

Pécsi Tudományegyetem
Bölcsészettudományi Kar
Nyelvtudományi Doktori Iskola
Alkalmazott Nyelvészeti Program

Aleksa Melita

**A horvát sztenderd nyelv számítógépes morfológiai
elemzésének alkalmazási lehetőségei német nyelvterületen**

Tézisfüzet

Témavezető: Dr. Barics Ernő
egyetemi docens

Pécs
2008

1. Bevezetés

Jelen dolgozat a sztenderd horvát nyelvre összpontosít, ugyanakkor párhuzamosan tárgyal egy másik nyelvcsaládból származó nyelvet, a németet. Mindehhez kiindulópontként a nyelvek morfológiai elemzését veszi alapul. A téma megközelítése előtt fontos azonban bizonyos fogalmakat tisztázni, illetve néhány kérdésre választ adni. Az egyik a morfológiai elemzés választásának indokát taglalja, a második pedig a német nyelvre mint kitűzött célcsoportra vonatkozik.

1.1. Miért esett a döntés egy horvát morfológiai elemző kidolgozására?

Az Európai Tanács által létrehozott *Közös Európai Referenciakeret*¹ megjelenése után számottevő igény keletkezett a piacon olyan tananyagok kidolgozására, amelyek a horvát mint második és idegen nyelv tanítását segítik, mindamelllett összhangban lesznek a Referenciakeretben leírt útmutatóval. A Referenciakeret horvát nyelvű kiadását részletesebben elemezve láthatjuk, hogy a nyelvtani kompetencia leírásánál a hangsúly többnyire a hagyományos nyelvtani szintekre kerül (Jelaska 2005a: 22). Ezt a tényt támasztja alá néhány Horvátországban elkezdett, illetve már befejezett projekt is, amelyek közül említést érdemel a *Hrvatski kao strani jezik: razvojna gramatika i rječnik* (MZOŠ 130738)² [A horvát mint idegen nyelv: nyelvtani fejlődés és szókincs] és a *Hrvatski kao drugi i strani jezik* (MZOŠ 0130438) [A horvát mint második és idegen nyelv] munkák. A projektek eredményei különböző tananyagok, egy morfológiai szótár és igeragozási útmutató, valamint a horvát alapszókincshez összegyűjtött, 700 szóból álló lista³. Mindezen eredmények már 2001 óta mutatják a projektek célját, azaz, hogy olyan tananyagok készüljenek, amelyek a horvát mint második és idegen nyelvet tanulók nyelvtani kompetenciájának fejlődését segítik elő. Megerősítésként a horvátot mint idegen nyelvet oktatók véleménye szolgál, akik többek között azt állítják, hogy a „meglévő nyelvtankönyvek nem felelnek meg sem az oktatók, sem a diákok igényeinek”⁴ (Cvikić 2005a: 320). Többször is hangsúlyozták, hogy szükség lenne nagyobb mennyiségű horvát mint idegen nyelv tanítását segítő taneszköz, nyelvtan és nyelvtani gyakorlókönyv, valamint az általános és szaknyelvet feldolgozó tankönyv, szótár, házi olvasmány kiadására (Cvikić 2005a: 320).

¹ Council of Europe (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press, a továbbiakban „Referenciakeret”

² <http://www.croatiana.org/croatiana--projekti-hrvatskikaostrani.htm>, 2008 február 4.

³ A pontos adatok és a kiadványok horvát megnevezése a dolgozatban olvasható.

⁴ Eredeti szöveg: „Postojeće gramatike za izvorne govornike ne odgovaraju ni potrebama predavača ni potrebama učenika.” (Fordítás MA).

Logikus folytatásként olyan, a horvátot mint második és idegen nyelvet segítő taneszközök megtervezését kezdeményeztük e dolgozat keretein belül, amelyek a horvát sztenderd nyelv automatikus morfológiai elemzésének felhasználásával valósíthatók meg.

A dolgozat a számítógépes morfológiai elemző két alkalmazási lehetőségét tárgyalja. Az első egy nyelvészeti jellegű alkalmazás, vagyis a nyelvtani paradigmák konkrét szemléltetése német anyanyelvűek számára. Ez az alkalmazás tulajdonképpen egy — a közeljövőben kidolgozandó — horvát számítógépes morfológiai szótár. Mindazonáltal az elemző révén létrehozott formalizmus a paradigmák újszerű megközelítését és elsajátítását mutatja be. A morfológiai szótár sajátossága, hogy a többi, már meglévő szótárhoz képest magában foglalja a Vladimir Anić *Veliki rječnik hrvatskoga jezika*⁵ [Nagy horvát értelmező szótár] összes lemmáját, és nem csak nyelvtani leírásokon alapul, hanem korpuszelemzés, illetve a nyelvhasználat által pontosított leírásokat is tartalmaz.

A morfológiai elemzés második alkalmazási lehetősége a számítógépes fordítást illetve a különböző nyelvtanulást segítő eszközök létrehozása, amelyek horvát, német és magyar⁶ viszonylatban is megvalósíthatók (vö. 9. fejezet). Mivel jelen dolgozat célja azon nyelvészeti problémák és dilemmák szemléltetése, amelyek a morfológiai elemző kidolgozása közben jelentkeztek, az elemző további alkalmazási lehetőségei csupán elméleti szinten kerülnek ábrázolásra.

1.2. Miért esett a választás a német anyanyelvűekre?

A fent említett morfológiai elemzésen alapuló eszközök elkészítése egy konkrét célcsoportot kívánt meg, ezért a német anyanyelvűekre a határon túli horvátok száma alapján esett a választás. A Horvát Külügyminisztérium adatai szerint (Prilog 1 — 1. függelék) a határon túli horvátok száma az angol és a német területen a legmagasabb. Figyelembe kell venni azt a tényt is, hogy a horvátok kivándorlása a német nyelvterületre már 1918-ban megkezdődött⁷, illetve hogy már léteznek hasonló taneszközök angol anyanyelvűek számára, valamint azt a

⁵ Anić, V. (2000). *Veliki rječnik hrvatskoga jezika*. Zagreb: Novi Liber, a továbbiakban RHJ.

⁶ A magyar nyelvre való taneszközök kidolgozására Barić (2006) tanulmánya alapján nagy igény mutatkozik.

⁷ <http://www.mvpei.hr/hmiu/tekst.asp?q=osi001>, 2008 február 4.

hipotézist, hogy a nem horvát anyanyelvűek számára⁸ a legnagyobb gondot a horvát nyelvtan helyes elsajátítása jelenti. Többek között Németországban is számos helyen oktatják a horvátot második illetve idegen nyelvként (Jelaska 2005b). Mindezek alapján megalapozottnak látszik a fent említett eszközök kifejlesztésekor a német anyanyelvűekkel kezdeni. Mivel tudomásom szerint nem létezik az elvárásoknak megfelelő segédeszköz erre a célcsoportra, jelen dolgozat keretein belül született a már fent említett döntés.

1.3. A morfológiai elemzési eljárás kiválasztásának indokai

A morfológiai elemzés kiindulópontként való kiválasztásához két ok vezetett. Először egy olyan eljárást kellett kidolgozni, amely produktuma könnyen átalakítható és implementálható bármilyen, nemcsak fordítást támogató rendszerbe, hanem más, a nyelvelsajátítást segítő eszközbe is. A választás az unifikációs alapú HUMOR⁹ rendszerre esett. A második indok magában az eljárás előnyeiben rejlett, pontosabban abban a tényben, hogy az elemzésen alapuló eszközök olyan paradigmákat és nyelvtani információt tartalmaznak, amelyek korpuszvizsgálat és tényleges nyelvhasználat alapján lettek kidolgozva. Ez a szintézisen alapuló eszközökről nem minden esetben állítható (vö. 2. fejezet).

2. Horvát nyelvű tananyagok vizsgálata

Mivel e dolgozat mellékcélja a horvát morfológiai szótár kidolgozásához szükséges lépések leírása, ezért a horvát nyelvű tananyagok vizsgálata arra szolgál, hogy átvilágítsa a horvátot oktató tananyagokban szereplő szókincset. A vizsgálat betekintést nyújt a szókincs használatába, elemzi a nyelvtan és segédkönyvekben szereplő leírások pontosságát, egyértelműségét, és a horvátot

⁸ A hipotézis egy kérdőíves felmérésen alapul, amelyet a Horvátországi Oktatási és Művelődési Központ horvát szakos tanárai töltöttek ki, továbbá egy szociolingvisztikai vizsgálaton is alapul, amelyet 2004-ben végzett a szerző. A kutatás eredményei azt mutatják, hogy a tanulók 85%-a magyar anyanyelvű, és csak 24% használja a horvátot a mindennapi kommunikációban (Aleksa 2007a). A Horvátországi Oktatási és Művelődési Központ tanárai személyes véleménye szerint a tanulók többsége B1-es és B2-es nyelvi kompetencia szinttel rendelkezik. A kompetenciaszint pontos megállapításához azonban további kutatások szükségesek. E dolgozat keretein belül a kidolgozandó anyagok ezért elsősorban a B1 és B2, valamint a magasabb nyelvi szinteket célozzák meg.

⁹ High-speed Unification Based Morphology, a MorphoLogic által kidolgozott morfológiai elemzőrendszer.

mint második illetve idegen nyelvet tanulók¹⁰ szempontjából való használhatóságát.

2.1. A tananyagok vizsgálatának elvei

A tananyagok vizsgálata két részletben történt meg. Az első részben — a dolgozat céljainak megfelelően — a német nyelvterületen használt német tankönyvek elemzése szerepel, a másodikban pedig a nyelvtan- és segédkönyvek vizsgálata történik. Mindkettő a *Kurikulum hrvatske nastave u inozemstvu*¹¹ [A horvát nyelv külföldön való oktatásának keretterve] elveit veszi figyelembe.

A dolgozat célkitűzéseinek megfelelően a tankönyvek vizsgálata a következő kérdésekre kísérel meg választ adni:

1. Mely szavak és szófajok szerepelnek leggyakrabban az elemzett tankönyvekben, valamint ezek melyik helyen találhatóak a horvát nemzeti korpuszon alapuló gyakorisági szótárban?
2. Milyen nyelvtani és lexikai kollokációk (vö. 2.1.3. fejezet) fordulnak elő leggyakrabban az elemzett tankönyvekben, és ezek hogyan viszonyulnak gyakorisági szempontból a referens korpusz (vö. 6.3. fejezet) vizsgálatának eredményeihez?

A nyelvtan- és segédkönyvek vizsgálata a következő kérdésekre próbál válaszolni:

1. Milyen mértékben járul hozzá a paradigmák és a morfológiai leírások szemléltetése, továbbá a tankönyvekben leggyakrabban előforduló ragozható szavak és kollokációs egységek ragozási paradigmáinak leírása a horvát nyelv sikeres elsajátításához? A kérdést a következő szempontok alapján válaszoljuk meg:
 - a) a leírások egyértelműsége;
 - b) a leírások alkalmazhatósága.
2. Megfelelő információt nyújtanak-e a nyelvtankönyvek a tankönyvekben leggyakrabban előforduló szavak paradigmáit illetően?

A feltett kérdésekre korpuszvizsgálat segítségével adtuk meg a választ.

¹⁰ A fent említettekhez méltóan a tanulók alatt e dolgozatban a közép- illetve magasabb szintű nyelvi kompetenciával rendelkező diákok értendők.

¹¹ A *Kurikulum hrvatske nastave u inozemstvu* a horvátországi Oktatási Minisztérium honlapjáról származik: <http://public.mzos.hr/Default.aspx?sec=2116>, 2008. június 6., a továbbiakban: Kurikulum.

2.1.1. Korpuszvizsgálat

A szógyakoriság kiszámítására a *dtSearch*¹² programot használtuk, illetve a kollokációk vizsgálatát az *Ngram Statistics Package*¹³ segítségével bonyolítottuk le. A továbbiakban a két program rövid ismertetése következik.

2.1.2. dtSearch

A szavak gyakorisági vizsgálatához a dtSearch program bizonyult a leghatékonyabbnak. A korpusz gyors indexelése révén statisztikai vizsgálatokra is megoldást kínál. Hiányosságai közé tartozik az, hogy a horvát szövegek lemmatizációja egyelőre nincs megoldva. Emiatt egy olyan horvát lemmatizálót kellett használni, amelyet a szerző a MorphoLogic eszközeinek segítségével dolgozott ki.

2.1.3. NSP

NSP — azaz az Ngram Statistics Package — egy, a Perl nyelven alapuló, Ted Pedersen kutatócsoportja által kialakított, ingyenesen hozzáférhető számítógépes programcsomag, amelynek segítségével nagy terjedelmű korpuszok vizsgálhatók. Összetett matematikai eljárások alapján számos hasznos információt nyújt a szókapcsolatokról, amely nagy segítséget nyújthat a nyelvészeti kutatómunka különböző területein (pl. nyelvtanítás, gyakorisági és frazeológiai szótárak szerkesztése).

A program elsősorban nem csak közvetlen szókapcsolatokat fedez fel egy elméletileg korlátlan mennyiségű szavakból álló szövegben, hanem olyan két vagy több szó közötti kapcsolatokat is vizsgál, amelyekben a szavakat esetlegesen néhány nyelvi egység (szó vagy mondattani jel) választja el egymástól. Az NSP működésének leírását összegezve azonban két dolog mindenképpen megemlíendő. Elsőként az eredmények relevánsága szempontjából a szókapcsolatok keresését lemmatizált szövegeken végezzük el,

¹² <http://www.dtsearch.com/customDevelop.html>, 2008. április 29.

¹³ <http://www.d.umn.edu/~tpederse/nsp.html>, 2008. április 29.

és csak ezután kezdjük el a kollokációkeresést¹⁴. Az ilyen módon kinyert eredmények így a tényleges nyelvhasználatot tükrözik. Ugyanis a szótári alakok kapcsolatát vizsgálva a kollokációk gyakoriságának és előfordulási valószínűségének kiszámítása relevánssá válik. Mivel szabadon elérhető horvát nyelvű lemmatizáló programok számomra nem ismertek, e dolgot megírásánál a korábban említett horvát lemmatizálót használtam fel.

2.2. A német nyelvterületen fellelhető horvát tankönyvek vizsgálata

Jelen fejezet célja a 2.1. pontban megtárgyalt kérdések alapján, korpuszvizsgálat segítségével elemezni a német nyelvterületen található nyelvkönyvekben lévő szókincset, valamint az eredményeket összehasonlítani az alapszótár összeállítási elveivel. A vizsgálat a továbbiakban megkísérel kitekinteni azokra a nyelvtani információkra, amelyek szükségesek a német nyelvterületen fellelhető horvát tankönyvekből tanuló diákok számára, és amelyek nem találhatók meg a Kurikulum által definiált kiadványokban. A korpuszt két horvát nyelvű, német nyelvterületen használatos könyv alkotja, amelyeket eddig még nem vontak be hasonló kvantitatív és kvalitatív jellegű korpuszvizsgálatba (Blagus 2005a és 2005b, Cvikić 2005b: 221): a Drilo, S. *Kroatisch, TL 1, Lehrbuch für Anfänger* (2006)¹⁵ c. tankönyve, valamint a *Kroatisch Lehrbuch für Fortgeschrittene* (1996)¹⁶ c. kiadványa.

A vizsgálat eredményei a következők: a KA és KF összesen 7591 különböző szót tartalmaz, amelyek 3833 lemmára vezethetők vissza. A dolgozatban közölt 1. és 2. ábra a tankönyvekben fellelhető szófajok és szavak felosztását szemlélteti. A 3. ábra azokat a lemmákat tartalmazza, amelyek megjelennek a könyvekben, de ma már nem a horvát, hanem inkább vagy kizárólag a szerb szókészlet részei (erről bővebben a 6.2. fejezetben). A tankönyvek szókincse — néhány kivételtől eltekintve — összhangban van a horvát gyakorisági szótárban

¹⁴ A kollokáció mint fogalom meghatározása során több nyelvből származó különböző definícióval találkozhatunk. A Bakos Ferenc (2002: 338) *Idegen szavak és kifejezések szótárának* definíciója szerint a kollokáció „gyakran együtt használt szavak szókapcsolata; állandósult szószerkezet / szókapcsolat”. A német Duden szótár a kollokációt a nyelvi elemek tartalmi kombinálhatóságának, illetve különböző tartalmak egy logikai egységen belüli összeadásának tekinti. Ezzel szemben a horvát értelmező szótárból származó definíció alapján a kollokáció egy, a szavak közötti kötelező, nyelvtan által nem meghatározott szókapcsolat. Jelen munka a kollokációt egy, a nyelvtanilag nem meghatározott szókapcsolatnak tekinti, amely két vagy több szó között alakul ki, és ahol nagy valószínűséggel fennáll a lehetősége annak, hogy összetevői (a lexikai egységek) mindig egymás mellett vagy egymás közelében helyezkednek el.

¹⁵ A továbbiakban: KA.

¹⁶ A továbbiakban: KF.

feltüntetett sorrenddel. Ha a morfológiai elemző, illetve szótár kidolgozásánál az elsőbbséget ezek a szavak kapják, a szöveg megértése és feldolgozása jelentősen javulhat. A *t-test* eredménye (Prilog 3. — 3. függelék) azt mutatja, hogy tokeneként a kollokációkban az igék szerepelnek a legnagyobb számban, utánuk a főnevek következnek. Mindazonáltal feltételezhetjük, hogy nagyobb igény lenne a nyelvtankönyvekben szereplő igei és főnévi paradigmára vonatkozó információk kikeresésére. A *Pointwise Mutual Information* elemzés (Prilog 4 — 4. függelék) pedig a melléknévi paradigma fontosságára utal. Többek között azt bizonyítja, hogy a tényleges nyelvhasználatban gyakran fordulnak elő melléknévi kollokációk is. A szavak szintaktikai viszonyáról további magyarázattal az NSP által lebonyolított tri- és 4-gram elemzés szolgál, amelyet az 5. és 6. függelék tartalmaz. Ezen túlmenően a szavak szintaktikai tulajdonságainak leírása és elemzése jelen dolgozat keretein kívül esik.

2.3. A horvát nyelv oktatásában használt nyelvtankönyvek elemzése

Az elemzendő nyelvtankönyvek kiválasztása a horvát Oktatási Minisztérium által engedélyezett kiadványokon¹⁷, valamint a horvát szakos tanárok (akik a horvátot nem horvát anyanyelvűeknek tanítják¹⁸) által kitöltött kérdőív válaszain alapul. A végső választás a reprezentativitás elv felhasználásával a következő kiadványokra esett:

- J. Silić–I. Pranjković (2005) *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*¹⁹,
- E. Barić et al. (1995) *Hrvatska gramatika*²⁰,
- D. Raguž (1997) *Praktična hrvatska gramatika*²¹,
- a *Gramatički tezaurus hrvatskoga jezika v 1.2. c.* elektronikus nyelvtani szótár²²,

¹⁷ Az elemzésbe bevett kiadványok a *Kurikulum hrvatske nastave u inozemstvu c. dokumentum* 8.1.3 pontja alatt vannak felsorolva (<http://public.mzos.hr/Default.aspx?sec=2116-a>, 2008. június 8.). Mivel a többi feltüntetett kiadványt nem használják a horvát nyelv tanítására a B1-es és B2-es szinteken, ezért az elemzésben nem szerepelnek.

¹⁸ A dolgozat jellegzetességeit és a német nyelvterületen dolgozó horvát tanárok a szerző számára való elérhetetlenségét figyelembe véve, a kérdőívet csak olyan horvát szakos tanárok töltötték ki, akik magyar anyanyelvű diákoknak horvátot mint idegen nyelvet tanítanak.

¹⁹ A továbbiakban: GHJ.

²⁰ A továbbiakban: HG.

²¹ A továbbiakban: PHG.

²² A továbbiakban: GT.

- és a *Hrvatski morfološki leksikon* horvát morfológiai lexikon internetes verziója²³.

2.3.1. Igék

Az igei paradigma vizsgálata után az a következtetés szűrhető le, hogy a horvátot mint második vagy idegen nyelvet tanuló diákok számára egy újszerű megközelítés szükséges, amely elsősorban a korpuszelemzésen alapul. Mindezt bővebben a 7.4.2. fejezet tárgyalja.

2.3.2. Melléknevek

A mellékneveket érintő leírások elemzésével a fentihez igen hasonló következtetésre juthattunk. Felmerül azonban a kérdés, hogy vajon mekkora a tényleges melléknévi lemmák száma a RHJ-ben, ugyanis a szótárban felsorolt melléknevek esetén következetlenségek vehetők észre (Aleksa 2007b). Az újszerű melléknévi paradigma szemléltetése a dolgozat 7.3. fejezetében tekinthető meg.

2.3.3. Főnevek

Ha az elemzett segéd- és nyelvtankönyvekben, valamint Anić (2000) szótárában tüzetesebben megvizsgáljuk a főnévi paradigmát, először a szükséges információk hiánya illetve nem egyértelműsége tűnik fel, utána pedig a grammatikai nem problémája, amelyről a 8.4.2. fejezetben lehet bővebben olvasni.

2.3.4. Névmások és számok

A nyelvtan- és segédkönyvekben szereplő névmások és számnevek paradigmájáról szóló információk vizsgálata után arra a következtetésre jutottunk, hogy a nem horvát anyanyelvűek számára egy formalizmus általi megközelítés konkrét információkat nyújthat, és minden morfológiával kapcsolatos kérdésre konkrét választ ad. A nyelvi formalizmus a HUMOR

²³ A továbbiakban: HML.

leírásán alapul, amely tulajdonképpen a számítógépes morfológiai szótár alapja is.

2.3.5. A Vladimir Anić *Veliki rječnik hrvatskoga jezika* c. horvát értelmező szótár vizsgálata

A vizsgálat alapján az a következtetés vonható le, hogy a nem anyanyelvűek számára nem elegendők a RHJ-ban lévő információk, valamint az, hogy egy más jellegű tanulói szótár megszerkesztése kívánatos, amely — ahogyan az Prószycki (1997) tanulmányában olvasható — dinamikus keresési lehetőségeket nyújt. Egy ilyen jellegű szótár létrehozásához mindenképpen a jelen dolgozat által felvázolt morfológiai szótár tekinthető alapnak.

3. Morfológiai elemzés a morfológiai szótár kidolgozásának szolgálatában

3.1. Alapelvek a horvát sztenderd nyelv morfológiai szótárának kidolgozásához

A horvát morfológiai szótár kidolgozása előtt felmerült néhány kétely és kérdés, amelyeket szükséges volt megválaszolni. Először is, mivel a horvát számítógépes morfológiai szótár a horvát sztenderd nyelv morfológiai elemzésének egyik alkalmazási lehetősége, fontosnak bizonyult bővebben megindokolni a szótár kidolgozásához kiválasztott eljárást, tulajdonképpen az elemzés és nem a szintézis használatát.

A tankönyvek vizsgálata másfajta megközelítést tett szükségessé, ezért fontos volt másmilyen jellegű, a tényleges nyelvhasználathoz illeszkedő taneszközök kidolgozásáról elgondolkodni, amelyek eltérnek a hagyományos, szintézisen alapulóktól. Tadić (1994: 31) dolgozatában megemlíti azt a tényt, hogy szintézissel mint eljárással többnyire a kutatói hipotézisek tesztelhetők, és célja alapvetően nem az alakok felismerése. A horvát morfológiai elemző kidolgozása során, a szövegek közel 100%-os felismerésére törekedve, a lexikai adatbázis nemcsak a RHJ-ból származó szavakat tartalmazza, hanem a tesztkorpusz vizsgálatának eredményeit is figyelembe veszi. Ily módon egy kibővített, nyelvhasználat által pontosított alakban alapot nyújt a morfológiai szótár kidolgozásához (vö. 4.2. fejezet).

A morfológiai szótárral kapcsolatban a második felmerülő kérdés a nyelvi változat kiválasztásának problémáját érinti.

3.2. A morfológiai szótár számára kiválasztott nyelvi változat

A horvát nyelvi változat kérdése elsősorban a sztenderd nyelvi változat korlátozására, illetve a szótárba az egyéb változatokból származó lemmák felvételére vonatkozik. Tulajdonképpen két lehetőség kínálkozik: az első a morfológiai elemző sztenderd nyelvi alap közvetlen felhasználása. A másik pedig, a német nyelvterületen található horvát tankönyvek elemzését kiindulópontként felhasználva (vö. 2.2. fejezet) a horvát nyelv egyéb változataiból származó lemmák feldolgozása.

A döntés végső soron csak a horvát sztenderd nyelvi változat felhasználására esett, mivel a Kurikulum a többi nyelvváltozatot csupán receptív készségek szintjén kezeli.

4. HUMOR mint egyfajta morfológiai elemző

A HUMOR, azaz a High-Speed Unification based Morphology olyan unifikációs alapú, a MorphoLogic által kidolgozott morfológiai elemző program, amelyet több nyelv morfológiai elemzésére fejlesztettek ki (Prószéky–Kis 1999). A program univerzalitását sokoldalú használata is bizonyítja: a HUMOR az alapja például a MobiDic és MobiMouse szótárkezelő alkalmazásnak, valamint a MetaMorpho programnak, beépíthető különböző helyesírás ellenőrzőkbe és a nyelvek morfológiai elemzésére is kiválóan használható (vö. 9. fejezet). Demóverziója 1992-ben készült, és ma már a működő lengyel, német, angol, magyar, valamint számos, kidolgozás alatt álló változaton túl, a jelen dolgozat keretében a horvát változat is elkészült.

4.1. A HUMOR működése

Röviden összefoglalva a morfológiai elemzés a szavak stem-re és term-re való felbontásán alapul. Stem-nek tekintjük a szó azon részét, amelynek írása a ragozás során nem változik meg. Például a *noga* 'láb' szó esetében stem-nek a *no-* részt tekintjük, mivel a ragozás során megjelenik a *nozi* változat is, szótőnek ugyanakkor a *nog* tekintendő. Ugyanebből adódik az, hogy a horvát term, amely alatt egyszerűsítve a szó végződését értjük, hosszabb és változékonyabb, mint a hagyományosan értelmezett horvát toldalék. Mivel a HUMOR leírása nem a hagyományos grammatikai kategóriákkal operál, fontos volt a program

kidolgozásánál új fogalmakat bevezetni a szótő, illetve a rag²⁴ helyett (Prószéky–Kis 1999). A HUMOR részletes leírásáról, elemzési eljárásának működéséről, illetve a magyar és horvát verzió működésének jellegzetességeiről a dolgozat e fejezetében olvasható bővebben.

4.2. A morfológiai elemző struktúrája

A morfológiai elemző struktúrája több részre osztható. A technikai, ún. *engine* egységen kívül, még egy adatbázis is tartozik hozzá. Nyelvészeti szempontból a lexikai alap (a HUMOR-féle szótár) a legfontosabb. A horvát nyelv esetében az Anić (2000) szótárának teljes készletét magában foglalja. A lexikai alap azonban nem egy meghatározott, véges számú elemből áll össze, hanem szükség esetén bármikor új lemmákkal bővíthető.

4.3. A fogalmak definiálása

A morfológiai elemzés működésének részletesebb leírása előtt fontos néhány fogalmat meghatározni.

1. Mivel maga a rendszer a szóközt elemzési határnak tekinti, és a szót grafikus szinten értelmezi (pl. az összetett igeidők felismerése ez által problémát okozhat), a szócsoporthoz tartozó elemek az első fázisban csak akkor ismerhetők fel, ha a lexikai adatbázisban lemmák részeként szerepelnek. Pl. a visszaható névmás *sebe/se* együtt szerepelhet a *kupati* 'fürdet' igével a *kupati se* 'fürdik' jelentésben is. Mivel a *kupati* ige nem vonz kötelezően visszaható névmást, az ige a lexikai adatbázisban így kétszer szerepel: *kupati*-ként, illetve *kupati se*-ként. Ezzel a szóköz problémája részben megoldható.

Mivel a jelen tanulmány csak a morfológiai analízis problémáival foglalkozik, eltekintünk a szintaktikai, fonetikai és pragmatikai aspektustól. Mindazonáltal amennyiben a HumorESK-et, vagyis a HUMOR szintaktikai tudással gazdagított változatát használjuk fel különböző programokban, ezzel növelhetjük a szöveg számítógép általi felismerését.

2. A hagyományos nyelvtantól eltérően, a HUMOR rendszer keretein belül a szavak szófajhoz való tartozását a paradigmája határozza meg (vö. 7.

²⁴ Jelen dolgozatban a *rag* kifejezés alatt, ha másképpen nincs kiemelve, a horvát *nastavak*, illetve *sufiks* megfelelőjét értjük, ami megfelel a magyar *toldalék* fogalmának és magában foglalja a *képzőt*, a *jelet* és a *ragot*.

fejezet). Pl. a melléknév kategóriája minden melléknévi paradigmával rendelkező szót tartalmaz.

3. A HUMOR rendszer keretein belül kidolgozott megoldások hasznosságát egy pilotkutatással tervezzük a jövőben felülvizsgálni. A morfológiai szótár kidolgozása jelenleg az utolsó fázisnál tart (vö. 8. fejezet).
4. Az elemzésnél a horvát összetett betűk (*dž*, *lj* és *nj*) egy egységnek tekintendők.

5. Az agglutináló, flektáló és magasan flektáló nyelvekre kidolgozott egyéb morfológiai elemzők áttekintése

A HUMOR szerzői (Prószekey–Kis 1999: 261) többször is hangsúlyozták a rendszer kétszintű morfológiai unifikációs alapját (Koskenniemi 1983), amely a német GERTWOL leírásában is szerepel (Haapalainen–Majorin 1994). Ez a fejezet részletesen tárgyalja a német nyelv specifikumait és a GERTWOL által nyújtott megoldásokat, valamint összehasonlítja ezeket a horvát, magyar és lengyel nyelvleírási problémák jellegzetességeivel. Mivel a morfológiai elemző egyik jövőbeli alkalmazása a horvát–német, illetve horvát–magyar fordítóeszközök kidolgozása, jelen fejezet bővebb információt nyújt a két rendszerről, és a fentiekben említett nyelvcsoporthoz tartozó nyelvek feldolgozásáról.

6. A horvát morfológiai elemző kidolgozása

6.1. A nyelvtörténet és nyelvpolitika kérdése

A modern horvát nyelv csak a XX. század 90-es éveiben alakult át egy erős nyelvtisztító folyamat révén, az 1990 előtti művek szerbhorvát nyelven íródtak. A sztenderd horvát nyelv kodifikálása 1835-ben kezdődött el, majd a folyamat megszakadása után a szerbhorvát nyelv kialakulása következett. Emiatt a horvát nyelvnek három fontosabb változata létezik: az 1835 előtti, a szerbhorvát nyelvi változat 1990-ig, illetve a modern horvát nyelv. Mivel a lengyel (a HUMOR egyetlen kidolgozott szláv nyelvi változata, amely a latin írást használja) lehetővé teszi a XVIII. századi szövegek elemzését is (Wołosz 2005), felmerül a kérdés, hogy a horvát változatnak is kell-e ilyen lehetőséget biztosítani. Ha az esetlegesen felmerülő igénynek eleget kívánunk tenni, a meglévő szókészlet jelentős bővítése válna szükségessé, és figyelembe kellene venni a különböző nyelvtani és helyesírási szabályokat, valamint a szerb nyelvből származó lexémákat, ami az adott esetben a program elnevezésében a horvát jelző használatát megkérdőjelezi. A végső döntés a horvát sztenderd nyelvi változatra esett a szókészlet bővítésének lehetőségével, ha a korpuszvizsgálat egy szónak

illetve a szó változatának gyakori használatát mutatja, és ezáltal jelentősen javulhat a szövegfelismerés.

6.2. A nyelvváltozat kérdése

A 3.2. fejezetben tárgyaltakhoz hasonlóan, a morfológiai elemző kidolgozásánál is felmerült a nyelvváltozat kérdése. A dilemma az elemzőnek az írott változatra való korlátozása, és az esetleg beszélt változat felvétele között állt, valamint a helyi és egyéb nyelvi változatok tipikus lexémáinak hozzáadására vonatkozott. A végső döntés az elemző írott változatára esett, a többi változathoz származó lexémák az elemző lexikai adatbázisába való bevételi lehetőségének fenntartásával, amennyiben ennek szükségességét a korpuszvizsgálat alátámasztja.

6.3. A tesztkorpusz

Mivel számunkra a teljes Horvát Nemzeti Korpusz (HNK) nem volt elérhető, a rendszert egy, a HNK mintáján készült tesztkorpuszon tettük próbára. A korpusz öt alkorpuszból áll össze, amelyek a következő részekre oszthatók:

1. a legfontosabb horvát irodalmi művek (dráma, próza és líra);
2. efemer anyagok és nyomtatott írások;
3. elektronikus médiából származó szövegek (internetes oldalokról, blogokból);
4. szakirodalomból származó szövegek;
5. különböző tankönyvekből származó szövegek.

A műfajok kiválasztásának indokairól, illetve a korpusz összeállításának kérdéseiről bővebb információ a dolgozat e fejezetében található.

7. A morfológiai elemző használata a horvát sztenderd nyelv oktatására és tanulására

A 2.3.5. fejezetben leírtak szerint igény támadt a horvát nyelv oktatását és tanulását segítő különböző eszközök kidolgozására. A CALL (Computer-Aided Language Learning) (Prószéky 1997) keretén belül, a Prószéky által összefoglalt igényeknek megfelelően, olyan eszközök kidolgozása vált szükségessé, amelyek a tanulóknak dinamikus módon, megfelelő információval szolgálnak szavak paradigmáiról, valamint összekötik a horvát lemmákat a

tanulók anyanyelvéből származó lemmákkal is. Ennek következtében ebben a fejezetben, a jelen dolgozat célkitűzéseit figyelembe véve, a horvát nyelv számítógépes, német anyanyelvűek részére kidolgozott morfológiai szótárának felvázolására kerül sor. A továbbiakban a felmerült kérdésekről és problémákról esik szó.

7.1. A horvát ragozható szófajok feldolgozásának problémái és ezek megoldási lehetőségei

A morfológiai elemző kidolgozása során több, a ragozható szófajt (horvátul *promjenjiva vrsta riječi*) érintő probléma keletkezett. A jelen dolgozat csak a legfontosabbakat szemlélteti, valamint azokat a megoldásokat mutatja be, amelyek jelentősen eltérnek a hagyományos nyelvtani reprezentálástól, és segítséget nyújtanak a horvát mint második és idegen nyelv oktatásában.

7.2. Főnevek

7.2.1. Főnévi paradigma

Ha a nem horvát anyanyelvű tanuló a nyelvtan- és segédkönyvekben szereplő információkra támaszkodik, az alábbi adatok szükségesek számára, hogy sikeresen generáljon egy helyes horvát főnévi paradigmát:

- az adott hímnemű főnév élő vagy élettelen fogalmat ábrázol-e;
- melyik ragozásról van szó (*-a, -e* vagy *-i deklinacija*),²⁵
- a főnév alapja

A HUMOR rendszer által kialakított ragozási mátrixhoz (a dolgozat 24. ábrája) nem szükséges a fent említett információ, illetve eljárás. Csupán az alábbi, minden egyes főnévnél feltüntetett információ segítségével létrehozható a helyes főnévi paradigma:

- a szó stem-je
- az adott hímnemű főnév élő vagy élettelen kategóriához tartozik-e;
- a főnév 11-es (nominatív egyes szám), 12-es (genitív egyes szám), 21-es (nominatív többes szám) és 22-es (genitív többes szám) term-je

²⁵ A főnevek egyes ragozási csoportokba való besorolása a főnév genitív alakja szerint történik. Más szóval, a HG szerint a végződéseket ahhoz az alaphoz illesztjük, amelyet általában úgy állapítunk meg, hogy a genitív egyes számból elhagyjuk a végződést (Barić et al. 1995: 104). (A horvát szöveg: „nastavci se dodaju na osnovu, koja se, u pravilu, dobije ako se u gen. jedn. izostavi nastavak”.)

Így pl. a *more* 'tenger' vagy a *dječak* 'kisfiú' lemma paradigmájához szükséges információk a következők lennének. A függőleges vonal a stem-nek a term-től való elválasztását jelenti.

mor|e sr.r (nž., -e, -a, -a, -a)
dječa|k m.r (ž., -k, -ka, -ci, -ka)

Ha a pilotkutatás a 24. ábrában feltüntetett főnévi mátrix hasznosságát bizonyítja, szükségessé válik egy ilyen információval gazdagított tanulói szótárt is összeállítani.

7.2.2. Nyelvészeti problémák

A főnévi paradigmát illetően a legfontosabb nyelvészeti problémakörök közé az esetek száma és a vokatívusz eset meghatározása tartozik. Mivel a főnévi, továbbá a melléknévi paradigmánál a datívusz és lokatívusz esetek kivétel nélkül egyformának bizonyultak morfológiai szinten, a morfológiai szótárban hét eset helyett elegendő hatot ábrázolni. A vokatívusz problémáját, valamint az esetek további kérdéseit bővebben a 8.4.1-es fejezet tárgyalja.

7.3. Melléknevek

A hagyományos nyelvtani leírás szerint a horvát nyelvtan a mellékneveket először szemantikai tulajdonságuk, majd határozottsági tulajdonságuk alapján csoportosítja, majd három nemben ragozza (egyes és többes számban), élő vagy élettelenként, a melléknevet követő főnévtől függően. Ebből következik, hogy pl. a *crven* (piros) leíró melléknévnek 93 különböző, az azonosakkal együtt összesen 123 ún. ragozási cellája létezik. Fontos azonban hangsúlyozni, hogy az említett ragozási alakok csak az írott szöveg elemzésére vonatkoznak. A beszélt nyelvben a hangsúlyt is figyelembe véve a különböző ragozási cellák száma tovább bővül. A HUMOR rendszer számára kidolgozott melléknévi mátrix (a dolgozat 27. ábrája) és minden melléknév tizenhat különböző alakja segítségével (a dolgozat 26. ábrája) létrehozható a teljes melléknévi paradigma. A melléknevek stem-jeik illetve term-jeik alapján 21 ragozási csoportba sorolhatók be (a dolgozat 30. ábrája). Döntőnek minősülnek az 1-es és az 5-ös (szótári) alakok, amelyek párban való előfordulása kizárólag a leíró melléknevek tulajdonsága. Kivételt képeznek a főnévből képzett melléknevek, amelyeknek a ragozási alakjai azonosak a leíró melléknevekével, viszont a paradigmájuk kizárja a határozott ragozást.

7.3.1. Melléknevek fokozása

Mivel a HUMOR leírásában a melléknevek közép- illetve felsőfokot képező ragjai különböző term-eket és stem-eket alkotnának (ami jelentős mértékben bonyolítaná a melléknévi ragozás rendszerezését), így ezek a fokozott melléknevek külön lemmaként szerepelnek a szótárban, és természetesen utalnak az adott melléknév szótári alakjára.

7.3.2. Nyelvészeti problémák

A bemutatott melléknévi rendszer leírása nemcsak számítógépes kezelés szempontjából hasznos, hanem a nem anyanyelvűek számára is nagy segítséget nyújt a horvát nyelv elsajátításában, mert elsősorban nem a szavak szemantikai tulajdonságát veszi alapul, hanem azok morfológiai felépítését. Továbbá nincs szükség mind a 93 ragozási cella megtanulására, hanem csak a melléknév tizenhat különböző alakját kell elsajátítani.

Az első előbukkanó probléma a hím- és semlegesnemű alakokkal nem rendelkező melléknevek ragozását (pl. *trudna* 'terhes'), majd a második a nem fokozható melléknevek fokozását illeti. Ezeket a kérdéseket a dolgozatban alaposabban górcső alá vesszük. Gondot okoznak a helyesírási eltérések is. Például a *mađarski* 'magyar' melléknévnek a Babić – Finka – Moguš (1995) *Hrvatski pravopis* c. horvát helyesírási szótár szerint két írásmódja létezik: a *mađarski* és a *madžarski* (1995: 277). Ezeket azonban Anić értelmező szótára nem tünteti fel, és csupán az első említett alakot közli. Megoldásként mindkét alak külön lemmaként szerepel a HUMOR szótárában.

7.3.3. A horvát melléknévi paradigma számítógépes feldolgozásának összehasonlítása más nyelvek melléknévi paradigmájának feldolgozásával

A kidolgozott nyelvi formalizmus hasznosításának hasonló példái más szláv nyelvekben is megtalálhatók. A dolgozat 31. ábrája az orosz példát ábrázolja, ám hasonló megoldás a lengyel nyelvből is idézhető. Az említett nyelvekből származó szótári megoldások a horvát morfológiai szótár elkészítésénél is alkalmazhatók.

7.4. Igék

7.4.1. Igei paradigma

7.4.2. Igei paradigma a morfológiai elemzõn belül

A 2.3.1-es fejezetben lévõ igei paradigma szemléltetése alapján arra a következtetésre juthatunk, hogy az igei paradigma a horvát nyelvten egyik legtöbbet tárgyalt tétele. A szemléltetések és magyarázatok változatossága és ellentétes volta (vö. 2.3.1-es fejezet) arra utal, hogy szükség lenne egy új paradigmaleírásra, amely konkrét válaszokkal szolgál majd a tanulóknak. Anić (2000) szótárában 12 413 ige szerepel, amelyeket a hagyományos nyelvten hét igeragozási csoportba, ezek alatt további osztályokra bontja. Az igeragozást illetõ nyelvészeti problémák elsõsorban a múlt igeidõk alakjainál (*aorist*, *imperfekt*), az igei fõnév (*glagolska imenica*) és igei melléknév (*glagolski pridjev*) kategorizálásánál, valamint ezek fõnévi, melléknévi illetve igei paradigmához való besorolásánál jelentkeznek. A HUMOR számára javasolt megoldás a korpuszvizsgálat alapján megpróbál ezekre a kérdésekre válaszolni. Az igeiket stem-jeik és term-jeik alapján 96 ragozási csoportba sorolja. A csoportokat képviselõ paradigmák az Anić (2000) szótárában meglévõ összes ige helyes alakjának generálását biztosítják, miként az a 35. és a 36. ábráról leolvasható.

7.4.3. További problémák

Külön nehézséget okoz, hogy a HUMOR rendszer számára a szóköz vagy a szóköz értékû karakter mindig az elemzési határt jelenti. Az ebbõl adódó problémák a HumorESK esetleges horvát változatának használatával valamennyire orvosolhatók. A horvát morfológiai rendszer leírásánál figyelembe kellett venni azt is, hogy sok olyan szó létezik, amelynek ragozása egyedi, és nem követ egy-egy típust. Ezeket a szavakat mind külön lehetett csak a HUMOR szótárába felvenni.

8. A horvát sztenderd nyelv számítógépes morfológiai szótára

8.1. A szótár struktúrája

8.2. Technikai megoldások

8.3. A morfológiai szótár további alkalmazási lehetõségei

A felhasználók igényeit és a már érintett problémákat figyelembe véve, a tervezett morfológiai szótár több táblából álló összetett adatbázis szerkezetének köszönhetõen a következõ lehetõségeket nyújtáná:

1. A felhasználó kiválasztaná a keresett szót a horvát szólistából, vagy az idegen nyelvű listából, amely a szó horvát megfelelőjére automatikusan hivatkozna.
2. Az idegen nyelvű szólistán kívül a felhasználó választhatna több nézet közül. A lemmákat ábécé sorrendben soroltathatja fel, vagy *a tergo* sorrendben.
3. A lemma kiválasztása után a felhasználónak az egyes lemma, illetve paradigmára vonatkozó információk rendelkezésére állnának:
 - a) „úgy ragozódik mint az x”
 - b) az y-dik kategóriához tartozik
 - c) szükséges nyelvtani információk (nem, szám, stb.)
 - d) a szó változatlan részét (stem-et) úgy alkotjuk, hogy a szó végéből z-számú karaktert vágunk le, és ehhez adjuk utána a term-eket
 - e) a teljes paradigma
 - f) a lemma illetve a szóalak kontextuson belüli használata
 - g) a szavak elválasztásának módja

A szótár kibővíthető különböző egyéb információkkal, mint pl. kollokációkkal. Ez azonban egy jövőbeli projekt célja.

8.4. Nyelvi kérdések és megoldások

8.4.1. Esetek

A szótár kidolgozása során több olyan kérdés merült fel, amelyre megfelelő megoldást kellett találni. Ezek a megoldások azonban megkérdőjelezzik az eddigi hagyományt és egyes tételek újszerű megközelítését kívánják meg. Az elemző által használt formalizmus minden bizonnyal meg tudná könnyíteni a kérdéses tételek bemutatását és tanítását nem horvát anyanyelvű diákoknak. Az első kérdés, amely részletesebb kifejtést igényel, az esetek kérdése. A második pedig a grammatikai nemre vonatkozik.

A datívusz és a lokatívusz eseteken kívül az egyik legproblémásabb eset a vokatívusz (a probléma részletesebben a dolgozat e fejezetében olvasható). A nehézségek voltaképpen a nyelvtankönyvekben lévő információk mennyiségében és minőségében tükröződnek. Mivel az egyes szavak helyes vokatívusz alakját illető kérdésre még a korpuszvizsgálat sem tud elfogadható megoldást adni (pl. a vokatívusz eset alakja mindig egybeesik a paradigma valamelyik szóalakjával és ez által csak (fél)ig manuálisan emelhető ki a korpuszból), így a kérdés további kutatást igényel egy jövőbeli projekt keretein belül.

8.4.2. A nemek problémája

A szavak neme szintaktikai kategória, amit a Barić et al. (1995: 101) *Hrvatska gramatika* c. nyelvtankönyv is állít. Mivel a lengyel nyelvészek több nyelvtani nem létezését bizonyították a lengyelben (Saloni et al. 2007: 9), felmerült a kérdés, hogy vajon a hagyományos horvát grammatikai nemek felhasználásával a diákok képesek lennének-e grammatikailag helyes mondatokat generálni. Mivel a válasz egyértelműen negatívnak bizonyult, több újszerű nyelvtani nem kategória létrehozása vált szükségessé, és egyben elkerülhetetlenné. Ezek a nemek a hagyományos nyelvtani nemek alkategóriáinak tekinthetők, és segítségükkel a számítógép, illetve a nem horvát anyanyelvű tanulók képesek lennének helyes mondatokat generálni.

Ezen felfogás szerint kilenc nyelvtani nem létezik a horvát nyelvben, amelyet az alábbi mondatba a főnevek behelyezése bizonyít:

Iako [*biti*] ovdje [imenica] [odnosna zamjenica] volim, ne vidim [pokazna zamjenica]

Habár [*van*] itt [főnév] [vonatkozó névmás] szeretek, nem látom [mutató névmás]

[*dva*] [imenica].

[*kettő*] [főnév].

A mondatban használt számnév, illetve a vonatkozó névmás alakja bizonyítja az egyes nemek létezését, ahogyan az a dolgozat 44. ábrájáról is leolvasható.

9. A morfológiai elemzés további alkalmazása

9.1. Nyelvtanulásra és -oktatásra kidolgozott eszközök

9.1.1. Egynyelvű és többnyelvű szótárak

9.1.2. Magyar nyelv tanulására kidolgozott eszközök

9.2. Fordítástámogató rendszerek

A HUMOR rendszer unifikációs alapja lehetőséget kínál többféle rendszerbe való implementálásra. Jelen fejezetben a MorphoLogic által kidolgozott, a HUMOR illetve HumorESK rendszert használó eszközök bemutatására is sor került. Fontos azonban megemlíteni, hogy a bemutatott eszközök célja nem reklám, csupán szemléltető, informatív jellegű. A bemutatandó eszközök kiválasztásánál a szakmai elismerés döntött. Ebben a fejezetben a következő eszközök lettek bemutatva: MobiDic (*multi-dictionary environment*), MobiMouse (*context-sensitive instant comprehension tool*), Słownik gramatyczny języka polskiego, Rječnik hrvatskoga jezika na CD-romu, DUDEN Großwörterbuch der deutschen Sprache, DigiMorf, Helyesen, Elragoz

és a *Metamorpho (MetaMorpho Dictionary)*. A fejezet részletesen kitér az eszközök előnyeire és hátrányaira is.

10. Kitekintés

Jelen dolgozat két szükségletre világított rá. Az egyik praktikus jellegű — a horvát mint második illetve idegen nyelv tanításához szükséges segédeszközök kidolgozását jelentette —, a másik pedig elméleti, és a horvát nyelvtanításban a nem horvát anyanyelvűeknek szóló tételek újszerű feldolgozását és ábrázolását foglalja magában. A dolgozat korlátait figyelembe véve nem minden probléma és megoldás került bemutatásra, a hangsúly csak a legfontosabb, újszerű nyelvészeti feldolgozás témakörébe tartozó tételekre helyeződött. A jövőbeli tervek közé tartozik a morfológiai elemzés egyik alkalmazási lehetősége, a már említett, és több tudományos konferencián bemutatott kollokációs szótár kidolgozása is.

Hivatkozott irodalom

- Aleksa, Melita (2007a). Die Problematik des bilingualen Erstspracherwerbs bei kroatisch- ungarischen Schülern: Eine Untersuchung der schriftlichen Korpora in kroatischer Sprache In: Balaskó Maria – Szatmári Petra (szerk.) *LE 59: Sprach- und Literaturwissenschaftliche Brückenschläge – Vorträge der 13. Jahrestagung der GESUS in Szombathely, 12.-14. Mai 2004 Szombathely*. München: Lincom Verlag, 209–217.
- Aleksa, Melita (2007b). A horvát melléknévi rendszer nyelvtani leírása és automatikus számítógépes kezelése. In: Heltai Pál (szerk.) *MANYE XVI. Nyelvi modernizáció. Vol 3/2*. Pécs-Gödöllő, 1257–1263.
- Anić, Vladimir (2000). *Veliki rječnik hrvatskoga jezika*. Zagreb: Novi Liber.
- Babić, Stjepan, Božidar Finka, Milan Moguš (1995). *Hrvatski pravopis*. Zagreb: Školska knjiga.
- Bakos Ferenc (2002). *Idegen szavak és kifejezések szótára*. Budapest: Akadémiai kiadó.
- Barić, Ernest (2006). *Rode, a jezik?*. Pečuh: Biblioteka Nova.
- Barić, Eugenija et al. (1995). *Hrvatska gramatika*. Zagreb: Školska knjiga.
- Blagus, Vlatka (2005a). Odabir riječi u udžbenicima hrvatskoga za strance. In: Jelaska, Zrinka. (szerk.) *Hrvatski kao drugi i strani jezik*. Zagreb: Hrvatska sveučilišna naklada, 261–273.
- Blagus, Vlatka (2005b). Pregled udžbenika hrvatskoga za strance. In: Jelaska, Zrinka. (szerk.) *Hrvatski kao drugi i strani jezik*. Zagreb: Hrvatska sveučilišna naklada, 226–234.
- Cvikić, Lidija (2005a). Hrvatski kao drugi i strani jezik: stanje i potrebe. In: Jelaska, Z. (szerk.) *Hrvatski kao drugi i strani jezik*. Zagreb: Hrvatska sveučilišna naklada, 311–329.
- Cvikić, Lidija (2005b). Pregled priručnika hrvatskoga kao drugoga i stranoga jezika. In: Jelaska, Z. (szerk.) *Hrvatski kao drugi i strani jezik*. Zagreb: Hrvatska sveučilišna naklada, 219–225.
- Drilo, Stjepan (1996). *Kroatisch Tl.2. Lehrbuch für Fortgeschrittene*. Heidelberg: Julius Groos Verlag.
- Drilo, Stjepan (2006). *Kroatisch Tl.1. Lehrbuch für Anfänger*. Tübingen: Julius Groos Verlag.
- Haapalainen, Mariikka, Ari Majorin (1994). *GERTWOL: Ein System zur automatischen Wortformerkennung deutscher Wörter*. Lingsoft, Inc., <http://www.lingsoft.fi/cgi-pub/gertwol>, 2008.08.06.
- Hrvatski morfološki leksikon, <http://hml.ffzg.hr/hml/>, 2008.10.01.
- Jelaska, Zrinka (2005a). Jezik-znanje ili sposobnost. In: Jelaska, Z. (szerk.) *Hrvatski kao drugi i strani jezik*. Zagreb: Hrvatska sveučilišna naklada, 11–24.
- Jelaska, Zrinka (2005b). *Hrvatski kao drugi i strani jezik*. Zagreb: Hrvatska sveučilišna naklada.
- Koskeniemi, Kimo (1983). *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production*. Publication No. 11. Helsinki: University of Helsinki.

Kurikulum hrvatske nastave u inozemstvu:

<http://public.mzos.hr/Default.aspx?sec=2116>, 2008.10.1.

Prószéky Gábor (1997). Language Technology in the Service of CALL. In: Kohn, János; Bernd Rüschoff; Dieter Wolff (szerk.) *New Horizons in CALL (Proceedings of EUROCALL 96)*., 53–64. Szombathely: Berzsényi Dániel College

Prószéky Gábor, Balász, Kis (1999). A unification-based approach to morpho-syntactic parsing of agglutinative and other (highly). inflectional languages. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*. Maryland, USA: College Park, 261–268. http://www.morphologic.hu/h_pgpub.htm, 2006.08.02.

Raguž, Dragutin (1997). *Praktična hrvatska gramatika*. Zagreb: Medicinska naklada.
Saloni, Zygmunt, Włodzimierz Gruszczyński, Marcin Woliński, Robert Wołosz (2007). Grammatical Dictionary of Polish. Presentation by the Authors. *Studies in Polish Linguistics*, volume 4, 5–27.

Silić, Josip, Ivo Pranjković (2005). *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*, Zagreb: Školska knjiga.

Tadić, Marko (1994). *Računalna obrada morfologije hrvatskoga književnog jezika*. Ph.D. Thesis, Manuscript. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu. <http://www.hnk.ffzg.hr/mt/>, 2008.10.01.

Wołosz, Robert (2005). *Efektywna metoda analizy i syntezy morfologicznej w języku polskim*. Warszawa: Akademicka Oficyna Wydawnicza EXIT.

Publikációs jegyzék

- Aleksa, Melita–Wołosz, Robert (2005). Development of a New Implementation of the Unification-Based Morphological Analyzer HUMOR for the Croatian Language. In: *Papers in Computational Lexicography COMPLEX 2005*. Budapest: Linguistics Institute, Hungarian Academy of Sciences, 9–17.
- Aleksa, Melita (2005). A HUMOR rendszer adaptálása a horvát nyelvre. In: Klaudy Kinga és Dobos Csilla (szerk.) *A világ nyelvei és a nyelvek világa. MANYE XV*. Pécs–Miskolc, 2006. Vol. 2/2, 212–216.
- Aleksa, Melita (2006a). Automatic Morphological Analysis of the Croatian Language: The Verbal, Adjectival and Nominal Inflections within the Morphological Parser HUMOR. In: *CESCL1, Proceedings of the First Central European Student Conference in Linguistics*. Budapest: Nyelvtudományi Intézet (<http://www.nytud.hu/cescl/proceedings.html>).
- Aleksa, Melita (2006b). Der kroatische HUMOR: Überlegungen zu einer computergestützten morphologischen Analyse der flektierenden Sprachen. In: *Jezikoslovlje 7.1–2*, Osijek: Filozofski fakultet, 141–152.
- Aleksza Melitta (2007). A horvát melléknévi rendszer nyelvtani leírása és automatikus számítógépes kezelése. In: Heltai Pál (szerk.) *MANYE XVI. Nyelvi modernizáció. Vol 3/2*. Pécs–Gödöllő, 1257–1263.
- Aleksa, Melita (2008a). HUMOR als Basis für maschinelle Übersetzungsprogramme: Morphologische Analyse der kroatischen und deutschen Adjektivformen In: Karabalić, Vladimir – Omazić, Marija (ur). *Istraživanja, izazovi i promjene u teoriji i praksi prevođenja – Explorations, Challenges and Changes in Translation Theory and Practice – Theorie und Praxis des Übersetzens: Alte Fragen und neue Antworten*. Osijek: Sveučilište J.J.Strossmayera – Filozofski fakultet, 219–232.
- Aleksa, Melita (2008b). Zu einer computergestützten Bearbeitung des kroatischen Deklinationssystems. In: Karin Pittner (szerk.) *Beiträge zu Sprache und Sprachen 6. Vorträge der 16. Jesus-Tagung in Bochum*. München: Lincom Europa, 131–141.

A kutatási témával kapcsolatos előadások

Problematika izrade računalnog programa za morfološku analizu hrvatskoga standardnog jezika

VII. Međunarodni kroatistički skup
Pécs, 2004. december 10.

Automatische Suche nach phraseologischen Einheiten in kroatischen und polnischen Texten

„Phraseologie disziplinär und interdisziplinär”. Internationale EUROPHRAS Tagung, veranstaltet vom Germanistischen Institut an der Universität Veszprém (Ungarn) und der Europäischen Gesellschaft für Phraseologie
Veszprém, 2006. június 11.

Pridjevska paradigma u računalnoj obradi hrvatskoga književnog jezika

VIII. Međunarodni kroatistički znanstveni skup
Pécs, 2006. október 13.

Zur Erstellung eines korpusbasierten kroatischen Kollokationswörterbuches

„Phraseology global – areal – regional”, EUROPHRAS Tagung
Helsinki, 2008. augusztus 15.