

SZOFTVER-KRITIKA

ROHONYI ANDRÁS

A konkordanciáról egy magyar program ürügyén

A program neve: ConTour

Változat: 1.0

Fejlesztő: Vlaskovits Dóra

Hardver-igény: IBM kompatibilis PC

(Minimum: AT 80286; javasolt: AT 80386)

Operációs rendszer: MS-DOS 3.3

Memóriaigény: 1 MB

A szoftver jellege: shareware

A program szerzője a program leírásában Samuel Beckettől kölcsönzi mottóját: „Szavak – ez mindenünk”. A ConTour egy olyan nyelvészeti segédprogram (konkordancia-szoftver), amely érdekes és hasznos eszköz lehet nemcsak nyelvészek, hanem mindenki számára, aki nyelvet tanít, tanul vagy tudatosan használ. „Bárki számára izgalmas lehet egy szövegben előforduló szavak gyakoriságának és szöveggörnyezetének vizsgálata. Esszéket és egyéb dokumentumokat elemezhetünk vele”, írja Vlaskovits Dóra, majd szerényen hozzáteszi: „Ehhez hasonló szoftver eddig csak angol nyelven készült, mely a mostani programmal ellentétben nem volt alkalmas magyar ékezetes betűk, ezáltal magyar nyelvű szövegek kezelésére.”

A ConTour program önmagában, mint a fenti adatokból is látható, igen szerény igényű, egy egyetemi hallgató szakdolgozatának részeként készült. Szerzője, aki nyelvszakos tanulmányai mellett az idén felsőfokú programozó képe­ sítést szerzett, jelenleg többek között egy számítógépes nyelvészeti kurzust hallgat és tervezi a program továbbfejlesztését, hogy az nagyobb terjedelmű állományokat is kezelni tudjon. A program jelenlegi formájában maximum kétezer sornyi szöveget képes kezelni, amelyben legfeljebb tízezer szó vagy írásjel (ezek között maximum kétezer különböző) fordulhat elő.

A vizsgálni kívánt állomány egyszerűen kiválasztható könyvtárstruktúra segítségével. A program képes a szavak szűkebb (120 karakteres) és bővebb (700 karakteres) szöveggörnyezetének a megjelenítésére, valamint egy szó többszöri előfordulása vagy több kiválasztott szó esetén a sorok rendezésére különböző szempontok alapján, illetve annak a vizsgálatára, hogy a kiválasztott szavak közelében egy adott másik szó előfordul-e. Tömören közli a szövegre vonatkozó mennyiségi információkat is.

Mielőtt azonban részletesebben bemutatnánk a program működését, érdekesnek látszik – szoftver-kritikában talán szokatlan módon – rövid kitérőt tenni a konkordancia-programok általános jellemzésére. Tekintettel arra, hogy a hazai nyelvoktatási és nyelvtanulási gyakorlatban a konkordancia-programok ismerete és használata korántsem nevezhető általánosnak, e rövid cikk végén néhány bibliográfiai utalást is teszünk a téma irodalmára.

A *Modern nyelvoktatás* olvasói Prószéky Gábor „Nyelvtechnológia 2000” című cikkéből már tájékozódhattak a számítógépes nyelvészet terén folyó kutató- és fejlesztőmunka hazai eredményeiről, az úgynevezett korpusz-nyelvészet lehetőségeiről és feladatairól.

A konkordancia nem más, mint egy (vagy több) szó, szórész vagy szókombináció előfordulásainak a kigyűjtése a szövegek környezetével együtt, egy elméletileg (és hovatovább gyakorlatilag is) tetszőleges méretű szövegtörzsből. A konkordancia leggyakoribb megjelenítési módja az úgynevezett KWIC (keyword in context) formátum. Az európai nyelvészeti kutatásokban leggyakrabban használt, legismertebb konkordancia-programok a *Longman Mini-Concordancer (LMC)* és az *Oxford MicroConcord (MC)*. Az *LMC* 40-50 ezer szót tud kezelni egyidejűleg, ami gyakran kevésnek bizonyul. A *MicroConcord* program jóval nagyobb, a sajtóból és tudományos ismeretterjesztő művekből összeállított kb. 2 millió szavas korpuszra épül és input-korlátok nélkül használható, bár az output itt is korlátozott.

Mire való a konkordancia?

Mire lehet felhasználni egy konkordancia-programot a nyelvoktatásban? Kézenfekvő, hogy a konkordancia alkalmas lehet a szókincs bővítésére, de hasznos lehet nyelvtani, sőt irodalmi, stilisztikai stúdiumokban is. Itt csak utalhatunk a kontraszt, a kollokáció, a szemantikai viszonyok elemzésében rejlő lehetőségekre, példaként említve a homonímia vagy a poliszémia eseteit. A szavak jelentését a kontextus adja meg igazán.

A Grazi Egyetemen működő Bernhard Kettemann professzor konkordancia-vizsgálatok segítségével kimutatta például, hogy a gyakorlati angol nyelvhasználatban a feltételes mód képzésének a szabálya, az úgynevezett *if-clause rule* korántsem a nyelvtanítási kánonok szerint érvényesül: a valóságban az esetek egyharmadában mind a főmondatban, mind a mellékmondatban jelen idejű formák szerepelnek. Hasonlóan érdekes megállapításra jutott a függő beszéd elemzése során: az általa vizsgált angol szövegekben az igeidők szigorú egyeztetése az esetek egynegyed részében nem érvényesült. Más konkordancia-vizsgálatok bebizonyították, hogy az írott angol nyelvben a *simple past* és a *present perfect* aránya tíz az egyhez: figyelembe véve, hogy a beszélt nyelvben valószínűleg ennél is jóval ritkább a *present perfect* használata, elgondolkodhatunk azon, hogy vajon a mi nyelvtanítási gyakorlatunkban (legalábbis kezdő fokon) helyes-e az, hogy a *present perfect* megtanítására olyan sok időt fordítunk.

A szövegek számítógépes statisztikai vizsgálata megmutathatja a mondatok átlagos hosszát és a különböző szótagszámú szavak gyakoriságát, amiből a stílusra is következtethetünk. (Ilyen szempontból érdekes lenne például Joyce és Hemingway összevetése.) A konkordancia-programok felhasználására irodalmi

művek tanulmányozásához álljon itt egyetlen példa. Kettemann egy múlt század végi amerikai novella elemzésében a *he* és *she*, illetve a *his* és *her* szavak konkordanciája segítségével igen pregnánsan tudta illusztrálni a nemek közötti megkülönböztetést. A nő elsősorban a háztartással, a testi szépséggel, lelki, érzelmi megnyilatkozásokkal összefüggő szövegkörnyezetben jelenik meg, míg a férfi esetében a fej, az ész, a gondolatok, elvont fogalmak, gondok, tervek dominálnak.

A ConTour program lehetőségei és korlátai

Ez a kis (485 kB méretű) konkordancia-program a jelenlegi formájában természetesen nem alkalmas a fentiekben említett példákhoz hasonló vizsgálatok elvégzésére, elsősorban az általa elemezhető szövegállomány méretének korlátozottsága miatt. A program csak DOS szöveges (CWI-kódkészletű) állományok kezelésére alkalmas, ezért a különböző szövegszerkesztőkkel készült dokumentumokat először konvertálni kell ilyen formátumba. (A továbbfejlesztéshez a szerző tervezi saját konvertáló funkció beépítését a programba.)

A CONTOUR.EXE fájl indítása után megjelenik a felső szalagmenü, amely a program futtatása során mindvégig jelen van, de nem minden helyzetben érhető el az összes menüpont. A képernyő alján található üzenetsorban a program futása során különböző információk jelennek meg, amelyek segítséget adnak az éppen használt funkcióhoz.

A „FÁJL/BEOLVAS” menüpont kiválasztásával betölthető a memóriába az elemzendő szöveges állomány. A képernyőn megjelenő ablakban kell megadni a vizsgálni kívánt szövegfájl nevét és elérési útvonalát. A „FÁJL” menüben lehet beállítani a monitor típusát (színes vagy fekete-fehér) és a kiválasztott nyelvet (angol vagy magyar), és itt van lehetőség az aktuálisan megjelenített adatok mentésére is, tetszőleges célállományba, az elérési útvonal megadásával.

Az „INFO/MEGJELENÍT” menüpont választásával statisztikai információkhoz jutunk a betöltött szöveges állományról. Megtudjuk, hogy az állomány hány sorból, illetve hány karakterből áll. Megjelenik a szavak és az írásjelek száma, és az is, hogy ezek közül hány különbözik, továbbá a szavak előfordulási arányszáma is.

Ebben a cikkben például a ConTour elemzése szerint 1327 szó és 318 írásjel van, ezek közül 779, illetve 10 különböző, a „konkordancia” szó pedig 15-ször fordul elő, az alábbi kontextusokban:

TVER-KRITIKA__ConTour: egy magyar konkordancia -program __A program neve: ConTou
gy olyan nyelvészeti segédprogram (konkordancia -szoftver), amely érdekes és hasz
tlan módon – rövid kitérőt tenni a konkordancia -programok általános jellemzésére
si és nyelvtanulási gyakorlatban a konkordancia -programok ismerete és használata
lehetőségeiről és feladatairól. _A konkordancia nem más, mint egy (vagy több) sz
zóleges méretű szöveggörpuzsból. A konkordancia leggyakoribb megjelenítési módja
gyakrabban használt, legismertebb konkordancia -programok a Longman Mini-Concord
t itt is korlátozott.__Mire való a konkordancia?__Mire lehet felhasználni egy ko
ncia?__Mire lehet felhasználni egy konkordancia -programot a nyelvoktatásban? Kéz
yelvoktatásban? Kézenfekvő, hogy a konkordancia alkalmas lehet a szókincs bővíté
gyed részében nem érvényesült. Más konkordancia -vizsgálatok bebizonyították, hog
Joyce és Hemingway összevetése.) A konkordancia -programok felhasználására irodal
korlátai_Ez a kis (485 kB méretű) konkordancia -progam a jelenlegi formájában te
özül 779, illetve 10 különböző, a "konkordancia " szó pedig 15-ször fordul elő, a
asoljuk a szerzőnek, hogy a "nagy" konkordancia -programok (pl. MC) mintájára épí

A „LISTA” funkció megadja a szövegben szereplő összes szót, kívánság szerint ábécérendben, a végződés szerinti ábécérendben vagy az előfordulás gyakorisága szerint. Kifogásolható, hogy a szavak két oszlopban jelennek meg a képernyőn, ami nehezíti az áttekintést. A szavak melletti számok jelzik, hogy azok hányszor fordulnak elő a szövegben. A listáról kiválasztott szavak kijelölése után megjeleníthető a szavak szöveggörnyezete. A kijelöléshez a <SZÓKÖZ> billentyűt kell használni, ami egyes levelezőprogramokban (pl. Pegasus Mail) szokásos, de a Norton Commanderhez szokott felhasználónak inkább az <INSERT> billentyű alkalmazása lenne kézenfekvő. A „LISTA” menüpontból megkaphatjuk az írásjeleket is.

Az egész programban talán a legfontosabb a „LISTA/SABLON” menüponttal elérhető funkció, ahol a listázandó szavak helyettesítő karakterek felhasználásával adhatók meg. A * tetszőleges számú karaktert, míg a ? egyetlen karaktert helyettesít. A szerző példáit használva a program leírásából: *a*=alma*, *atom*, *anonym stb.*, *??a?=ágot*, *akar*, *csak*, *igaz stb.*, *?z*=azok*, *szavak*, *szem stb.* A sablon segítségével megjeleníthető az egyes szavak ragozott alakjainak a konkordanciája, ami a magyar nyelv morfológiai gazdagsága miatt különösen fontos.

A „SZÖVEGKÖRNYEZET” menüpont alatt akár a szólistából, akár az „ÚJ SZÓ” beírásával megkaphatjuk a keresett konkordanciát. Nem szerencsés megoldás, hogy a bővebb szöveggörnyezetet csak a nyilak használatával, az adott sorokat balra vagy jobbra fokozatosan eltolva kaphatjuk meg. A szöveg így nehezen olvasható, nem jelenik meg egységben. Jó viszont, hogy a szövegből eltüntethető az eredeti sorvégek egyébként zavaró jelzése.

A „RENDEZ” funkcióval a szokásos szempontok szerint rendezhetjük a sorokat. A „SZOMSZÉDSÁG” menüpont a szókapcsolatok konkordanciáját adja meg.

A ConTour program kicsiben lehetővé teszi szinte mindazokat a vizsgálatokat, amelyekre egy nyelvtanárnak vagy nyelvtanulónak szüksége lehet. A program hibái a korlátaiból fakadnak: érdemes elvégezni azt a továbbfejlesztést, ami majd lehetővé teszi nagyobb korpuszok feldolgozását is anélkül, hogy szükség lenne az állományok előzetes konvertálására.

Javasoljuk a szerzőnek, hogy a „nagy” konkordancia-programok (pl. MC) mintájára építsen be a programba egy komoly magyar nyelvű korpuszt, az Interneten elérhető újságok, periodikák és egyéb dokumentumok anyagából.

IRODALOM

- Barlow, M. (1995), „Corpora for Theory and Practice”, *International Journal of Corpus Linguistics* 1 (1).
- Higgins, J. & T. Johns (1984), *Computers in Language Learning*. London: Collins ELT.
- Higgins, J. (1991), „Which Concordancer? A Comparative Review of MS-DOS Software”, *System* 19/1&2, pp. 91–100.
- Johns, T. & P. King (szerk. 1991), *Classroom Concordancing*. Az ELT Journal különszáma, Vol. 4. Centre for English Language Studies, the University of Birmingham.
- Kettemann, B. (1994), „Concordancing in Stylistics Teaching”, *Festschrift zum 30jährigen Bestehen des Instituts für Anglistik der Universität Salzburg*, Salzburg.

- Kettemann, B. (1995), „On the Use of Concordancing in ELT”, *TELL&CALL* 1995/4, pp. 4–15.
- Kettemann, B. (1996), „Concordancing in English Language Teaching”. World Wide Web Document. URL <http://gewi.kfunigraz.ac.at/institute/anglistik/project/concord1.html>
- Murison-Bowie, S. (1993), *MicroConcord Manual. An Introduction to the Practices and Principles of Concordancing in Language Teaching*. Oxford: Oxford University Press.
- Owen, C. (1996), „Do Concordances Need to Be Consulted?” *ELT Journal* 50, pp. 219–224.
- Prószéky Gábor (1996), „Nyelvtechnológia 2000”, *Modern Nyelvoktatás* 1996/1–2, pp. 31–34.
- Sinclair, J. M. (1991), *Corpus Concordance and Collocation*. Oxford: Oxford University Press.
- Stevens, V. (1991), „Classroom Concordancing: Vocabulary Materials Derived from Relevant, Authentic Text”, *English for Specific Purposes* 10, pp. 10–15.
- Tribble, C. & G. Jones (1990), *Concordances in the Classroom*. Harlow: Longman.

The

LIBRA BOOKSHOP

LANGUAGE BOOKS

*The W I D E S T selection of
English-Language teaching materials*
LONGMAN • NELSON • OXFORD UP
CAMBRIDGE UP • PENGUIN FIT

Custom Orders / available
1085 Budapest, Kölcsey u. 2.
Tel./Fax: 313–7630,
266–0065/102

Open: Mon–Fri: 10–18
Sat: 9–13

EXPANDED *Fiction and Literature
section*

Also featuring

LANGENSCHIEDT and **DUDEN** books

*Everything you need
for your German studies*

