

PÁLFY MIKLÓS

Számítógépes szótárírás és szójelentéstan

A címszó kiválasztás és a szócikktagolás elvei: változások a lexikális struktúrákról kialakult hagyományos felfogásunkban

„A számítógépes szótárakat többféleképpen csoportosíthatjuk. Megkülönböztethetjük a számítógéppel olvasható szótárakat (machine-readable dictionaries) és a számítógépesített szótárakat (computerized dictionaries) [...]. [Ez utóbbi] olyan adatbázis, amely jól tükrözi a szótár belső szerkezetét, speciális adatbázis-kezelő szoftverrel van ellátva, amelynek segítségével hatékonyan kereshetünk a szótár bármely részében.”¹

1. Az informatika (inkább az informatizáltság) jelentősége tehát elsősorban abban rejlik, hogy a szótárak *adatbázisként* jönnek létre. Az adatbázis-jelleg pedig azt jelenti, hogy a szótár struktúráját meghatározó kódokkal a szótár elektronikus változatában adatok (és általunk tetszőlegesen összeállított adathalmazok) között összefüggéseket deríthetünk fel, illetve keresztreferenciákat kereshetünk és kaphatunk.

2. Az adatbázisok különösen jelentős szerepet töltenek be a kétnyelvű szótárírásban: *egy kétnyelvű szótárnak ugyanis szinte szükségszerűen mindkét irányban kódoló és dekódoló funkciót kell ellátnia*, ami azzal jár, hogy a szócikkek tagolásában az egyes jelentések elhatárolására jelentősen megnő a különböző kiegészítő információk száma: ezek lehetnek szinonimák, árnyalatok, jelentésszűkítések, analógiák, körülírások, definíciók; de erre a célra a különböző stílusréteg-minősítések és a használati értékre vonatkozó utalások is felhasználhatók, akárcsak a szak- vagy alkalmazási terület jelzése, vagy a tipikus szerkezetek megadása. Mindezek hatékony kezelése csak adatbázisban képzelhető el. A két elv – egyrészt tehát a kódoló és dekódoló funkció együttes alkalmazása, másrészt az adatbázis-jellegű szerkesztés – nemcsak kiegészíti, hanem egyenesen feltételezi is egymást.

3. Mind elméleti, mind gyakorlati szempontból fölmerül ezek után a kérdés: az informatika, vagy helyesebben: az informatizáltság miként módosítja a lexikális struktúrákról kialakult eddigi felfogásunkat? Változtat-e mindez a szójelentéstan alapfogalmaink (főleg a szinonímia és a poliszémia) értelmezésén? Hogyan látjuk a fogalom és a jelentés viszonyát?

Ebben a kis tanulmányban azt vizsgáljuk, hogy a fent mondottak miként befolyásolják, illetve határozzák meg a kétnyelvű szótárak, a kétnyelvű lexikális párhuzamok struktúráját. Illusztrációként a Szegeden, a József Attila Tudományegyetem Francia Tanszékén 1998-ban elkészült új francia–magyar elektronikus kéziszótár adatait használjuk fel. A magyarországi szótárak közül ez az első francia szótár, amely a szerkesztés kezdetétől egy sajátos „nyelvtan” által definiált adatbázis koncepciójának felhasználá-

sával készült. Megpróbálunk választ adni arra a kérdésre is, hogy az informatizáltság mennyiben módosít(hat)ja általános szójelentéstan szemléletünket.

4. A munkálatok elején, a szerkesztési útmutató kidolgozásakor lehetetlen volt minden körülményt előre látni. A szoftver állandó módosítása s a már elkészült anyag újra és újra átnézése nagy idővesztéssel járt. Ezért különbséget kellett tennünk a szócikkek szerkezeti alapelemei és járulékos, vagy másodlagos elemei között. Az utóbbiak kódolása manuális, ez előzőeké billentyűmakrók segítségével történt.

Ami a szócikkek tagolását illeti, erre nézve a következő újításokat vezettük be.

Mivel a célnyelvi megfelelők „ömlesztett” megadása zavarba ejtő, ezeket differenciáltan kellett megadni. Ezen két dolgot értek:

a) Egyrészt a célnyelvi megfelelők közötti szemantikai affinitásnak, illetve az affinitás hiányának a jelzését. (Az „affinitás” kifejezéssel szándékosan kerülni kívánom a „szinómia” terminust.) *Nyilvánvaló ugyanis, hogy más az értéke azoknak a megfelelőeknek, amelyek egy adott kontextusban többé-kevésbé fakultatív módon használhatók, s más azoké, amelyek között nem áll fenn felcserélhetőség, jóllehet szemantikailag közel állnak egymáshoz.* Másképpen adjuk meg tehát a megfelelőket, ha köztük lehetséges a (para)szinómia, s megint másként, ha ez a lehetőség nem áll fenn. (Az előbbi esetet törtvonallal, az utóbbit pontosvesszővel jelöljük.)

b) Másrészt differenciálás az is, hogy az eltérő megfelelők bizonyos esetekben egyetlen szócikktagolási egységen („jelentésen”) belül maradnak, más esetekben pedig külön „jelentés” veszünk a számukra föl attól függően, hogy ezek az eltérő megfelelők azonos fogalmi körbe tartoznak-e, vagy sem. (Tehát mind fogalmi, mind szemantikai szempontból azonos mező: törtvonal; azonos fogalmi, de eltérő szemantikai mező: pontosvessző; mind fogalmi, mind szemantikai szempontból eltérő mező: külön „jelentés”. Ha a több jelentés között nem érződik már az összefüggés, homonimákról beszélünk.)

Talán furcsának látszik, hogy fogalmi összetartozás alapján döntünk a jelentések elkülönítése során, ám gondoljuk meg: a megfelelők eltérése gyakran csak szóhasználati, vagy szókapcsolati kérdés. Úgy is fogalmazhatunk, hogy *a szócikkek tagolása során a fogalmi mezőknek elsőbbsége van a szemantikai mezőkkel szemben.* (Lásd a példákat.)²

Példák

Fogalmi körülírások

daltonien, -ienne adj n mf 1 (*absence de perception de certaines couleurs*)

színvak 2 (*confusion de certaines couleurs*) szintévesztő

daltonisme nm 1 (*absence de perception de certaines couleurs*) színvaktság 2 (*confusion de certaines couleurs*) szintévesztés

Felcserélhető és fel nem cserélhető ekvivalensek

1 bloquer (1) (*regrouper plusieurs choses*) összevon/tömbösít vmit: ~ **ses jours de congé** összevonja a szabadnapjait; **elle a bloqué ses cours sur sept semaines**

hét hétre tömbösítette az óráit; **l'assemblée nationale procède au vote bloqué** *Jur* az országgyűlés az egész törvénycsomagra szavaz

2 bloquer (1) *I v tr 1 (immobiliser qn/qc)* veszteglésre kényszerít vkit/vmit; eltorlaszol vmit: **nous avons été bloqués une heure sur l'autoroute** egy óráig rostokoltunk *fam* az autópályán; **le navire était bloqué par les glaces** a hajó a jég fogságába került **2 (arrêter qc)** meghúzza vmit; megszorít vmit: ~ **son frein** behúzza a kéziféket; ~ **un écrou** megszorít egy csavart **3 (~ qn)** beszorít vkit vhová: ~ **qn contre le mur** falhoz szorít vkit **4 (serrer)** leblokkol vmit; megakaszt vmit; kiakaszt vmit: **les roues ont été bloquées** leblokkoltak a kerekek; **les négociations sont bloquées** megakadtak a tárgyalások; **bloque la porte** akaszd/támaszd ki az ajtót! **5 Fin (geler)** befagyaszt; zárol: ~ **les prix/salaires** befagyasztja az árakat/a fizetéseket; ~ **un crédit** befagyasztja a hitel folyósítását; ~ **un compte en banque** zárol egy bankszámlát; **dépôt bloqué pour un mois** egy hónapra lekötött betét **6 ~ qn Psych** leblokkol vkit: **après cet échec, j'étais complètement bloqué** ez után a kudarc után teljesen leblokkoltam **7 Mil** blokád alá vesz **8 Sport ~ le ballon** lestoppolja a labdát **II v pron 1 (roue/machine; frein/porte) (kerék/gép)** megakad; (fék/ajtó) beakad **2 Psych (sujet qn)** (vki) leblokkol

Általánosított (rejtett) példák és egyedi (konkrét) példák

boisage nm 1 Constr (action; revêtement) fával burkolás: **le ~ du salon** a nappali fával burkolása **2 (d'une galerie de mine) (bányavárat)** aládúcolás **3 (matière)** dúcolat

boisement nm Agr (d'une rue; d'un terrain déboisé) (utca fásítás; irtás) (újra)erdősítés/erdőtelepítés: ~ **par plantation/semis** fásítás palántáról/magról **boisé, -e adj (pays) (tájék)** erdős/fás

1 boiser (1) *v tr Agr (un terrain)* (be)fásít/erdősít

2 boiser (1) *v tr 1 Constr (un mur) (falat)* (be)lambériáz **2 (une galerie de mine) (bányaváratot)** aládúcol

Más példát is tudnék hozni az előbbi tétel illusztrálására: a példákban törtjellel választ-hatunk el szemantikailag távol álló, de fogalmilag közeli szavakat (kizárólag a helynyerés céljából), a törtjel azonban nem használható, ha a szavak között nemcsak szemantikai, hanem fogalmi távolság is van: a fogalmi összefüggés tehát itt is fontosabb a szemantikainál.

- A szótárnak a francia felhasználót is ki kell szolgálnia, számára be kell töltenie a kódoló funkciót. Ezért az egyes jelentések könnyebb azonosítása érdekében gyakran francia nyelvű szemantikai, vagy fogalmi utalások³ előzik meg a magyar megfelelőket. (Lásd a *daltonien* példáját.) Mivel magyar nyelvű utalásokra is nagyon gyakran van szükség, s mivel ezek legtöbbször a francia nyelvű utalások ekvivalensei, ezeknek a használatát kiterjesztettük, s így indirekt példaként is felfogható párhuzamokat adunk. *Ezek a többletinformációk teszik a francia–magyar elektronikus szótárt újszerűvé a franciául tanuló magyarok számára is.*

Ezeket a párhuzamokat nemcsak generalizált, hanem indirekt, rejtett példának is tekintjük, s funkcionálisan szembeállítjuk a direkt, vagy egyedi példákkal. *A generalizált példák tehát többnyire fogalmi összefüggések alapján működnek, és elsősorban didaktikai célúak, az egyedi példák viszont főleg szemantikai értékűek, és gyakran a direkt megfelelőket helyettesítik.*

- A „fogalmi összefüggések” és „árnyalt értékű példák” ilyen szembeállításával tulajdonképpen a klasszikus, saussure-i dichotómiához jutunk vissza: *a fogalmi összefüggések paradigmatis megjelentésével áll szemben az egyedi példák diszkurzív jellege.*
- Különleges szerephez jut tehát a szinonímia jelensége: a kiinduló nyelvben ott szerepelhet az általános példák megadása céljából, ám a fogalmi összefüggések lehetősége miatt helyette gyakran analógiákat találunk – a célnyelvben (a magyarban) pedig lényegében a címszókiválasztást (poliszémia : homonímia) és szócikktagolást, valamint a jelentéstagolást szolgálja. Az előbbi esetben elég tág értelemben szerepel tehát, az utóbbi esetben viszont a lehető legszűkebben.

5. Az informatizáltság diktálta szerkesztési elvek tehát (adatbázis-jelleg és mind kódoló, mind dekódoló funkció) nagyjából a következő jelentős dichotómiák érvényesülését eredményezhetik a kétnyelvű lexikális párhuzamok szerkezetében:

- A szinonímia egyrészt (a francián belül) igen tág, az analógia felé mutató, másrészt viszont (a magyar megfelelők között) igen szűk értelmezése.
- A fogalom elsőbbsége a jelentéssel szemben.
- Generalizált példák : egyedi példák; vagyis: szótári jelentés : diszkurzív jelentés.

Az új francia–magyar elektronikus szótár szerkezetét a következő oppozíciók határozzák meg:

- a címszókiválasztás során: homonim felfogás : poliszém felfogás a célnyelvi megfelelők közötti szemantikai affinitások függvényében (szűken értelmezett szinonímia);
- a szócikktagolás során: fogalmi mezők : szemantikai mezők a kiinduló nyelven belüli szemantikai affinitások függvényében (tágra értelmezett szinonímia);
- s végül az egyes jelentéseken belül: generalizált példák : egyedi példák (fogalomkörü összefüggés : jelentésárnyalás).

Szótárunk koncepciója tehát a fő szójelentéstanai jelenségeknek rendszerszerű szemléletén, a következő nagy oppozíciókon alapul:

- homonímia : poliszémia
- poliszémia : szinonímia
- szinonímia : analógia
- analógia, illetve asszociációs mezők : generikus mezők

a) Ami a homonímia és a poliszémia kérdésére adott lexikográfiai válaszunkat illeti, ez így foglalható össze: az egyes jelentések árnyalt megadása érdekében mindenféle-

képpen figyelembe kell venni a célnyelvi megfelelők közötti szemantikai affinitást. Ez a szempont egyszersmind azt is jelentheti, hogy – a „B” nyelv felől nézve – „A” nyelv „a” szava polisizém, ha a $b_1 \dots b_n$ megfelelők között az affinitás fennáll; ennek hiányában célszerű, ha az „A” nyelven belül $a_1 \dots a_n$ homonimákat különböztetünk meg.

Mindez kiegészülhet a kiinduló nyelv fogalmi mezőinek, valamint szóképzési paradigmáinak a figyelembevételével: eszerint csak akkor hajlunk a homonim felfogás felé, ha – az előbb említetteken túl – a kiinduló nyelv jelentései eltérő fogalmi körökhöz is tartoznak, illetve ha ezekhez más és más szóképzési paradigma járul.

b) A polisizémia és a szinonímia viszonyában a *langue* és a *parole* dichotómiája tükröződik. A paradigma, azaz a szubsztitúciók tengelyén jelennek meg azok a kommutációs lehetőségek, amelyeknek az összefüggései jelentik a szinonímiát. (Itt tanulmányozhatjuk az antonímia jelenségét is.) Ezzel szemben a szintagmatikus tengelyen, a kombinációk tengelyén ragadhatók meg a szó kapcsolódási lehetőségei, azaz a polisizémia.

A polisizémiának és a szinonímiának ezt az összefüggését igazolja a lexikai hibák zöme: vagy nem vagyunk tisztában egy szó lehetséges kontextusaival, vagy rosszul választunk több szó között egy adott kontextusban. Akár a polisizémiáját, akár a szinonimáit nem ismerjük egy szónak: a végeredmény ugyanaz. A gondosan válogatott példaanyag ezért szerves része egy kéziszótár szócikkeinek.

c) A kiinduló nyelv szűken értelmezett szinonim viszonyait az egyedi példák kell, hogy megvilágítsák. Az egyes jelentések elhatárolása során azonban szerepet kap a tágan értelmezett rokonértelműség is: a szemantikai affinitások mellett a fogalmi összefüggések felmutatása, amit a francia szójelentéstan analógiának nevez.

d) Az analógiával, az asszociációs mezőkkel (a fogalmi mellérendeltséggel) szembeállíthatók ugyan a generikus mezők: a hipo- és hiperonímia (fogalmi alá- és fölérendeltség); a lexikográfiai gyakorlatban azonban ez az oppozíció nem jelenik meg.

6. Vizsgálataink körét tovább tágítva, a szóképzés legáltalánosabb összefüggéseit véve figyelembe talán nem túlzás, ha úgy véljük: az oppozíciókban való gondolkodás a lexikális makrostruktúrák világában is szükségszerű:

- grammatika : szóképzés
- jel : mező
- szó : lexéma
- morféma : lexéma

a) A francia nyelvtan és a francia szóképzés oppozíciója véleményünk szerint főleg abban áll, hogy míg a nyelvtan elsősorban strukturális, azaz minőséget tükröző szabályok halmaza viszonylag kevés lexikálisan kezelendő „kivétellel”, addig a szóképzés (a szóalkotás viszonylagos szegényessége miatt) főleg numerikus, azaz mennyiséget tükröző adatokkal jellemezhetjük: ezek többnyire statisztikai jellegűek, és a szóképzés „láthatatlan” struktúráját alkotják. Ilyen jellemzők a gyakoriság, a megoszlás, az előfordulási arányok, a sztochasztikus (valószínűségi alapon fennálló) összefüggések.

b) Ahogyan a jelen belül összetartozik a jelölő és a jelölt, úgy elválaszthatatlan egymástól a jel és a mező: itt is, valamint

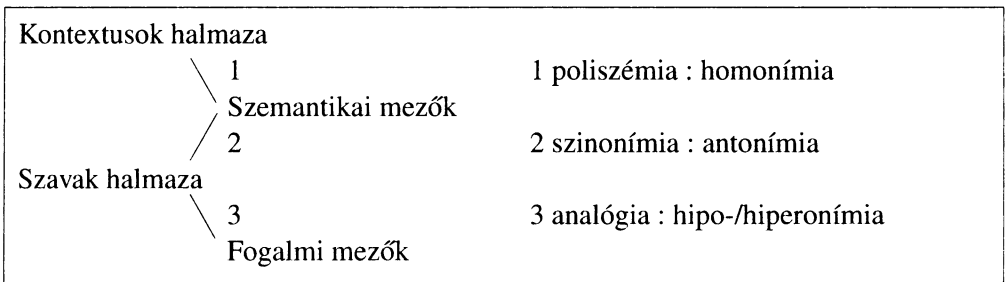
c) a szó és a lexéma oppozíciójában is a tényleges és a virtuális valóság ellentmondását s egyben egységét látjuk.

d) Ami a morféma és a lexéma két eltérő felfogását illeti: úgy véljük, hogy felesleges a morfémát a jelölővel, a lexémát pedig a jelölttel azonosítani, függetlenül attól, hogy ezek grammatikai, vagy lexikális egységek-e. Ezzel a felfogással ugyanis nemcsak a grammatika és a lexikon oppozícióját rejtjük el, hanem burkoltan tagadjuk a nyelvi jel egységét is.

7. A szójelentéstan alapjelenségei a következő általános összefüggést mutatják:

| | | | |
|------------------------|------------------------|------------------------------|--|
| Kontextusok halmaza | | Szavak halmaza | |
| Szemantikai mezők | | Fogalmi mezők | |
| poliszémia : homonímia | szinonímia : antonímia | analógia : hipo-/hiperonímia | |

Másképpen ábrázolva:



Míg lexikográfiai szemléletű vizsgálódásaink elején a szinonímia az analógiával alkotott oppozíciót, addig a fenti ábrák azt mutatják, hogy a szinonímia egy általános megközelítésben az antonímiával áll ellentétben. Mindez arra utal, hogy az antonímia, amely – a kétnyelvű lexikográfiában való hasznosíthatatlansága miatt – kívül rekedt vizsgálódásaink körén, valószínűleg túlmutat a szorosán vett jelentéstani kereteken: összekapcsolja a jelentéstant a stilsztikával (a tagadás tagadása, pl. litotes), a logikával (a tagadás tagadása; komplementaritás és kölcsönösség), a filozófiával (hübrisz, heteronímia) valamint egy sajátos képviséggel, ami már a pszichológia és a mítoszok felé mutat.

8. A szókészletet jellemző adatok jó része statisztikai jellegű, és a szókészlet „láthatatlan” struktúráját alkotja. Ilyen jellemzők a gyakoriság, a megoszlás, az előfordulási arányok, a sztochasztikus (valószínűségi alapon fennálló), ezen belül az asszociációs összefüggések. (Asszociációnak nevezzük a két minőségi ismérv közötti sztochasztikus kapcsolatot.) Ezekre nézve több számítást is elvégeztünk:

a) Először arra voltunk kíváncsiak, hogy a korpusz nagysága hogyan befolyásolja a főbb igei kategóriák arányait, másrészt azt kutatjuk, hogy van-e összefüggés az egyes igei kategóriák (illetve az ezekből képzett halmazok), s az ezeket képviselő igék szemantikai adatai (az értelmezések, a nyelvi szint, valamint a szakterület és a tipikus szerkezetek jelöltsége) között.

Vizsgálatunk eredménye a következő:

- az igei kategóriák, illetve az ezek interszekciójából létrejövő igei csoportok elkülönítéséhez és meghatározásához kb. 8–10 ívnyi középszótári korpuszra van minimálisan szükség akkor, ha 1 ív átlagosan 2000 adatot tartalmaz: mintegy 200 címszót, 800 lexikográfiai és 1000 egyéb adatot;

- a szemantikai jelölések (szinonimák, jelentésszűkítések, analógiák, körülírások, definíciók) mennyisége meghaladja a stílusréteg-minősítések, a szakterületre való utalások és a tipikus szerkezetek számát együttvéve;
- a százalékos arányok vizsgálata alapján kimondhatjuk: ha egy ige több kategóriába tartozik, ez a körülmény nem növeli meg nagy valószínűséggel az ige kiegészítő szemantikai adatainak várható számát.

b) A másik vizsgálat arra kereste a választ, hogy kimutathatók-e sztochasztikus, asszociációs kapcsolatok a leggyakoribb metanyelvi jelek: a grammatikai kategóriák és a szemantikai minősítések között. A sztochasztikus (valószínűségi) kapcsolat átmenetet képez a kapcsolat hiánya (függetlenség) és a teljes meghatározottság (függvény-szerű kapcsolat) között.

Mind a grammatikai kategória, mind a szemantikai minősítés olyan minőségi ismérv, amelynek a meghatározása a legkevesebb szubjektív mozzanatot tartalmazza. A leggyakoribb szófajok: az igék és a névszók, ezen belül a tárgyas és a nem tárgyas igék, illetve a főnevek és a többi névszó, valamint a két leggyakoribb minősítés: a használati terület és használati érték összefüggéseit vizsgáltuk.

Eredményeink a következők:

- Az igéken belül (tárgyas igék : nem tárgyas igék) nem tudunk kimutatni asszociációs kapcsolatot a két kategória valamelyike és a szemantikai minősítések valamelyike között.
- A névszókön belül határozottan elkülönülnek a főnevek a többitől, s a szakterületre való utalások is elkülönülnek a nyelvi szintre utaló adatoktól: feltűnő, hogy a főneveknél az előbbi adatok száma az utóbbiaknak több mint háromszorosa.
- Ami az igék : névszók szembeállítását illeti: az asszociációs kapcsolatok itt a legmarkánsabbak. A névszók közötti szemantikai minősítések száma az igék között találhatóéknak a háromszorosa; a névszókön belül ugyanez az arány a szakterületre, illetve a nyelvi szintre utaló adatok között. Ez akkor is érdekes, ha tudjuk: a szakterületi jelölések viszonylag magas számát nemcsak a szemantikai megkülönböztetésre nagyon alkalmas mivoltuk, hanem az a tudatos törekvésünk is magyarázza, hogy ennek a kódjelnek a segítségével könnyebb legyen a komplex keresés a szótár elektronikus változatában.

A szócikktagolás elveinek táblázatos összefoglalása

Elvek

- Mind fogalmi, mind szemantikai szempontból azonos mező: törtvonal (ilyenkor a magyar megfelelők között is van szemantikai affinitás).
- Azonos fogalmi, de eltérő szemantikai mező: pontosvessző (ebben az esetben a magyar megfelelők között sincs paraszinonímia).
- Mind fogalmi, mind szemantikai szempontból eltérő mező: külön jelentés.
- Ha a jelentések között nincs már összefüggés: homonímia.

Táblázatos áttekintő formában

| A lexikális párhuzamok típusai | Fogalmi mezők a franciában | Szemantikai mezők a franciában | Szinonímia a magyarban | Optimális ábrázolás a francia–magyar elektronikus szótárban |
|---|-----------------------------|--------------------------------|------------------------|---|
| magyar 1 / francia \ magyar 2 | azonos | azonos | van, vagy lehet | törtvonal |
| magyar 1 / francia 1 / 2 \ magyar 2 | | | | |
| magyar 1 / 1 francia 2 \ magyar 2 | eltérő, de van összefüggés | eltérő | nincs | külön jelentés |
| magyar 1 / francia 1 _____ francia 2 \ magyar 2 | | | | |
| magyar 1 / francia 1 _____ francia 2 \ magyar 2 | eltérő és nincs összefüggés | | | külön szócikk |

JEGYZETEK

1. Pajzs Júlia: Számítógép és lexikográfia. MTA Nyelvtudományi Intézete, 1990, *Linguistica, Series A, Studia et dissertationes* 4., p. 9.
2. Pálffy Miklós: Begriff und Bedeutung in der Artikelstruktur eines neuen Wörterbuches (Französisch-Ungarisch). *Acta Germanica* 5.: *Studien zur zweisprachigen Lexicographie Deutsch-Ungarisch* I, pp. 123–127, JATE, Szeged
3. Ün. „irányítószók”: lásd Mollay Erzsébet: *A holland–magyar kéziszótár szerkesztési elvei*. Budapest, ELTE Néderlandisztikai Műfordító Műhely, 1998.