Weöres Sándor költői nyelvének számítógépes feldolgozása

Nagy János[1], Alexin Zoltán[2]

[1] Szegedi Tudományegyetem, JGYTFK, Magyar nyelvi tanszék, Szeged, Boldogasszony sgt. 6., e-mail: nagy@jgytf.u-szeged.hu
[2] Szegedi Tudományegyetem, TTK, Informatikai Tanszékcsoport, Szeged Árpád tér 2., e-mail:alexin@inf.u-szeged.hu

Kivonat A cikkben bemutatásra kerül egy Weöres Sándor költői nyelvének számítógépes feldolgozásával foglalkozó projekt. A kutatás 1999 és 2002 között folyt, eredményeképpen kb. háromezer számítógépes oldalnyi korpusz jött létre, elemezve a HuMor magyar szóalaktani elemző programmal. Eredmény az adatbázis-formátum, amely lehetővé teszi a korpusz eltárolsát és a gépi feldolgozás megalapozását. Rendelkezésre áll Weöres szövegeihez a metrikai-ritmikai struktúra leírása. A kutatás távlataiban meghatározható a magyar nyelvű lírai textusok eltárolására alkalmas adatbázis-formátum; leírható a hangzás formalizálható adatainak digitális metakommunikációja; megadható a szükséges szoftverek száma és munkaerő szükséglete, – és megrajzolhatók a számítógépes elemzés korlátai.

Kulcsszavak: számítógépes szövegreprezentáció, természetesnyelv-feldolgozás, poétikai jellemzők adatbázisa

1. Bevezetés

Weöres Sándor költői munkásságának számítógépes feldolgozása 1999-ben kezdődött el a Szegedi Tudományegyetem Juhász Gyula Tanárképző Főiskolai Kar Magyar nyelvi tanszékén. [3]

Az előfeltételek között szerepel az az évtizedes filológiai kutatás, amely Weöres költészetének modern feldolgozását végezte, s 1991 és 1999 között sikeresen megvédett kandidátusi disszertációt, két monografikus elemzést (1996, 1998), számos tanulmányt és előadást eredményezett (Nagy L. János).

A számítástudományi előfeltételeket Alexin Zoltán kutatói műhelye, az Informatikai tanszékcsoport biztostotta, amely elismert magyar nyelvtechnológiai centrum. A tanszékcsoport kutatói a számítógépes szövegreprezentáció területén értek el eredményeket[1][2]. A nevükhöz fűződik a Szeged Korpusz elkészítése. A korpuszhoz használt szövegreprezentációs technológia teremti meg az alapot Weöres Sándor verseinek számítógépes feldolgozásához.

[3] A szerzők köszönetüket fejezik ki az Országos Tudományos Kutatási Alapnak: az ismertetésre kerülő kutatást az **OTKA T 029 431** sz. pályázttal támogatta.

A kutatást tanácsaival támogatta a nemzetközi hírű szövegkutató Petőfi S. János és a filológus akadémikus Fónagy Iván. Az alábbiakban a ismertetjük a kutatás fontosabb célkitűzéseit és eddigi eredményeit.

2. A projekt fontosabb célkitűzései

A munkaterv Weöres Sándor költői nyelvének számítógépes feldolgozását tűzte ki célul. Ehhez az alapcélhoz számos részcél kapcsolódott.

A *filológia* területén először azt kellett tisztázni, mit is jelent a "Weöres Sándor költői nyelve" terminus. A négyéves kutatási ciklus második felére, az MTA Irodalomtudományi Intézetében tartott előadás [4] diszkussziójában vált nyilvánvalóvá, hogy

- a virtuális kritikai kiadás létrehozása az első lépés, azaz az *Egybegyűjtött írások* és az *Egybegyűjtött műfordítások* szolgálnak az 'editio princeps' alapjául, s ezek korpuszához adódnak a Weöres halála után kiadott munkák, illetve a költő életében keletkezett, de az *Egybegyűjtött írások* és *Egybegyűjtött műfordítások* anyagába be nem került szövegek (részletesebben l. a [5] 3. fejezetében);
- a változatok (az *Egybegyűjtött írások* és *Egybegyűjtött műfordítások* előtt, illetve után publikált, kisebb-nagyobb eltérésekkel közölt szövegek) elemzése akkor valósulhat meg, ha ezt (az elbírálás folyamatában lévő) a célt is támogatja a kutatásfinanszírozás, – kétségtelen, hogy az igazi virtuális kiadás csakis eképpen valósítható meg;
- a változatok és az 'editio princeps' összevetése, a tanulságok adatokon alapuló, szolid összegzése feltétlenül szükségessé teszi a metrikai-ritmikai elemzések elvégzését, – kétségtelen, hogy ennek finanszírozása külön is szükséges (l. az előző bekezdést);
- várható bár, hogy a továbbiakban napvilágra kerülnek olyan szövegek, amelyek eddig kiadatlanok voltak, – jelentős számú ilyen textus előkerülése azonban nem valószínű.

Összefoglalóan elmondható, hogy a nálunk számítógépen (és CD-ROM-on) tárolt Weöres-szövegtár 'editio princeps'-e a lehetőségek szerint teljes; míg a variánsokról ezt nem állíthatjuk.

A számítógépes feldolgozás alapvető munkafolyamatai következők voltak:

- mindenekelőtt szkenneléssel elérhetővé tettük a szöveganyagot a számítógép számára: a Recognita program használatával a kötetek betűtípusaihoz is alkalmazkodhattunk ("első változat");
- a szkennelt anyagot ellenőriztük és javítottuk, majd az eljárást megismételtük, – jellemző, hogy a metrikai-ritmikai elemzés is segített néhány rövid-hosszú magánhangzó értékének meghatározásában a poétikai funkció(k)nak megfelelően ("második változat" és "harmadik változat");
- a javított és ellenőrzött textus lexikai anyagát a HuMor magyar nyelvi morfológiai elemző programmal elemeztettük, ezzel létrejött az elemzett változat;

- az elemzett anyagon gyakorisági vizsgálatot végeztünk nagyobb szövegegységenként, hogy összehasonlítható eredményekhez jussunk, azaz a lírai, a drámai és a prózai szövegeket külön-külön és együttesen is adatoltuk;
- az így nyert, gyakoriság szerint rendezett szóanyagon kvül megmaradt lexikai korpusz további finomítási feladatokat ad, a nyelvek szerinti tagolásban külön-külön tároltuk a képzelt, illetve az idegen nyelvű részhalmazt;
- a további célokat – eredeti célkitűzéseink szerint is – kb. hároméves munkával érhetjük el.

Jelenleg világosan látható, hogy a kutatási célt érdemes kissé kiterjeszteni. Eszerint az egyik bővítés az összehasonlítható metrikai-ritmikai adatbázis létrehozását, a másik a virtuális kritikai kiadás megteremtését eredményezheti. A két feladat össze is függ: az egyik a másik nélkül kevésbé jelentős. Egy-egy vers lexikai szövetének a változataiban magától értetődően vannak jelen a kicserélődő szavak, szórészek morfémáinak a hangzásban, a ritmusban bekövetkező variánsai. Nem is szólva azokról a (rövidebb) versekről, amelyeknek a létrejötte is közvetlenül összefügg dallamokkal, dallamtöredékekkel, – a kompozíciójuk pedig akár zenei műformákkal.

Ha van egyáltalán olyan költőegyéniség, akinek a lírájában a poétikai textus elengedhetetlenül, lényegében tartalmazza metrikai-ritmikai hangzását, azt a hangzást, amely csakis a jelentés csorbítása révén változtatható, – nos, Weöres Sándor éppen ez a poéta. Praktikus szempontból pedig: kár volna a megteremtett szövegbázistól kínált lehetőséget kihasználatlanul hagyni.

Ami pedig azt a kérdést illeti, hogy a Saussure utáni kor areferenciális gondolkodásmódja a véletlen, a szórtság, a játék ,,végtelen szöveg" fogalmával operál, s a retorikai megközelítésnek ad elsőbbséget a szemantikaival szemben a nyelvészeti kutatásban, – a Weöres Sándor-i ,,röpülő vers" éppen avantgárd gyökereivel, totalitásra törekvésével kínálkozik mintául. Mintául, amely számos különböző megközelítést visel el.

2.1. A tárgykörben kidolgozott elméletek, módszerek és eljárások

A tárgykörben alkalmazott nyelvészeti és poétikai elméletek, módszerek a korpusznyelvészetben ismertek: a *filológiai* feladatokban Gáldi Lászlóra név szerint is hivatkoztunk. [4] A metrikai-ritmikai elemzésben a szimultán verselésnek sajátos, Weöres Sándor által (Marsall Lászlónak küldött) levelekben megírt módszereit igyekeztünk követni. [5] A filológiai eredmények elérésében kulcsszerepük volt azoknak a hallgatóknak, akik speciálkollégium keretében ritmizálták a szöveganyagot; az adott metrikai-ritmikai struktúrák segítségével lehetőség nyílik a számítógépes elemzésben az igen távlatos, Weöres Sándor költészetén messze túlnyúló vizsgálatokra. Ezek a szövegfüggetlen struktúrák összevetéseiben ragadhatnak meg olyan belső összefüggéseket, amelyek a lexikai textus közbeavatkozása nélkül analizálhatják a formai egyezéseket és különbözőségeket.

[4] A Petőfi-szótár előszavában kövonalazott elvekről van szó, Gáldi László 1972. I: 7–18

[5] L. még Szuromi Lajos: *Szimultán verselés* Akadémiai Kiadó, 1990.

A *számítástechnikai* feldolgozás számára alapkövetelmény a biztonságra törekvés, az ellenőrizhetőség, az eljárások folyamatos tesztelése, az adatok pontossága. Ennek eredményeképpen meggyőző eredmény a gyakorisági elemzés, finomításán tovább dolgozunk (remélhetőleg az újabb OTKA-támogatás segítségével is). A további feldolgozás számára is nélkülözhetetlen a folyamatos kontroll: ebben az elemzésben is számos indexálási megoldás alapján lehetséges a számítógépes adatrendezés.

A költői nyelv feldolgozására magyar nyelvre kidolgozott adatbázis formátum nem áll rendelkezésre. A világban elérhető adatbázisok közül megvizsgáltuk azokat, amelyeket eredetileg angol nyelvű feldolgozásra fejlesztettek ki. A számunkra és a feladatnak leginkább megfelelőt, a TEI (Text Encoding Initiative)[6] konzorcium által ajánlott formátumot választottuk ki. Ennek a magyar nyelvű textus igényeihez adaptálása, a szükséges változtatások elvégzése áll a folytatás középpontjában.

3. A retorikus nyelvhasználat Weöres Sándor költészetében

Weöres Sándor tipikus XX. századi költő, azzal a retorikus sajátsággal, hogy hatásával kiszakítja olvasóját a mindennapok prózai világából: ,,emberfölötti áramok"-at alkalmaz. Ezek az emberfölötti áramok a vers hangzásvilágának hatásaként érvényesülnek, ahogyan maga Weöres fogalmaz: ,,a vers tartalmilag fogalmi, formailag auditív művészet". A költő törekvéseiben jelen van a totalitásra törekvés is: a teljes univerzumot kívánja megragadni, az emberiség történelmét, kultúráját a tájak és a korok egységében, ugyanakkor mindenfajta kötöttségtől függetlenül.

Nagy L. János harmadik Weöres-könyvében [7] [5] a retoricitás klasszikus és modern fogalmából indul ki. Legfontosabb elemzései a chiasztikus struktúrákat, a költői hangzásokat, a szövegvariációkat és a többrészes kompozíciókat kutatják. Az értelmezések a verselés metrikai-ritmikai szerkezetének, az értelmi és ritmikai hangsúlyoknak a jelentésképző szerepét kutatják.

Nagy L. János meghatározott elméleteket alkalmaz: a funkcionális stilisztikát, a szemiotikai textológiát, a szimultán verselést. Az értelmezés tényeihez és a következtetésekhez a szintaktikai, szemantikai és pragmatikai oszcillációk segítségével jut el. A formailag is megragadható azonosságok és különbségek szolid alapot adnak annak megközelítéséhez, ami a számítógépes elemzésre támaszkodva, de csakis a számítógép nélkül, a befogadói intellektus interpretáló tevékenységében ragadható meg.

---

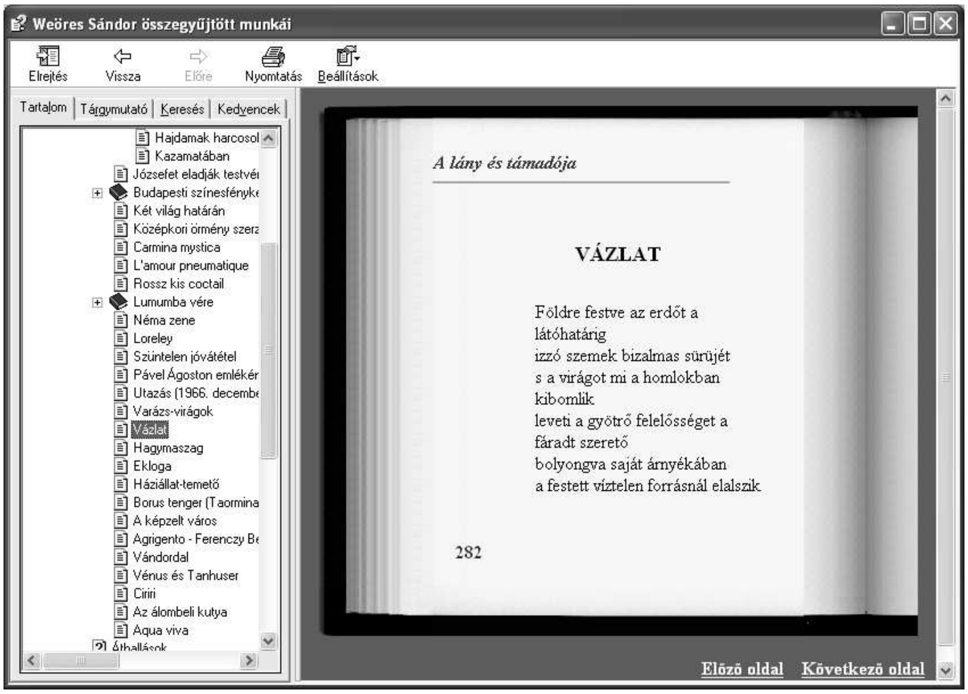
[6] A TEI konzorcium honlapja: http://www.tei-c.org

[7] Ezt a könyvet a kiadó és a szerző Weöres 90. születésnapjának tiszteletére szánta, 1913. június 22.

4. Eredmények és további lehetséges kutatási irányok

Az eddigi eredmények a XX. századi költői nyelv feldolgozásában korábban hasznosított eljárásokra épülő módszerekkel születtek, ilyen módon sokoldalúan öszszehasonlítható és felhasználható tanulságaik vannak. A további célok eléréshez annak függvényében juthatunk el, amilyen mértékben támogatja az OTKA (és más lehetséges finanszírozó) a terveket, pl. az informatikai és adatbeviteli tevékenységek honorálásával.

Az összegyűjtött korpuszhoz Papp Endre programtervező matematikus hallgató diplomamunkája keretében Alexin Zoltán irányításával egy webes megjelenítő programot készített. Ennek egy képernyője látható az 1. ábrán.



1. ábra. A Weöres korpusz webes megjelenítő modulja

Az eredmények alapján az OTKA pályzat kezdetén megfogalmazott célok jelenleg is érvényesek, azaz a folytatásban is követendők:

a) Weöres Sándor 'editio princeps'-ének gyakorisági szótára befejezéséhez szükség van kb. kétévi munkára és egyévi ellenőrzésre; a tevékenység finanszírozására újabb OTKA-pályázatot nyújtottunk be;
b) a szövegvariánsok digitalizálása megteremti a virtuális kritikai kiadás textusát, ezzel az 'editio princeps' kiegészül a szövegkörnyezettel;

c) a metrikai-ritmikai elemzés adatait számítógépbe visszük (ez a tevékenység csak hallgatók bevonásával képzelhető el); erre irányul a pályázat második főiránya;
d) a számítógépes feldolgozás fentebbi lehetőségei közül a kiválasztott adatbázist adaptáljuk a magyar nyelvű szövegek igényeihez; az adatokat bevisszük és a feldolgozásra előkészítjük: erre irányul a pályázat harmadik célcsoportja.

A 2002-ben lezárult OTKA vizsgálatban sajátos feldolgozási problematika jelent meg. Weöres Sándor ugyanis abban is kitűnt kortársai közül, hogy gyakran használt az írásaiban az élő (és holt) nyelvek szókincsén kívül képzelt nyelvű lexikai anyagot. Ilyen módon a lexikai-morfológiai elemzés a feldolgozáskor nyilvántartotta ugyan ezeket az adatokat, de megelégedett a regisztrálásukkal, nem elemezhette a szóalakok felépítését.

Ennek a lexikának van "eredetiben és fordításban" egyaránt közölt része, pl. az 1944-es *Barbár dal* (alcíme: *Képzelt eredeti és képzelt fordítás*). Az 1960-as évek végén ennek a két változatnak az egymásra vonatkoztatása segítségével a jelenleg Párizsban élő *Nyéki Lajos* elkészítette a "képzelt eredeti" szótárát is. Ugyancsak egymás mellett közölte Weöres az 1946-os *Arany kés forog* (korábbi címe: *Örvényben, naur glainre iki kezdetű*) kétféle textusát, ennek azonban szótárszerű feldolgozása nem készült. A két említett vers publikálása abban tér el egymástól, hogy a laptükörben a *Barbár dal* nyomtatott formában a képzelt nyelvű anyag mellé teszi a képzelt fordítást, az *Arany kés forog* pedig a magyar textus mellé helyezi a képzelt nyelvű sorokat.

BARBÁR DAL
(Képzelt eredeti és képzelt fordítás)

Dzsá gulbe rár kicsere
áj ni musztasz emo
áj ni mankütvantasz emo
adde ni maruva bato! jaman!

Ole dzsuro nanni he
ole csilambo ábábi he
ole buglo iningi he
lünlel dáji he! jaman!

Vá pudd shukomo ikede
vá jimla gulmo buglavi ele
vá leli gulmo ni dede
vá odda dzsárumo he! jaman!
(1944)

Szél völgye farkas fészke
mért nem őriztél engem
mért nem segítettél engem
most nem nyomna kő! ajaj!

Könnyemmel mosdattalak
hajammal törölgettelek
véremmel itattalak
mindig szerettelek! ajaj!

Földed tüskét teremjen
tehened véres tejet adjon
asszonyod fiat ne adjon
édesapád eltemessen! ajaj!

Képzelt ősi kultúra képzelt ősi nevei jelennek meg a *Mahruh veszése* című szövegben (1952): bevezetést, 101 négysorost és a *Negyven király éneke* című, hat tömbre tagolódó befejező verset tartalmaz. Weöres a képzelt lantost *Rou Erou-*

nak, azaz *Bíbor Lángnak* nevezi. A négysorosok névanyagából néhány példa: *Kartiabh-todarh* (1.), *Lulobh* (4.), *Zölülügh* (11.), *Jöhpruk* (50.), Nuantompank (94.) stb. A tájnevek és személynevek egyszer-kétszer fordulnak elő mindössze.

Avantgárd szövegeiben Weöres gyakran tördeli morfémákra, morfémarészekre a szavakat; a hangzási kísérletekben számos nonszensz textust is ír: *Tapéta és árnyék* (1963), *Egérrágta mese* (1960), *Az áramlás szobra* (1944), *Az égő szótár* (1967), *A megmozdult szótár* (1949-53), *A levegő morzsái* (1978). (L. Károlyi Amy: Weöres Sándor és az avantgárd. *Holmi*, 1990: 1003.) A *Kilencedik szimfónia* (1955-60) *Finale* tételében az 1956-os forradalmat leverő orosz páncélosokat idézi: *"füstöl a tábornok / csikorgó szeke- / recsegő kere- / köhögő mene- / kürtöl a kolosszus / a kolosszusnak a"*. A *Táncdal* (1942) az egyik legismertebb hangzó nonszensz (*panyigai panyigai panyigai / ü panyigai ü*).

A *Wawiri* (1942) című gyűjtemény darabjaihoz hasonlít Weöres *Hangcsoportok* című háromrészes kompozíciója (1941), részei: *Puha, forró hangok (Ange amban ulanojje...)*; *Gyors, gyöngyöző, vidám hangok (Vikulili hejriri sziggaga...)* és *Áradó, sugárzó hangok (Khúnái áfháiszthái mengoh...)*. A francia *écriture automatique* ('automatikus írás') módszerével is jó néhány Weöres-szöveg született, pl. In memoriam Devecseri Gábor (jellemző az alcím: *Három repedezett szikla- vagy felhőgomoly-alakzat).* A második rész (*Post Mortem*, 1971) egyik sora: *"ílyenkor elborít geometriai a az"*.

Összefoglalóan megállapítható, hogy a poétikai elemzésben pontosabb képet nyerhetünk majd ezeknek a lexikai elemeknek a szerepéről is. A hangzó lexikai elemekre is érvényes Weöres egyik nyilatkozata: "Mindig ugyanúgy viadalban, viaskodásban voltam vele, mint mindennel, ami érdekelt ... hogy milyenek a hangzói, milyen a struktúrája, ez érdekelt, ez foglalkoztatott." (A költészet hivatása. Tardos Júlia rádiós beszélgetése Weöres Sándorral. 1972. június 25. In: Domokos Mátyás (szerk.): *Egyedül mindenkivel.* Szépirodalmi, 1993: 216.)

4.1. A távlatok a következő *filológiai* eredményekkel kecsegtetnek

- Létrehozható az első olyan magyar nyelvű verskorpusz, amely a lexikával együtt tárolja a metrikai-ritmikai, hangzási annotációkat.
- Ez lehetőséget nyújt a lexikai és a metrikai-ritmikai oldal együttes megjelenítésére egy alkalmas szoftverrel.
- Mindez könnyen kezelhetővé, kereszthivatkozásokkal is nyitottá teszi az eddig mintegy 3000 számítógépes oldalnyi anyagot.
- A változatok és az 'editio princeps' lexikai és ritmikai tényezői közelebb hozzák a befogadóhoz és a kutatóhoz a költői alkotófolyamat fázisait.
- A kutatás egyértelműen és pontosan meghatározhatja a szövegváltozatok közötti különbségeket.
- Meghatározható, melyek azok a poétikai jegyek, amelyeket érdemes Weöres költészetében figyelembe venni; – ugyanakkor melyek azok, amelyeket más költők esetében is figyelembe kell venni; - és melyek azok, amelyek vizsgálata nem érdemel figyelmet.
- Meghatározható, hogy pontosan mely poétikai jegyek (és értelmezésük) kutatható számítógppel, melyek pedig nem: azaz meghúzhatók a gépi elemzés kompetenciájának határai.

A korpusznyelvészet jelentős informatikai lehetőségeiből:

- Megteremthető az első olyan magyar verskorpusz, amely szabványos adatbázis formátumban tárolja egy nemzetközi hírű magyar költő összes textusát.
- Nagy valószínűséggel megtervezhetővé teszi magyar verskorpuszok számára az optimális adatbázis-formátumot: adatbázis-változatokat segítene elő.
- Szoftverek kszülhetnek, amelyek támogatják a rögzítési, feldolgozási munkát.
- A korpusz létrehozása és a szükséges szoftvermennyiség és -minőség tapasztalatokat kínálna az adott célú korpuszok előállításának optimalizálására.
- Országosan terjeszthető CD-ROM változatban nemzeti kulturális értéket teremtene (beleértve angol, német stb. nyelvű ismertetését).

Hivatkozások

1. Alexin Z., Csirik, J., Gyimóthy, T., Bibok K., Hatvani, Cs., Prószéky, G., Tihanyi, L.: *Manually Annotated Hungarian Corpus*, in Proceedings of the Research Note Sessions of the 10th Conference of the European Chapter of the Association for Computational Linguistics EACL'03, Budapest, Hungary, 53–56 (2003).
2. Csendes, D., Hatvani, Cs., Alexin, Z., Csirik, J., Gyimóthy, T., Prószéky, G., Váradi, T.: *Kézzel annotált magyar nyelvi korpusz: a Szeged Korpusz*, Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2003), Szeged, Magyarország, 238–245, (2003).
3. Nagy L. János: *Költői szótár a XXI. században*, előadás a Magyar Alkalmazott Nyelvészeti Konferencián, Pécs, Magyarország, 2001. április 5. p. 8.
4. Nagy L. János: *Filológiai problémák Weöres Sándor költői nyelvében*, előadás az MTA Irodalomtudományi Intézetben, Budapest, Magyarország, 2001. április 11. p. 24.
5. Nagy L. János: *A retorikus nyelvhasználat Weöres Sándor költészetében*, Akadémiai Kiadó, Budapest, 2003. 284 oldal + bibliográfia
6. Nagy L. János: *Poésie musicale ou musique poétique? Sur le compositions*, S. Weöres, F. Sebő: *Revue d'Études Francaises*, Budapest, **3** 181–195, (1998)
7. Nagy L. János: *A szintaxis poétikájához*, In Schaeffner A., Bene K., Madácsy P. (szerk.): *Új tendenciák a komparatisztikában IV., Nouvelles tendences en littérature comparée IV.* JGYTFK, Szeged-Amiens, 291–304, (2004)
8. Nagy L. János: *Chiasme: a semiotic outline, Sprachteorie und germanische Linguistik.*, Kossuth Egyetemi Kiadó, Debrecen, **1** 19–35, (2004)