

PRÓSZÉKY GÁBOR

A szótári világ átalakulási tendenciái az internet megjelentével

Changing tendencies of the dictionary world in the age of internet

There is a well-known competition between printed dictionaries and electronic dictionaries. The paper argues that there is another sort of competition: between freely available dictionaries and dictionaries for fee. The average user of our age tends to choose free solutions even if its quality is weaker than other dictionaries s/he has to pay for. This tendency has been introduced by the internet, and its analogy does not exist in the non-virtual, 3D-world. Commercial language technology companies are in a rather strange position: they have to offer free versions of their valuable developments. This way, they themselves are competitors to their other marketable dictionary products. In the case of less spoken languages (e.g. Hungarian) the final solution would be – according to our vision – the establishment of a non-profit institution developing high-quality computational lexicography solutions for free. With the help of this, even existing but never published high-quality dictionaries for less supported language pairs could be published, and the best dictionaries would reach a wider public, because their publication would not burden the publisher with the difficulties of profit-making.

Azzal a fontos megállapítással kezdjük szótártechnológiai áttekintésünket, hogy a papírszótárak és az elektronikus szótárak nem igazán összehasonlíthatók, különösen össze nem mérhetők, bár mindkét formában piaci szereplőként kell kezelnünk őket. Nem elfelejtve az elektronikus világnak a könyvvilágra gyakorolt hatását, azt vehetjük észre, hogy napjaink papír alapú szótárai a saját területükön elsősorban a többi papírszótárral versenyeznek, az elektronikus szótárak pedig az elektronikusakkal. Amikor ugyanis olyan kontextusban jelenik meg egy papírszótár, ahol számítógép is működik, ott valóban kevésbé használják, ezzel szemben, ha például fotelban ülve újságot olvasunk, többnyire nem kelünk fel, hogy elindítsunk gépünkön egy elektronikus szótárat. Mivel minden elektronikus eszköz egyre erősebben van jelen a piacon, kimutatható, hogy papírszótárakból egyre kevesebbet adnak el. Ezt a jelenséget mindig az elektronikus szótárak számlájára írják, jóllehet van, aki eleve nem vesz papírszótárat, mert az elektronikus éra szülte; akik pedig megszokták, hogy papírszótárat vásárolnak, azok közül továbbra is vannak, akik vennének, ha anyagi lehetőségeik ezt könnyebben engednék. Sokakat, akik valamilyen okból úgy gondolják, hogy papírszótárra van szükségük – például, mert „az iskolában mondta a tanár, hogy a gyerekeknek kell szótár” – könnyen megtéveszti a kínálat. Ott vannak ugyanis a sokszor legújabbnak látszó szótárak, speciális reklámszövegekkel (például „legújabb, regisztrációs kiadás”), amelyek általában a korábban, sőt, sokszor jóval korábban kiadott szótáraknak a reprintjei,

tehát olyan olcsó, de elavult termékek, amelyek devalválják a szótárpiacot. A papír-szótár konkurenciája sokszor nem az elektronikus szótár, hanem a gyenge szótár, így a papírszótárak közt kell keresni a konkurenciát is, meg a nehéz helyzet kialakulásának az okát is.

Az elektronikus szótárak használói

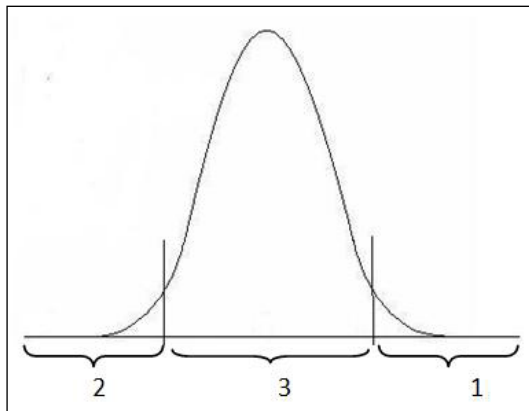
Az elektronikus szótárak világában más a helyzet. Amióta van elektronikus szótárkiadás, a felhasználók megoszlása a haranggörbét követi (1. ábra). Ennek alapján az alábbi három nagy kategóriába sorolhatók a vásárlók:

1. Létezik olyan vásárló, aki minden körülmények között megvesz szinte minden lehetséges elektronikus változatot, vagy azért, mert anyagi lehetőségei megengedik, vagy azért, mert a szótárhasználat megszállottja, esetleg az elektronikus eszközök elkötelezett használója; a tapasztalat szerint ez a réteg nem is mindig használja, amit megvesz, de kell neki.

2. Van egy másik szélsőség is: aki soha nem vesz meg semmit, általában azért, mert meg tudja szerezni más – a vásárlástól eltérő – úton is. Ezt mondhatjuk úgy is, hogy „ellopja”,

de ez a kifejezés durva, hiszen ugyanezeknek az embereknek a nagy része garántalan nem venné el más tulajdonát a „háromdimenziós világban”, ám az elektronikus környezetben feloldva érzi magát a gyakorlatban betartott erkölcsi szabályok alól. Ez és az előző kategória adja a vásárlói haranggörbe széleit.

3. A haranggörbe közepén hosszú ideig csak azok a szótárhasználók voltak, akik megérezték vagy megértették, hogy elektronikus környezetben elektronikus szótárra van szükségük: ilyen vagy olyan nyelvűre; a gyerekek vagy nekik; otthon vagy munkahelyen (az utóbbira elsősorban természetesen csak akkor, ha meggyőzték a főnöküket). E cél érdekében hajlandók voltak pénzt adni a kínált szótárért, hiszen nem volt más elektronikus megoldás, és ők nem „lopós fajták”.



1. ábra

Az utolsó tizen-egynéhány évben azonban megjelentek az internetes szótárak. Mivel a haranggörbe közepén szereplők nem „feltétlen elkötelezettjei” a szótárhasználati ügynek, elmondható róluk, hogy csak olykor-olykor van szükségük szótárra. Az internet megjelenéséig más módon, mint vásárlással nem tudtak hozzájutni számukra használható tartalomhoz, tehát elektronikus szótárát is vásárolni kényszerültek. Ma viszont teljesen spontán módon használni tudják az ingyenes – ám nem feltétlenül a legjobb minőségű – internetes szótárakat, illetve szótárnak mondott egyszerű adatbázisokat. Ez körülbelül azt jelenti, hogy az a terület, amely korábban csak vásárlással volt megtalálható, az most egyszerű, ingyenes internetes eszközök használatával helyettesíthető. Ezen a területen tehát az így kialakult internetes szótáraknak egyetlen

igazi ellenfelük van: az elektronikus (de nem feltétlenül internetes) szótárak, amelyek bizony pénzért vásárolhatók csak meg, pontosan úgy, mint papír-társaik.

Ingyen vagy pénzért?

Az elektronikus világban tehát a pénzért beszerezhető szótárak állnak szemben az ingyenesen hozzáférhető szótárakkal. A jelenség természetesen nemcsak a szótárvilágban tipikus: így jártak az enciklopédiák, a lexikonok, de akár a napilapokat is említhetnénk. Az interneten ugyan lehet pénzért is szolgáltatni, amit a jó minőség talán meg is kívánna, de különösen itt, Közép-Kelet-Európában a fizetős internetes szolgáltatások használatának nem alakult ki igazi kultúrája. Magának a világhálón történő vásárlásnak nem az az eleme a kritikus, hogy tulajdonképpen mennyibe is kerül az illető szolgáltatás, hanem az, hogy egyáltalán fizetni kell valamiért, ami az interneten van. Ha egyébként bármilyen fizetős szolgáltatásnak a világban létezik olyan alternatívája, hogy „nem kell fizetni érte”, az emberek többször is meggondolják, hogy elég érvük van-e a fizetős lehetőség választására. Ráadásul, ha az ingyenes szolgáltatás nem használhatatlanul rossz, csak esetleg „kevésbé jó”, akkor hamar meg is születik a döntés az ingyenes változat használata mellett. Sőt, sokszor ideológia is gyártódik hozzá: „nem is kell ez a professzionális szótár nekem annyira, hogy fizetnem kelljen érte, hiszen nekem csak néha van szükségem szótárra, arra meg jó ez az egyszerűbb is”. Szótári területen is kialakul tehát az a jelenség, ami az internetnek gyakorlatilag minden területén: nem baj, ha valami nem igazán jó, csak ne kérjenek érte pénzt, mert annyira azért nem kell. Valamilyen furcsa értelemben az internetről „dekódolható” lett egy olyan üzenet, hogy „nem baj, ha valami nem a legjobb, csak ingyen legyen”. Ahhoz a fizetős szolgáltatás tartalmának vagy ezerszer jobbnak kell lennie az ingyenesnél, vagy nagyon elkötelezettnek kell lennie a felhasználónak az adott témakör irányában, hogy hajlandó legyen fizetni a jobb vagy több tartalomért. Ha van is ilyen terület az interneten, az nem igazán az elektronikus szótárak területe.

Ebben a világban a papírszótárak értelemszerűen nincsenek jelen, mert papírszótárát eltulajdonítani – mint már utaltunk rá: a háromdimenziós világban még mindig működő etikai korlátok miatt – nem szokás. A papír alapú „univerzumban” viszont nem feltétlenül veszik meg az értékes, de olykor nem olcsó szótárakat, aminek alapvetően a potenciális vásárlók anyagi helyzete, vagy csak az alacsonyabb szótári műveltség is az oka.

A szótárak mérete

Egy másik érdekes kérdés: a szótár mérete. Sokan azt gondolják, hogy a legjobb szótár mindig a legnagyobb, tehát igényük sokszor erre irányul (ha anyagi helyzetük ezt lehetővé teszi). Mások ezzel szemben legfeljebb azt mondják ellenérvként, hogy „úgy-sincs arra a hatalmas szóanyagra szükséged”. Azt a szótárkritikákban ugyan elolvashatnák, hogy például mennyire modern egy szótár tartalma, azaz mai életünkben mennyi mindenre használható, illetve nem használható a benne levő tartalom, de sajnós, nyugodtan kijelenthetjük, hogy a szótárkritikákat a potenciális szótárvásárlóknak igen kis százaléka olvassa. A hazai szótárak minőségének autentikus jelölését egyébként az MTA Szótári Munkabizottsága fontos feladatának érezte, ezért 2007-ben elindította a „Kiváló magyar szótár” díjat, melynek kitüntetettjei a felhasználók szá-

mára is egyértelműen üzenhetik, hogy díjazott műveikben valóban autentikus a tartalom.

A szótár méretének mint egyetlen – de legalábbis elsődleges – mérőszámnak a gyakori használata arra sarkallta a kiadókat, hogy minél nagyobb számot közöljenek, hiszen az van előrébb a marketingversenyben, aki nagyobbat mond. Ez a „számbevendés” az elektronikus változatok világában még inkább jelen van, hiszen ott még sokkal nehezebb ellenőrizni egy darabszámra vonatkozó állítás valódi tartalmát. Egy könyv alakú szótár vastagsága, papírja, tipográfiája irányadó lehet a szótár mérete tekintetében, ám ez az elektronikus változat esetén egyáltalán nincs így. Az elektronikus szótárak méretei nehezen ellenőrizhetőek, így az lett az elsődleges marketingfogás, hogy az adott szótár a legnagyobb. Legalábbis – jó esetben – van olyan megszámlálásra épülő ismerv, amely szerint ez így is van. Ez az ismerv szótártani szempontból sokszor nem meghatározó, ám hatásos a vásárlók körében. Ráadásul akárhány szócikk is van egy papírszótárban, az elektronikus változatban lehet ennél sokkal többet is mondani, mert – épp a kiinduló szótárnak az esetleges nem jogszerű felhasználása miatt – nem szívesen közlik az eredeti számokat. Nyilván valamilyen technikával ezeket a számokat át lehet alakítani olyanná, hogy ezen az alapon a plágiumnak még a gondolata se jelenhessen meg. Ennek a „számmissztikának” egy másik hozadéka, hogy a felhasználót könnyen félre lehet vezetni, hiszen – mint már említettük – a számokat igen nehéz ellenőrizni, ugyanis a legtöbb elektronikus szótár felületén nem látszanak egyidejűleg a szócikkek.

Az általános szótárak és a szakszótárak viszonya

Az elektronikus szótárak esetében a gyors és könnyű kipróbálhatóság segíthet eldönteni, hogy egy adott szótár ad-e releváns találatokat azokon a területeken, ahol a későbbiekben dolgozni szeretne a felhasználó. A papírszótárak világában a boltban állva, némi lapozgatás után kell ezt a döntést meghozni, ami nagyobb rizikóval jár, mint a konkrét szövegek konkrét szavain kipróbálható számítógépes eszköz. Nyilván a gépi kipróbáláshoz is kell egy kis gyakorlat. Például a szakfordítók számára egy szótár akkor igazán jó, ha a szakterületük terminusait jól lefedi. Ha ez egy nagyon általános szakterület, akkor az úgynevezett nagyszótárak is tartalmazhatják az elvárt nyelvi egységek nagy részét. A nagyon ritka szakszavakat persze még a nagyszótárak sem tartalmazzák. Ilyenkor jön szóba a valódi szakszótár, amely bizony sok szakterületen könyv alakban nem is létezik, és sokszor csak egy szűk szakembergárdának vagy épp az adott szakfordítónak a saját szógyűjtése helyettesíti.

Könnyen észrevehető tehát, hogy általában a hétköznapi használat és a szaknyelv határa adja a két jól ismert szótárméret, a középszótárak és a nagyszótárak különbségét. Kijelenthető az is, hogy egy nyelv ténylegesen használatos szavait egy jó középszótár mindig tartalmazza, a nagyszótárak pedig valójában a különböző szakirányok (biológia, mérnöki tudományok, jog stb.) olyan kifejezéseit is felajánlják, amelyek a mindennapi, nem szakirányú nyelvhasználatban nagyon ritkán fordulnak elő. Valójában ezeknek a szavaknak és kifejezéseknek a megtalálására alakult ki a szakszótárak világa, ám a határaikat nagyon nehéz megtalálni, a kritikus kifejezéseket jól fordítani pedig csak épp azok a professzionális szakfordítók, szaktolmácsok tudják, akik számára – épp a szaknyelvben való jártasságuk miatt – ezek a szótárak sokszor kevésbé

szükségesek. Ezzel valójában azt írtuk körül, hogy a szakszótárak előállítására viszonylag drága és munkáigényes, mert profikat kell megfizetni a létrehozásukhoz, ám eladásukat épp a szaknyelv szűk használati területe teszi nehezzé. Így alakult ki az, hogy a szakszókincs egy részét nem külön szótárban, hanem a nagyközönség anyagilag tehetősebb, de nem szakirányú érdeklődésű részével „fiztetik meg”, amikor ezek nagyobb hányadát az általános közép-szótárakhoz adva nagyszótárakat hoznak létre. Magyarul: a szótárak mérete nem az elektronikus korban kezdett el marketingfogás lenni, csak a két szótártípus különbségéből adódóan az elektronikus szótárak esetében a méretek hangoztatásával még nagyobb az esély a felhasználó félrevezetésére.

Nyelvtechnológiai újdonságok: kipróbálhatóság és tövesítés

Az elektronikus szótárak felhasználói sok esetben persze csak az adott nyelv iránt érdeklődők vagy az azon a nyelven időnként olvasók, nem pedig professzionális felhasználók. Így fordulhat elő, hogy a szótárak – felszínes, de a vásárlás szempontjából meghatározó – megítélése az első egy-két keresendő szó megtalálásával korrelál: ha megvannak a szótárban, a felhasználó jó szótárnak minősíti; ha nincsenek, akkor rossznak. Ráadásul ez sokszor még azzal a következménnyel is jár, hogy akiknek csak időnként lenne szükségük a szótárra, az egyébként sikeres teszt után mégsem veszik meg. A pénz gyors kiadása helyett még az internetes szótárak előtti korban is sokan úgy döntöttek ilyenkor, hogy a néha megkeresendő egy-egy szóhoz megfelel nekik az egy-két szóra teljes találatot adó, de hosszabb használat esetén már csak ún. demó módban futó bemutató verzió is. Ma, amikor ezeket a találatokat ingyen is megkaphatja az ember valahonnan az internetről, még nagyobb a kísértés: miért is kellene megvásárolni az elektronikus szótárat, ha az interneten ott van ingyen – majdnem ugyanez?¹

Az elektronikus szótárak készítőinek időközben újabb kihívásokkal kellett szembenéznük, hiszen ha a felhasználó nem a szótári alapalakot írja be a keresőablakba, hanem azt, amit a szövegben olvas, azaz az aktuális, esetleg toldalékos formát, akkor az olyan szótárakban, amelyek nincsenek kombinálva szótövesítő modullal, nem lesz találat. A demonstráció céljára gyártott szótárverziók készítőinek tehát két szélsőség között kell döntenük: vagy nem lesz elég jó a demóváltozat, és akkor nem veszik meg a teljes programot se (mert félnek, hogy az is keveset tud majd); vagy nagyon jó lesz a demóváltozat, és így az is kielégíti a nem-professzionális felhasználók igényeit (aminek következtében szintén kisebb az esély a vásárlásra). A papírszótáraknál ilyesmi nincsen: ha valami nem jó, akkor az nem jó, és ezt rögtön láthatja a szótáron a felhasználó, hiszen a teljes változatot a kezében tartja.

A márkanév

A papírszótárak világában egyébként maga a „márkanév” is nagyon erős tud lenni. A tapasztalatok azt mutatják, hogy ma Magyarországon még mindig sokan azt kérdezik egy angol szótár esetében, hogy az az Országgh-szótár-e, a németről, hogy a Hallász-e, a franciáról, hogy az Eckhardt-e? Évtizedeken át ugyanis csak ezek a nevek fémjelzték az adott nyelvi szótárkiadást hazánkban. Az „igazi” márkanevek persze

¹ Itt viszont a „majdnem”-re érdemes hangsúlyt tennünk.

bárhol a világban nagyon számítanak a szótárak kiválasztásánál, de ezek általában a kiadók nevei, tehát az valóban számít, hogy az Oxford, a Longman, a Langenscheidt, a Robert, a Zingarelli vagy bármely más, nagynevű kiadó adta-e ki az adott szótárat. A hazai szótárkiadásban hosszú idő telt el, míg egyáltalán meg tudtak jelenni az előzőkben említettektől eltérő szótáriskolák és a hozzájuk tartozó új lexikográfusnevek, akiket az igényes szótárforgató közönségnek legalább egy része ismer is. Ma már elmondható, hogy hazánkban egyre szélesebb körben ismertek például a Lázár–Varga-féle angol, a Hessky-féle német, vagy akár a Bárdosi-féle, akár a Pálffy-féle francia szótárak. Ezeknél az említett műveknél természetesen nem az az érték, hogy új neveket hoztak be a köztudatba, hanem az, hogy a szótári tartalom mögötti munka gyökerei nem a korábról jól ismert és fentebb emlegetett elődöknél keresendők, hanem egy mai szótárkészítő generációnak a modern világ sok szempontból megváltozott körülményeihez jól alkalmazkodó, ugyanakkor alapos lexikográfiai hátterében.

Változatok és másolatok

Sajnálatos módon – és elsősorban a szerkesztők említése nélkül – a korábbi nagynevű szótáraknak rengeteg változatuk van közkézen. Igaz, ez elsősorban az elektronikus világban tipikus. Mivel meglehetősen nagy jogi és lexikográfiai apparátus kell ahhoz, hogy „hivatalosan” egy „átvételtől származó” szótárat egy piacon levő másikhoz „elég hasonlóknak” tudjon minősíteni valaki, sokan élnek ezzel az – illegális – lehetőséggel. Nagyon szerteágazó kritériumrendszert kell figyelembe venni ilyenkor, és még inkább így van ez az elektronikus területen. A szócikkek belső szerkezete kulcskérdés a lexikográfiában, ám az átalakított – vagy pontosabban: a szerzett, a másolt vagy egyszerűen csak az eltulajdonított – szótárak esetében ez azért kulcskérdés, mert a másolatnak minősíthető szótár általában ezt a struktúrát szünteti meg, egyszerűsíti le – épp a hagyományos és az elektronikus környezetben való használat különbözőségére hivatkozva. Az elektronikus kiadások nagy részében a szócikkszerkezet nagyon sokszor nem látszik, ugyanis tipikusan szó–szó, kifejezés–kifejezés, esetleg egy kifejezés–több kifejezés alakú, adatbázis-szerű párok jelennek meg a képernyőn. Ezek a szó-, illetve kifejezéslisták sem tipográfiai elrendezésükben, sem a szócikkeik sorrendjében nem követik a kiinduló „forrásanyagot”, így egy hagyományos szótárral való esetleges azonosításuk ezen az alapon meglehetősen nehézkes. És akkor a feldolgozott szókincs-ről magáról még nem is beszéltünk. Az elektronikus szótárban ugyanis nem „sorban” keresünk a szemünkkel, hanem amit a beviteli ablakba beírtunk, azt a szótárprogram segítségével valahol, valahogyan, a felhasználók által nem ismert módon megtaláljuk. Minden valamirevaló szótár copyright-bejegyzésében ugyan ott szerepel, hogy „sem egészben, sem részben nem másolható”, de hogy mit jelent a „részben”, azt ebben a kontextusban nagyon nehéz értelmezni. Az említett példában a hasonlóságra könnyen utalni képes tipográfiát és az egymás utáni elrendezést vették csak ki az eredeti anyagból, ám ezek után azt is igen nehéz megmondani, hogy ez még mindig az eredeti szótáranyag-e. Annak ellenére így van ez, hogy a szócikkekben szereplő karaktersorozatokat erősen hasonlítanak az eredetiben szereplő karaktersorozatokra. Az „átalakításkor” sokszor ezek egy részének az „egyszerűsítése”, vagy egyszerűen a kihúzása is megtörténik, és akkor a maradék szóanyag esetében szinte lehetetlen az eltulajdonítás tényét a fennmaradó címszókészletre visszavezetni. Azt az állítást ugyanis nehéz

védni, hogy például egy fenti módon létrejött angol–magyar szótárban csak az eltolajdonítás miatt lett a „table” magyarul „asztal”, vagy a „horse” éppen „ló”. Az igazi kérdés tehát az, hogy hol kezdődik az a nyelvi ismeret, ami egy adott szótárban úgy szerepel, hogy a felhasználó e szótár nélkül nem tudna hozzájutni? Mivel minden felhasználónak más a nyelvi szintje, azt nagyon nehéz egyértelműen kimondani, hogy mi az, ami még más szótárakkal való hasonlítás nélkül is elfogadható, azaz általánosan ismertnek tekinthető; és mi az, ami csak azért van benne és úgy van benne egy szótárban, amiért és ahogyan azt egy adott lexikográfus egy másik szótárban odatette.

Fordítóprogram vagy szótárprogram?

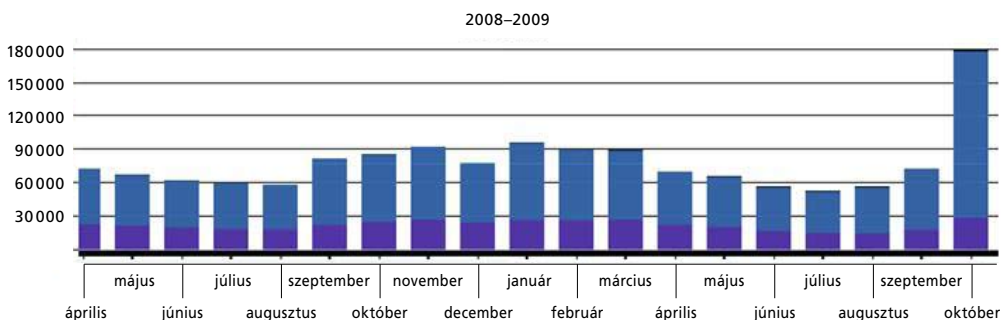
Az elektronikus, de különösen az internetes szótárak tehát a fenti okokból szépen lassan – ha tetszik: módszeresen – leszoktatják a felhasználót a strukturált szócikkek értelmezéséről. Magunk² is tapasztaltuk ezt, hiszen az általunk üzemeltetett www.webforditas.hu weboldalon a gépi fordítóprogram mellett többek között megtalálható egy szótárprogram is³. Ennek a fordítómodulhoz képest eleinte különösen alacsony volt a látogatottsága. A feltételezett okok közt nyilván a sok más konkurens internetes szótár szerepelt, és csak később vettük észre, hogy sokan nem is sejtik a különbséget a beadott mondatok elvileg legjobb fordítását megadni szándékozó fordítóprogram és a szavak, kifejezések összes fontos ekvivalensét megadni képes szótárprogram között. Ennek következtében sokan rájöttek, hogy még a másik fülre való átkattintás is fölösleges, azaz a fordítóprogram helyett a szótárprogram felületét mutató ablakba való átlépést sem kell megtenni: ha a fordítóprogramnak egy-egy szót írunk be mindössze fordítandóként, akkor ez „automatikusan” szótárprogramként fog üzemelni. Valóban, semmi nem zárja ki, hogy egy-egy szövegben egyszavas egységek is legyenek, például rövid válasz egy kérdésre. A belső kimutatások azt bizonyították, hogy igen sok betűköz nélküli, rövid fordítandó anyag került a fordítóprogramhoz, amit – feltehetően – szótári lekérdezés céljából írogattak be a felhasználók az idők folyamán. Mivel napi több tízezer látogatóval van dolgunk, az arányszám meglehetősen nagy szótári érdeklődést látszott takarni. Ezért azt a megoldást választottuk 2009 őszén, hogy ha már a felhasználó nem tudja, mi a különbség a fordítóprogram és a szótár között, akkor segítünk neki: a korábban elkészített szótárablak fog jelentkezni akkor is, ha a fordítóablakba írta be az érdeklődő a szótári lekérdezésnek minősíthető, rövid, nem mondatszerű kérést.⁴ Azt kellett látnunk, hogy szeptemberről októberre azonnal többszörösére növekedett a szótárprogram használóinak száma (2. ábra), másrészt – a visszajelzések alapján – ezeknek a használóknak egy része a számára bonyolult, hagyományos lexikográfiai elvek alapján felépített szócikk-szerkezetet nehézkesnek, nehezen értelmezhetőnek, sőt egyenesen zavarónak tartotta a fordítóprogram által korábban visszaadott egyetlen megoldáshoz képest. A fordítóprogram ugyanis minden esetben a legjobbra törekszik (ami sajnos, sokszor nem esik egybe az ember által legjobbnak tartott fordítással), ezért mindig csak egyetlen kimenettel szolgál:

² A MorphoLogic nyelvtechnológiai vállalkozás kutatói.

³ <http://www.webforditas.hu/szotar.php>

⁴ Ezt egyébként meglehetősen nehéz automatikusan eldönteni, mert ami rövid, az nem feltétlenül szótári lekérdezés, ami hosszabb, az meg nem feltétlenül mondatként fordítandó nyelvi egység – de az automatikus megoldásoknál néhány százalék hiba mindig marad az emberi döntésekhez képest.

a szerinte legjobbal. Ez történik természetesen akkor is, ha egyetlen szó a bemenet. Ilyenkor ezt próbálja meg a maga módján a „legjobban” lefordítani, ami egy többjelentésű szó esetén a lehetséges további jelentések teljes „lenyelésével” jár. Ez tehát a tapasztalatok alapján kevésbé zavarta a felhasználók egy részét, mint a pontosságra, kiegyensúlyozottságra törekvő, a szó lehetséges jelentéseinek mindegyikét felsoroló szócikk megjelenése. Ez utóbbi bonyolultságának értelmezése nagyobb problémát jelentett számukra, mint az esetleg pontatlan fordítás. Az üzenet nemcsak a szótári világ, hanem egész jelenünk egyik komoly társadalmi problémája, mely az igénytelenség növekedésére és a némi intellektuális befektetést igénylő válaszoknak (esetünkben: egy hagyományos szócikk elolvasásának és értelmezésének) a lehetséges elkerülésére utal.



2. ábra

A szótári szerkezet és a szavak környezete

Kétségtelen, hogy a mai szótárak szerkezete a Gutenberg-galaxis melléktermékeként jött létre, és mint ilyen, sok – papírvastagsághoz, ólombetűhöz, nyomdászai szokáshoz kapcsolódó – esetlegesség is van benne. Azért alakultak ki például a szóbokrok, mert tipográfiaileg jól lehet bizonyos nyelvi hasonlóságokra rávezetni az ember szemét, ugyanakkor jól lehet rövidíteni a hasonló egységeket, ami papírmegtakarítással is jár. Az elektronikus szótárak szívesen keverik az ékezetes betűket vagy a nagybetűt és a kisbetűt, mondván, hogy majd a felhasználó kiválasztja a kapott találatok közül azt, ami őt érdekli. Így lehetséges az, hogy internetes szótárak a „ló” szóra a „ló”, a „lokál”, vagy szerencsétlenebb esetben még a „malom” szót is találatul adhatják, hiszen ezek ékezetlen változatai mind tartalmazzák keresőkérdésünk ékezetlen változatát, a „lo”-t. Torz, de rövidebb találatihalmaz-lista esetén ez valóban nem nagy gond. Le kell viszont szögezni, hogy – a terjedőben levő fenti keresési gyakorlat ellenére – a hosszabb keresgéssel sem könnyen fellelhető megfelelők nem szükségszerű velejárói az elektronikus szótárak használatának. Sőt, épp ellenkezőleg, a dinamikus lehetőségekkel, nyelvi elemzőprogramoknak a szótárhoz illesztésével új világ jelenhet meg a szemünk előtt: az a világ, ahol beírt keresőkérdéseink nyelvi értelmezésén keresztül a nyelvtechnológiai tudással is rendelkező számítógépes program a számunkra az adott helyzetben legrelevánsabb találatokat is képes megjeleníteni. Ezáltal persze nyelvfüggővé válik a szoftvereszköz, aminek az a hihetetlen következménye a globalizálódó világban, hogy más nyelvi piacokon – sokszor nem olcsó – egyedi átalakításra, nyelvi adaptáció-

ra van szükség a program más nyelvekkel való használhatóságához. Ez pedig nem igazán előny a „gyorsan és olcsót” titkos szlogent a szemei előtt tartó világunkban, és további költségekkel is jár. A minőség mint harmadik tényező csak az előző kettővel – a sebességgel és az árral – együtt jelentkezik, méghozzá úgy, ahogy a sokak által jól ismert reklámszöveg mondja: „Jó minőségben dolgozunk, olcsók vagyunk és gyorsak: Ön bármelyik kettőt választhatja!”

A hagyományos szótáraknak van még egy igen fontos tulajdonságuk: a megértéshez szükséges nyelvi információ egy részét „a felhasználó fejében helyezik el”, azaz arra építenek, hogy a felhasználó ember, méghozzá olyan, aki valamennyire jártas a szóban forgó nyelvek legalább egyikében. Sőt, egy kicsit még a felhasználó kulturális háttérére is gondolnak, ezért a nyelv használati régiójában elvárható világismeretet természetesen feltételezik. Ezzel szemben a számítógépes szótárak esetében ennek az „intelligenciának” egy részét a gépnek kellene ismernie. Persze a végfelhasználó itt is az ember⁵, ám közben van egy dinamikusan működő „értelmező” modul, ami a papírszótárak esetében nincs, ott csak egy statikus nyomtatott anyag áll előttünk. Ezt a dinamikus felületet a számítógép kezeli, de a programok, amelyek arra valók, hogy a számítógép felületén megjelenő szótárak a felhasználó számára jól működjenek, sokszor nem veszik még figyelembe azt a nem elhanyagolható ténytet, hogy a megjelenítendő adatok nyelvi adatok. Más szóval: a szótári program nem olyan, mint egy áruházi termékeket tartalmazó adatbázis, hanem a kiinduló szöveggörnyezetre is nagyon érzékeny – nyelvi – tartalmat bemutatni képes eszköz. Összefoglalva: mára az informatikusok vették át azt a meghatározó szerepet, amit korábban – egészen más módon – a nyomdászok birtokoltak a szótárak formájának kialakításban.

Üzlet vagy kulturális misszió?

A kimutatások szerint egyre kevesebb papírszótárra van szükség, mert a szótáreladás egyre kevésbé kifizető üzlet. Sok kiadó már csak az angol nyelvben bízik, hogy tudniillik minden nyelvre és minden nyelvről szükséges lesz olyan szótár, amelyik az adott nemzeti nyelv és az angol között adja meg a szavak és a kifejezések ekvivalenseit. A többi nyelv egyre inkább csak másodlagosan van jelen a szótárkiadásban. Erre mutat az 1. táblázat is, amely a különféle keresőprogramokból a – meglehetősen sok nyelvre történő fordítást felajánló – webfordítás.hu oldalra érkező, az angolon kívül más nyelvet nem említő húszt leggyakoribb keresőkérdés listája⁶: a felhasználók közel 90%-a ezeknek a kérdéseknek a segítségével jut el a weboldalra.

1. angol–magyar	45,81%
2. angol–magyar szótár	9,56%
3. web	6,63%
4. angol szótár	3,15%
5. online fordítás	2,68%
6. angol	2,53%

⁵ Bár ma már egyre több géppel olvasható szótár további számítógépes programok számára és nem az embernek készül.

⁶ <http://www.alex.com>: High Impact Search Queries for webforditas.hu (2011.03.05.)

7. program	2,45%
8. magyar–angol szótár	2,22%
9. angol–magyar fordító	2,21%
10. magyar–angol fordító	1,57%
11. fordítás	1,44%
12. angol fordító	1,31%
13. translate hungarian to english	1,05%
14. magyar fordítás	1,05%
15. online	1,02%
16. web fordítás	0,80%
17. hungarian to english	0,66%
18. translate	0,64%
19. angol–magyar fordítás	0,64%
20. fordítás magyarról angolra	0,61%
	88,03%

1. táblázat

Nyilván ezek a jelenségek a kisebb nyelvek és a gazdaságilag nehezebben működő régiók esetében látszanak még markánsabban. Érdeemes elgondolkozni például azon, hogy mikor adtak ki utoljára román–magyar, lengyel–magyar vagy magyar–szlovák szótárt, amelyek ebben a földrajzi régióban nyilvánvalóan iszonyú fontosak lennének, de a kiadásuk üzleti alapon – amint a gyakorlat is bizonyítja – képtelenség. Így aztán sok, kulturális szempontból fontos nyelvpárra egyáltalán nincs modern, a kor jelenségeinek nyelvi vetületét jól követő szótár. Mivel a szótárkiadás irányítása ma már egyre több helyen nem a kulturális tárcák vagy a tudományos akadémiák kezében van, nagy az esély arra, hogy a jövőben a modell megváltoztatása nélkül, azaz pusztán üzleti alapon egyre kevésbé születnek ilyen szótárak.

Arról is érdemes beszélni, hogy vannak olyan nyelvpárok (például a magyar és az albán vagy a hindi), amelyek szótárai gyakorlatilag készen vannak. Készítőik nem üzleti alapon, hanem – ha tetszik – megszállottságból, a nyelv és a kultúra szeretetétől vezérelve hozták létre őket, méghozzá nagyon igényesen, komoly lexikográfiai munkával. Kiadásuk magas szintű lexikográfiai tartalmuk ellenére sem várható, mert senki sem akar üzleti bukást, márpedig a magyar és az említettekhez hasonló bármilyen más nyelv szótárának kiadása igen kevésbé ígér gyümölcsöző üzletet.

Terv – vagy esetleg megoldás?

A fenti problémákat kizárólag egy alapvetően csak lexikográfiai tevékenységet végző, lexikográfiai tartalmakat létrehozó, illetve a meglévő tartalmakat modern eszközökkel közkinccsé tevő intézménnyel lehetne megoldani. Ennek az intézménynek nem akadémiai kutatóhelynek kellene lennie, hiszen ezt a területet nem kutatni, hanem elsősorban művelni kell, értelemszerűen az elérhető legkorszerűbb technikai eszközökkel. A „szótárlétrehozás” természetesen szigorúan tudományos alapokon történne: további amatőr megoldásoknak helye nincs. Egy ilyen intézmény működését a mai kornak megfelelően kizárólag professzionális számítógépes nyelvészek biztosíthatják,

valamint az a hihetetlen mennyiségű, folyamatosan létrejövő lexikográfiai produktum, amit az előző időszak létrehozott, de aminek az eredményeivel a felhasználók sajnos nem találkozhattak az üzleti alapú publikálás szűk keresztmetszete miatt. A technológiai eszközök magas szintű használata teszi ezeket az eredményeket versenyképessé. Ilyen jellegű tevékenységgel egy-egy szótár esetében már próbálkozott néhány vállalkozás, de szisztematikus munkát ebben az irányban csak egyetlen nyelvtechnológiai műhely végzett eddig: a MorphoLogic. A most kereken húsz éves vállalkozás kialakította a rendszert a mai igényeknek maximálisan megfelelő szerkezetű szótárak építésére, a meglévők nyelvészeti elvek mentén történő feljavítására, átalakítására, valamint ezek elektronikus és papír formában történő publikációjának támogatására. Ezt a rendszert kár volna újból létrehozni. Ám – mint minden vállalkozás – ez is piaci szereplő, ezért tevékenységének eredményét, a létrejött szótárakat üzleti alapon tudja csak piacra hozni. Lehetséges volna azonban, hogy az internet által fokozatosan ingyenessé tett világban a jó minőségű lexikográfiai tartalom alanyi jogon járjon például minden magyar anyanyelvű állampolgár számára. Ehhez a háttérben egy olyan állami szándék volna szükséges, amelyik arra törekszik, hogy

- a magyar és más nyelvek közötti szótárak létrejöjjenek,
- ha megvannak, publikussá váljanak, és
- a közvetlen haszonszerzés ne terhelje a lexikográfia világát.

Ennek a tervnek a támogatásával lehetővé válna, hogy egy olyan vállalkozás, mint az ez évben kereken húsz éve nyelvtechnológiai cégeként működő MorphoLogic erre irányuló tevékenységét a köz javára fordítsa, és lexikográfiai működését ne a piac szabályai, hanem a regionális kulturális igények határozzák meg. Az erre a feladatra alkalmas stáb tehát már létre is jött: a kérdés csak az, hogy a felvázolt működéshez szükséges háttér mikor teszi lehetővé, hogy megjelenjenek azok a jó lexikográfiai minőségű elektronikus – és igény szerint természetesen papír alapú – termékek, amelyek elnyomnák a végfelhasználók számára hasonlóképpen ingyenes, ám kevésbé jó, de napjainkban még jelen lévő más próbálkozásokat.

IRODALOM

- Oszetzky Éva – Pálfy Miklós – Borbás László (2004): *Szótárszerkesztés, szótárkiadás*. Szeged: Grimm Kiadó.
- Pálfy Miklós (1999): Számítógépes szótárírás és szójelentés. (A címszókiválasztás és a szócikktagolás elvei: változások a lexikális struktúrákról kialakult hagyományos felfogásunkban.) *Modern Nyelvtanítás* 5/2–3, pp. 18–25.
- Prószéky Gábor (2007): A számítógépes nyelvészet hatása a nyelvleírásra. In: Talyigás Judit (szerk.): *Mozaikok a hazai telematika eredményeiből: Gordos Géza 70. születésnapjára*. Budapest: MIL-ORG Kft., pp. 83–87.
- Prószéky, Gábor – András Földes (2005): Between Understanding and Translating: A Context-Sensitive Comprehension Tool. *Archives of Control Sciences* 15(4), pp. 637–644.
- Prószéky Gábor – Tihanyi László (2009): Webfordítás.hu: egy internetes nyelvtechnológiai szolgáltatás tanulságai. In: Vincze Veronika (szerk.): *A VI. Magyar Számítógépes Nyelvészeti Konferencia előadásai*. Szeged: SZTE, pp. 19–23.