

Az ördög a részletekben, mármint a szelekcióban rejlik – egy korpuszalapú kollokációs szótár, a SZÓkapTÁR margójára

Hollós Zita

1. Bevezetés

A NÉMET SZÓkapcsolatTÁR (röviden SZÓkapTÁR) típusát tekintve egy német–magyar, adatbázis- és korpuszalapú¹ kollokációs szótár. A munka 90%-a 2005–2008 között, egy négyéves OTKA kutatási projekt keretében valósulhatott meg. A szótárkonceptiót megalapozó részletes elméleti alap kutatások megtalálhatóak a *Lexicographica Series Maior* sorozat 116-os kötetében (Hollós 2004).

A projekt munka legelején berendeztem – a szótárprojektnek otthont adó Károli Gáspár Református Egyetemen – egy szótári műhelyt, ahol egy számítógépes konfiguráció, nyomtatott és elektronikus szótárak, adatbázisok² támogatták az 5–10 diákmunkatárs munkáját³. Az adatbázis programozási munkálatai⁴ javarészt elkészültek az első évben, de a szótárszerkesztő program optimalizálása és továbbfejlesztése még évekig folyt. Az adatbázis Windowshoz hasonló adatbeviteli felülete – az interdiszciplinaritás jegyében – csak a lexikográfiai és programozási ismeretek hathatós egyesítésével jöhetett létre. Ennek eredményeképpen

¹ A SZÓkapTÁR első kotextuspartner-szelekciójához az adatok a lipcsei Deutscher Wortschatz korpuszából származnak. Szeretnék köszönetet mondani a lipcsei projekt vezetőjének, Prof. Uwe Quasthoffnak, hogy éveken keresztül rendelkezésünkre bocsátotta a szükséges adatokat, aminek eredményeképpen megszülethetett az első korpuszalapú kétnyelvű kollokációs szótár.

² A CCDB-nek (Kookkurrenzdatenbank), mely egy, az IDS-korpusz alapján létrehozott kookkurenciaadatbank, korlátlan lekérdezése csak a projekt vezetőjének, Cyril Belica-nak a külön engedélyvel és segítségével valósulhatott meg, ezért külön köszönet illeti meg őt ezen a helyen.

³ Itt szeretném megemlíteni a két diákmunkatárs-csapatom tagjainak nevét: 2005–2007: Koronka Anna, Németh Gábor, Németh Réka, Riesz Ábel, Tompa Ágnes; 2008–2009: Baranyi Csilla, Barna Ágnes, Illés-Tóth Viktória, Nagy Orsolya, Nyíri Csilla. Külön köszönöm Tompa Ágnes, Riesz Ábel, Baranyi Csilla és Illés-Tóth Viktória közreműködését, akik a diákszerződések lejárta után is önzetlenül támogatták a projektet kitarató és odaadó munkájukkal. Nélkülük nem jöhetett volna létre a SZÓkapTÁR.

⁴ Szeretném megköszönni Bunda Péter programozó munkáját, aki nélkül a SZÓkapTÁR adatbázisa nem jöhetett volna létre. Önzetlen és kitarató munkája, amelynek eredményeképpen egy professzionális többfunkciós és felhasználóbarát munkafelülettel ellátott adatbázis született, több konferencián elismerést aratott.

az adatbázis egyes elemei (pl. Megjegyzés-opciók) éveken át biztosították az ún. *workflow*-t, amit a havi szerkesztői munkaülések és jegyzőkönyvek is támogattak.

2009 nyarára az adatbázisba bekerült adatok⁵ készülségi foka lehetővé tette a végső szerkesztési fázist, melynek keretében kiegészítési, ellenőrzési és egységsítési munkálatok zajlottak. 2011 nyarára kigenerálhatóvá vált az adatbázisból a végső, lektorálható kézirat. 2012 tavaszára megtörtént a lektori javaslatok⁶ szerkesztői ellenőrzése és az elfogadott javaslatok és javítások átvezetése a szótári adatbázisba. 2012 nyarára pedig az utolsó anyanyelvi ellenőrző munka is elkészült, így az adatbázisból bármikor kigenerálható a majdani nyomtatott szótár szerkeszthető dokumentuma, akár egyszerű szöveg akár xml-formátumban.

Jelen cikk egyfajta folytatása a Hollós 2007a és 2007b cikkeknek. Előbbiben a SZÓKAPTÁR és a korpusznyelvészet kapcsolatáról számoltam be röviden, utóbbiban a lexikográfiai folyamatot mutattam be a SZÓKAPTÁR példáján keresztül, az akkori – folyó munkálatok – eredményeinek figyelembe vételével. Azóta befejeződtek a SZÓKAPTÁR munkálatai, így ebben a cikkben tudom először röviden bemutatni az elkészült – az egyszerűség kedvéért kollokációs szótárnak nevezett – SZÓKAPTÁR főbb szótári jellemezőit (2. fejezet) és egyidejűleg összegzem a szótárban alkalmazott integratív kollokációfogalom lényegét (3. fejezet). Ezek után az olvasó már valamelyest képet alkotott a szótár elméleti háttéréről, így meginvitálhatom egy kalandos „utazásra” a kotextuspartnerek világába, hogy betekintést nyerhessen a szelekció rejtelseibe (4. fejezet).

Ha az olvasó valóban velem tart a felfedezőúton, megtapasztalhatja a közmondások igaz voltát, melyek akár mottójául is szolgálhatnak bármely korpuszalapú kollokációs szótárnak:

*Az ördög a részletekben rejlik
Qual der Wahl*

(német *Der Teufel steckt im Detail*)
(magyar *Nehéz a választás*).

⁵ Megkülönböztetett köszönet illeti Jörg Dötsch-öt, aki német anyanyelvű munkatársként nemcsak az embert próbáló első kotextuspartner-szelekciót, hanem az első kettő és a negyedik lektorálást is komoly alapossággal végezte. Nélküle nem születhetett volna meg a SZÓKAPTÁR.

⁶ A lektorálást Zádor Éva végezte, a lektori munkát a Grimm Kiadó finanszírozta. Ezúton is köszönet Zádor Éva odaadó munkájáért, a több mint 600 oldalas, igen nehéz és sűrűn adatolt kézirat gyors és alapos lektorálásáért.

2. A SZÓkapTÁR főbb szótári jellemzői

A szótárkonceptiót részletesen bemutató disszertációmban következőképpen definiáltam a szótár típusát:

deutsch–ungarisches, monoskopales, **korpuserorientiertes**, polyfunktionales, dennoch **produktionsbezogenes syntagmatisches Lernerwörterbuch** mit einem primären Wörterverzeichnis der Basen und einem sekundären der Kollokatoren und zusätzlich mit einem ungarisch–deutschen Register zum Basis-Wörterverzeichnis. (Hollós 2004: 174, kiemelés H.Z.)

Szeretnék a fenti idézetben kiemelt, a szótártípust leginkább jellemző attribútumokhoz rövid magyarázatokot fűzni:

– német–magyar

Mivel a német nyelv tipikus kollokációinak és szókapcsolatainak a feltérképezése a cél, valamint a két nyelv tekintetében akkoriban csak a német nyelv rendelkezett megfelelő nagyságú és fejlett elemző eszközöket integráló elektronikus korpuszokkal, ezért a német a szótár forrásnyelve. Nem beszélhetünk azonban a magyar esetében hagyományos értelemben vett célnyelvről, mivel a szótár első-sorban nem dekódolási funkciót lát el. A kotextuspartnerek⁷ fordítási javaslatai azok egymástól való elkülönítését szolgálják a nyelvi produkció számára. Ezt támogatják az őket kiegészítő stilisztikai-pragmatikai és grammatikai adatok is. A kollokációk és szókapcsolatok lefordítása megkönnyíti a gyakran közeli jelentéssel bíró vagy szinonim kifejezések szemantizálását⁸, illetve lehetővé teszi, hogy kevésbé haladó nyelvtanulók is sikerrel használhassák a szótárt.

– korpuszalapú, helyesebben korpuserorientált

Megbízható nyelvi adatokat a német szavak tipikus együttes előfordulásáról (co-occurrence) csak a tényleges nyelvhasználat vizsgálatával kaphatunk. Ehhez „csupán” megfelelő nagyságú és összetételű (kiegyensúlyozott) elektronikus és annotált korpuszokra van szükség, melyek rendelkeznek összetett, statisztikai korpuszelemző eszközökkel is. Az IDS-korpusz és a lipcsei *Deutscher Wortschatz* (Német Szókincs) projekt korpusza⁹ ilyenek. A SZÓkapTÁR esetében

⁷ A kotextuspartner fogalmát a kollokátor- és kombinációs partner-fogalom fölé rendelt fogalomként használom. A fogalom motivált, mivel itt a közvetlen nyelvi kontextuson, az ún. kotextuson van a hangsúly és a partner mindenkor elmélettől függő státusza irreleváns.

⁸ Lehetséges alternatíva lenne, hogy németül adunk leírásokat, de ennek közérthetősége főleg középhaladó (B2) nyelvtanulóknál megkérdőjelezhető. Magyar nyelvű célcsoportnál az ekvivalensek megadása a legcélravezetőbb és több haszna is van. Ezen a helyen csak az interlingvális kollokációs jelenségek szisztematikus feltérképezésének a lehetőségét szeretném megemlíteni.

⁹ Később Lipcsei korpuszként utalok rá

az elsődleges korpusz a lipcsei, mivel onnan kaptuk a szótárprojekt részére a tetemes mennyiségű korpuszadatokat: statisztikai szignifikációs szólisták formájában mind a csaknem 2300, a szótárban szereplő szóhoz.

– **nyelvi produkcióra irányuló** (azt segítő) **szintagmatikus tanulószótár**

A szótár célja (közép)haladó nyelvtanulók nyelvi produkciójának segítése, mégpedig elsősorban a lexika szintű szintagmatikus kapcsolatok leírásával¹⁰. A szócikkek tartalmaznak grammatikai szintű szintagmatikus adatokat is (pl. kötelező és fakultatív valencia mindhárom lemmatípus – igei, melléknévi és főnévi – esetében, kötelező valenciarealizációk igei lemmatípus esetén, különböző releváns fonológiai és morfológiai adatok).

A fenti jellemzőket feltétlenül ki kell még egészíteni egy fontos szótári karakterisztikával, mely a szótárkonceptió konkrét megvalósításakor mint a lexicográfia iránt támasztott új igény jelentkezett:

– **adatbázis-alapú**

Az adatok adatbázis-alapú tárolása lehetővé teszi az ún. *Cross-media-Publishing* gyakorlatát. Ez azt jelenti, hogy az adatmodellálás koncepcióját az adatok prezentációs formájától (nyomtatott vagy digitális szótár) függetlenül tervezzük. A külön erre a szótárra létrehozott XML-alapú adatbázis rendelkezik egy a célnak megfelelő, Windowshoz hasonló adatbeviteli munkafelülettel, melynek különböző pozíciói legördülő menüválasztásos vagy szabad adatbeviteli opcióval vannak ellátva. Egyes menük értékei a programban előre definiáltak (pl. szófaj vagy főneveknél a névelő/névelőkombinációk), mások a szerkesztő által bővíthetők, mint amilyen a nyelvtani adatok pozíciója az egyes szófajoknál (pl. főnévi lemmajeleknél: <tbsz n>, <ritk tbsz>, <csak tbsz>) vagy az egyes kollokátortípusoknál (pl. főnévi lemmajelnél melléknévi kollokátor esetében: <csak jel>, <ált jel>, <csak áll>). Az adatbázis egyes elemei – mint amilyen pl. a szócikk létrehozásának dátuma, illetve készültségi foka (nem kész / kész), a szerkesztői státusz (szerkesztés alatt / beszerkesztett), valamint a Megjegyzés-opciók az adatok rögzítésének különböző szintjein – nagymértékben támogatták az ún. *workflow*-t és hatékonyan segítették a projekt résztvevői közötti kommunikációt.

A következőkben a SZÓKAPTÁR első bázis-szószedetének adatait szeretném röviden ismertetni és néhány rövid magyarázattal kiegészíteni:

Összes kotextus száma: 44 622

Kotextus alatt a szeméma-specifikus kollokációk és szabad szókapcsolatok értendők.

Összes kotextuspartner száma: 9933

¹⁰ A szintagmatika lexikai és grammatikai szintű megkülönböztetése a brit kontextualizmusból, a korpuszalapú kollokációkutatás kezdeti időszakából származik.

Kotextuspartner mint átfogó fogalom alatt a bázis szószedetben szereplő kollokátorok és kombinációs partnerek értendők, melyek listája megtalálható a szótár német kotextuspartner-regiszterében. Négyhasábos nyomtatott formátumban körülbelül 200 A4-es oldal.

A regiszter eredeti elképzelése annyiban változott, hogy nem a magyar szemémákhoz készült regiszter – bár ennek lehetősége továbbra is fennáll –, mivel a szótár célcsoportjánál (elsősorban középhaladó, haladó nyelvtanulók és germanisták) feltételezhetőek ezek az ismeretek, hanem helyet adott egy új, német kotextuspartner-regiszternek, mely tartalmazza az összes kotextuspartner bázis-szócikkbeli és szeméma-specifikus előfordulását, függetlenül a kotextuspartner státuszától: vajon kollokátor vagy kombinációs partner-e. Előbbi kapcsolódása a lemmajelhez kollokációt, utóbbié szabad szókapcsolatot eredményez.

Bázis-szócikkek száma: 2267

A bázis-szószedet lemmáink a száma: 1279 főnévi lemma, 579 igei lemma, 362 melléknévi lemma és 47 határozószói lemma. Hagyományos kéthasábos, nyomtatott formátumban több, mint 600 A4-es oldal.

A bázis-szószedet mind adatainak mennyiségében, mind minőségében oly mértékben túlnötte önmagát, hogy a második kollokátor-szószedet kidolgozására már nem kerülhetett sor. Az adatbázis viszont rendelkezik azokkal a programkomponensekkel (pl. német regiszter), melyek lehetővé teszik ennek megvalósítását, vagy egy második kiadásban, vagy akár önálló szótárként is.

3. A SZÓkapTÁR integratív kollokációfogalma

Mielőtt az adatszelekció rejtelseibe betekintést nyernénk, elengedhetetlen tisztázni, milyen kollokációfogalom képezi a SZÓkapTÁR alapját. Meglátásom szerint a különféle kollokációfogalmak előnyeit ötvöző, integratív kollokációfogalom a leginkább megfelelő erre a célra.

Alább négy nyelvészeti részterület idevonatkozó eredményeit szeretném röviden felvázolni – ezeket az integratív kollokációfogalom hivatott egyesíteni:

- I. **Idegen nyelvi didaktika** (bázis és kollokátor megkülönböztetése, kollokációs struktúratípusok, interferencia)
- II. **Szemantika** (idiomatikusság)
- III. **Korpusznyelvészet** (szignifikáns kookekurenciák)
- IV. **Szintaxis** (valencia)

I. A francia **idegen nyelvi didaktika** jelentős felismerése, hogy a kollokációt alkotó szavak (autozsemantikumok) nem egyenlő rangú partnerek, hanem egy szemantikailag autonóm bázishoz kapcsolódik egy kollokátor, mely a bázishoz affin. Utóbbi jelentésének meghatározásához elengedhetetlen a bázis. Az affinitás fogalmát úgy definiálják, hogy ez a szavak abbéli hajlandósága, hogy

közösen szerepeljenek bizonyos kontextusokban (vö. Hausmann 1984: 398). A brit kontextualizmus hagyományát követve a hat kollokációs struktúratípus – példákkal illusztrálva – a következő (a kollokáció bázisa vastagon szedve):

- SUBS + SUBS *ein Stoß Wäsche* (egy kupac/stósz mosnivaló)
- ADJ + SUBS *stimmgewaltiger Sänger* (erős/telt hangú énekes)
- SUBS + VERB *(ein) Eis lecken* (fagyit nyal)
- ADV + VERB *sich schlagartig ändern* (hirtelen/egy csapásra megváltozik vmi)
- ADV + ADJ *landläufig bekannt* (általánosan ismert)
- ADV + ADV *gänzlich anders* (egészen másképpen/másként)

Ezt kiegészíti még az interferencia aspektusa, ami alapján definiálható az interlingvális kollokációfogalom, melyre terjedelmi okokból jelen tanulmányban nem tudok kitérni. A SZÓKAPTÁR-ban az interlingvális kollokációkra – mint amilyen a klasszikus példa: német *sich die Zähne putzen* – magyar *fogat mos* (németül szó szerint fordítva **Zahn waschen*) – egy piktoqram (♠) hívja fel a figyelmet. A szótári adatbázisban jelenleg 6620 különböző interlingvális kollokáció (KOLL_{inter}) található.

II. A **szemantika** az idiomatikusság fogalmának meghatározásával járul hozzá az integratív kollokációfogalom létrejöttéhez. Egyes kollokációk nem idiomatikusak (pl. német *triftiger Grund* – magyar *nyomós ok*). Mégis a kollokációkhoz soroljuk ezeket és ennek megfelelően a frazeológiai egységek perifériáján foglalnak helyet, mivel relatíve stabilak és a főnévi elem korlátozott mértékben helyettesíthető, mint ahogy ezt korpuszvizsgálatokkal korábban már kimutattam (vö. Hollós 2004). Egyes kollokációk gyengén idiomatikusak (pl. német *die Ampel überfahren* – magyar *áthajt a piroson* – a német változat szó szerinti fordításában: *áthajt a lámpán*), hisz az *Ampel* lexémának nincs olyan jelentése/szemémája, amely a piros közlekedési lámpajelzésnek felelne meg. Egyes kollokációk részben idiomatikusak. Ide sorolom azokat, melyek egy frazémát tartalmaznak, pl. egy határozói funkciójút (pl. német *sich zu Tode langweilen* – magyar *halálra unja magát*). Itt a határozói frazéma számos igéhez kapcsolódhat a németben, ezért nem a teljes kifejezésnek van frazeológiai státusza. Hogy mely kollokációk tartoznak ehhez a csoporthoz, egyenként kell korpuszvizsgálatokkal leellenőrizni, mivel a német egy- és kétnyelvű szótárak, főleg a határozói funkciójú frazémák szótári alakjának megadásakor inkonzekvensek és sokszor pontatlanok is. A rendelkezésre álló szótári adatok, itt elsősorban a frazémák variációs lehetőségeire, de a külső valenciára is utalnék, nem korpuszstatisztikai vizsgálatok eredményeit tükrözik, ezért objektív voltuk igencsak megkérdőjelezhető.

III. Az utolsó gondolatmenet pedig már átvezet a következő ponthoz, a **korpusznyelvészethez**.

Mára lehetővé vált az automatikus korpuszlekérdező eszközökkel nagy mennyiségű, megbízható és a tényleges nyelvhasználatot tükröző adatokat nyerni. Bár ezek automatikus felhasználása lexikográfiai célokra, nyelvész közre-

működése nélkül, véleményem szerint nem lehetséges, mégis, a szótárkészítési gyakorlat e nélkül mára már nem képviselhető. Főleg a statisztikailag szignifikáns kookkurencialisták kincseshányái a frazémáknak és kollokációknak. Ha eddig a kollokációkat legalább kéttagú, a nyelvidaktika által meghatározott struktúratípusok alapján nem, gyengén vagy részben idiomatikus, szemantikailag autonóm bázis és ahhoz tartozó affin kollokátor kapcsolataként értelmeztük, akkor ki kell egészíteni a jellemzők sorát a korpusznyelvészeti felismeréssel, mely szerint az intralingvális kollokációk ezzel egyidejűleg szignifikáns kookkurenciaként is definiálhatóak¹¹. Az intra- és interlingvális kollokációk fogalmával jórészt lefedhető a nyelvtanulók számára fontos tipikus szókapcsolatok széles skálája. Ami ide nem kerülhet be, arra a SZÓKAPTÁR koncepciójának kidolgozásánál egy plusz kategóriát alkottam, a kombinációk kategóriáját <KOMB>. Ezek szinte kivétel nélkül mind – interferencia szempontjából – jelölt, többnyire háromszavas kapcsolatok, gyakran variációs lehetőségekkel¹².

IV. A **szintaxis** mint a kollokációfogalom egyik aspektusa – legjobb tudomásom szerint – egyelőre nem található meg a szakirodalomban. Ez főleg annak köszönhető, hogy az illusztrációra használt példák a nem problematikus esetek közül származnak (egy, legfeljebb két valenciával rendelkező igék, bázisként vagy kollokátorként), vagy leegyszerűsített példák, ahol az esetleges kötelező/fakultatív vonzatok nem szerepelnek. Valóban nehéz feladattal áll szemben, aki a valencia aspektusát is integrálni szeretné a kollokációfogalomba. Ezáltal számos kéttagúnak tűnő kollokáció „hirtelen” többtagúvá (három- vagy négytagúvá) válik, nem beszélve a valenciarealizációkról, amelyek egyfajta második szintű variációs lehetőséget implikálnak a már meglévő mellé, ami az intralingvális kollokáció esetében a bázis behatárolt variációs/szubsztitúciós lehetősége. Ennek megfelelően a kollokáció sémája nem egyszerűen bázis+kollokátor [B+K], hanem bázis+kollokátor+kötelező/fakultatív valencia [B+K+V/(V)]. A kötelező valencia ilyen esetekben többnyire grammatikailag és szemantikailag behatárolt, valamint szótári megjelenítése is problematikus.

Két példán szeretném illusztrálni a jelenséget (a bázist vastagon kiemeltem, a valenciát aláhúztam a példákban):

jm <Gäste, Mann ...> **Kaffee servieren** [**B**_{SUBS}+K_{VERB}+V_{Dat}]
etw^{Akk} <Jacke, Mantel, Gepäck ...> *unentgeltlich abgeben* [**B**_{VERB}+K_{ADV}+V_{Akk}].

A kollokációk valenciatulajdonságainak vizsgálata nemcsak azok nyelvidaktikai és lexikográfiai hozadéka miatt sürgető, hanem az intralingvális kollokációk variációs lehetőségeinek a feltérképezéséhez is elengedhetetlenül szükséges.

¹¹ A kollokációk megkülönböztetése intra- és interlingvális típusokra, illetve ezek kombinációi és definíciói megtalálhatóak: Hollós 2004: 23–89.

¹² Az első kutatási eredmények megtalálhatóak: Hollós 2010.

Miután a SZÓKAPTÁR főbb szótári jellemzőit és a szótár elméleti alapját képező integratív kollokációfogalmat dióhéjban bemutattam, szeretnék rátérni a kotextuspartnerek szelekciós folyamatának illusztrálására.

4. A belső szelekciós folyamat a kotextuspartnerek tükrében

Bevezetésként szeretném röviden bemutatni a korpuszadatok kigenerálásával, rögzítésével és adatbázisba történő beolvasásával kapcsolatos legfontosabb „technikai” tudnivalókat.

Ahhoz, hogy a kotextuspartner-adatok tartalmilag egyezők legyenek, tehát a megfelelő lemmajelhez a megfelelő szófajú és pozíciójú kotextuspartnereket tartalmazza a generált, statisztikai lista, meg kellett találni a lehetséges keresési algoritmusok közül a legcélravezetőbbet. Ez a gépi nyelvfeldolgozással foglalkozó szakemberek feladata volt, miután megismerték a készülő szótár főbb paramétereit. Ahhoz viszont, hogy a szótárprogram beolvasást támogató komponense a szelektált adatokat automatikusan az adatbázisba importálhassa, az adatokat nemcsak mindig ugyanabban a formátumban kell kigenerálni és rögzíteni, hanem a szemémaspecifikus kézi szelekciót is mindig ugyanabban a formában kell végezni (ezt alább példákkal fogom szemléltetni), különben a beolvasóprogram hibát jelez és nem tudja beolvasni az adatokat, –. A formai hibák kézi javítása után ugyan újra be lehet olvasatni az adatokat anélkül, hogy a már bent lévő partnerek újra bekerülnének (egyszóval: a program blokkolja az adatok duplikálását), de ez mégis fárasztó és idegölő feladat. Tömeges előfordulása nagyon lelassítja a szótárkészítői munkát.

Az **atmen** lemmajelhez generált statisztikai lista segítségével szeretném megmutatni, milyen formátumban állnak rendelkezésre az adatok, amiből szeméma-specifikusan, kézi szelekcióval kerülnek kiválogatásra a megfelelő kotextuspartnerek:

atmen	schwer	A	311	49
atmen	frei	A	210	43
atmen	selbständig	A	59	10
atmen	richtig	A	56	16
atmen	unter Wasser	A	47	8
atmen	leicht	A	43	10
atmen	tief	A	40	10
atmen	besser	A	37	13
atmen	ruhig	A	35	8
atmen	allein	A	20	8
atmen	wieder	A	19	14
atmen	flach	A	15	3

atmen	gleichmäßig	A	15	2
atmen	normal	A	11	3
atmen	schlecht	A	11	4
atmen	heftig	A	11	2
atmen	Häufig	A	9	2
atmen	gemeinsam	A	9	2
atmen	kurz	A	8	2
atmen	kontrolliert	A	7	2
atmen	dagegen	A	6	4

A fenti lista szemlélteti az első problémát, amivel szembesülhetünk, ha korpuszstatisztikai adatokkal dolgozunk: a generált lista nagyon kevés határozói kotextuspartnert tartalmaz és ezeket az *atmen* ige két jelentéspontja alá kell beválogatni.

A következő lista már a válogatás eredményét mutatja. A „>” jellel választjuk ki a partnert; a kotextuspartner mögötti szám jelzi, melyik jelentéshez kerül. Ha nem áll a kotextuspartner mögötti szám, az azt jelenti, hogy minden jelentéshez passzol (jelen esetben kettőhöz):

atmen	>schwer 1	A	311	49
atmen	>frei 1	A	210	43
atmen	selbständig	A	59	10
atmen	richtig	A	56	16
atmen	unter Wasser	A	47	8
atmen	leicht	A	43	10
atmen	>tief	A	40	10
atmen	besser	A	37	13
atmen	>ruhig	A	35	8
atmen	allein	A	20	8
atmen	wieder	A	19	14
atmen	>flach 1	A	15	3
atmen	gleichmäßig	A	15	2
atmen	normal	A	11	3
atmen	schlecht	A	11	4
atmen	heftig	A	11	2
atmen	Häufig	A	9	2
atmen	gemeinsam	A	9	2
atmen	kurz	A	8	2
atmen	>kontrolliert	A	7	2
atmen	dagegen	A	6	4

Ha csupán a kookkurencia-listából kiválogatott partnerek kerülnének be a szócikkbe, eléggé szegényes lenne az adatmennyiség, ezért ezt a helyzetet és a

lektori szelekció esetleges hiányait és hibáit kiküszöbölendő, az első körben egy- és kétnyelvű szótárakból (LGWDaF, PGWDaF ill. HESS-NM, HESSIKE) származó adatokkal is kiegészülnek a kotextuspartner-blokkok. Az alábbi szócikkben (Szócikk 1) szürke háttér jelzi a szócikk azon tartományát, ahol az adatok a korpuszból származnak:

atmen atmete, hat geatmet, ige

1: (Mod) lélegzik (vhogyan)

<ADV> (Mod)-realizációként, lásd Csoport 1,2 is **schwer** nehezen * átv is **frei** szabadon * **flach** \emptyset gyen-
gén, erőtlennül * unregelmäßig szabálytalanul * **feifend**
sípólva * **kaum** alig

<KOMB> (Mod)-realizációként **durch den Mund**
a. szájon át lélegzik * **durch die Nase a.** \emptyset orron (át)
lélegzik * **durch Kiemen a.** \emptyset állat kopolytúval lélegzik

2: Akk, vál belélegzik, belélegez vmit Akk Luft, Sauerstoff, Rauch<ADV> lásd Csoport 1,2

1,2:

<ADV> 1. jel-hez (Mod)-realizációként

tief mélyet¹, mélyen² * **ruhig** nyugodtan * **kontrolliert** \emptyset odafigyelve * vál **mühselig** nehezen, nehézkesen
* **selbständig** önállóan * **gleichmäßig** egyenletesen

Ritka az az eset, amikor nincsenek közös partnerek, de csak egy meghatározott számú közös partner esetén tekint el a koncepció a partnerek megismétlésétől, hiszen a közös partnerek különböző jelentéscsoportokhoz tartozhatnak és fordításuk is jelentésenként eltérő lehet. A szótárprogramban a határt 5-nél húztuk meg, tehát ha 5 kotextuspartner ugyanazon 2 vagy 3, ritkán 4 jelentés alatt is szerepelne, akkor a közös kotextuspartnerek (≥ 5) a szócikk végére, külön csoportba kerülnek, ahogy ez a mellékelt szócikkben is jól látható. A *tief* kotextuspartnernél az indexált ekvivalensek az **atmen** lemmajel 1. és 2. jelentésére utalnak. A jelentésekhez a német határozószó *tief* különböző magyar ekvivalensei kapcsolódnak (1. *mélyet lélegzik* vs. 2. *mélyen belélegez vmit*).

A harmadik szelekciós lista és a hozzá tartozó szócikk jól szemlélteti azt az ideális esetet, amikor szinte minden partner a korpuszból származik. Jelen esetben csak egy plusz partner (*geistig*) került be a szócikkbe a HESS-ből, de ez éppen egy interlingvális kollokációt eredményez (\emptyset):

arm	viel	A	199	76
arm	hochverschuldet	A	112	16
arm	>besonders	A	105	48

arm	>bitter 1	A	93	20
arm	verschuldet	A	91	16
arm	>relativ	A	85	25
arm	gerade	A	77	43
arm	>wirklich 3	A	73	31
arm	>vergleichsweise 1,2	A	67	18
arm	ganz	A	64	48
arm	meist	A	53	27
arm	zugunsten	A	46	9
arm	ohnehin	A	40	15
arm	eben	A	38	21
arm	>extrem	A	38	12
arm	geistlich	A	34	4
arm	ziemlich	A	33	12
arm	>absolut 1,2	A	32	8
arm	eher	A	31	19
arm	andere	A	28	26
arm	zumeist	A	28	10
arm	deutlich	A	26	10
arm	genauso	A	25	8
arm	vier	A	24	13
arm	für Kinder	A	22	5
arm	wesentlich	A	21	6
arm	zwar	A	18	11
arm	künstlich	A	17	4
arm	chronisch	A	16	4
arm	ungleich	A	16	3
arm	zahlreich	A	15	4
arm	wirtschaftlich	A	15	4
arm	zehn	A	15	9
arm	zu viel	A	14	4
arm	gleich	A	14	10
arm	heute	A	14	8
arm	traditionell	A	13	3
arm	lieb	A	12	5
arm	ebenso	A	11	6
arm	verzweifelt	A	11	3
arm	äußerst	A	11	4
arm	materiell	A	11	2
arm	betroffen	A	10	4
arm	geistig	A	10	3
arm	keineswegs	A	10	4

arm	notorisch	A	9	2
arm	automatisch	A	9	3
arm	sogar	A	9	5
arm	einstmals	A	8	2
arm	paar	A	8	5
arm	doppelt	A	8	3
arm	>furchtbar 1,3	A	8	2
arm	hoffnungslos	A	8	2
arm	>schrecklich 1,3	A	8	2
arm	zwangsläufig	A	8	2
arm	zweifellos	A	8	2
arm	heutzutage	A	7	2
arm	oft	A	7	5
arm	regelrecht	A	7	2
arm	wieder	A	7	11
arm	afrikanisch	A	7	2
arm	häufig	A	6	3
arm	nur	A	6	17
arm	dauerhaft	A	6	2
arm	einmal	A	6	7
arm	gleichermaßen	A	6	2
arm	manchmal	A	6	3
arm	meistens	A	6	2
arm	weltweit	A	6	3
arm	vermeintlich	A	5	2
arm	eigentlich	A	5	4
arm	freiwillig	A	5	2
arm	lang	A	5	4
arm	lebt	A	5	2
arm	wenig	A	5	5
arm	fünf	A	5	5
arm	unbedingt	A	5	2
arm	einig	A	5	2

arm ärmer, ärmst-, mn

1: *föl anyagi javakban szegény*

+Subs Mensch, Kind, Leute, Familie, Land

+Verb leben, sterben, geboren sein

<ADV>

bitter mélységesen * **relativ** aránylag * **vergleichs-**
weise viszonylag * **extrem** roppant(ul) * biz **furchtbar**

szörnyen, rettenetesen * biz **schrecklich** borzasztóan
2: csak áll, an+Dat *szűkölködő szegény* vmiben an+Dat
Bodenschätze, Wälder, Kalorien, Vitamine, Nährstoffe
<ADV>

besonders különösen (is), rendkívül * **relativ** arány-
lag * **vergleichsweise** viszonylag * **extrem** nagyon
kimondhatatlanul * biz **absolut** abszolút * **geistig** a. ǂ
szellemileg sivár

3: *sajnálatra méltó szegény, szerencsétlen ördög*
+Subs Teufel, Kind, Kerl, Hund
<ADV>

besonders rendkívül * **wirklich** valóban, igazán

Ugyan az eredetileg kiválogatott közös partnerek némelyikét a második lektorálásnál törölni kellett, de így is jól látható, hogy vannak kisebb számban közös partnerek az 1. és 2. jelentés alatt (*relativ, vergleichsweise, extrem*) és a 2. és 3. jelentés alatt is megtalálható egy (*besonders*), de ezek fordítása részben jelentésként eltérő. Fontos itt azt is megemlíteni, hogy mind az **atmen** lemmajelhez tartozó valenciarealizációk, mind az **arm** lemmajelhez egyes jelentésekben felsorolt főnévi és igei partnerek, valamint a 2. jelentés alatt szereplő valenciarealizációk elsősorban a már említett szótárakból származnak.

Utolsóként szeretném szemléltetni a többjelentésű főnevek esetében felmerülő szelekciós problémákat. Ez a munkafolyamat talán a legnehezebb, mivel a szelekciós szempontok közül egyszerre többet kell szem előtt tartani. Nehéz lenne hierarchiát felállítani közöttük, mivel hol az egyik, hol a másik, hol pedig egyszerre több érvényesül. Mindeközben figyelni kell a célcsoport igényeire is, akiknek körét elsősorban a (közép)haladó (B1–C2) nyelvtanulók és a germanista egyetemi/főiskolai hallgatók teszik ki.

Az alábbi szempontok képezik – a felvázolt elméleti háttér figyelembevételével – a mindenkori szelekció alapját¹³:

- relatív gyakoriság: a szavak közös előfordulásának statisztikailag szignifikáns volta¹⁴;
- anyanyelvi kompetencia: a kifejezés jellemző, de mégsem „banális” volta a német nyelvben;
- stílári szempont: egyszeri, illetve választékosabb kifejezések, melyek gyakran valamely speciális fogalmi egységet fejeznek ki;

¹³ A felsorolt szempontokat a következő publikációból vettem át: Hollós 2007a. Ott példákkal is illusztráltam az egyes szelekciós szempontokat.

¹⁴ Az itt használt szignifikáció-fogalomról pontos információk találhatóak a lipcsei projekthez kapcsolódó publikációkban: <http://asv.informatik.uni-leipzig.de/publications/books>

- interkulturális szempont: a német nyelvi diskurzusra jellemző szavak/kifejezések felvétele;
- kontrasztív szempont: a szó vagy szókapcsolat magyartól eltérő fordítása, illetve problematikus fordítása magyarról németre;
- mennyiségi szempont: a szótári ábrázolhatóság mennyiségi keretei¹⁵;
- nyelvdidaktikai szempont: nyelvtanulók számára hasznos, a szócikkben lexikálisan még nem lefedett nyelvi tartalom kifejezésére szolgáló szó vagy szókapcsolat¹⁶.

Példaképpen álljon itt a **Glas** lemmajelhez tartozó szótári jelentésszerkezet és a Lipcsei korpuszból generált szelekciós lista, mely bár hosszúnak tűnhet, mégis főnévi lemmajeleknél aránylag rövidnek számít. Szeretném megkérni a kedves olvasót, tegyen egy kísérletet és próbálkozzon meg a szótári jelentésszerkezet segítségével az alábbi kotextuspartner-listából kiválogatni azokat a partnereket, melyeket véleménye szerint a SZÓKAPTÁR-ba, a **Glas** egyes jelentéseihez fel kellene venni!

A szelekciónk eredményeként első körben 27 kotextuspartner került be az adatbázisba: 16 melléknév (*kratzfest* 2x, Σ: 17) és 8 ige (*klirren, spiülen* 2x, Σ: 10), melyeket a program automatikusan beolvasott a SZÓKAPTÁR **Glas** szócikkének megfelelő jelentéspontja alá. A beolvasás után a következő állapotban állt rendelkezésünkre a szócikk:

Glas, fn

1: tbsz n, anyag üveg ;

<ADJ>

**getönt * kugelsicher * kratzfest * matt * entspiegelt *
wärmedämmend**

<VERB>

splittern * klirren * schleifen

2: üveg- pohár ;

¹⁵ Nyomatott szótárakban a szócikkek hosszúsága és ezzel összefüggő áttekinthetősége nagymértékben meghatározza a szótár felhasználóbarát voltát. Mivel a SZÓKAPTÁR először nyomtatott szótárként fog megjelenni, fontos szempont, hogy a szócikkek ne legyenek túl terjedelmesek.

¹⁶ Ez a szempont valóban utolsóként valósult meg: a második diákmunkatárs-csapatom feladata volt ennek a szempontnak a figyelembevételével kiegészíteni az adatbázisban szereplő lemmajelek kotextuspartner-blokkjait további új partnerekkel a DU2 és a CCDB alapján. Ennek a – részben ellenőrző funkciót is betöltő – munkafázisnak köszönhető, hogy a SZÓKAPTÁR a plusz kotextuspartnerek szelekciójánál – az egy- és kétnyelvű tanulószótárakon kívül – egy egynyelvű speciális szótárra (DU2) és egy másik szignifikációs kookkurencialistára (IDS-korpusz alapján létrehozott CCDB) is támaszkodhatott.

<ADJ>

halbvoll * mundgeblasen * bauchig

<VERB>

klirren * nippen * heben * leeren * spülen

3: lekváros v befőttes- üveg ;

<ADJ>

groß * dekorativ

<VERB>

spülen

4: ált tbsz, szemüveg lencséje ;

<ADJ>

dick * geschliffen * verspiegelt * kratzfest

*** hauchdünn * randlos**

<VERB>

herunterfallen

Az egyes kotextuspartner-blokkokat természetesen – ahogy erre már korábban utaltam – még kiegészítettük az egy- és kétnyelvű tanulószótárak, illetve második körben a DU2 és az IDS-korpuszból létrehozott kookkurencia-adatbank (CCDB) segítségével, hogy az esetleges szelekciós hiányok és hibák arányát minimalizáljuk. Végül az első és a második lektorálás során az anyanyelvi munkatárs ellenőrizte a beolvasás és a kiegészítések eredményét, illetve az ismételt kiegészítéseket. Ekkor számos partner törlésre került (pl. *kratzfest*: 4. jelentésben vagy *klirren* 1. jelentésben, valamint számos, szótárból beválogatott kotextuspartner is), vagy a már bent lévő kotextuspartner másik jelentésponthoz lett átirányítva (pl. *geschliffen*: 4. jelentésből 1. jelentésbe). Így kerültek be végül az adatbázisba a következő plusz kotextuspartnerek:

1. jelentéshez: *feuerfest* (DU2), *bersten*, *einschneiden*, *behandeln* (LGWDaF és HESS);

2. jelentéshez: *handgeschliffen* (lektor), *anstossen*, *zerbrechen*, *springen*, *abetzen*, *vollgießen*, *nachfüllen*, *ergreifen*, *ansetzen*, *aufstellen*, *ausschütten*, *umkippen*, *umwerfen*, *herunterwerfen* (LGWDaF és HESS);

3. jelentéshez: *dicht*, *aufmachen*, *wiederverwerten* (lektor);

4. jelentéshez: *konkav*, *konvex* (LGWDaF), *dunkel* (CCDB), *herausfallen*, *brechen* (lektor)

és a 2. jelentés <KOMB>-blokkjában található két kombináció a HESS-ből.

Végül a 3. anyanyelvi lektorálásnál, amit egy kétnyelvű – magyar szülők gyermekeként Németországban felnövő – munkatárs, Zádor Éva végzett, ismét bővült a kotextuspartnerek száma. Így kerültek be a *bruchsicher*, *blasen*, *springen* (1. jel.) és *graviert* (2. jel.) kotextuspartnerek. Ebben a lektorálási folyamatban ritka volt a törlés, de annál gyakoribb a kiegészítés, így a szótár adatbázisa viszonylag nagymértékben nőtt:

- Összes kotextus száma: **44 622 49 213** (~10%-os növekedés)
- Összes kotextuspartner száma: **9933 10 393** (<5%-os növekedés)
- Bázisszócikkek száma: **2267 2269**

Végül álljon itt a kész szótári szócikk a **Glas** lemmajelhez (szürke háttérrel jelöltem a NEM a Lipcsei korpuszból származó adatokat):

Glas das, des Glas(e)s, die Gläser, fn

1: tbsz n *anyag üveg*

<ADJ>

getönt színezett * **kugelsicher** golyóálló * **schusssicher** golyóálló * **kratzfest** karcálló * **matt** matt, tompa(fényű) * **entspiegelt** † tükröződésmentes * **wärmedämmend** hőszigetelő * **geschliffen** csiszolt * **unzerbrechlich** *szinte* törhetetlen, összezúzhatatlan * **feuerfest** tűzálló * **bruchsicher** törésálló

<VERB>

splittern^{NOM} széttörik, szilánkokra törik * **blasen**^{AKK} fúj * **schleifen**^{AKK} (meg)csiszolja * *vál* **bersten**^{NOM} darabokra törik * **einschneiden**^{AKK} *méretre* bemetszi, bevágja * **mit etw**^{Dat} **behandeln**^{AKK} kezeli vmivel * **springen**^{NOM} (meg)reped

2: *üveg-* pohár

<ADJ>

halbvoll félig teli * **mundgeblasen** † fúj * **bauchig** öblös, hasas * **handgeschliffen** kézi csiszolású, kézzel csiszolt * **graviert** gravírozott

<VERB>

die Gläser klirren^{NOM} (meg)csörrennek a poharak * **am G. nippen** † kortyol/hörpint a pohárból * **sein G. auf jn/etw heben**^{AKK} emeli poharát vkire/vmire * **leeren**^{AKK} (ki)üríti * **spülen**^{AKK} elmosogatja, elöblíti * **mit seinem G. auf jn/etw anstoßen** koccint a poharával vkire/vmire * **zerbrechen**^{NOM} eltörik * **springen**^{NOM} elpattan, megpattan * **absetzen**^{AKK} leteszi, lerakja * **austrinken**^{AKK} kiissza * **vollgießen**^{AKK} teletölti * **nachfüllen**^{AKK} utántölti * **ausschwenken**^{AKK} kiöblíti * **ergreifen**^{AKK} megragadja, megfogja * **ansetzen**^{AKK} szájához emeli * **aufstellen**^{AKK} *miután felborult* felállítja * **ausschütten**^{AKK} kiönti *ami benne van* * **umkippen**^{NOM} felborul * **umwerfen**^{AKK} feldönti * **herunterwerfen**^{AKK} leverli

<KOMB>

das G. bis auf den Grund leeren † fenéig üríti a poharát

* **das G. auf einen/in einem Zug leeren** † egy húzásra kiissza/kiüríti a poharát

3: *lekváros v befőttés- üveg*

<ADJ>

groß nagy(méretű) * **dekorativ** dekoratív, mutatós * **csak**

áll dicht † *szivárgásmentesen zárva*

<VERB>

spülen^{AKK} (el)mosogat, kiöblít * **biz** **aufmachen**^{AKK} kinyit

* **wiederverwenden**^{AKK} újra felhasznál

4: ált tbsz *szemüveg lencséje*

<ADJ>

dick vastag * **kratzfest** karcálló * **entspiegelt** †

tükröződésmentes * **hauchdünn** lehetővékony * **randlos**

keret nélküli * **fiz konkav** konkáv, homorú * **fiz konvex**

konvex, domború * **dunkel** sötét

<VERB>

herunterfallen^{NOM} leesik * **herausfallen**^{NOM} az egyik

kiesik * **brechen**^{NOM} eltörik

Végül szeretnék rávilágítani a szelekción folyamat egyik problematikus pontjára. A Lipcsei korpuszból beválogatott kotextuspartnerek sorrendje megfelel a lemmával közös előfordulásuk statisztikailag szignifikáns voltának. Mihelyt azonban a kotextuspartnerek kiegészülnek szótárbeli adatokkal, egyszerűen nincs már lehetőség ezeknek a plusz adatoknak a statisztikai sorban elfoglalt helyének megállapítására. Egyenként ugyan visszakereshetőek, feltéve, hogy szerepelnek a listában, de a hosszú listák láttán és a plusz partnerek számát tekintve ez gyakorlatilag lehetetlen. Ennek ellenkezőjét példázza azonban a **Glas** lemmajel 2. jelentése alatt szereplő *blasen* kotextuspartner helye. Elvileg a szürke összefüggő blokk legvégén kellene lennie, mert nem a Lipcsei korpuszból került be az adatbázisba, hanem a 3. lektorálási fázis során. Szerkesztőként feltűnt azonban nekem a szókapcsolat tipikus volta és pusztán kíváncsiságból megpróbáltam megkeresni a statisztikai listában (lásd Lista 4). Meg is találtam az első 20 legszignifikánsabb ige között, a 19. helyen. Választanom kellett, meghagyom-e legutolsó kotextuspartnerként (8. helyen), vagy felléptetem a programban, a <VERB>-blokkon belül, a megfelelő helyre (2. helyre). Utóbbit választottam. Ez az eljárás azonban egyedinek számít. A következő táblázatban összegyűjtöttem néhány példát erre a nem szokványos és szabályos sorrendcserére:

Lemma jelentés	kotextpartner	3. lektorálás utáni helye	végleges (statisztikailag korrekt) helye
Hals 1	brechen	8. (utolsó előtti)	3.
haltbar 3	wissenschaftlich	4. (utolsó)	2.
Hemd 1	blütenweiß	15.	5.
	kariert	16.	1.
	gestreift	17.	2.
	gebügelt	18.(utolsó)	3.
Straße 2	eisglatt	16. (utolsó előtti)	2.
	regennass	14.	4.
Täter	unbekannt	13. (utolsó)	1.
Tier	quälen	19. (utolsó)	6.
unwohl	ziemlich	8. (utolsó)	3.
	höchst	6.	4.
verantwortlich 2	voll	7. (utolsó)	2.
Wohnung	möbliert	10.	6.
	eingrichtet	11.	9.
	bezahlbar	12. (utolsó)	4.
	mieten	13.	5.
	vermieten	14. (utolsó)	6.

Meg kell azonban barátkoznunk azzal a ténnyel, hogy a kotextusblokkok elején még érvényes szignifikációs sorrend, a kotextuspartner-blokk egy bizonyos pontjától (amikor kiegészülnek a korpuszadatok újakkal) már nem tartható. Innentől a kotextuspartnerek beviteli sorrendje pusztán esetleges. A szótárhasználat számára azonban ez nem nyomon követhető, tehát nem tudhatja, honnan nem érvényesül már a szignifikációs sorrend. Nagymértékben profitál ugyanakkor a plusz adatokból, hiszen a lemmajel egyes jelentéseihez tartozó tipikus kotextuspartnerek, illetve kotextusok/kollokációk összességében jól lefedik az adott szó disztribúciós lehetőségeit. Ha tekintetbe vesszük, hogy mennyi hiányt és milyen értékes, sokszor elsősorban interlingvális szempontból releváns adatot sikerült pótlólag a szótárak, a CCDB és a lektori kiegészítések segítségével a szócikkekbe integrálni, akkor talán ez a kis sorrendi „szépséghiba” utólag megbo csátható.

5. Összegzés

A SZÓKAPTÁR – tudomásom szerint – az első kétnyelvű, korpusz- és adatbázis-alapú szintagmatikus szótár, mely a jövőbeli publikációs médiumtól relatíve függetlenül született, így egyszerre többféle formában is megjelentethető.

Az adatbázis lehetővé teszi az adatok aktualizálását és bővítését, valamint a szótár más komponensekkel történő kiegészítését (második szövszedet, indexek stb.). Ily módon új szótárak alapját képezheti.

A SZÓKAPTÁR-t leegyszerűsítve korpuszalapú kollokációs szótári adatbázisnak is nevezhetjük, de annyi bizonyos: nyomtatott formában új szótártípust képvisel majd a magyarországi lexikográfiai gyakorlatban.

Bibliográfia

Korpuszok

CCDB = Belica, Cyril: Kookkurrenzdatenbank CCDB - V3.2. Eine korpuslinguistische Denk- und Experimentierplattform. Internet-Adresse: <http://corpora.ids-mannheim.de/ccdb>

IDS-Korpora. Institut für Deutsche Sprache, Mannheim. Größe der öffentlich verfügbaren Korpora im Juli 2002: mehr als 1736 Mio. Textwörter. Internet-Adresse: <http://corpora.ids-mannheim.de>

Leipziger Deutscher Wortschatz. Größe: 500 Mio laufende Wortformen. Internet-Adresse: <http://wortschatz.informatik.uni-leipzig.de/>

Szótárak

DU-2 = *Duden. Stilwörterbuch der deutschen Sprache*. Die Verwendung der Wörter im Satz. 7., völlig neu bearb. u. erw. Aufl. v. Günther Drosdowski unter Mitarbeit folgender Mitarbeiter der Dudenredaktion: Wolfgang Eckey, Dieter Mang, Charlotte Schrupp, Marion Trunk-Nussbaumer. Duden. Bd. 2. Mannheim [usw.] 1988 (Der Duden in 10 Bänden).

HESS-NM = Hessky Regina 2005. *Német–magyar kézisztár*. Harmadik, javított kiadás. .

HESSIKE = Hessky Regina/Iker Bertalan 2005. *Német–magyar, magyar–német tanulószótár*. Szeged: Grimm Kiadó

LGWDaF = *Langenscheidt. Großwörterbuch Deutsch als Fremdsprache. Das einsprachige Wörterbuch für alle, die Deutsch lernen*. Neubearbeitung. Hrsg.: Dieter Götz, Günther Haensch, Hans Wellmann. In Zusammenarbeit mit der Langenscheidt-Redaktion. Berlin [usw.] 2003.

PGWDaF = *PONS Großwörterbuch Deutsch als Fremdsprache*. [Bearb. v. Susanne Balhar, Frank Mönkemöller, Elke Hippe, Silke Löhle, Martina Lacatena, Lydia Quaas, Ulrich Saßmann, Julia Weißmüller. Infokästen: Noemi Zach. Illustrationen Sven Palmowski. Barcelona [usw.] 2004.

Szakirodalom

- Hausmann, Franz Josef 1984. Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. In: *Praxis des neusprachlichen Unterrichts* 31. 395–406.
- Hollós Zita 2004. *Lernerlexikographie: syntagmatisch. Konzeption für ein deutsch–ungarisches Lernerwörterbuch*. [Tanulói lexikográfia: szintagmatikusan. Német–magyar tanulószótár koncepciója.] Tübingen. (Lexicographica. Series Maior 116)
- Hollós Zita 2007a Kollokációk és korpusznyelvészet a SZÓKAPTÁR tükrében. [Kollokationen und Korpuslinguistik im Spiegel von KOLLEX.] In: Heltai Pál (Szerk.): *Nyelvi modernizáció. Szaknyelv, fordítás, terminológia*. A XVI. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai. Gödöllő, 2006. április 10–12. 2007. Vol. 3/1. Pécs – Gödöllő, 229–234.
- Hollós Zita 2007b. A lexikográfiai folyamat a SZÓKAPTÁR, egy korpuszalapú kollokációs aktív szótár példáján keresztül. Projektbeszámoló. [Der lexikographische Prozess am Beispiel des KOLLEX, eines korpusbasierten produktionsorientierten Kollokationswörterbuchs. Projektbericht.] In: Magay Tamás (Szerk.): *Félmúlt és közeljövő*. Budapest: Akadémiai Kiadó, 237–256. (Lexikográfiai füzetek 3)
- Hollós Zita 2010. Auf dem schmalen Grad zwischen Kollokationen und festen Wortverbindungen. [A kollokációk és kötött szókapcsolatok szűk mezsgyéjén.] In: *Feste Wortverbindungen und Lexikographie. Kolloquium zur Lexikographie und Wörterbuchforschung*. Hrsg. von Peter Ďurčo. New York 2010, 85–96. (Lexicographica. Series Maior 138)