

ABKAROVITS ENDRE

ÚJ IRÁNYZATOK A SZÓTÁRÍRÁSBAN ANGLIÁBAN

Abstract: The paper gives first a short survey of monolingual dictionaries in England after the war. Hornby's Advanced Learner's Dictionary dominated the market for a long period. The first real rival was Longman's dictionary in 1978. A fundamental change in the approach to lexicography was brought about by Collins COBUILD dictionaries, which were based on their own corpus containing 20 million data at the time of the first edition in 1987, and 200 million when the new edition was published in 1995. The database enables lexicographers to introduce new features like the frequency bands of the new COBUILD dictionary. The paper also tries to illustrate the usefulness of computer-assisted research through the example of a group of verbs of emotion and compares how they are described by various reference books.

Előzmények

Angliában a háború óta számos egynyelvű szótár jelent meg, részint az anyanyelvűek, részint az angolt idegen nyelvként tanulók (learners' dictionary) számára. Ez utóbbi fajtáról kívánok először egy rövid történeti áttekintést adni, megjegyezve, hogy tapasztalatom szerint ezt a fajtát is sűrűn forgatják angol anyanyelvűek, természetesen elsősorban nyelvtanárok.

Az egynyelvű, az angolt mint idegen nyelvet tanulók részére készített szótárak sorát A. S. Hornby szótára nyitotta meg 1948-ban, „A Learner's Dictionary of Current English” címmel, mely később számos kiadásban és utánnomásban „Oxford Advanced Learner's Dictionary” címmel szinte

egyeduralkodó volt a piacon. Természetesen más szótárak is léteztek, de a „Hornby” fogalom volt.

Ezek a szótárak a haladó szinten tanulókat célozták meg, hiszen a szótár használata már feltételezett egy bizonyos szókincset, s a nyelvtani fogalmak ismeretének egy megfelelő fokát. Készítői számoltak azzal, hogy a nyelvet tanuló addigra a kétnyelvű szótárak segítségével a szótárhasználat területén egyfajta jártasságra tett szert. Kezdetből próbáltak a kétnyelvű szótárakhoz képest valami többletet nyújtani, olyan gazdag információkat adni, melyek még a legkitűnőbb szótárakban (melyek közé bizonyára odasorolható Ország László szótára is) sem voltak megtalálhatók.

Angliában több kiadó között indult meg a rivalizálás ezen a téren, melyet minden bizonnyal az angol nyelvet világszerte tanulók és beszélők hatalmas száma által teremtett piac tett lehetővé. Sokáig az A. S. Hornby neve által fémjelzett oxfordi szótár járt az élen, melyet a szerző saját leíró nyelvtani rendszerére épített. Ebből a későbbiekben hagyomány lett. Szinte minden kiadó a saját nyelvészei által használt vagy kitalált terminológiát alkalmazta, vagy legalábbis igyekezett ahhoz közelíteni. Így történt ez a Longman szótárak esetében is, ahol Randolph Quirk és szerzőtársainak nyelvészeti befolyása érvényesült, de így van ez a Collins COBUILD szótárainál is, akik a szótárírással párhuzamosan saját nyelvészeikkel nyelvtankönyveket is írtak. Természetesen az eltérések mellett minden szótáríró törekedett arra, hogy az általánosan elfogadott nyelvtani elnevezések közül minél többet beépítsen, hogy ezáltal a szótár használójának munkáját megkönnyítse. Ezért aztán sokszor vitatható félmegoldások születtek. Bár például a szófajok területén, többek elnevezése körül nem lehet bonyodalom, de olyan egyszerű szavak, mint „my”, „your” stb. esetében sem lehetünk biztosak, hogy egy adott szótár éppen determinánsnak, birtokos névmásnak vagy melléknévnek fogja-e hívni őket. (Néha még ugyanannál a kiadónál is más az elnevezés a számítógépes adatbázis kódolásában és a szótárak terminológiájában. Például a fenti szavak a COBUILD kódrendszerében (tagging) melléknévként, a szótárakban determinánsként szerepelnek.)

Az egynyelvű szótárak egyik fő célja, hogy a címszavakat a lehető legegyszerűbben magyarázzák meg. Az oxfordi szótár első komoly vetélytársaként az 1978-ban először megjelent Longman Dictionary of Contemporary English például arra törekedett, hogy 2000 alapszó segítségével magyarázza meg a többi szót, s ha mégis bele kellett venni a jelentésmeghatározásba más szót, azt eltérő betűtípussal nyomták. Alapelve volt, hogy egy szó jelentését mindig nálánál gyakoribb szavakkal magyarázzák meg.

A másik fő jellemzőjük ezeknek a szótáraknak a szavak nyelvtani viselkedéséről adott információ nyújtása saját kódrendszer segítségével. Ez az első Hornby szótárak esetében, de még a Longman-szótár első kiadásai-

ban is, szinte megjegygezhetetlen jelrendszert jelentett. A használónak állandóan a rövidítéseket tartalmazó táblázatokhoz vagy a bevezetőhöz kellett lapozni, hogy kitalálja, hogy például egy ige után milyen lehetséges igeveves szerkezetek vagy mellékmondatok következhetnek. Bár a Longman-szótár már egyszerűbb jeleket használt, a rövidítés nem juttatta a tanuló eszébe automatikusan, hogy milyen szó helyett áll. Különösen az egy betűből és egy számból álló kombinációk voltak nehezen felismerhetőek, de már volt egy törekvés arra, hogy a betű lehetőleg annak a szónak a kezdete legyen, amit jelölt. (Például: „T” = transitive verb.) Igazi fejlődést ebben a tekintetben csak az oxfordi szótár új kiadása jelentett 1989-ben, amikor A. S. Hornby halála után az új szerkesztő, Tony Cowie, újfajta, úgynevezett „transparent” rövidítéseket vezetett be. Erről az akkor nagy újdonságról – még a szótár megjelenése előtt – a kiadóval akkor folytatott levelezésem alapján, talán először számolhattam be hazai konferencián. Az akkor még nagy lépésnek számított, hogy egész nyelvtani szerkezetek rövidítése vált megjegyyezhetővé azáltal, hogy a kezdőbetűk összerakásából keletkeztek a kódok. (Például: „Tsg” = transitive verb + 's genitive + gerund.) Tulajdonképpen abban az időben talán ezekre a nyelvtani struktúrákra vonatkozó információk számítottak az egynyelvű szótárak egyik legfőbb erényének, túl azon, hogy anyanyelvűek írták, hiteles(nek tűnő) példákkal szolgáltak, s még a szömeghatározások is megerősítették a korábbi nyelvtudást.

Persze az egynyelvű szótárak hasznossága mindig is függött attól, hogy a szóhasználó mennyire volt kész a szótárban található mindenféle információ befogadására. Ezzel kapcsolatban nem kedvezőek a tapasztalataim. Nagyon sok tanítványom volt, aki ugyan már a gimnáziumban megvette az oxfordi szótárt, de soha el nem olvasta a bevezetőjét, soha nem figyelte sem a nyelvtani, sem más vonatkozásban (stílus, regiszter stb.) használt rövidítések jelentését. Legtöbbször kizárólag a szó jelentésére voltak kíváncsiak, már a kiejtésre sem figyeltek az esetek jó részében, a további információkat pedig végképp nem keresték.

Pedig ezek az információk egyre szaporodnak, de tartok attól, hogy az ebben az évben megjelent új Collins COBUILD szótár 30 oldalas bevezetőjét csak kevesen fogják elolvasni, enélkül pedig a szótár teljes gazdagsága nem tárul fel előttünk. Ezt a veszélyt feltehetően a szótárírók már korábban is érzékelték, mert a Longman-szótár például kétféle ismertetőt is adott: egy rövidebbet az elején és egy részletesebbet a végén. Feltehetően azért, hogy a kevésbé igényesek legalább a rövidebbet elolvassák, míg a nyelv iránt mélyebben érdeklődők esetében talán a hosszabb leírás áttanulmányozása is remélhető volt.

S bár szinte reménytelen feladatra vállalkoztak a szótárkészítők, hiszen a legkülönbözőbb anyanyelvi háttérrel, a legváltozatosabb nyelvtani ismeretekkel, szótárhasználattal kapcsolatos felkészültséggel rendelkező

nyelvtanulók számára készítették ezeket a műveiket, mégis hallatlanul népszerűek lettek ezek a szótárak, világszerte elterjedtek. Minden valószínűség szerint más nyelvekhez képest rendkívül gazdag választék alakult ki, egyre javuló minőséggel, egyre több információval. A német nyelv vonatkozásában például sokáig nem lehetett hasonlóan jó szótárt találni. Újabban a Langenscheidts Grosswörterbuch Deutsch als Fremdsprache közelíti meg talán az angol szótárakat, mely nem véletlenül – más szótárirók tapasztalata mellett – elsősorban a Longman-szótár modelljét követi. (Bár ahogy Howard Jackson tanulmányában kimutatja, egyes vonatkozásokban attól nehezebben kezelhető, más területeken viszont még gazdagabb információt ad.)

A számítógépes adatbázisra épülő szótárak megjelenése

Forradalmian új megközelítést jelentett a Collins COBUILD szótárának megjelenése 1987-ben, mely John Sinclair professzor szellemi irányításával jött létre, akkor még egy 20 milliós adatbázisra épülve.

A korábbi szótárak szerzői saját intuícióikra hagytak, többnyire csak saját nyelvérzékük alapján döntöttek el, hogy például egy nyelvtani struktúra helyes, elfogadható-e. Két dolog azonban az, hogy valami nyelvtanilag helyes, elfogadható, és az, hogy az angolul beszélők ténylegesen használják-e. Már korábban más nyelvkönyvekben is felmerült az a probléma, hogy a szerző valamiről azt állította, hogy az helyes alak, csak éppen az angolok nem használják. (Például Alexander „Living English Structure”-ében.) Ezek a nyelvészek ilyenkor természetesen csak saját tapasztalatukra, emlékezetükre, nyelvérzékükre támaszkodhattak, nem állt rendelkezésükre olyan hatalmas adatbázis, amelyből megbízható következtetéseket vonhattak volna le. Sokszor talán az egyéni érdeklődésükön múlt, hogy egy-egy nyelvtani jelenségről mennyire részletes információt adtak. Számomra például többször csalódást okoztak az igenevek leírásában a Longman kiadónál megjelent, Randolph Quirk és szerzőtársai által írt és a jelenleg forgalomban lévő legterjedelmesebb nyelvtankönyvek („A Comprehensive Grammar of the English Language” (1985) és „A Grammar of Contemporary English” (1972)), míg az ugyanennél a kiadónál megjelent, L. G. Alexander által írt jóval kisebb nyelvtankönyv ugyanezekről a dolgokról sokkal részletesebb információt kínál.

Hasonló a helyzet a szótárakkal. Jóllehet feltételezhetnénk, hogy egy szótárból több információt kapunk egy szó nyelvtani környezetére vonatkozóan, mint egy rövidebb nyelvtankönyvben vagy egy vékonyka szótárban (BBI), ez távolról sem biztos, hogy így van.

Ezt illusztrálандó a következő táblázattal szeretném bemutatni, hogy néhány érzelmet kifejező ige után milyen igeneves szerkezetek használatát

mutatják a leggyakrabban használt szótárak és nyelvtankönyvek. Négy pozitív érzelmet (vagy preferenciát) és négy negatív érzelmet kifejező igét választottam ki erre a célra.

	A	B	C	D	E	F
love	1,2,3	1,2	1,2,3,4,5	1,2,3	1, 2, 3	1,2,3,4,5
ike	1,2,3,4	1,2,3,4	1,2,3,4,5	1,2,3	1, 2, 3	1,2,3,4,5
enjoy	2	2	2	2,5	2	2,5
prefer	1,2,3	1,2,3	1,2,3	1,2,3,5	1, 2, 3	1, 2, 3
hate	1,2,3,4	1,2	1,2,3,4,5	1,2,3,4,5	1, 2, 3, 4	1,2,3,4,5
dislike	–	2	2,4,5	2,5	2	2,4,5
loathe	2	2	2	2	2	2
detest	2	2	2,4,5	2,5	(2)	2,4,5

Az ige után következhet: 1 = főnévi igenév, 2 = gerundium, 3 = tárgy + főnévi igenév 'to'-val, 4 = tárgy + gerundium, 5 = birtokos eset + gerundium

A vizsgált könyvek: A = Collins COBUILD English Dictionary, (1995); B = Collins COBUILD English Grammar, (1990); C = Oxford Advanced Learner's Dictionary, (1989); D = The BBI Combinatory Dictionary, (1986); E = Longman Dictionary of Contemporary English, (1990); F = Longman English Grammar (1988).


A Collins COBUILD szótár 1995-ös kiadásának legfőbb erénye, hogy adatbázisa tízszer nagyobb, mint az első kiadásé. Ez természetesen sokkal biztonságosabb következtetések levonására ad lehetőséget. Megőrizték a szótár néhány olyan jellegzetességét, mely már az első kiadást is megkülönböztette más szótáraktól. A szerzők továbbra is egész mondatos jelentésmeghatározásokat használnak, szemben a legtöbb szótár helykímélő, rövidebb definícióival. A szótár továbbra is csak a mai angol nyelvet tükrözi, az utóbbi 10–15 év adataiból felépített korpuszra támaszkodik. Példáit nem a szótárírók találták ki, hanem az adatbázisban tárolt rengeteg példa közül tudatosan választották ki a minden szempontból legtipikusabbakat. Még az első kiadás legtöbb példáját is frissekre cserélték. Megőrizte a szótár azt a sajátosságát is, hogy a nyelvtanra, szóhasználatra vonatkozó információit egy külön oszlopban („extra column”) hozza. Ugyanakkor sajnos néhány olyan jellegzetességet is megőrzött, amit már korábban sem fogadtak használói tetszéssel. Ilyen az a tulajdonsága, hogy a szavak jelentését nem a hagyományos módon, szófajonkénti csoportosításban tünteti fel, hanem a különböző szófajú jelentéseket keverve. Az új kiadásban csak néhány szó esetében vezették be a 'superheadword' fogalmát, amellyel első-sorban alakilag egybeeső, de egymással például etimológiailag kapcsolatban nem levő szavak jelentéseit próbálják szétválasztani.

A szótár talán legnagyobb újdonsága a szavak gyakoriságára vonatkozó információ („frequency band”). Ez egy ötfokozatú skála, mely a besatírozott „gyémántok” számától függően fejezi ki a szó gyakoriságát.

Nézzük meg a már fentebb említett nyolc ige esetében, hogy mit talál-tam én a COBUILD birminghami adatbázisában.

	Összes előfordulás	Előfordulás a beszélt nyelvben	Milliónkénti előfordulás a beszélt nyelvben
like	267398	30662	3492.9
love	62322	2339	266.5
enjoy	24309	1344	153.1
prefer	12931	311	35.4
hate	9877	588	67.0
dislike	2563	69	7.9
loathe	1186	24	2.7
detest	347	11	1.3

Az új kiadásban ez a következő módon tükröződik:

like		hate	
love		dislike	
enjoy		loathe	
prefer		detest	

Hangsúlyozni kell, hogy a szótár a gyakoriságra vonatkozó mutatót a szóra nézve általánosságban adja meg, nem pedig a különböző jelentésekre, stílusokra vagy eltérő nyelvtani környezetekre elkülönítve. Mint a saját táblázatomból is kiderül, bár a „prefer” ige általánosságban gyakoribb, mint a „hate”, a beszélt nyelvben ez fordítva van. (Külön érdekesség, mely egy másik tanulmány témája lehet, hogy a pozitív érzelmeket kifejező igék miért gyakoribbak, mint a negatív érzelmeket kifejezők.)

Távlatok

A számítógépes adatbázisokra építő szótárírás szinte beláthatatlan lehetőségeket hordoz magában. Ma már a világ számos helyén léteznek eltérő nagyságú és rendeltetésű korpuszok. A szótárkészítésben legnagyobb hagyománnyal rendelkező kiadók sem vonhatják ki magukat az új lehetőségek alkalmazása alól. Ennek újabb jele, hogy az Oxford Advanced Learner's Dictionary idén megjelenő új kiadását is úgy hirdetik már a prospectusok, hogy az több egyetem és más intézmény összefogásával létrejött,

100 milliós adatbázison alapszik. (Jellemző, hogy a kiadók igyekeznek hangzatos elnevezéseket adni adatbázisuknak. Például: COBUILD: The Bank of English, Oxford: British National Corpus.)

Az adatbázisokban tárolt adatok mennyisége napról napra növekszik. (A COBUILD új szótára 200 millió adatra épül, de valójában már 500 millió adat van a számítógépükben!) Minél nagyobb egy korpusz, annál biztonságosabb következtetéseket lehet levonni belőle. Nem véletlen, hogy a kiadók igyekeznek hangsúlyozni, hogy az ő korpuszuk a legnagyobb, a legrepresentatívabb, a valós nyelvet leginkább tükröző.

A fentiekben leírtakból is következően számos területen van lehetőség a továbblépésre. Valószínűleg megnövekszik a gyakoriságra vonatkozó mutatók jelentősége. Jelenleg még csak a szó egészére található ilyenek a COBUILD szótárban. Várható, hogy előbb-utóbb a szó különböző jelentéseire, lehetséges nyelvtani környezetére stb. vonatkozóan is megjelennek ilyen mutatók. Ez óriási munkát jelent majd a szótárak összeállítóinak, de ezek a lehetőségek már ma benne vannak a számítógépes adatbázisokban. (Magam is végeztem kutatásokat arra vonatkozóan például, hogy a fenti igék esetében a korpusz milyen gyakoriságot mutat infinitív és gerundiumos szerkezetekkel kapcsolatban, s ezek eredménye számos esetben meglepő, a hagyományos nyelvtankönyveknek ellentmondó. Például – ahol van egyáltalán választási lehetőség – jóval gyakoribb az infinitív, mint ahogy talán a legtöbb nyelvtanár gondolná. Csak egyet lehet érteni azzal, amit Biber–Conrad–Rippen egy másik jelenséggel, a főnévi bővítményekkel kapcsolatban kimutat, hogy a nyelvtankönyvek szerzői számára a múltban nem annyira egy szerkezet előfordulásának gyakorisága, hanem nehézsége és/vagy taníthatósága számított. A lexikográfia területén pedig az jellelheti az egyik fő veszélyt, ha a nyelvet homogénnek tekintjük, s nem vesszük figyelembe, hogy például egy szinonimacsoport tagjai közül egyik szó az egyik regiszterben, a másik egy másikban gyakoribb. Ezért is lenne szükség például a COBUILD szótár gyakorisági mutatóinak további finomítására.)

Hasonlóképpen nagy fejlődés várható a szótárak kezelhetőségét, a rövidítések megjegyezhetőségét, a szóra vonatkozó egyéb információk gazdagodását illetően is. A kiadók és kutatási műhelyek közötti rivalizálás is csak termékenyítőleg fog hatni erre a folyamatra.

Bibliográfia

- Allan, W. Stannard: *Living English Structure*. Longmans, 1959.
- Alexander, L. G.: *Longman English Grammar*. Longman, 1988.
- Benson, Morton–Benson, Evelyn–Ilson, Robert: *The BBI combinatory dictionary of English*. John Benjamins Publishing Company, Amsterdam/Philadelphia, 1986.
- Biber, Douglas–Conrad, Susan–Reppen, Randi: Corpus-based Approaches to Issues in Applied Linguistics. In: *Applied Linguistics*, Vol. 15, No. 2, Oxford University Press, 1994.
- Collins COBUILD English language dictionary*. Ed. in chief John Sinclair, Collins, London /Glasgow, 1987.
- Collins COBUILD English Dictionary*. New Edition. Ed. in chief John Sinclair, HarperCollins Publishers, London, 1955.
- Collins COBUILD English Grammar*. Collins, London/Glasgow, 1990.
- Francis, Jill–Sinclair, John: 'I Bet He Drinks Carling Black Label': A Riposte to Owen on Corpus Grammar. In: *Applied Linguistics*, Vol 15. No. 2, pp 190 - 200, Oxford University Press, 1994.
- Hornby, A. S.: *Oxford Advanced Learner's Dictionary of Current English*. Fourth Edition. Chief Editor: A. P. Cowie, Oxford University Press, Oxford, 1989.
- Jackson, Howard: Learners' Dictionaries in Contrast: Langenscheidt and Longman. In; *Fremdsprachen Lehren und Lernen*, Vol. 24, 1955. (megjelenés előtt)
- Longman Dictionary of Contemporary English*. Ed. by Tom McArthur, Longman, 1978.
- Quirk, R. et al.: *A Grammar of Contemporary English*. Longman, 1972.
- Quirk, R. et al.: *A Comprehensive Grammar of the English Language*. Longman, London/ New York, 1985.