

HORVÁTH JÓZSEF

## Modern nyelvoktatás a JPU korpusz internetes gyűjteményével

Több mint háromszáz angol szakos hallgató vett részt abban a projektumban, melynek során a kilencvenes évek végére Pécsen összegyűjtöttük az akkori egyik legnagyobb európai angol nyelvi tanulói korpusz anyagát. Az intézményt akkor még Janus Pannonius Tudományegyetemnek hívtuk, így a gyűjtemény ennek angol megfelelője alapján a *JPU korpusz* nevet kapta. Igen izgalmas volt a vállalkozás: a csaknem félmillió szavas szövegtest lehetőséget teremtett arra, hogy a korpusznyelvészet objektív módszereivel leírjuk a diákíráások jellemzőit. Vizsgálni lehetett többek között azt, hogy az orosz szakos tanárok átképzésében részt vevők angol nyelvű esszéiben milyen arányban található helyes és hibás határozott névelői használat, de azt is, hogyan különböznek a korpusz egyik legnagyobb alcsoportját képező tanulmányok bevezető és záró mondatai.

Számomra a korpusz elveinek kialakításakor, a gyűjtéskor és a folyamatos elemzéskor mégis talán az volt a leghasznosabb, hogy ez a gyűjtemény valamennyi ideig bizonyonnyal megmarad a magam, de mások számára is: a diákoknak, akik hozzájárultak esszéik, történeteik és tanulmányaik felhasználásához, és kollégáknak, akik szintén érdeklődtek tanulói korpuszok iránt. (Lásd például Kiszely, Lehmann és Szirmai munkásságát.) Újra el lehet olvasni a szövegeket, korábban észre nem vett jellemzőkre bukkanhatunk, kapcsolatot teremthetünk a diákok és tanárok generációi között.

A JPU korpusz kialakítását és tartalmát annak idején már ismertettem (2000), egy másik írásban pedig bemutattam, hogyan lehet ötvözni a korpusznyelvészet és az íráspedagógia elveit (2001). Most, tíz évvel az után, hogy a gyűjtemény nagy része összeállt, újra visszatérek hozzá, hogy feltárjam, mik azok a lehetőségek, melyeket a modern nyelvoktatás számára a gyűjtemény internetes változata biztosít, részben önmagában, részben egy bárki számára elérhető eszközrendszer, a Thomas Cobb által kialakított *Compleat Lexical Tutor* segítségével. Remélem, talál ezek között majd olyat néhány angoltanító olvasó, melyet ő is kipróbálhat – a többi nyelvszakos kolléga pedig egyéb nyelveken rendelkezésre álló tanulói korpusz alapján alkalmazhatja őket. Végül azt is megmutatom, hol tart ma a JPU korpusz gyűjtésével megkezdett kutatói és oktatói gyakorlat.

### A tanulói korpusz haszna

A tanárokat mindig is érdekelte, miről és hogyan írnak a diákjaik: a tanulói korpusz tehát nem új keletű találmány. Ami új, az a komputeres technológia által nyújtott se-

gítség és az, hogy az eredményeket széles körben lehet ismertetni: ezáltal csökkenhet a tanulói csoportok elszigeteltsége.

A belgiumi angol korpusznyelvészeti központ, a Centre for English Corpus Linguistics volt a kilencvenes években a tanulói korpuszok úttörője – az általuk megfogalmazott elvek bizonyultak a legtávolabbi hatásúnak. McEnery, Xiao és Tono, továbbá Nesselhauf, Ooi, valamint Wang és Sun munkáiban részletes leírás található arról, hogyan gyűjthetünk tanulóktól nyelvi adatokat, ezeket miképp lehet korpuszá formálni és azt elemezni. Az európai angol tanulói korpuszok közül kiemelkedő jelentőségű az *International Corpus of Learner English* lengyelországi komponense (lásd Kaszubski munkáját).

A JPU korpusz kialakításának és használatának ma talán az a legnagyobb haszna, hogy bárki számára szabadon elérhető: az egyes szövegek önmagukban is olvashatók, élvezhetőek és elemezhetőek, de csoportosíthatjuk és ennek alapján vizsgálhatjuk őket. Mindezek mellett modern nyelvoktatási gyakorlatok alapját is képezhetik. Megismerhetjük a szövegek által, milyen gondolatok foglalkoztatták a diákokat, mit olvastak, milyen sorozatokat néztek a tv-ben, milyen egyetemi előadásokból mit tanultak: mit vettek észre a világban és mit tettek hozzá. A gyűjtemény szabadon felhasználható: kereshető, nyomtatható és kutatható.

## A JPU korpusz internetes gyűjteménye

A JPU korpusz jelenleg az interneten található gyűjteménye (<http://joeandco.blogspot.com>) a teljes korpusz nagy része, 221 szöveg. Valamivel több mint 289 ezer szóból áll. A legtöbb szöveg tanulmány, de található köztük esszé és novella is. A mostani netes publikálás célja az volt, hogy minél egyszerűbb, átláthatóbb felületen legyen elérhető az anyag, ezért a blog formátum mellett döntöttem, mivel ezt tanárok és diákok is egyre nagyobb számban ismerik. Egyetemi oktatói munkámban a hároméves képzésben folyó olvasás- és íráskészség tanításában használom ezt az anyagot órákon és kiegészítő célokkal.

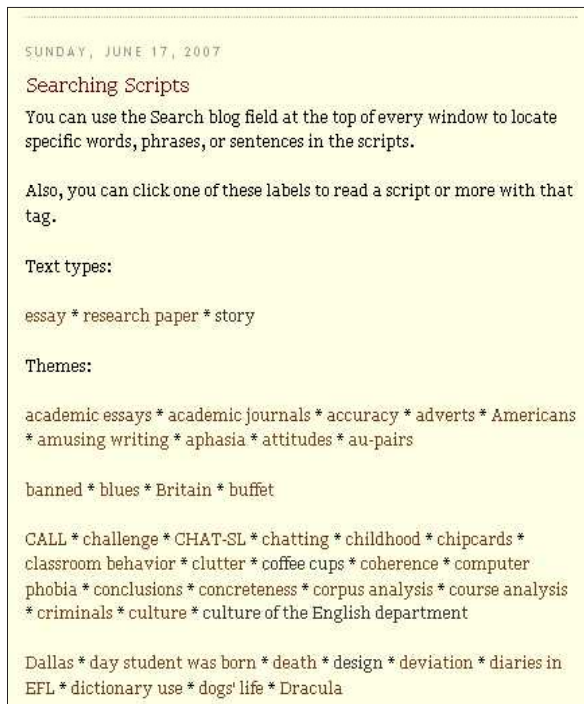
Az 1. ábra a keresési lehetőségeket illusztrálja.

A kereső mezőben például megkérdezhetjük, szerepel-e a korpuszban utalás Jockey Ewingra, és ha igen, mit írtak róla annak idején a diákok. De kattinthatunk a listán látható valamelyik témára is, és így az összes azzal a címmel megjelölt szöveg teljes terjedelmében olvasható, tovább vizsgálható, menthető, nyomtatható, kutatható. Ezenkívül lehetőség van a három alkorpusz strukturált keresésére és felhasználására is. (Ennek módját a korpusz blogján ismerheti meg az olvasó.)

Az egyik alkorpusz a nyelvgyakorlat órákra írt esszék gyűjteménye, melyek közül ehhez a cikkhez az L 220 F kódú szöveget választottam ki. Ez arról szól, hogyan merülnek fel Graham Greene regényeiben és novelláiban életének élményei. Így kezdí az esszéjét a hallgató:

Graham Greene is the kind of writer whose novels and short stories are influenced by his own experiences in life. He does not write directly about his life but his attitude to the phenomena of the world and the things that happen to him can be felt in the ways he makes his stories. The role of childhood experiences, the unpleasant side of life and escapism are important aspects in Greene's life. This essay will examine how and why he deals with them in his works too.

Már ezek a bevezető gondolatok is érdekesek lehetnek a mai olvasó számára – nemcsak tanároknak, hanem diákoknak is, akik közül néhányan talán nemcsak tartalmi és retorikai, hanem nyelvgyakorlati, szókincsfejlesztő vagy más célra is fel szeretnék használni ezt és a többi írást.



1. ábra. A JPU korpusz témáinak részlete a blogon

Nézzük ennek a példának az alapján, hogyan kapcsolható össze a JPU korpusz és az az eszköztár, amelyet Thomas Cobb hozott létre, a Compleat Lexical Tutor (CLT).

## A CLT eszközei és a JPU korpusz

Rövid idő alatt, könnyen elsajátíthatóak a CLT interneten elérhető tutoriális, tanári és kutatói segédleteinek alapjai. A tutoriális eszközök az érdeklődő diákokat segítik abban, hogy plasztikus képet nyerjenek szókincsükről a produktív és reprodukív szóismeretet mérő tesztekkel. A tanári segédletekkel többek között saját feladatlapokat készíthetünk. Sok eszköz angol és francia változatban is rendelkezésre áll, és más nyelveken írt szövegeket is fel lehet tölteni. A JPU és más tanulói korpuszok legjobban a kutatói eszközökkel működnek együtt.



2. ábra. A CLT emblémája

Valamennyi eszközt érdemes kipróbálni és diákjainkkal megismertetni – lássunk ezek közül néhányat.

### **Gyakorlatok konkordanciával**

A konkordancia olyan korpusz-alapú eljárás eredménye, melynek segítségével egy szó vagy kifejezés közvetlen kontextusát vizsgálhatjuk. Bármely számítógépes korpusz feltölthető egy speciális programra és ez automatikusan létrehozza a konkordancia sorait. Ezek többnyire nem teljes mondatok, csak részletek, de lehetőség van a nagyobb kontextus vizsgálatára is.

A konkordancia létrehozásának egyik módja a CLT-n az, hogy egy saját magunk vagy a diákunk által írt szöveget töltünk fel. A Graham Greene műveit elemző esszé idézett bekezdésében például megvizsgálhatjuk, hogyan alkalmazza a szerző a *he* névmást: mit is mond Graham Greene-ről. Ehhez nem kell mást tennünk, mint kijelölni a kívánt szöveget a korpusz blogján, átváltani a CLT saját szöveg feltöltését lehetővé tevő oldalára ([www.lexutor.ca/concordancers/text\\_concord](http://www.lexutor.ca/concordancers/text_concord)), beszúrni a bekezdést a szöveglapba és beírni a kulcsszó rovatba a *he* szót. A 3. ábra az eredményt mutatja meg, középen a nagybetűvel kiemelt kulcsszóval:

This essay will examine how and why **HE** deals with them in his works  
by his own experiences in life. **HE** does not write directly about h  
to him can be felt in the ways **HE** makes his stories. The role of

### **3. ábra. A „he” konkordanciája a CLT segítségével**

Konkordanciát azonban nemcsak ilyen beillesztett szövegekből, hanem a CLT által elérhetővé tett korpuszokból is kaphatunk. Cobb tizenhét angol nyelvű korpuszt vett fel a konkordanciával támogatott eszköztárba ([www.lexutor.ca/concordancers/concord\\_e.html](http://www.lexutor.ca/concordancers/concord_e.html)), melyek közül a két legnagyobb a Brown Corpus teljes gyűjteménye és a British National Corpus (BNC) egy-egymilliós írott és beszélt nyelvi mintája. Itt találhatjuk meg a JPU korpuszt is (lásd a 4. ábrát).

Akik tehát már elkezdtek megismerkedni azzal, hogy milyen témák foglalkoztatták a pécsi diákokat a kilencvenes években, és hogy ezekről milyen remek angolsággal (de természetesen nem mindig hibátlanul) értekeztek, most arra is sort keríhetnek, hogy mind a 221 szöveget elemezni kezdik a konkordancia eszköz segítségével. A keresett szó vagy kifejezés megfelelő mezőjébe való beírása után fókuszálhatnak például arra, milyen kontextusban jelent meg Jockey Ewing neve vagy akár Graham Greene-é.

Az a cél is megvalósítható, hogy az egyes korpuszok konkordanciáit összevegyük – ehhez nem kell mást tennünk, mint kiválasztani ezeket és kinyomtatni vagy elmenteni a kapott konkordanciákat. Nem kell tehát mindig a gép előtt ülve, a képernyőn olvasni az eredményeket.

### **Gyakorlatok szókincs-profillal**

A CLT-n számos mód van arra, hogy elemezzük a korpuszok vagy azok egyes szövegeinek szókincsét. A Vocabulary Profile linkre kattintva találjuk meg a British National Corpus szógyakoriságát alapul vevő profilkészítőt, mely intuitív, igen szemléletes módon ad információt a szövegekről.



4. ábra. Az online konkordancia korpuszlistája

A BNC százmillió szóból áll. Ezeket a szavakat a gyakoriság alapján ezresével csoportosítja a profilkészítő, mely a feltöltött szöveg szavainak gyakoriságát összeveti a BNC ezres csoportjaiéval. A Graham Greene írásait elemző bekezdés szavaira az 1. táblázatban közölt gyakorisági statisztikát kapjuk.

Freq. Level	Families	Types	Tokens	Coverage%
K1 Words :	42	48	76	87.36
K2 Words :	5	5	5	5.75
K3 Words :	2	2	2	2.30
K4 Words :	1	1	1	1.15

Plusz: Graham és Greene (2), melyek a BNC listán kívüli szavak

1. táblázat. A Greene-bekezdés szókincs-profilja a BNC csoportjai szerint rendezve

K1 a leggyakoribb ezer szót jelöli, ide tartozik a szövegből 76 szó (token), ezek teszik ki a bekezdés 87 százalékát. K2 a második ezres szócsoporthoz jelöli, és így tovább. De nemcsak számokkal, hanem színekkel is meg lehet jelölni a szavakat. A profilkészítő egyik nagy előnye az, hogy a vizuálisan tanulóknak lehetővé teszi, hogy saját szövegüket vagy másokét ebben a formában is megvizsgálhassák. A K1 szavak világoskék, a K2-k zöldek, a K3-ak citromsárgák, a K4-ek okkerek. A Greene-szöveg részben zöld, de leginkább világoskék a képernyőn és a színes nyomtatáson.

Mit kezdhet a színes szólistával vagy a többi szemléletes megjelenítéssel a tanár és a tanuló? Többek között tanulmányozhatja, mennyire azonosak vagy eltérők az egyes bekezdések szókincs-profiljai. Felmérheti, hogyan függ ez össze a téma kifejtésének retorikájával: mennyire tudatos a speciálisabb, kevésbé gyakori szavak alkalmazása inkább a szöveg elején, közepén vagy a végén. Hogyan alkalmazható, illetve fejleszt-

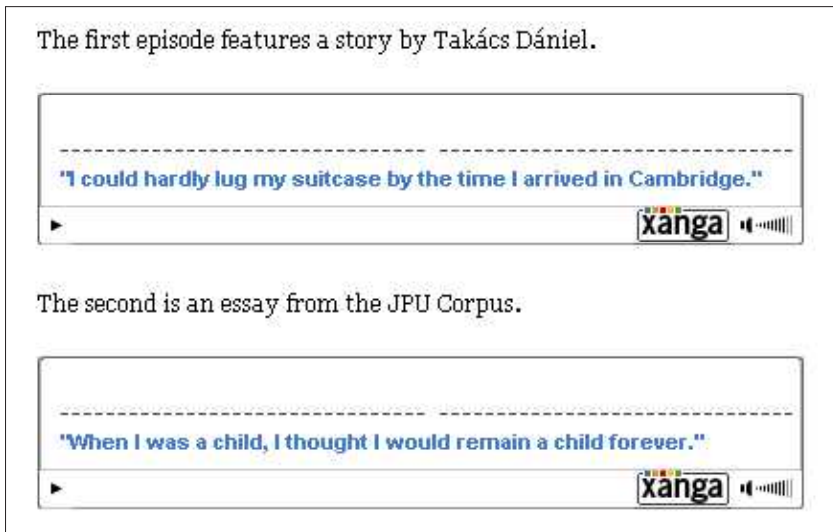
hető ez a tudatosság? Jó-e, ha csupa világoskék a szövegünk? Mennyire kell arra törekednünk, hogy több legyen benne a citromsárga, az okker és a többi szín?

A JPU korpusz valamennyi írását fel lehet tölteni a szókincsprofilra, hogy így valóban képet alkothassunk a szövegekről. De magunk is kialakíthatunk kisebb alkorpuszokat, hogy azokat intenzíven vizsgálva tanulhassunk és taníthassunk. És kit ne érdekelne, az ő saját szövege milyen színes – és színes-e pusztán azáltal, hogy a BNC csoportjai alapján az.

### JPU után Take Off

A JPU korpusz blogja és Thomas Cobb eszköztára sok más lehetőséget is rejteget. Az én jelenlegi gyakorlatomban többek között azt, hogy kidolgoztam egy olyan internetes tananyagot, mely a korábbi évek kutatói és oktatási tapasztalataira építve nyújt teret az angol nyelvi olvasás- és íráskészségek fejlesztésének az egyetemi alapképzésben. Take Offnak neveztem el ezt a tananyagot.

A Take Off egyik szolgáltatása a JPU korpuszt író diákok és a mai hallgatók virtuális párbeszédének elősegítése. Egymásnak üzennek azokban az MP3-as formátumú hangzó anyagokban, melyek egy-egy diák szövegét teszik közzé. A Take Off blogján is meg lehet hallgatni ezeket a szövegeket, vagy le lehet őket tölteni hordozható eszközökre.



5. ábra. A Take Off két epizódja egy mai diák írása és egy JPU korpusz-szöveg alapján

És mivel a JPU korpusz és a Take Off is blogként van kialakítva, tanárok és diákok is hozzászólhatnak, folytathatják a párbeszédet a modern nyelvoktatás és nyelvtanulás egy lehetséges útjának behatóbb megismerésével.

## IRODALOM

- Horváth József (2000): A JPU korpusz kialakítása és tartalma. *Modern Nyelvtanítás* 6/2–3. pp. 42–9.
- (2001): *Advanced writing in English as a foreign language: A corpus-based study of processes and products*. Pécs: Lingua Franca Csoport.
- Kiszely Zoltán (2006): Magyar középiskolások angol fogalmazási stratégiái. *Nyelvinfo* 14/2. pp. 8–12.
- Lehmann Magdolna (2003): The lexis of writing and vocabulary size: The relationship between receptive knowledge and productive use. In: Andor József – Horváth József – Nikolov Marianne (szerk.): *Studies in English theoretical and applied linguistics*. Pécs: Lingua Franca Csoport. pp. 172–81.
- McEnery, T. – R. Xiao – Y. Tono (2006): *Corpus-based language studies: An advanced resource book*. New York: Routledge.
- Nesselhauf, N. (2004): Learner corpora and their potential for language teaching. In: Sinclair, J. (szerk.): *Teaching and language corpora*. Amsterdam: John Benjamins. pp. 125–52.
- Ooi, V. (2002): From Shakespeare to Hungarian EFL writing: Using WWW corpora to motivate student learning. In: Tan, M. (szerk.): *Corpus studies in language education*. Thailand: Assumption University. pp. 163–77.
- Szirmai Monika (2005): *Bevezetés a korpusznyelvészetbe: A korpusznyelvészet alkalmazása az anyanyelv és az idegen nyelv tanulásában és tanításában*. Budapest: Tinta.
- Wang, L. – K. Sun (2005): Current developments in learner English corpus in and outside China. *Computer-Assisted Foreign Language Education*. pp. 19–24.

### A JPU korpusz és a cikkben szereplő többi internetes forrás címe

- Centre for English corpus linguistics:  
<http://cecl.fltr.ucl.ac.be/Cecl-Projects/Iccl/icle.htm>
- Cobb, Thomas (2007): *The compleat lexical tutor*. Version 5.  
<http://www.lextutor.ca>
- Horváth József (2007): *Take off: Tips and tasks for readers and writers*.  
<http://take-o-f-f.blogspot.com>
- The JPU corpus* (2007):  
<http://joeandco.blogspot.com>
- Kaszubski, Przemek (2003): *The PICLE corpus*.  
[http://ifa.amu.edu.pl/~kprzemek/PICLE\\_search.php](http://ifa.amu.edu.pl/~kprzemek/PICLE_search.php)

### A CLT három eszközeinek címe

- Konkordancia a CLT gyűjteményeivel:  
[http://www.lextutor.ca/concordancers/concord\\_e.html](http://www.lextutor.ca/concordancers/concord_e.html)
- Konkordancia saját szöveggel:  
[http://www.lextutor.ca/concordancers/text\\_concord](http://www.lextutor.ca/concordancers/text_concord)
- Szókincs-profil:  
<http://www.lextutor.ca/vp/bnc>