuszban a variációt és a hasonlóságot ragadhatjuk meg egy-egy vizsgált jelenség kapcsán.

A következő fejezetben pontosan ezt kísérli meg a szerző: a *go and VERB* 'megy és csinál valamit' sémára illeszkedő szekvenciákat vizsgálja meg, és ezeknek a frazális egységeknek a jelentését próbálja azonosítani. Ezt alapvetően Sinclair (1996, 1999) szellemében, azt továbbgondolva teszi. Sinclair – valamint (Teubert, 2010) is – azt hangsúlyozza, hogy a parafrázisok és az intertextus vizsgálata együttesen ad ki egy olyan technikát, amely a szemantikai elemzést lehetővé teszi. Ezek a jelentés tanulmányozásának analitikus eszközei. A jelentés bizonyítékát két tényező adja: a) az intertextusban azonosítjuk az ismétlődő szakaszokat, amelyek formailag azonosak vagy nagyon hasonlóak, valamint b) a közvetlen ko-textust vizsgáljuk, és azonosítjuk az együttesen előforduló szegmentumokat, amelyek szemantikailag hasonlóak.

Mielőtt a szerző bemutatná az elemzését, egy teljes fejezet erejéig foglalkozik a korpusznyelvészeti módszereivel és az indukció problémakörével. Ez első olvasásnál nagy kitérőnek tűnt, valójában viszont a tanulmány legfontosabb részéről van szó, és valóban szükség van rá ahhoz, hogy a szerző későbbi elemzését jól meg lehessen érteni.

Francis (1993) óta kétféle módszert szokás megkülönböztetni a korpuszvizsgálatokban. Az egyik a korpuszalapú, amellyel egy kialakult elméletet tesztelünk, és az adatok alátámasztják vagy cáfolják majd az elméletet. A másik a korpuszvezérelt, amely kifejezetten azt hangsúlyozza, hogy előzetes elmélet nélkül kezdjük a kutatást, introspekcióval nem feltárható jelenségek

kimutatása a cél, és tisztán az adatból kiindulva, indukció által jutunk eredményre.

A szerző úgy véli, hogy a korpuszvezérelt módszer esetében túlhangsúlyozzuk az indukciót (és jellemzően objektívebbnek tartjuk a korpuszalapúnál). Viszont egy jelenség vizsgálatánál tulajdonképpen sosem a semmiből indulunk ki. Van egy kutatási kérdés, amelyet körüljárunk, tesztelünk – ez pedig sok tekintetben inkább deduktív.

Stubbs szerint tehát eltúlzott a különbségtétel a kétféle módszer között, mivel itt nem is annyira módszertani kérdésről van szó, mint inkább akadémiai politikáról. A korpuszalapúság folytonosságot hangsúlyoz a korábbi munkákkal, a korpuszvezéreltség pedig egyfajta paradigmaváltást akar képviselni. Stubbs azt javasolja, hogy hagyjuk is el ezeket, és korpuszvezéreltség helyett a megfigyelhető *sequence*, korpuszalapúság helyett az elméleti *order* legyen csak megkülönböztetve. Hozzáteszi, hogy tapasztalata szerint a kutatást inspiráló ötletek rendszerint amúgy is különböző (elméleti és empirikus) források keveredéséből származnak.

Ebben a fejezetben felteszi a kérdést: Végül is mit csinálnak a korpusznyelvészek? Tényleges választ nem ad rá, ehelyett érdekes párhuzammal sugallja a megfejtést az olvasónak. Egy ábrát mutat (Darwin, 1845), amelyen a Galapagos-szigetvilág néhány pintyfaja látható. A madarakat egységesen profilból, balra néző fejjel ábrázolták, így a hasonlóságok és különbségek is láthatók. Stubbs ezután mutat egy korpuszból kinyert konkordancia listát, ahol a vizsgált szó egyes előfordulásai szerepelnek a középső oszlopban, sorban egymás alatt. A különféle bal- és jobboldali szövegrészek ehhez igazítva jelennek meg. Zárásképpen azt írja a szerző, hogy a korpuszadatokban a minták felfedezése három tényezőtől függ: a megfigyelő nézőpontjától, érdeklődésétől és tapasztalatától. Úgy vélem, ez egy nagyon jó (és a kutatók részéről ritkán bevallott) észrevétel.

A tanulmány végén a szerző bemutatja a *go and VERB* sémához adott elemzését, amely lényegében Sinclair modelljébe illeszkedik, kisebb módosításokkal. Úgy véli, a neo-Firth-i hagyományon belül Sinclair volt a legeredményesebb abban, hogy a jelentés-egységeket modellezze. A modellnek négy paramétere van: az első és a második a lexikai és grammatikai formát, a harmadik és negyedik a témát és a beszédaktust érinti.

A *go and VERB*-bel kapcsolatos adatok beleilleszthetők ebbe a modellbe (Stubbs részletes elemzést ad, én most csak a lényegyet emelem ki). A fentebb említett négy paraméter a következőképpen alakul:

1. A kollokációk szintje. Ez lényegében szavak lineáris egymásutánja, a nyers adat gyakorisági vizsgálatát jelenti. A *go and VERB* séma esetében nincsenek erős kollokációk.

2. A kolligációk szintje. Ez nyelvtani preferenciák vizsgálatát jelenti, vagyis absztraktabb, mint az első szint. A vizsgált esetben azt látni, hogy a *go and VERB* jellemzően valamilyen diskurzusjelölő (pl. *and* 'és', *but* 'de', *so* 'így', *then* 'akkor') után áll, és rendszerint múlt időben jelenik meg.

3. A téma szintje. Ez az adott szerkezet szemantikai preferenciáit érinti, és eléggé absztrakt, mivel intuitív értést, szövegértelmezést igényel. A *go and VERB* nagyon leegyszerűsítve akkor használatos, amikor a beszélő olyasmiről beszél, ami jól kezdődött, aztán rosszul folytatódott. Tipikus a *went and spoiled everything* 'fogta magát és mindent tönkretett' szekvencia.

4. A beszédaktus (szemantikai prozódia) szintje. Ekkor a kutató értelmezi a beszélő szándékát: milyen céllal tette az adott kijelentést, milyen az attitűdje, stb. Tehát a konkrétól elértünk a legabsztraktabb szintig. A *go and VERB* esetében Stubbs a következőt mondja: akkor használt, amikor a beszélő meglepetésnek vagy bosszúságnak ad hangot valaki viselkedése miatt.

A modell tehát végeredményben az *order*-rel, és nem a *sequence*-szel foglalkozik. Nem induktív általánosítás megfigyelt tényekből, mivel a tulajdonságait deduktív módon bontja ki a szerző. Azt vizsgálja, hogy ez mennyire hangolható össze más jelenségekkel, illetve megmagyarázhat-e más jelenségeket. Stubbs azt találja, hogy a modell nagyon sok más szemantikai munkában is jól alkalmazható. A tanulmány egyébként rengeteg hivatkozást tartalmaz, a szerzőnek széles rálátása van a tárgyalt problémákra. A tudományos stílust néha üdítően személyes hangvétel váltja fel, amely segít abban, hogy közérthető legyen a cikk.

A kötet tanulmányairól összességében elmondható, hogy logikus felépítésűek, jól követhetők. Mindegyik bővelkedik a megértést segítő ábrákban, táblázatokban. A kötet megjelenése olvasóbarát, a formai kivitelezése követi a *Studies in Corpus Linguistics* sorozatban megszokott megoldásokat. A kötetet olyan nyelvészeknek – és elszánt nyelvészhallgatóknak – ajánlom, akik korpuszon alapuló kutatást terveznek, vagy folyamatban lévő kutatásuk módszertani problémáira keresnek inspiráló megoldási lehetőséget. Ez a kötet nemcsak a lexikon vizsgálatához, hanem morfológiai, szintaktikai jelenségek kvantitatív elemzéséhez is jó ötleteket adhat.

## Irodalom

**Darwin, Ch. R.** (1845) *Journal of Researches into the Natural History and Geology of the Countries Visited during the Voyage of H.M.S. Beagle round the World, under the Command of Capt. Fitz Roy, R.N.* 2. kiadás. London: John Murray.



- Firth, J. R.** (1951) *Modes of meaning*. Indianapolis: Bobbs-Merrill.
- Francis, G.** (1993) A corpus-driven approach to grammar: Principles, methods and examples. In: Baker, M., Francis, G. & Tognini-Bonelli, E. (ed.): *Text and Technology*. Amsterdam: John Benjamins. pp. 137–156.
- Halliday, M. A. K.** (1961) Categories of the theory of grammar. *Word* 17/3. pp. 241–292.
- Palmer, F. R.** (1964) 'Sequence' and 'order'. In: Stuart, C. I. J. M. (ed.): *Report of the Fifteenth Annual Round Table Meeting of Linguistics and Language Studies*. Washington DC: Georgetown University Press. 123–130.
- Sinclair, J.** (1996) The search for units of meaning. *Textus* 9/1. pp. 75–106.
- Sinclair, J.** (1999) The lexical item. In: Weigand, E. (ed.): *Contrastive Lexical Semantics*. Amsterdam: John Benjamins. 1–24.
- Teubert, W.** (2010) *Meaning, Discourse and Society*. Cambridge: CUP.

**BUSZI** = Budapesti Szociolingvisztikai Interjú <http://buszi.nytud.hu/a-buszi-rol>

**ÓMK** = Ómagyar Korpusz: <http://omagyarkorpusz.nytud.hu/>

**MNSZ** = Magyar Nemzeti Szövegtár: <http://clara.nytud.hu/mnsz2-dev/>

**MTSZ** = Magyar Történelmi Szövegtár: [http://clara.nytud.hu/mtsz/run.cgi/first\\_form](http://clara.nytud.hu/mtsz/run.cgi/first_form)

**TMK** = Történelmi Magánéleti Korpusz <http://tmk.nytud.hu/>

KALIVODA ÁGNES

Pázmány Péter Katolikus Egyetem, PhD hallgató

MTA-NYTI Nyelvtechnológiai Osztály, segédmunkatárs

MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoport, segédmunkatárs

[kalivoda.agnes@mta.nytud.hu](mailto:kalivoda.agnes@mta.nytud.hu)