

# Absztrakt szöveg: az elektronikus szövegek alapjai

- szerző: Karsai Róbert
- verziószám: 7
- utolsó módosítás: 2003.05.29.
- webcím: <http://magyar-irodalom.elte.hu/robert/szovegek/>
- e-mail: robert@magyar-irodalom.elte.hu

## Tartalomjegyzék

- Bevezető
- Az írás és tárgyai
- Absztrakt szöveg
- Szegmentálódás
- A szegmentumok felismerése
- Struktúra vagy kép?
- Szövegrögzítés
- Karakterhelyesség
- Nyelvi technológia
- Gépszerűtlen helyesírási kérdések
- Írásjelek
- Tipográfia
- Összefoglalás
- Jegyzetek
- Felhasznált irodalom

## Bevezető

Az írás az élőszó, a beszéd absztrahálásának eredménye, amely tartóssá, térben és időben kötetlenebbé tette a szövegeket. Dolgozatommal arra szeretném felhívni a figyelmet, hogy az írott és nyomtatott szövegek, illetve elektronikus társaik között érdemes egy újabb elvonatkoztatási szintet feltételezni, és arra, hogy ez az újabb szint nem a számítógépes szövegek megjelenése óta létezik, a gépi feldolgozás csak a szövegekről való differenciált gondolkodásmódunk formalizálásában nyújt segítséget. A gyakorlati jellegű fejezetek egy része a magyar szövegek elektronikus rögzítésének apróbb fonátságaira és az egységes magyar előírások hiányára figyelmeztet.

## Az írás és tárgyai

Számunkra az írás: betűírás, amelynek egységei legtöbbször nyelvspecifikus fonémákat jelölnek. Természetesnek tartjuk, hogy a szó hangsorból áll, a hangsor elemeit pedig egy-egy fonéma realizációjának tekintünk, külön jellel írjuk, ennek felismerése azonban nem egyszerre történt az írás felfedezésével [1].

Az írás előzményeiről keveset tud a tudományos világ, a civilizált népek népszokásait, illetve az írást nem ismerő afrikai, ausztráliai és csendes-óceáni természeti népek körében végzett kutatások nyomán azonban már eddig is többféle beszéd utáni, írás előtti gondolatközlési módszerre derült fény, ilyen a tulajdonjegy, a rováspálca, a hírnökbót, a csomójelek, kagylófüzérék, de efféle eszköznek tekinthető a régi magyaroknál a véres kard is [2]. „*Ösztönös rajzok vezették vezeték el az embert lassan ahhoz a felfedezéshez, hogy a rajzokat gondolatai közlésére is felhasználhatja*” [3]. A kialakuló írás feltételei: a gondolatközlési szándék és a beszédet helyettesítő jelek alkalmazása.

A valódi írás a *képirással* (piktográfia) veszi kezdetét, ennek legfőbb jellemzője, hogy konkrét tárgyakat ábrázol, amelyek jelölhetnek egy-egy szót, akár egész mondatot, és nincsenek nyelvhez kötve, mindenki úgy olvashatja és értelmezheti őket, ahogy tudja, a saját nyelvén. Ennek kései példái a dakoták csigavonalban bölénybőrre rajzolt Téli Krónikái a 17. századból. A képirás jelei tehát konkrét, kézzelfogható és látható tárgyak képei, a *fogalomírás* (ideográfia) túllép ezen, elvonatkoztatásra készítet [4], szimbolikus ábrái többet jelentenek a konkrétumoknál, használata közmegegyezést feltételez, a nyelvektől még mindig függetlenül. „*[A sumérok] már az i.e. 4. évezred utolsó századaiban ismerték az [ék]írást, [...] eleinte konkrét tárgyakat ábrázoltak nagyon leegyszerűsített stilizált formában, [...] az egyes jeleket azonban többféle értelemben is felhasználták, a szem és a víz jele együtt a sírást, könnyezést fejezte ki*” [5]. A sumér írás jellegzetes hordozói az agyagtáblák voltak. A történet következő fontos állomása a *szó- és szótagírás*. Lassanként az egyes jelek az egyes szavakkal a gyakori használat következtében szorosán összeforrtak, sőt egy idő után már nemcsak a jelentésre, de a szavak hangalakjára is utaltak (tehát egy adott nyelvre jellemző hangsor helyett állhatnak), az ilyen jelhasználatot szokás szóírásnak nevezni. A szóírással lehetőség nyílt arra, hogy több szótagú hosszabb szavakat rövidebb szavak jeleinek soraként, képrejtvényszerűen, szótagírással írjanak le [6]. A i. e. 2900-tól adatolható egyiptomi hieroglif írás különös keveredést mutat, ugyanis „*szóírás volt, de jelei közül néhányat szótagok, helyesebben mássalhangzócsoporthoz, sőt egyes hangok jelölésére is felhasználtak*” [7]. A hieroglif írás főleg kőbe vésve, illetve papirusztekercsen maradt fenn. A képirást, fogalomírást, szó- és szótagírást használó népeknek nehézséget okozott a világ, a környezet változó természete, az új tárgyak, új földrajzi helyek, új uralkodók neveinek leírása. A problémát további jelek bevezetésével, illetve *akrofóniával* is meg lehet oldani. Az akrofónia a szóírás jeleinek alkalmazása azok kezdőhangjainak hangértékében, az akrofónia útján jutott el az ember az egyes hangokat jelölő *betűíráshoz*. Az első betűírásokban megfigyelhető akrofóniát jól mutatja az i. e. 13-14. században kialakult föníciai írás, „*nyelvükben az ökör neve alef, és jelét az »a« hang írására használták, a házat béth-nek mondták és jelével írták a »b« hangot*” [8]. Mivel a nyelvek ugyanazt a közlendőt más és más hanganyaggal fejezik ki, a

betűírás pedig szoros kapcsolatban áll a nyelv hanganyagával, az így leírt szöveg nyelvspecifikus lesz.

A számítógépen előállított, mágneses vagy optikai módszerrel tárolt, karakterkódokból felépülő írás első pillantásra nem jelent újdonságot. A nyomtatástól kezdve az írás elveszti egyedi jellegét, de a lényeg változatlan. Annyit mindenesetre érdemes megjegyeznünk, hogy a számítógép nem magát az írást tárolja, hanem a jel jelét (karakterkódokat). Az, hogy a gépen tárolt kódokhoz mit rendelünk, pragmatikai kérdés, a használaton, a közmegegyezésen, de leginkább a számítógépre telepített operációs rendszeren, és az azt előállító programozókon múlik. Tekintsünk most ismét végig az írás történeke rövid vázlatán! Eddig az írás tárolása és megjelenése egyetlen tárgyhöz volt kötve, minden leírt közlendő/szöveg tárgy volt a szó leghétköznapibb értelmében (is). Rajz és barlang, rovás és pálcá, papír és betű, mikrofilm: egység. Ha az írás megkopik, vele kopik a hordozója, elválaszthatatlanok. Alkot-e egységet a számítógépen tárolt szöveg képe és hordozója?

## Absztrakt szöveg

@ 64	P 80	` 96	p 112
A 65	Q 81	a 97	q 113
B 66	R 82	b 98	r 114

A szöveg külön adattípus a számítógép számára, olyan adat, amely karakterkódokból építkezik. A *karakterek* az írásunkban általánosan használt grafikus jelek lényegét kifejező helyettesítő jelek, ezek gyakran megegyeznek a betűkkel, de a magyar ábécére vonatkoztatva (gondoljunk csak többjegyű betűinkre) nem minden esetben. Karakter áll a számjegyek, egyéb jelek (pl. kereszt, csillag) és írásjelek (vessző, pont, kérdőjel) helyén is [9]. A karakter a gép számára egy szám alakjában testesül meg, a Ő jelölésére manapság például a 213-as szám szolgál. Egy nyelv szövegeinek leírásához szükséges karakterek képét és a hozzájuk

rendelt számkódokat úgynevezett *kódtáblákban* [10] szokás definiálni. Kódtáblából végtelenül sok van, a pillanatnyilag érvényben lévő magyar szabvány szerint elfogadott magyar kódtáblák: az ISO 8859/2, az IBM 852, és az EBCDIC. A mellékelt ábrán egy kódtábla részlete látható [11], amelyben a jelölt karakterek egy lehetséges képe és a hozzájuk rendelt számkód van feltüntetve. Bár az alapok egyszerűek, egy karakterhez egy számkódot rendelünk, a gyakorlatban széttartás figyelhető meg, a különböző kódtáblák más-más számkódokat párosítanak ugyanahhoz a karakterhez, az adott ország, nyelv vagy a használt szoftver fejlesztőjének szükségleteihez igazodva. Egy kódtábla karakterei alkotják az adott kódtábla *karakterrepertoárját*. Ugyanazt a karakterrepertoárt akár többféle kódtábla is leírhatja, az egyikben az Ő-t a 213-as számhoz [12], a másikban esetleg a 336-oshoz rendelve [13]. Nemzetközi szinten, a számítógépes hálózatok szintjén, bizony könnyen belezavarodhat az ember a különböző kultúrák különbözőképpen kódolt szövegeinek visszafejtésébe. Ezt megkönnyítendő születtek a karakterek átviteléről és az információcsere kódjairól szóló nemzetközi szabványok, amelyek akár felületes ismertetésétől is megkímélem az olvasót, hiszen érthető, logikus bevezetést sokan írtak már a témában, jó áttekintést találhatunk Prószéky Gábor és Kis Balázs könyvében [14]. A

magyar szövegek írásához szükséges kódtáblákra, helyesebben azok hiányosságaira még kitérek a későbbi fejezetekben.

Fontosnak, kiemelendőnek tartom azonban a karakter jelölőjének (karakterkód) és jelöltjének (karakter képe) határozott elkülönítését. A *karakterkód* – például az Ő karakteré, 213 – a sok leírt Ő közös jellemzői helyett áll, egy elvonatkoztatással létrehozott entitás, egy szám, egy írásunkban általánosan használt jelet jelölő jelölő. Ezzel szemben a *karakter képe*, Ő, az absztrakt jel individualizációjával keletkezik, a számhoz betűt, típust, méretet, színt rendelve.

Az írásjelek efféle tárolása fellazítja a szövegkép vagy szövegfelület és annak hordozója közt eddig fennálló megbonthatatlan viszonyt. A hatékony működés érdekében alapelemeire választottuk szét a folyamatot, modularizáltuk az írást. Az elektronikus szöveg egy olyan rendszer, amelyben külön alrendszer tárolja a kódolt karakterszekvenciát (pl. mágneslemez, CD, DVD), külön alrendszer felelős a reprezentációhoz szükséges adatok biztosításáért (pl. betűtípusok vagy akár hangminták) és külön alrendszer állítja elő az érzékszerveinkkel felfogható kimenetet (hangot, képet). Bármelyik lecserélhető, változtathatunk a központi szerepű operációs rendszeren, amely összefűzi az alrendszereket; ha megtelt, akkor nagyobb kapacitású eszközre cserélhetjük a tárolásért felelős alrendszert; megváltoztathatjuk a szöveg betűtípusát vagy hangmintáját; a kimenetet pedig leolvashatjuk a képernyőről, vagy a frissen nyomtatott papírlapról, és közel az idő, amikor megkérhetjük a gépet: olvassa fel, hiszen a karakter kódját nem muszáj képpel összekapcsolnunk, a karakterkép az absztrakt fogalom (a karakter) egy lehetséges, de nem kizárólagos inkarnációja.

## Szegmentálódás

A szövegek megjelenésének, írásmódjának és tagolásának története szorosan összefügg az olvasás történetével, az olvasó sohasem magával a művel találkozik, hanem szöveget hordozó tárgyakkal, amit (f)elolvas. Az ókori Görögországban az olvasás felolvasást jelent, ahogy azt Jesper Svenbro az olvasás kifejezésére használt igék jelentésébe ('szétoszt', 'hozzátenni valamihez a kimondást', 'hang révén való felismerés') kapaszkodva kimutatja [15]. Az olvasás ekkor még bonyolult, lassú művelet, amit csak nehezít a folyamatos írás, a scripto continua, és a helyesírási szabályok hiánya. Mintha mondandónkat kizárólag a kiejtés elve alapján, szóközök és írásjelek nélkül íránk egybe. A csendes olvasás, a görög dráma egyes szövegeivel is bizonyítható módon, az i. e. 5. századtól jelenik meg, kiváltságos dolog, amit csak az „írás szakemberei” gyakorolnak. Az írások sorba rendezésének, a címadás rendszereinek és a szövegeket tagoló paragraphosz-jelek használatának szabályait a hellenizmus korában az alexandriai filológusok dolgozták ki.

A Római Birodalom évszázadai alatt új hordozó veszi át tekercsek helyét, a kódex. Bár ez a forma összekapcsolt füzetek, táblácskák, zsebkönyvek formájában ősidők óta ismert volt [16], csak az i. sz. 3-5. századtól mondható általánosnak a használata. A kódex átértelmezi a könyv fogalmát, a tekercs mindig egy lezárt, önálló egységet jelentett: egy könyvet, akkor is, ha a teljes munka több könyvből állt. A kódex jóval takarékosabb, olcsóbb, az anyag

mindkét oldalára írtak, ezért egy mű több könyvét tartalmazhatta, sőt több művet is, így lassan megváltoztatta a könyv fogalmát. A tekerecs folyamatosságát tehát a lapokra tördelt írás váltja fel, az ennek eredményeképpen a kialakuló szövegrészeket megjelölték, például a kezdőbetűk megnagyobbításával, így könnyebb volt a szövegben keresni [17]. Rómában az i. sz. 1. századig használták a szavakat elválasztó pontokat, a latin szövegekben mégis elterjedt a görög világban bevett scripto continua, majd a kódexek térhódításával megjelennek a szövegek egyéni befogadását segítő központosított kódexek.

A középkori olvasók és írók anyanyelve már nem a latin, nem az írás nyelve, ezért a szövegek könnyebb értelmezése érdekében több újítást is bevezettek. Korlátozták a ligatúrák számát, szóközöket iktattak be, a forrásokból átemelt uniciális és rusztikus írást megőrzik a minuszkulával másolt szövegekben, kialakítva az idézés konvencióját [18]. A könnyebben érthető, anyanyelvű szövegek azonban körülbelül a 13. századig scripto continuával íródnak [19]. A középkor első századaiban egyre nagyobb figyelmet szenteltek a csendes olvasásnak, már Szent Benedek Regulájában található erre vonatkozó utalások [20].

Hatalmas változások történtek a skolasztika korában. Az olvasás igazán tudatos tevékenységgé válik, a cél már nem a bölcsesség, hanem a tudás. Gyakorlati igényeket kielégítő újdonságok jelennek meg, amelyek tovább könnyítik a keresést és a használhatóságot: a tartalomjegyzék, konkordanciák, index, fejezetcím [21]. Az olvasás már nem lassú, elmélyült tevékenység, nem a szövegen való végighaladás, hanem csak szemelvények olvasása. A scriptoriumoktól lassan az egyetemi körökig terjed a csendes olvasás.

A könyvnyomtatás korára többé-kevésbé kialakulnak azok a szegmentumok, amelyekből mai szövegeink is állnak, már ismertek a mai könyv szokásos kellékei, amelyek segítik a könyvben való tájékozódást, bizonyos szöveghelyek megtalálását és a csendes olvasást, noha a mai értelemben vett nyomtatott könyvről csak a 16. század után beszélhetünk. A könyvnyomtatás igazi újdonsága az, hogy a kéziszajtó addig soha nem látott mennyiségű példányszám előállítására képes.

Ami a kezdetben az olvashatóságot és a könnyebb szövegkezelést szolgálta, annak egy része mára elvárás, a szöveghez szorosan kapcsolódó vagy ahhoz hozzátartozó önálló szövegrész lett. Ha hiányzik valamilyen megszokott egység, az feltűnő, zavaró, nem illeszkedik a szövegről kialakított felfogásunkhoz. Még ha nem ismerjük a szerzőt, az sem akadályoz meg minket abban, hogy a szöveg fölötti szerzőnyi helyre azt írjuk: ismeretlen szerző. Ha a szövegnek nincs címe vagy nincs felismert címe, automatikusan kitöltjük a cím számára fenntartott helyet a kezdősor zárójelbe tett szavaival. A szövegben minden új szakasz (legyen az versszak vagy bekezdés) új sorban, esetleg új oldalon kezdődik. Pusztán a szövegtér hatására egymással nem összefüggő, de nyelvileg és tipográfiaiilag jólformált szavak, mondatok sorozatát a szövegtérrel függően (pl. rövidebb sorokból álló csoportok egymásutánját versként) kezeljük, értelmet keresünk benne.

## A szegmentumok felismerése

A szöveg mélyén rejlő képet észrevétlenül, lassan felejtettük/felejtjük el, valahogy úgy, ahogy az író igénk eredeti ótörök jelentését. A szó jelentésszűkülése szinte kézzelfoghatóvá teszi a képtől a szövegig a bejárt utat. Az író a régiségben grafikai tevékenység, az első magyar nyelvű könyvből, a Jókai-kódexből már adatolható, jelentése: 'vonalat húz, fest, ír' [22]. Ma levelet, receptet, verset, dolgozatot: szöveget tudunk írni, nem értjük ki az a képiró. A szöveggép és ehhez kapcsolódóan a szöveg, mint tárgy szerepe ma talán csak a tudományos igényű szövegfeldolgozásban nélkülözhetetlen [23].

Hártó Gábor arra hívja fel a figyelmet [24], hogy a „szöveg grafikai mozzanata elfeledett”. Az írott szó „redukálhatatlanul grafikai” jellegét a „deviáns” tipográfiával, a képverssel mutatja be. A „nem deviáns” tipográfiában, a „normális” szövegekben a betűknek számos jegyben kell egyezniük, hogy a szöveggép rendezetlensége ne szúrjon szemet az első pillanatban. Vannak kivételes szövegrészek, például a cím, alcím, idézet, mindegyiknek sajátos, a szöveg más részeivel lehetőleg harmonizáló képet kell mutatnia. A „normális” szöveg betűi függőlegesen állnak, a sorok egymással és a lap szélével párhuzamosan futnak. A „deviáns” szöveg kihasználja, hogy – bizonyos zavaró tényezők ellenére is – az egymáshoz közeli írásjelek sorozatát szövegnek tekintjük:

*„ez a reflex olyan erős, hogy még a sík felületet sem tartjuk lényegesnek hozzá képest (vö. tejeszacskó), sőt azt sem, hogy a szöveg sorai egyenes vonalakat alkossanak [...] ezzel a tehetetlenséggel úgymond vissza lehet élni, és pontosan ez az, amit a képvers tesz.”*

Az előző fejezetben felvázoltuk a „nem deviáns” szöveggép kialakulásának történetét a grafikus eszközökkel, térhasználattal nem tagolt szövegtől (scripto continua) a szegmentumokra (cím, fejezetcím, bekezdés, idézet stb.) bontott nyomtatványokig. Láttuk, hogy a rendezett tagolás a szöveg kezelésének egyszerűségét hivatott segíteni. Kétségtelen, Hártónak ebben igaza van, a leírt szöveg az érzékelés szempontjából képi információ. Gondolkodásunk szempontjából viszont nem az, az írásjelek a közvetlen képi ábrázolástól a betűírásig absztrahálódtak, és szerintem a grafikus módszerekkel, elhelyezéssel kiemelt szövegdarabok sem csak a keresést és a tájékozódást könnyítik ma már. A szövegek tartalma mellett kialakult egy azzal párhuzamos, de mégis önálló rendszer, amely egy újabb absztrakciós lépés az írás történetében. A környezettől függő grafikus információt, a középre igazítást, a nagyobb betűméretet például fejezetcímként értelmezzük, noha a cím, más környezetben, más alakban is megjelenhet, kiadványról kiadványra változik, egyik tipográfiai műhely így, másik másképp alakítja, száz évvel ezelőtt is más betűkkel volt szokás a címek szedése és ma is. Mégis: felismerjük, tudjuk miről van szó. Megtanultuk a kultúránkra jellemző, síkon kialakított téreloszlást feldarabolni, pontosabban fordítva, azért alakítottuk ki a rendezett térhasználat hagyományát, hogy ne kelljen az olvasás során fáradtságos munkával nekünk feldarabolni a szöveget, az így kialakult látványt pedig képesek vagyunk szövegtípus-kompetenciánk segítségével általánosítani, majd besorolni.

A képpers talán nem is a szövegfelismerő reflexszel, nem grafikai mozzanattal él vissza, hanem azzal az általánosító gondolkodásmóddal, amelynek eddig csak grafikai megvalósulásait láthattuk.

Az absztrakció igénye hozta létre a tartalmi struktúrát, a szövegtípusokat nem grafikusán tároló elektronikus szöveget.

## Struktúra vagy kép?

A eddigi szövegfelfogás középpontjában, ha egyre rejtettebben is, de mindig a grafikus mozzanat áll, az a megszüntethetetlen tulajdonsága a szövegnek, hogy képe, tipográfiája van. Ezt a felfogást híven tükrözik a szövegtárolás eddigi eljárásai, szinte mindegy, hogy miről olvassuk az ilyen szöveget: tekercsről, kódexből, képernyőről. Nem változtat a lényegen a számítógépes szövegtárolás karaktereken alapuló elve sem, a „szöveg” adattípus bevezetésével inkább jobban elrejtjük a szöveg mélyén rejlő képet (de csakis akkor, ha a szövegről képként gondolkodunk). Szöveggént érzékeljük az interneten mindenhol látható színes, elmosódó vagy be- és elúszó reklámfeliratokat, egy számítástechnikával és az internettel még csak ismerkedő embernek külön el kell magyarázni, hogy ezek a feliratok a gép számára „kép” adattípusba tartoznak, ezért nehéz őket „igazi” szöveggént kezelni és felismerni [25]. Az egyszerű, grafikai alapelvekre vezethető vissza az emberiség történetében minden eddigi szövegtárolási eljárás, beleértve az olyan számítógépes módszereket is, mint a PDF, a PostScript, vagy az ebook különböző formái [26]. A lényeg: van egy felületünk, ezen – az érzékelés szempontjából – grafikai jeleket helyezünk el valamilyen rend szerint, az ilyen szöveg létrehozása során csak a megjelenés a fontos. Nem kivétel ezalól (noha a struktúrát is képes egy bizonyos szinten tárolni) a szövegszerkesztővel írt szöveg sem, hiszen nem a struktúra reprodukálására koncentrálnak, a WYSIWYG-elv [27] alapján kizárólag a megjelenítés a fontos.

De a számítógép emellett elérhetővé tesz egy olyan szövegábrázolási szintet, ami már nem is ábra, hanem az elvonatkoztatással, a szöveg szegmentálásával kialakult struktúrájának tárolása. Egy verset például a versszakok, azon belül a verssorok szerint, egy dolgozatot pedig fejezetcímek, bekezdések, idézetek, hivatkozások mentén tárolhatunk. A felosztási rendszer *nem normatív*, a szövegünk szegmentálása mindig a környezet és a felhasználás függvénye. Eddig is gondolkodtunk így szövegeinkről, e gondolkodásmód kézzelfogható bizonyítéka egy-egy igazi könyvsorozat, amelynek minden kötetében ugyanazon elvek szerint alakították ki az egyes szegmentumok tipográfiáját. A lényeges különbség az, hogy eddig csak gondolkodhattunk ilyen módon, ami ténylegesen a papírra került, az ennek az árnyéka volt. Mára elterjedtek azok a számítógéppel alkalmazható szövegleíró formanyelvek, amelyek segítségével úgy adhatjuk meg szövegeink szegmentumait, hogy közben nem hozunk létre, és nem is tárolunk képi információt. Az így létrehozott szövegnek többféle grafikus és nem grafikus reprezentációja is lehetséges, ugyanazon forrásból.

A karakter fogalmának bevezetésével a kép és a ténylegesen tárolt jel eltávolodott egymástól. A nyelv beszélt nyelvtana alapján a karakterkódokból álló szekvenciából

közvetlenül akár hangsort, értelmes hangzó szöveget is nyerhetünk. A számítógépen tárolt szöveg adattípus levetkőzte a grafikai mozzanatot. A szöveg szegmentumokban való tárolása pedig arra használható, hogy a szövegekről való gondolkodásmódunkat, a bekezdésekre, fejezetekre tagolást ne grafikusán, hanem elvonatkoztatással, egy szövegstruktúrát leíró mesterséges nyelven fejezzük ki. Vagyis még egy elvonatkoztatási szintet építettünk be a grafikus képtől való teljes elszakadáshoz, a szöveg szerkezetének megtartása mellett.

Példaként bizonyos SGML-alkalmazásokat [28], a szigorúan struktúráróló, megjelenítési információt nem tartalmazó HTML-t (ritka az ilyen a hálózatot felépítő sok terabájtnyi anyagban) vagy a Linux Documentation Project DocBook nevű leírónyelvét; XML-alkalmazásokat, esetleg a szövegeket szegmentumalapúan, struktúracentrikusan tároló adatbázist tudom említeni. Szemléltetésképp egy lehetséges, szótárak leírására szolgáló nyelven így fest egy képzeletbeli szótár szócikke:

```
<szocikk>
<szo>Pavarotti</szo>
<jelentes>Világhírű olasz tenor</jelentes>
</szocikk>
```

A tárolásnak ebben az esetben nem feladata semmilyen – az érzékelés számára – grafikus információ raktározása. Karakterkódokat és szövegstruktúrát tárolunk. Az ilyen szövegnek a kép csak egy lehetséges megvalósulása, de nem kizárólagos. Ma még főleg képi megvalósításokról tudunk, de a szöveg e tárolási mód segítségével elképzelhető bármilyen, a nyelv kettős tagoltságába beilleszthető érzékletként, például hangként (felolvasószoftverrel) vagy bőrérzékletként (braille).

## Szövegrögzítés

```
AAAA
aaaaaaa
aaaaaaaaaaaaa
aaaaaaaaaaaaa
aaaaaaaaaaaaa
aaaaaaaaaaaaa
aaaaaaaaaaaaa
aaaaaaaaaaaaa
aaaaaaaaaaaaa
aaaaaaaaaaaaa
```

Tételezzük fel, hogy van egy szövegcsoporthoz, a csoport mindegyik tagja hasonló felépítésű, a szövegeket az olvasók többsége hajlandó versként felismerni. Verseink nem véletlenül kerültek egy csoportba, versciklust alkotnak, amit most számítógépen szeretnénk tárolni. A képzeletbeli ciklus verseinek vannak közös, a forrásban grafikailag is jól elkülöníthető jellemzőik, ilyen a cím, az ajánlás, és az, hogy sorokból felépülő szakaszok, versszakok alkotják őket. Valahogy úgy festhetnek verseink, ahogy az a mellékelt ábrán is látható, bár nem mindegyik ilyen rövid. A cím kapitális és félkövér, az ajánlás kurzív, a versszakok között térköz. A gépi rögzítés során az első feladat a látható szöveg karakterkódokká (láthatatlanná) alakítása. Ha ezzel megvagyunk, bizonyára szeretnénk valahogy a forrásban következően hasonló, a ciklus versein áthatoló, szövegről szövegre ismétlődő, tipográfiaiailag megjelölt szegmentumokat is érzékeltetni. Alapvetően két lehetőségünk van, ez egyik a látvány reprodukálása, a másik a *felismert* szegmentumhatárok [29] reprodukálása.





```
<vers><cím>AAAA</cím><ajánlás>aaaaaaa</ajánlás><strófa><
sor>aaaaaaaaaaaa</sor><sor>aaaaaaaaaaaa</sor><sor>aaaaaa
aaaaaa</sor><sor>aaaaaaaaaaaa</sor></strófa><strófa><sor
>aaaaaaaaaaaa</sor><sor>aaaaaaaaaaaa</sor><sor>aaaaaaaa
aaa</sor><sor>aaaaaaaaaaaa</sor></strófa></vers>
```

A második nyelvnek számos előnye van. Ha van olyan programunk, amely helyesen kezeli az *ny2* elemeit, akkor csupán a rögzítési struktúra rendezett módja folytán olyan kvantitatív adatokhoz is hozzájuthatunk, amelyekre a rögzítés pillanatában még nem is gondoltunk. A teljes ciklus rögzítése után megkérdezhető mondjuk a versciklus összes strófáinak száma, vagy hogy átlagosan milyen hosszúak a versszakok. Akár rendezett listát is kérhetünk az ajánlásokban szereplő szövegekről vagy a versek címeiről, egyszóval: igazi többszempontú adatbázist hoztunk létre. Az *ny1* nyelv esetén mindezt csak nehezen, igazi programozói bravúrokkal tehetjük meg, hiszen a látványra koncentráltunk. Megpróbálhatjuk ugyan a versek ajánlásait kigyűjteni aszerint, hogy milyen szövegrészletek szerepelnek a *kurzív* elemen belül, de ha a vers a címében vagy bárhol a szövegben tartalmaz kurzivált részt, és ezt igazán nem tilthatjuk meg a szöveg képzeletbeli írójának sem, akkor azokat is megkapjuk a kérdezés eredményeképp. A rögzített adatok megjelenése *ny2* esetében is hasonlíthat a forráshoz, hiszen a *ny2* struktúraíró elemeihez könnyen lehet olyan tipográfiai szabályokat rendelni, amelyek az *ny1*-ben leírt megjelenést tökéletesen utánozzák. Az *ny2*-ben alkalmazott szerkezeti elemek természetesen nem használhatóak mindenféle vers tárolásához, *ny2* nem is előíró jellegű, pusztán az elképzelt korpusz adatai alapján konstruált lehetséges leírónyelv.

## Karakterhelyesség

Az elektronikus szövegrögzítés során lényeges szempont a karakterhelyesség: a szövegbe olyan a karakternek kell kerülnie, amilyenre szükségünk van, és nem olyannak, amelyiknek a csak a képe hasonlít rá. A gép számára az *O*, *Ó* és az *Ő* egyáltalán nem hasonló, nem egy alapjelből és annak mellékjeléből összeállított graféma. Ha ki szeretnénk használni a számítógép szótározásban, automatikus rendezésben, szóellenőrzésben, keresésben nyújtott, a karakterláncok (szövegek) kódszámainak összehasonlításán alapuló lehetőségeit, akkor karakterhelyes szöveget kell előállítanunk. A kódtáblák dokumentációjában épp ezért, hogy két karakter hasonló grafikus megjelenése ne lehessen megtévesztő, azt is feltüntetik, hogy pontosan mit jelöl egy adott karakter. Az alábbi rövid részlet a Unicode *General Punctuation* szekciójából (hex 2000-206F) való [30], és a leírások nélkül bizony nem is tudnánk mihez kezdeni ezzel a sok ponttal:

karakterkód	leírása
2024 (hex)	ONE DOT LEADER egy pontból álló sorkitöltő
2025 (hex)	TWO DOT LEADER két pontból álló sorkitöltő
2026 (hex)	HORIZONTAL ELLIPSIS írásjel: három pont
2027 (hex)	HYPHENATION POINT a szavakat elválasztó pont

A sorkitöltőt nem helyes a mondat végét jelző pont karaktere helyett használni, ahogy a

pontot (hex 2E) vagy a három pontot sem sorkitöltőként, mindegyiknek megvan a maga funkciója. A példa szándékolatlan szélsőséges, ám az igényes, nagy pontosságú kiadványokban betartandó.

A karakterhelyesség elve az egyik indok, hogy a Unicode-ban még nem létező, a tudományos életben, például nyelvészetben, irodalomtudományban és persze a mindennapi életben ma is használt speciális magyar jeleinket összegyűjtsük. Egy alapos igényfelmérés után haladéktalanul kapcsolatba kellene lépniünk akadémiai vagy akár állami szinten a Unicode fejlesztésén munkálkodó konzorciummal. A Unicode 1999-ben megjelent 3.0-ás verziójában a 16 biten váltójelek nélkül leírható 65536 karakterhelyből már 57709-et [31] lefoglaltak, és vészesen közel az utolsó pillanat, szinte most múlik, amikor még mi is kényelmesen beleférhetnénk az összes aktuálisan használt jelünkkel a 16 bitbe [32]. A feladat talán könnyűnek hangzik, pedig egészen biztosan hosszú, esetleg évekig tartó kutatómunkát és előkészítést fog igényelni, ha valaha is belekezdünk.

Irigylésre méltó az etruszk-kutatók helyzete, a Unicode 3.1-ről készült technikai beszámolóban [33] ugyanis az áll, hogy új szekció került a Unicode márciusban publikált friss verziójába, az *Old Italic* (hex 10300-1032F), amelyben az Appenini-félsziget ősi, történeti ábécéinek egy egyesített változata kapott helyet. Talán épp akkor foghattak hozzá e szekció munkáihoz, amikor egy hasonló, de magyar igényekre vonatkozó javaslat [34] is megjelent. A Unicode 3.1 az első olyan Unicode-változat, amely a 16 biten lehetséges 65536 (hex FFFF) karakterhely [35] fölé is helyez karaktereket, nem is keveset, az első körben 44946-ot, ezek között van a már említett *Old Italic* is, és ide kellene kerülniük a ma már nem használt, ám régi kéziratainkban, nyomtatványainkban még szereplő betűk karaktereinek, a Golden Dániel által javasolt *Magyar Nyelv Történeti Kódtáblájának*.

Valószínűleg 2003-ban fog megjelenni a Unicode 4-es verziója, előfordulhat, hogy már le is késtünk róla. A szabvány operációs rendszerekbe, programokba, fontokba való implementálása lassan halad, a legfrissebb előírások évek alatt kerülnek be a hétköznapi szoftverekbe. Milyen esélyeink vannak olyan karakterek használatára a közeljövőben, amelyek még a következő, a 2003-as Unicode-ban sem lesznek jelen?

## Nyelvi technológia

A nyelvi technológia olyan eszközök és eljárások halmaza, amely a nyelvi adatok rögzítését, feldolgozását, alkalmazását és általános hozzáférhetőségét teszi lehetővé [36]. A nyelvi technológia és a nyelv fejlődése szorosan összefügg. Korábban írásrendszerünk, és írásunk rögzítésének módja, például kódexeink összeállítása és másolása jelentette a nyelvi technológiát. Ma nyelvi technológiának számít a nyelv kódolásához szükséges kódtábla, a számítógépes szövegrögzítés, a karakterfelismerés (OCR), a szóellenőrző és nyelvhelyesség-ellenőrző algoritmus, a százmilliószövegszóból álló adatbázis és az azon alapuló modern számítógépes szótár, a fordító és kivonatoló algoritmus vagy az automatikus beszéd-előállítás.

*A nyelvi technológia fejlődésének egy-egy fontosabb állomása egy társadalom*

életében általában együtt járt a gazdasági fejlődésben bekövetkezett nagyobb változásokkal és nemegyszer a politikai rendszer megváltozásával is. Azok a népek, melyek nem tudtak ezzel a fejlődéssel lépést tartani, eltűntek a történelem színpadáról és velük együtt tűntek el – nemegyszer nyomtalanul – e népek nyelvei is. A nyelvi technológiák fejlődésében újabb lépést jelentett a könyvnyomtatás felfedezése, mely nélkül az újkorban bekövetkezett ipari fejlődés elképzelhetetlen lett volna. Tudjuk, nem minden írott nyelv jutott el erre a fokra. [...] Az a kultúra, amely nem tud ezzel a technológiai fejlődéssel lépést tartani, elveszíti tanulóképességét, rugalmasságát, ami nemcsak nyelvi-kommunikációs zavarokhoz vezethet [...] Arra azonban mindenképp rá kell mutatnom, hogy a nyelvi technológiák csak látszólag szorítkoznak technikai kérdések megoldására. Alap kutatás nélkül ugyanis e kérdések nem oldhatók meg [...] A nyelvi technológia persze közvetlenül nem érinti a kultúra, a szépirodalom nyelvét, közvetve azonban, ha nem tart lépést a világgal, veszélyeztetheti annak fennmaradását. [37]

## Gépszerűtlen helyesírási kérdések

Első olvasásra talán nem is tűnik fel Kis Ádám előadásának – *Gépszerű helyesírás* – hálózati változatában két apróság. A szöveg tanulságos, a helyesírási szabályzaton vezet végig az olvasót, azt vizsgálva, hogy a szótárakkal operáló számítógépes helyesírás-ellenőrzés [38] számára megvalósítható-e egy adott helyesírási szabály vagy nem. Agglutináló nyelvünkben egy-egy szótónak sok ezer alakja lehetséges, ezért az egyszerűbb szótárak helyett gépi morfológiai elemzésre, illetve keresőgráfokkal megadott szótárakra van szükségünk [39]. Az írás módjára közvetlenül nem vonatkozó szabályokat, továbbá a nem előíró szabályokat (bevezetéseket, összefoglalókat, elvi kérdéseket rögzítő pontokat) leszámítva gépszerűek (számítógéppel feldolgozhatók) azok a helyesírási szabályok, amelyek nem igényelnek morfológiai elemzést, a gépi morfológiai elemzésre alkalmas szabályok pedig potenciálisan gépszerűek. Az alapvetés után következik az önálló vagy csoportos szabálypontok gépszerűségi leírása. A szöveg néhány mondatával vitába szállnék, az első idézet az AkHSz. 3. szabálypontjának megítélése:

*„A teljesíthető (és gyakorlatilag teljesített) szabályok élén éppen a 3. szabály áll, amelyik a magyar nyelv betűkészletét írja elő. A személyi számítógépeken [...] alkalmazott 852-es kódkészlet tartalmazza az összes magyar betűt.”*

A harmadik ponthoz fűzendő mondanivalómat ki kell terjesztenem a szabályzat írásjelekkel foglalkozó fejezetére is (239-275.), hiszen már láttuk, a számítógépen a betű és az írásjel egyaránt karakter. Elsőként, eltekintve attól, hogy nagyon ritkán használjuk az említett IBM 852-es kódtáblát, érdemes felidézni Horváth Iván *Magyarok Bábelben* című könyvének néhány sorát:

*„Az 1992. decemberében kiadott MSZ 7795-3 számú Magyar Szabvány roppant engedékeny. Lehetővé teszi, hogy akár háromféle kódtáblával is éljünk, azzal, amelyet a Microsoft DOS (és például az IBM OS/2) használ [40], továbbá*

*azzal, amelyet a Microsoft Windows [41], valamint még egy nagygépes környezetben szokásos »EBCDIC« nevűvel is. (Mintha az egyméteres mellett az egyhuzas és a hetvencentis méterrúd használatát is engedélyznék.)”*

A gyakorlatban a Microsoft Windows ANSI 1250-es kódtáblájával szokás magyar szövegeket írni, újabban egyre jelentősebb szerepet kap e téren a Unicode átírási formátumainak (UTF-8, UTF-16) valamelyike is. Ez nem a felhasználók tudatos ellenállásának, hanem egy nagyvállalat, a Microsoft szoftverhonosító politikájának következménye, ők már rájöttek, hogy mi kell a magyarnak, mi magunk azonban még nem. A szabványban definiált három kódtábla (IBM 852, ISO-8859-2 és az EBCDIC) egyike sem tartalmazza a magyar idézőjelpár („”), a három pont (...) és a gondolatjel (–) karakterét [42]. Ezt a szöveget is a szabvány ellenére írom, UTF8-ban. Az írásjelek használatáról további részletekkel szolgál a következő fejezet. Az IBM 852-es kódtáblájának alkalmazása tehát nem az egyedüli, nem is az igazi megoldás a magyar szövegek tárolásához, és (az idézettel ellentétben) nem tartalmazza önálló jogon, karakterként az összes magyar betűt, csupán a magyar nyelvű szöveg képezéséhez szükséges karaktereket, ez jelentős különbség.

A szabályzat 10. pontja sorolja fel a magyar ábécé betűit. A *cs, dz, dzs, gy, ly, ny, sz, ty* és *zs* betűink a számítógépen csak karakterpárokkal, karakterhármassokkal ábrázolhatók, egyetlen kódtáblánkban sincsenek feltüntetve önálló karakterekként. Miért baj ez? Miért nem elég, ha több karakterből állítjuk össze őket, hiszen a végeredményen, a képernyőn, nyomtatón megjelenő *képen* úgysem látszik hogy egy, két vagy három karakterből áll egy betű [43]. Van egy alaplételem, amelyek hibátlan megvalósítása követelményként határozható meg minden magyar nyelvet kezelő számítógépes rendszer számára. Minden valamirevaló adatbázis, nyilvántartás, lista, index, tárgymutató előállításához nélkülözhetetlen a betűrendbe sorolás (14-15. pont). Ehhez, és az első idézethez is szorosan kapcsolódik a második idézet:

*„A betűrendbe sorolás szabályainak követése követelményként határozható meg minden számítógépes rendszer számára, és teljesítésének nincsen elvi akadálya”*

Szerintem fordítva. Elvi akadálya van, még mindig hordozza írásunk a kancelláriai és a huszita írás nyomait, de a gyakorlatban ezt ritkán (?) vesszük észre. Prószéky Gábor és Kis Balázs a könyvében [44] olvashatunk a magyar betűk írásmódjából következő akadályokról. A szóösszetételi határokon bizonyos esetekben számítógéppel még a betűk határait sem lehet egyértelműen megállapítani (pl. malacsült, nyolcszor, egészség). Ilyenkor gépi morfológiai elemzésre van szükség a szóösszetétel határának megállapításához. Ez túl költséges, idő- és számításigényes megoldás lenne nagyobb adathalmaz rendezése esetén. A magyar rendezési algoritmusok ezért nem használnak morfológiai elemzést, a több karakterből álló betűket (és az álbetűket is!) a rendezés idejére egyetlen speciális karakterrel helyettesítik, rendezik az adatokat, majd visszacserélik a helyettesítő karaktert a több karakterből álló eredeti kódokra. A kapott eredmény, főleg kevés adat esetén jó eséllyel helyes lesz... [45]

További két területen okoz még gondokat mostohán kezelt betűink használata. A számítógépes helyesírás-ellenőrzés egyik módszere a hibák modellezése karakterpárokkal és karakterhármassokkal [46]. A hibamodellek megadásával a gyakori elgépelések jól javíthatók (pl. em>me, aemly>amely; am>alm, alkamas>alkalmas). Két vagy három karakterből felépülő betűinket ilyenkor kiemelt figyelemmel kell kezelni. Az elválasztási algoritmusokban, ahol a szótaghatárok megtalálása alapvető feladat, szintén gondot okoznak többjegyű betűink. Ha az elválasztó algoritmus többjegyű betűink egyszerűsítetten kettőzött formájával (ssz) találkozik (már itt kezdődnek a problémák: valóban többjegyű betűről van-e szó?), át kell alakítania két többjegyű betűvé (szsz), karakterbetoldással.

A többjegyű betűk használatát leegyszerűsíthetné, ha a magyar kódtáblák önálló jogon tartalmaznák ezeket a betűket. Szükség van-e rájuk? Ez előzetes teoretikus döntést igényel [47]. Megéri-e átállni a ma már húszéves, széles körben elfogadott régi módszerről? A magyar karakterek kódolására vonatkozó szabványon előbb-utóbb egyéb okok (pl. hiányzó írásjelek vagy a háromféle kódolási konvenció megengedése) miatt is módosítani kell, mára vészesen elavult, már évekkel ezelőtt, 1998 februárjában [48] is elavult volt. A helyesírási szabályzatnak pedig tartalmaznia kellene az elfogadandó egységes magyar kódtábla helyesírási szempontból releváns karaktereinek meghatározását és a karakterek kódszámait.

## Írásjelek

Tipográfiai és helyesírási szempontból is elítélendő a helyesírási szabályzatban előírt írásjelek mellőzése szövegeinkben. Az előző fejezetben már kitértünk arra, hogy a szabványos magyar kódtáblák nem teszik lehetővé bizonyos írásjeleink, a három pont, az idézőjel, a gondolatjel, és a nagyköötjel megfelelő használatát, nem biztosítanak önálló karaktereket e jelek számára, így vagy a szabványt vagy a helyesírás szabályait kell követnünk, a kettő együtt nem lehetséges. Új karakterkódolási szabványunk elfogadása során minimális elvárás az AkHSz. vonatkozó pontjainak figyelembe vétele [49], a nemzeti kódtáblának tartalmaznia kell az írásjelekkel foglalkozó fejezetben (239-275. pont) megadott írásjeleket.

A hiányzó jeleket 8 bites kódtáblák esetén (a már kialakult szokásokat megtartva) érdemes az Amerikai Szabványügyi Hivatal (ANSI) 1250-es kódtáblájának karakterkódjaival tárolni, a 16 (vagy ennél is több) bites kódolás esetén pedig a Unicode, illetve ennek karakterpozíciói lehetnek mérvadóak. Az itt nem említett egyéb írásjelek, a pont, kérdőjel, felkiáltójel, vessző, pontosvessző, kettőspont, zárójel és a belső idézőjelpár karakterhelye megvan, ezek a gyakorlatban is különösebb gond nélkül használhatóak.

A nyitó idézőjellel és a berekesztő idézőjellel, azon kívül, hogy kirekedtek a szabványból, nincs probléma, egyértelműen meghatározható, hogy milyen kódot kell/érdemes hozzájuk rendelni. Ezekkel a vízszintes vonalakkal, a kötőjellel, gondolatjellel, nagyköötjellel viszont baj van, ahány szöveg, annyi szokás, ráadásul a szövegszerkesztő programok még egy újabb karakter, a feltételes elválasztójel [50] (soft hyphen) használatát is bevezetik.

Létezik a számítógépes kódtáblák viszonylag állandó részében (alsó 128 karakterhely) egy *hyphen* [51] nevű karakter, ezt szokás matematikai jelként, kötőjelként és elválasztójelként is használni, nemcsak nálunk, szinte mindenütt a világon. Tipográfiailag és helyesírásiilag is zavaró, ha a *hyphen* karaktert (nevezzük ezentúl kötőjelnek) a gondolatjel vagy a nagykötőjel helyett használjuk, akár össze is zavarhatjuk vele az olvasót:

*helytelen:*

- *Addig is - tette hozzá -, néha-né-  
ha benézhetnél...*

*helyes:*

– *Addig is – tette hozzá –, néha-né-  
ha benézhetnél... [52]*

A gondolatjelet tehát a gondolatjel karakterével kell jelölnünk. De hol található a gondolatjel karaktere? Az angol hagyományoknak megfelelő gondolatjelet *em dash* [53] néven lelhetjük meg a kódtáblák leírásaiban. Virágvölgyi Péter a *Tipográfia mestersége számítógéppel* című könyvében nem ajánlja az *em dash* (kvirtmínusz) magyar gondolatjelként való használatát [54], az csak az angolban használható megfelelően, szóközök nélkül, nekünk csak a rövidebb *en dash* [55] (félkvirtmínusz) szóközökkel való szedését javasolja. A nagykötőjel pedig a magyarban és az angolban egyaránt az *en dash*, (félkvirtmínusz), szóközök nélkül. Ezzel sérül ugyan a karakterhelyesség, de sajnos nincs igazi karakterünk a nagykötőjel számára. A kérdéses írásjelek tehát az alábbi táblázatnak megfelelő kódokkal jelölendők a 8 (ANSI 1250), illetve a 16 (Unicode) bites kódolások alkalmazása esetén, az MSZ 7795-3 negligálásával:

a karakter neve	kódja 8 biten	kódja 16 biten	képe
nyitó idézőjel (double low-9 quotation mark)	132 (dec) 84 (hex)	8222 (dec) 201E (hex)	”
berekesztő idézőjel (right double quotation mark)	148 (dec) 94 (hex)	8221 (dec) 201D (hex)	„
gondolatjel (en dash, szóközökkel)	150 (dec) 96 (hex)	8211 (dec) 2013 (hex)	–
nagykötőjel (en dash, szóközök nélkül)	150 (dec) 96 (hex)	8211 (dec) 2013 (hex)	—
három pont (horizontal ellipsis)	133 (dec) 85 (hex)	8230 (dec) 2026 (hex)	⋯

## Tipográfia

A kéziratos könyvek mindig egyediek voltak [56], gyakran illumináltak, és persze rendkívül drágák. Az ilyen könyvek nem egyenletesen oszlottak szét a társadalomban, szinte kivétel nélkül egyházi és nemesi tulajdonban voltak, vagyont, értéket képviseltek. Az írásbeliség kezdetén a kézirat külsőségei sokkal jobban befolyásolták az olvasót mint a

könyvnyomtatás hajnalán, 15. században, azelőtt a megjelenés a mű individualitásához tartozott [57]. A nyomtatás a maga óriási példányszámaival (kezdetben a néhány száz példány is rettentő sok volt) lerombolta ezt az egyediséget, tömegcikké alacsonyította a könyvet, hiába törekedtek az első nyomtatványok készítésénél a kézirat minél hívebb utánzására, hiába kellett ezeket az ősnymtatványokat is egyedileg illuminálni: a könyv hamarosan termékké vált. A kor emberének véleménye valószínűleg nem volt annyira szigorú, mint a könyvmásolóműhely-tulajdonos Vespasiano da Bisticci-é, aki *A XV. század kiváló embereinek életrajzai* című munkájában így nyilatkozik Federigo da Montefeltre urbinói könyvtáráról:

*„Abban a könyvtárban a könyvek mind kivételes szépségűek, mindet tollal írták, s egyetlen nyomtatott sem akadt köztük, hiszen szégyellte is volna”* [58]

A nyomtatott könyv története a tipográfia története, a kódexek egyediségét lassan felváltotta a könyvek vizuális reprezentációjára, a szöveg elrendezésére sokkal erősebben koncentráló szabályrendszer és hagyomány. Elvesztettük ugyan a tárgy/szöveg egyediségét, de ennek helyén művészet termelt. Ma, amikor már közhely arról beszélni, hogy a számítógépes szöveg valami más, új galaxis, mit veszünk? Nem vagyunk-e büszkéek saját könyvárunkra, néhány szépen szedett könyvsorozatra, egy-egy igazi csemegére? Nyomukba sem érhet a felszínen esetleg rendezetlennek tűnő hálózati kiadvány.

Ma mindenki tipográfus, aki szöveget állít elő a számítógépén: a DTP [59] a mindennapi tipográfiát, a pontosabban a tipográfia látványos hanyatlását hozta magával. Már magyarul is létezik a számítógépes tipográfia alapelemeivel foglalkozó, gyakorlati jellegű könyv [60], amely aprólékosan elmagyarázza, hogyan alakíthatjuk a szem számára kellemessé és elfogadhatóvá számítógéppel szedett szövegünk *képét*. Igen, megint a megjelenés és a struktúra szerinti szövegtárolás eltérő jellegébe botlottunk.

A számítógépen tárolt megjelenés-központú szöveg a nyomtatott szöveg közvetlen folytatása, előállításánál nekünk kell gondoskodnunk a tipográfiai konvenciók betartásáról, az oldalakra tördelésről, a helyes térkitöltésről, betűfokozatokról, összeálló betűtípusokról... Amikor szövegszerkesztő programmal dolgozunk ezt tesszük. Hiába biztosítanak a szövegszerkesztők strukturális elemeket (pl. cím, bekezdés), egyszerűen nem működik a dolog megfelelő hatékonysággal, Virágvölgyi Péter könyve arról szól, hogyan ne bízzunk meg a szövegszerkesztő és kiadványszerkesztő programok tipográfiai tudásában, mit és hol kell javítanunk, ha szövegünket lapra szánjuk. A megjelenés-központú szöveg képe mindig egy adott eszközhöz, tárgyhoz (pl. a számítógéphez csatolt nyomtatóhoz és a hozzá való papír méretéhez) viszonyítva tárolódik.

A struktúrát is tároló szövegek esetén a kép pusztán egy lehetséges reprezentáció, itt tehát *nem szabad* a szöveggel együtt a megjelenésre vonatkozó információt is (pl. betűtípus, betűméret, de a szavak elválasztása is ilyen) elraktározni. A szöveg tartalmával párhuzamos struktúra eleihez viszont kapcsolható megjelenítésre vonatkozó információ. Tipikus, meglehetősen hétköznapi példája ennek egy frissebb HTML-ajánlást [61] és a WCAG [62] elveit szigorúan követő weboldal, amelyben nem található vizuális reprezentációra



vonatkozó információ, de csatolható hozzá egy stíluslapnak nevezett fájl, amelyben a HTML nyelv szerkezetet leíró elemeihez tetszőleges színt, hátteret, betűtípust, méretet, pozicionálást stb. lehet rendelni. A stíluslap a tipográfus területe, de a szöveg képének kialakítása a reprezentációt végző program feladata. Ha papírra kerül egy ilyen szöveg, akkor a tipográfia egyrészt a megjelenítő algoritmusainak fejlettségén múlik. Azon, hogy milyen nyelvi technológia szolgálja ki a megjelenítés igényeit, hogy sorkizárás esetén például el tudja-e választani automatikusan a szavakat a sorvégen vagy nem; hogy érzékeli-e az egymás alá kerülő szóközökből kialakuló „utcákat”; hogy a betűtípushoz tudja-e igazítani a betűközöket, stb. Másrészt a megjelenítőn múlik a struktúra állandó elemeihez definiált stíluslap kezelése is. Egyáltalán nem biztos, különösen egy internetes szöveg esetében, hogy a világ különböző számítógépein létezik a stíluslap elkészítője által kedvelt betűtípus, ilyenkor egyfajta szórás figyelhető meg, a szöveghez a megjelenítőnek az aktuálisan rendelkezésre álló lehetőségekből kell kiválasztania az optimális megoldást (pl. egy groteszk betűcsaládot egy másik, de aktuálisan felhasználható groteszk családdal kell helyettesítenie). Lényeges különbség a megjelenést előtérbe helyező megoldásokhoz képest: a struktúrát tükröző szöveg képét nem tároljuk, a kép az elérhető eszközökhöz igazodva automatikusan jön létre.

Egy igazi tipográfus számára szentségtörésnek tűnhetnek az imént leírt szavak: a tipográfus programmal való helyettesítése és a szöveggép környezetfüggősége. A gép ma még nem nyújtja ugyanazt a minőséget. De ne feledjük, a megjelenítésért felelős algoritmusok egyre összetettebbek lesznek, a nyelvi technológiák alkalmazása folyamatosan fejlődik, a struktúra tárolásához és a különféle stíluslapok készítéséhez szükséges nyelvek pedig ma már mindenki számára elérhetőek [63].

## Összefoglalás

Szövegemben rendszerszintű összefoglalásra törekedtem, az egyes fejezetek (kevés kivétellel) tetszőleges mélységig bővíthetők, önálló monográfiák tárgyai, némelyiket már meg is írták. Mindegyik fejezetben annak az új szövegfelfogásnak a bemutatása volt a célom, amelynek legfőbb sajátága, hogy meghaladja az eddig elválaszthatatlanként kezelt egységet, a szöveghordozó és szöveggép azonosságát. A szöveg modulárisává válik a felfogás szerint, nem egység többé, több alrendszer, a tárolási, a reprezentációhoz információt adó és a reprezentációt végző alrendszer együttműködésének eredménye.

Az absztrakt szöveg mellett, helyesebben: benne, az eddigi szövegfelfogás is megfér, hiszen a hagyományos, redukálhatatlanul grafikai szöveg az absztrakt szöveg egy speciális, hordozóhoz kötött, grafikus elemekkel megvalósított esetének tekinthető.

## Jegyzetek

- [1] BÁRCZI Géza, *Fonetika*, Budapest, Tankönyvkiadó, 1951, 4
- [2] KÉKI Béla, *Az írás története*, 12-18
- [3] R. MOLNÁR Emma, *Hangtan*, 109
- [4] Például: „A Nap jele több népnél is a fényküllőkkel körülvett kör, de ezt az ábrát az idők folyamán már nemcsak a Napra mint égitestre, hanem a forróságra, hőségre is vonatkoztatták” (KÉKI, i.m.)
- [5] KÉKI, i.m.
- [6] Például: „Mintha a »silány« szavunkat csak úgy tudnók leírni, hogy egy sítalpat és egy lányt rajzolnánk egymás mellé” (KÉKI, i.m.)
- [7] KÉKI, i.m., 38.
- [8] KÉKI, i.m., 77.
- [9] Az egyszerűség kedvéért nem vezettem be külön a vezérlőkarakter fogalmát, a szövegben nem lesz rá szükség.
- [10] A kódtábla szinonimáinak tekinthető elnevezések: kódtáblázat, kódlap (code page), kódkészlet (code set), kódolt karakterkészlet (coded character set) vagy pusztán karakterkészlet (character set, charset). A szövegben a kódtábla elnevezést preferálok.
- [11] Forrás: UniBook, codepage viewer 2.0
- [12] ISO 8859-2
- [13] Unicode
- [14] PRÓSZÉKY, KIS, *Számítógéppel emberi nyelven*
- [15] Jesper SVENBRO, *Az archaikus és klasszikus Görögország: a csöndes olvasás feltalálása = Az olvasás kultúrtörténete a nyugati világban*, 44-70
- [16] Guglielmo CAVALLLO, *A volumentől a codexig: olvasás a római világban = Az olvasás kultúrtörténete a nyugati világban*, 89
- [17] *Uo.*, 93
- [18] Malcolm PARKES, *Olvasás, írás, interpretálás: a kora középkor szerzetesi gyakorlata = Az olvasás kultúrtörténete a nyugati világban*, 100-104
- [19] *Uo.*, 105
- [20] *Uo.*, 100
- [21] Jacqueline HAMESSE, *Az olvasás skolasztikus modellje = Az olvasás kultúrtörténete a nyugati világban*, 115-129
- [22] *A magyar nyelv történeti-etimológiai szótára*, II
- [23] HORVÁTH Iván, *Szöveg*, Egy->sok
- [24] HÁRTÓ Gábor, *A grafikai mozzanat a szövegben*
- [25] Az OCR (Optical Character Recognition) technológiát alkalmazó szoftverek épp erre, a kép adattípus szöveggé alakítására képesek, ilyen például a nemzetközi hírű, magyar fejlesztésű Recognita program.
- [26] A ebookról és különböző megvalósításairól ír PÁLFI Norbert, *Irodalom, szöveg, információ: könyv a G2 galaxisban, avagy az ebook helyzete és kilátásai* című dolgozatában.
- [27] What You See Is What You Get: minden modern szövegszerkesztő program alapelve, ami a képernyőn látszik, azt kapjuk kimenetként a papíron.

- [28] SGML-ben, XML-ben is elképzelhető a megjelenítést tükröző leírónyelv megalkotása, a HTML, leginkább a 3.2-es verziója részben ilyen nyelv. Az SGML-ről rengeteg bevezető található a hálózaton, újabban a szakkönyvek között is, az egyik legelső alaposabb magyar nyelvű leírás TURI László szakdolgozatában olvasható.
- [29] A hordozó tulajdonságából, a lapok méretének végességéből adódó, hibásan felismert szegmentumhatárookra kiváló példa VADAI István *Lenni vagy nem lenni* című írásának tárgyául választott Domonkos István-vers (kormányeltörésben) laphatárokhöz kötött furcsa metamorfózisa.
- [30] Unicode Code Charts (<http://www.unicode.org/charts>)
- [31] Az adat a Unicode Consortium weblapjáról való, *New in Unicode 3.0* (<http://www.unicode.org/unicode/standard/versions/Unicode3.0.html>)
- [32] A helyzet nem ennyire tragikus, vethetnénk közbe. A Unicode 6400 karaktert biztosít privát használatra (itt igazán elférnek speciális jeleink). A legfrissebb Unicode-verzióval (3.1) pedig elkezdődött az egyszerű 16 bites kódokkal el nem érhető (hex FFFF feletti) terület feltöltése, ahol további egymillió karakter számára van hely, úgy tűnhet: bőven van még időnk. Egyszerűbb megoldás lenne azonban a speciális magyar karakterek UCS2-be (az első 16 bitbe) való felvétele, ahol már csak korlátozott számú karakterhely áll rendelkezésre, tehát az FFFF alatti tartományba csak a ma is használt jeleink felvételét érdemes megpróbálni.
- [33] Technical Reports, Unicode Standard Annex #27, *Unicode 3.1* (<http://www.unicode.org/unicode/reports/tr27>)
- [34] GOLDEN Dániel, *TeXTeX*, ELTE szakdolgozat vonatkozó fejezete, a Magyar Nyelv Történeti Kódtáblájáról, a szöveg 1998-as. A szakdolgozat ezen fejezete bekerült a szerző és két társa által a Neumann-ház számára írt tanulmányba (GOLDEN Dániel, TÓTH Tünde, TURI László: *Virtuális örökkévalóság: objektumok a digitális könyvtárban*).
- [35] A Unicode a 16 biten lehetséges 65536 karakterhelyből 2048-at úgynevezett váltójeles karakterkódok számára tart fenn. DAVIES, BARBER, *Számítógép-hálózatok*, 205 (összefoglaló a váltójeles kódokról): „Általában ha egy  $N$ -bites kód  $M$  kombinációját váltóutasításnak használják,  $M$  váltóállás  $2^n - M$  kombinációval áll rendelkezésre. Az összes kombináció  $M(2^n - M)$ , amelynek maximuma akkor van, ha  $M = 2^{n-1}$ . Ez  $2^{2(n-1)}$  lehetőséget jelent a váltójelek nélkül megvalósítható  $2^n$  esethez viszonyítva; ez  $2^{n-2}$ -szer kedvezőbb. A váltójeles kódok hatékonysága azonban a kezelendő információ jeleinek sorozatától függ. Ha ritkán használnak váltójelet, akkor a hatékonyság jelentősen növekszik. A gyakori váltás a váltóparancsok nagy száma miatt hátrányos.”
- [36] Majdnem szó szerinti idézet KIEFER Ferenc, *Néhány gondolat a nyelvi technológiákról* című írásából. Az egész fejezet Kiefer gondolatainak megismétlése, szövegemhez igazítása, elhagytam azokat a részeket, amelyben az általa vezetett „projektumok” számára kér támogatást.
- [37] KIEFER Ferenc, *i.m.*
- [38] A számítógépes helyesírás-ellenőrzést a mai gyakorlatban ezt kétféle program valósítja meg, a szóellenőrző és a nyelvhelyesség-ellenőrző. Szövegemben csak a szóellenőrzéssel foglalkozom.

- [39] PRÓSZÉKY-KIS: *Számítógéppel emberi nyelven*, 95-133
- [40] IBM 852
- [41] A Windows régebbi változatai az Amerikai Szabványügyi Hivatal (ANSI) 1250-es kódtábláját használják, ez az MSZ 7795-3-ban is hivatkozott, a Nemzetközi Szabványügyi Hivatal (ISO) által elfogadott ISO-8859-2-es kódtáblának némileg módosított verziója. Az új Windowsok sokkal rugalmasabbak a kódtáblák kezelésében, mert a WGL4 nevű kódtáblát használják, amely a Unicode szűkített változata. A WGL4-es karakterreperatóárjából előállítható az ISO-8859-2-es kódban írt szöveg megjelenítéséhez szükséges összes jel. Az eredeti szövegben a szerző az ISO-8859-2-re utal.
- [42] A felsorolt karakterek megfelelői nem szerepelnek a szaküzletekben kapható magyar billentyűzeteken sem, ez is hozzájárul mellőzésükhöz.
- [43] Néha mégis látszik: pl. gyakran áll SZ az Sz helyén.
- [44] PRÓSZÉKY-KIS, *i.m.*, 36-40
- [45] Nem minden esetben kell a magyar ábécé szerint rendezni (16. pont): „Az olyan sajátos célú munkák (lexikonok, enciklopédiák, atlaszok és térképek névmutatói stb.), amelyekben magyar és idegen nyelvű szóanyag erősen keveredik egymással, rendszerint az úgynevezett általános latin betűs ábécét követik: a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z. Ebben a rendszerben mind a magyar, mind az idegen többjegyű betűknek minden egyes eleme külön, önálló egységnek számít, és a besorolás nincs tekintettel sem a magyar ékezetekre, sem az idegen betűk mellékjeleire. Az ilyenfajta munkák szavainak betűrendbe sorolása bonyolult feladat. A követendő eljárást a könyvtárügy és a szakirodalmi tájékoztatás (dokumentáció) területén hazai és nemzetközi szabványok írják elő, illetőleg belső szakmai útmutatók szabályozzák. Ezek hatálya azonban nem terjeszthető ki írásbeliségünk általános gyakorlatára.” (kiemelés tőlem)
- [46] PRÓSZÉKY-KIS, *i.m.*, 100-101
- [47] GOLDEN Dániel, TÓTH Tünde, TURI László, *Virtuális örököseink: objektumok a digitális könyvtárban*: „Előzetes teoretikus döntést igényel tehát, hogy a magyar nyelvű szövegek esetében egyszerűen írásjegyeket vagy a magyar ábécé betűit kívánjuk rögzíteni. (Az utóbbi esetben külön kódokat kellene alkalmaznunk a t, az y és a ty jelölésére; az előbbi esetben viszont semmi okunk sincs arra, hogy az ü kódjaként ne fogadjuk el az „u + [umlaut]” szintetikus kódolási formát.) Ám ezzel még koránt sincs vége a bonyodalmaknak: a 12. pont „rég, ma már egyébként nem használatos betűk”-ként határozza meg a következőket: áá, eé, eö, ew, oó, y, ch, cz, s, th, ts, w. A sor végén pedig fatális módon ez áll: stb...”
- [48] HORVÁTH Iván, *Magyarok Bábelben*, 116-119: „Irodalomtudományi informatikai kutatócsoportot szervezek. [...] A válság az 1970-es években alakult ki, ekkor jelentek meg a magyar piacon az első számjegyzérlésű fényszedőgépek [...]. Ekkor kellett volna Akadémiánknak szabályoznia a magyar nemzeti kódtáblát: meg kellett volna határoznia, hogy gépi vezérlésű folyamatokban melyik magyar írásjelnek melyik számjegy feleljen meg. A szabályozás elmaradt, és a '80-as években, a személyi számítógépek tömeges elterjedésének idején kialakult a káosz s az ebből fakadó mérhetetlen bosszúság és munkaerőpazarlás. [...] Egyik program ezt, a másik azt a kódtáblát támogatta, a nyomtató pedig esetleg egy harmadikat. A '90-es évekre a társadalom beletörődött a megváltoztathatatlanba. Megszoktuk a

kalapos „ű”-ket. Az 1993-as nemzeti szabvány immár megengedi háromféle kódtábla (a „852-es”, az „ISO-8859-2-es” és az „EBCDIC”) békés együttélését. Holott, természetesen, új, egységes kódtáblára lenne szükség Magyarországon. [...] Megígérjük, hogy 1999 végén letesszük az I. osztály asztalára azt a nemzeti kódtábla-javaslatot, amellyel kiegészíthető lesz az akadémiai helyesírási szabályzat.” A kutatócsoport létrehozása –természetesen– nem sikerült, az Akadémia nem reagált.

- [49] MSZ 7795-3:1992, 8. lap: „E szabvány kidolgozásában közreműködött az IBM Hungary Kft.”. Forrás: HORVÁTH Iván: *i.m.*, 115, további kommentárok ugyanitt, illetve a 117. oldalon.
- [50] PRÓSZÉKY, KIS, *i.m.*, 148, 150
- [51] kötőjel, választójel (ORSZÁGH-MAGAY: Angol–magyar nagyszótár)
- [52] A példa forrása: VIRÁGVÖLGYI Péter, *A tipográfia mestersége számítógéppel*, 30.
- [53] gondolatjel (ORSZÁGH-MAGAY: Angol–magyar nagyszótár)
- [54] VIRÁGVÖLGYI Péter, *i.m.*, 30: „Mivel a magyarban a gondolatjelet szóközökkel kell használni, ezért nekünk csak a rövidebb, az ún. félkvirtmínusz (en dash) felel meg. A kvirtmínusz szóközökkel tetézve helypazarló és esztétikailag káros, mert túl nagy »lyukat« támaszt a szedésben.”
- [55] elválasztójel (ORSZÁGH-MAGAY: Angol–magyar nagyszótár)
- [56] Kivétel is akad: a 14-15. század Itáliai cartolaiói nagy tételben szerezték be a könyvmásoláshoz szükséges anyagokat, írkokokat és illuminátorokat fogadtak föl, bizony előfordult, hogy spekulációs céllal több másolatot is főlhalmoztak egy-egy könyvből. További részletek: *Az olvasás kultúrtörténete a nyugati világban*, 211-218.
- [57] THIENEMANN Tivadar, *Irodalomtörténeti alapfogalmak*, 75-76: „Mai irodalmi gondolkodásunk fennakad azon, ami a középkori írásbeliség szempontjából magától értetődik, hogy pl. három jámbor költemény kézzelfogható erekye legyen. »Ha az asszony ezt a három éneket jobb kezébe fogja, mikor a szobába tér, nem szenved sokáig asszonyi fájdalmat.« (Wesle, Carl: Priester Wernhers Maria, 1927, 139. l.) Mivel a munka szellemi tartalma így teljesen felolvad a mű esetleges kéziratos alakjával, a modern szövegkiadásra törekvő philologus mindig nehéz feladat előtt áll, mikor a középkori kéziratot művet nyomtatott könyvvé akarja átváltoztatni. A nyomdafesték kiemeli a középkori művet a maga éltető eleméből, a kézirat-művészet intim és individuális formáiból, csak a szövegre irányítja a figyelmünket és elvont olvasásra készítet [...]. Akik a betűírást nem értették meg, megértették az illusztrátor képírását.”
- [58] Idézi Anthony GRAFTON, *A humanista olvasás = Az olvasás kultúrtörténete a nyugati világban*, 213. (Vespasiano da Bisticci, *Vite de uomini illustri del secolo XV*)
- [59] Desktop Publishing, asztali kiadványszerkesztés
- [60] VIRÁGVÖLGYI Péter, *i.m.*
- [61] HTML 4 vagy a még újabb XHTML-verziók
- [62] Web Content Authority Guide, a webes szabványok kialakításával foglalkozó szervezet, a W3C ajánlása különféle környezetekben jól reprezentálható HTML-oldalak készítéséhez.

- [63] Tartalmi struktúrát például ábrázolhatunk a már említett SGML, XML és bizonyos megszorítások mellett a HTML nyelvben is. Stíluslapokra is többféle megoldás létezik, az itt felsoroltakhoz illeszthető például a CSS vagy a DSSSL. További részleteket a W3C (Word Wide Web Consortium) honlapján (is) találhatunk (<http://www.w3c.org>).

## Felhasznált irodalom

- BÁRCZI Géza, *Fonetika*, Bp., Tankönyvkiadó, 1951
- *Cascading Style Sheets, level 2: CSS2 Specification*, W3C Recommendation, <http://www.w3.org/TR/REC-CSS2>
- DAVIES, D. W., BARBER, D. L. A., *Számítógép-hálózatok*, Bp., Műszaki, 1978
- *Developing International Software for Windows 95 and Windows NT*, Appendix H (Code Pages), <http://msdn.microsoft.com/library/default.asp>, (Microsoft Developer Network Library)
- ÉNEKES Ferenc, *A kiadványszerkesztés*, Bp., Novella, 2000
- ÉRSZEGI Géza, *Paleográfia = Bevezetés a régi magyar irodalom filológiájába*, szerk. HARGITTAY Emil, Bp., Universitas, 1996
- GOLDEN Dániel, *TeXText: számítógép és szöveg*, Bp., szakdolgozat (ELTE BTK), 1998
- GOLDEN Dániel, TÓTH Tünde, TURI László, *Virtuális örökkévalóság: objektumok a digitális könyvtárban*, Palimpszeszt, 1998, 4. sz., <http://magyar-irodalom.elte.hu/palimpszeszt/>
- GYURGYÁK János, *Szerkesztők és szerzők kézikönyve*, Bp., Osiris, 1997
- HÁRTÓ Gábor, *A grafikai mozzanat a szövegben*, Literatura, 1995/2
- HORVÁTH Iván, *Magyarok Bábelben*, Szeged, JATEPress, 2000
- HORVÁTH Iván, *Öt tárgy és öt mondat*, 2000, 1999/1, 48-50
- *HTML 4.01 Specification*, W3C Recommendation, <http://www.w3.org/TR/html401>
- KÉKI Béla, *Az írás története*, Bp., Gondolat, 1975
- *Kéziszedés*, szerk. HORVÁTH János, Bp., Műszaki, 1967
- KIEFER Ferenc, *Néhány gondolat a nyelvi technológiákról = A magyar nyelv az informatika korában*, szerk. GLATZ Ferenc, Bp., Akadémiai, 1999, 129-134.
- KIS Ádám, *Gépszerű helyesírás: az akadémiai helyesírási szabályzat és a számítógép*, VII. Országos Alkalmazott Nyelvészeti Konferencia, Bp., 1997 (hálózaton: <http://www.mek.iif.hu/porta/szint/tarsad/nyelvtud/gepscikk/>)
- *A magyar helyesírás szabályai*, 11. kiadás, 11. példaanyagában átdolgozott lenyomat, Bp., Akadémiai, 1994
- *A magyar nyelv történeti-etimológiai szótára*, Bp., Akadémiai, 1970, II.
- *Metamorphosis of the Book*, ed. Muro KENJI, Tokyo, Dai Nippon, 1999
- Neil BRADLEY, *Az XML kézikönyv*, Bicske, SZAK, 2000
- *Nyomdaipari ABC*, szerk. GARA Miklós, Bp., Műszaki, 1987
- *Az olvasás kultúrtörténete a nyugati világban*, szerk. Guglielmo CAVALLO, Roger CHARTIER, Bp., Balassi, 2000
- PÁLFI Norbert, *Irodalom, Szöveg, Információ: könyv a G2 galaxisban, avagy az ebook helyzete és kilátásai*, Bp., szakdolgozat (ELTE BTK), 2000
- PRÓSZÉKY Gábor, KIS Balázs, *Számítógéppel emberi nyelven*, Bicske, SZAK, 1999
- R. MOLNÁR Emma, *Leíró magyar hangtan*, Bp., Nemzeti Tankönyvkiadó, 1996
- Paolo SANTARCANGELI, *A betűk mágiája*, Bp., Európa, 1971
- SZIKSZAINÉ NAGY Irma, *Leíró magyar szövegtan*, Bp., Osiris, 1999
- THIENEMANN Tivadar, *Irodalomtörténeti alapfogalmak* (reprint), Pécs, Pannónia,

1985

- TURI László, *Számítógép az irodalomtudományban*, Bp., szakdolgozat (ELTE BTK), 1992
- *The Unicode Standard, Version 3.0, Online Edition*, <http://www.unicode.org/unicode/uni2book/u2.html>
- VADAI István, *Lenni vagy nem lenni*, Ex Symposion, 1994/10-12
- VIRÁGVÖLGYI Péter, *A tipográfia mestersége számítógéppel*, Bp., Osiris, 1999
- *Web Content Accessibility Guidelines*, W3C Recommendation, <http://www.w3.org/TR/WAI-WEBCONTENT>