

# **Félig kompozicionális főnév + ige szerkezetek: elméleti kérdések és számítógépes nyelvészeti elemzések**

A DOKTORI ÉRTEKEZÉS TÉZISEI

Vincze Veronika

Szegedi Tudományegyetem

2011. augusztus

Témavezető: Dr. Bibok Károly



Szegedi Tudományegyetem  
Nyelvtudományi Doktori Iskola  
Elméleti nyelvészet PhD-program



## 1. Bevezetés

A doktori értekezés témája a félig kompozicionális főnév + ige felépítésű szerkezetek elméleti és számítógépes nyelvészeti szempontú vizsgálata. A kutatást elsődlegesen magyar nyelvű anyagon végezzük, időnként más nyelvek (főként angol, orosz és spanyol) hasonló szerkezeteivel összevetve.

E szerkezetek hasonló felépítésűek, ám valójában nem alkotnak egységes kategóriát a kompozicionalitás szempontjából: a produktív pusztá főnév + ige komplexumok (*újságot olvas, levelet ír*) és az idiomatikus kifejezések (*csütörtököt mond, lépre csal*) mellett létezik egy harmadik kategória is, amelynek képviselői sem nem produktívak, sem nem idiomatikusak, és jelentésük nem teljesen kompozicionális, így azt meg kell adni a lexikonban. (1) alatt néhány angol és magyar példa látható. Mivel jelentésük ugyanaz, csupán glosszákat adunk meg:

(1) (a) magyar:

*előadást tart*

*virágba borul*

*lehetőség nyílik*

(b) angol:

*give a lecture*  
ad egy előadás

*come into bloom*  
jön ILL virág

*a possibility emerges*  
egy lehetőség felmerül

A szakirodalomban nincs egységes elnevezésük ezeknek a szerkezeteknek (Dobos, 1991; Dobos, 2001; Langer, 2004). A németben *Funktionsverbgefüge* (funkcióigés szerkezetek) a leggyakoribb megnevezés<sup>1</sup>, az angolban *complex verb structures* (komplex igei szerkezetek), *support verb constructions* (támasztóigés szerkezetek) vagy *light verb constructions* (könnyű

<sup>1</sup>Meg kell jegyeznünk ugyanakkor, hogy a német szakirodalmi hagyományok nem tekintik *Funktionsverbgefüge*-nek az alanyi főnévi komponenset tartalmazó szerkezeteket.

igés szerkezetek)<sup>2</sup> használatos, a franciában *constructions à verbe support* (támasztói-gés szerkezetek), az olaszban *costruzioni a verbo supporto* (támasztói-gés szerkezetek), a portugálban *construções com verbo suporte* (támasztói-gés szerkezetek) vagy *construções com verbo leve* (könnyű igés szerkezetek) elnevezéseket találunk. Míg a fenti elnevezések mindegyikében szerepel az igei komponens, addig az orosz terminus nem emeli ki egyik komponenst sem: *описательные выражения* (leíró kifejezések).

A magyarban a komplexumok megnevezésére a *körülíró szerkezetek* (Sziklai, 1986), *leíró kifejezések* (Dobos, 1991), illetve *funkcióigés szerkezetek* (Keszler (1992) nyomán) terminusok használatosak, de a – némi értékítélettől sem mentes – *terpeszkedő szerkezetek* kifejezés mind a Nyelvművelő kézikönyvtárban (Grétsy és Kemény, 1996, p. 571), mind pedig újabb szakirodalmi tételekben is előfordul (például Heltai és Gósy (2005) a terpeszkedő szerkezetek hatását vizsgálja a nyelvi feldolgozásra).

Ahogy a fenti példák is mutatják, a szakirodalomban használatos terminusok nagy része – különösen az idegen nyelvűek, az orosz kivéve – az igei komponensről nevezi meg a szerkezetet azt sugallva ezzel, hogy az a szerkezet feje. Mivel azonban az ige csak szintaktikai szempontból tekinthető a komplexum fejének, hiszen szemantikai szempontból a főnév funkcionál a kifejezés fejeként (l. alább), talán nem szerencsés egyik komponenst sem kiemelni a szerkezet megnevezésében. A pejoratív terminusokat és a szerkezet körülíró természetét<sup>3</sup> hangsúlyozó elnevezéseket szintén kerülendőnek tartjuk. Emiatt Langer (2004) nyomán *félig kompozicionális szerkezeteknek* fogjuk nevezni a főnév + ige komplexumoknak az értekezésben tárgyalt típusát.

## 2. Célkitűzés

A félig kompozicionális szerkezetek természete kettős. Egyrésztől, két szintaktikai összetevőből állnak (főnévi és igei komponens), másrésztől, egy szemantikai egységet képeznek. E kettősség határozza meg nyelvi jellemzőiket a nyelvtan minden szintjén:

- a szerkezet szintaktikai és szemantikai feje nem esik egybe, az ige lévén a szintaktikai

<sup>2</sup>Kisebb különbségek lehetnek a fenti terminusok használatában – például a szemantikailag üres támasztói-géket *light verb* megnevezéssel illeti Meyers et al. (2004), azaz a *support verb* terminus a *light verb* hipernimája.

<sup>3</sup>Véleményünk szerint a félig kompozicionális szerkezetek fontos többletjelentést (például aspektuális információt) képesek a cselekvéshez adni, így nem pusztán körülírják az igei variáns által jelölt szituációt.

fej, a főnév pedig a szemantikai fej (kettős fejűség);

- a szerkezet mint egész jelentése nem jósolható meg teljességgel részeinek jelentése alapján (félkompozicionalitás);
- az ige és a főnév (vagy főnévi osztályok) jelentéskomponenseinek legalább részlegesen meg kell egyezniük annak érdekében, hogy egy egységet formáljanak (félproduktivitás);
- a produktivitást és kompozicionalitást tekintve a félig kompozicionális szerkezetek a produktív főnév + ige szerkezetek és idiómák között helyezkednek el (fokozatosság);
- a félig kompozicionális szerkezeteknek létezhet igei párjuk, amely általában ugyanarra a tőre vezethető vissza, mint a főnévi komponens, és hasonló jelentéssel rendelkezik, mint a szerkezet (variativitás).

A félig kompozicionális szerkezetek kettőssége az alábbi lényeges, egymással összefüggő elméleti kérdésekhez kapcsolódik. Először: a szemantika és a szintaxis egy az egyben megfeleltethető-e egymásnak, azaz **homomorfizmus** áll-e fenn a két nyelvi szint között? Másodsor: miként kell értelmezni a **kompozicionalitás** fogalmát: abszolút (azaz egy szerkezet kompozicionális vagy sem) vagy relatív értelemben (vagyis egyes szerkezetek kompozicionálisabbak, mint mások)?

Amennyiben homomorfizmust feltételezünk a szintaktika és a szemantikai között, az az **erős kompozicionalitás** fogalmának felel meg (egy kifejezés részeinek jelentése és azok szintaktikai kapcsolódási módja **teljes egészében** meghatározza a kifejezés egészének jelentését), vö. Hale (1997). A kompozicionalitás elvének azonban nem minden nyelvi egység felel meg, például az idiómák jelentése nem számítható ki alkotórészeik jelentéséből és azok kapcsolódási módjából, azaz nincs egyértelmű leképezés szintaxis és szemantika között, így az idiómák nem kompozicionális nyelvi elemek. Az értekezésben amellet érvelünk, hogy a kompozicionalitás elvét követő kifejezések és a kompozicionalitás teljes hiányát mutató kifejezések pusztán a skála két végpontját képviselik, köztük még elhelyezkedhetnek olyan elemek, amelyekre kisebb-nagyobb mértékben érvényes a kompozicionalitás fogalma. A **félkompozicionalitás** fogalma tehát a következőképpen értendő: a szerkezet jelentése **többé-kevésbé** meghatározható

részeinek jelentéséből és azok kapcsolódási módjából (azaz nem teljesen kompozicionális és nem teljesen idiomatikus).

Az értekezésben leírt kutatás célja kettős. Először, az elméleti célok közé tartoznak:

- annak vizsgálata, hogy milyen mértékben hasonlítanak a félig kompozicionális szerkezetek a produktív szerkezetekre vagy idiómákra;
- a félig kompozicionális szerkezetek mint egység részletekbe menő szintaktikai és szemantikai elemzése több nyelvből származó adatok alapján, több elméleti kereten belül (például generatív és függőségi nyelvtanok);
- a szerkezet két komponense (a főnév és az ige) közti szintaktikai és szemantikai viszony elemzése;
- a félig kompozicionális szerkezetek és igei párjaik összevetése szintaktikai és szemantikai szempontból.

Másodszor, a félig kompozicionális szerkezetek számítógépes nyelvészeti kezelési módjait tekinti át az értekezés, külön hangsúlyt fektetve az elméleti eredmények számítógépes nyelvészeti alkalmazásokba való átültetésére.

Az értekezés fő célja, hogy az elméleti nyelvészeti eredményeket a lehető legnagyobb mértékben alkalmazza a számítógépes nyelvészet több részterületén, közvetlen kapcsolatot teremtve ezzel elmélet és gyakorlat között. Az értekezés két fő részét ez a törekvés kapcsolja egymáshoz.

### **3. Kutatási kérdések**

Lássuk ezek után egy kicsit részletesebb felsorolását a tárgyalás sorrendjében azoknak a kutatási kérdéseknek, amelyekre az értekezésben választ keresünk.

1. (a) Milyen viszonyban állnak a félig kompozicionális szerkezetek az idiómákkal és a produktív szerkezetekkel?
- (b) Hogyan jellemezhető a szerkezetek és igei variánsaik viszonya?
- (c) Milyen szintaktikai jellemzőkkel bírnak a félig kompozicionális szerkezetek?

- (d) Milyen szintaktikai viszony van a félig kompozicionális szerkezetek két tagja közt?
  - (e) Milyen szemantikai jellemzőkkel bírnak a félig kompozicionális szerkezetek?
  - (f) Milyen szemantikai viszony van a félig kompozicionális szerkezetek két tagja közt?
  - (g) Milyen lexikai reprezentációval írhatók le a félig kompozicionális szerkezetek?
2. A kutatás eredményei mennyire tekinthetők nyelvfüggőnek, illetve nyelvfüggetlennek?
3. Hogyan hasznosíthatók a kutatás eredményei a számítógépes nyelvészet különféle területein:
- (a) a félig kompozicionális szerkezetek automatikus azonosításában;
  - (b) a jelentés-egyértelműsítésben;
  - (c) az információkinyerésben és -visszakeresésben;
  - (d) a gépi fordításban?

## 4. A kutatás módszertana

A fenti kérdések megválaszolásának érdekében magyar és angol nyelvű korpuszokból származó nyelvi adatokat vetettünk elemzés alá. Ily módon az elméleti állításokat gyakorlati adatok segítségével lehetséges megerősíteni vagy megcáfolni. Hogy a kutatáshoz megfelelő mennyiségű nyelvi adatot gyűjtsünk, több témakörből származó szövegekben kézzel bejelöltük a félig kompozicionális szerkezeteket. Az annotációs elveket a korábbi elméleti eredmények határozták meg, egyúttal a korpuszokból gyűjtött adatok szolgálták a félig kompozicionális szerkezetek viselkedését leíró elméleti megállapítások alapjául. Ezáltal az adatgyűjtés módszertana is tükrözi az elmélet és gyakorlat összefonódását.

Az adatgyűjtéshez három, félig kompozicionális szerkezetekre annotált korpuszt hoztunk létre. A Szeged Treebankben (Vincze és Csirik, 2010), a SzegedParalellFX korpuszban (Vincze et al., 2010) és a Wiki50 korpuszban (Vincze et al., 2011) kézzel bejelölt szerkezetekből összeállítottunk egy adatbázist, amely 1524 magyar és 827 angol szerkezetet tartalmaz. Az értekezés szerzője felügyelte az annotálási munkálatokat, és maga is részt vett az annotálásban. A létrehozott korpuszok egyrészt felhasználhatók valós nyelvhasználati adatok gyűjtésére, másrészt

azon számítógépes nyelvészeti alkalmazások tanítására és kiértékelésére, amelyek természetes nyelvi szövegekben kívánják azonosítani a félig kompozicionális szerkezeteket.

## 5. Elméleti eredmények

Az értekezés első része elméleti szempontok alapján tárgyalja a félig kompozicionális szerkezeteket, aminek eredményei az alábbiakban foglalhatók össze.

### 5.1. Félig kompozicionális szerkezetek mint a többszavas kifejezések egy altípusa

A félig kompozicionális szerkezetek a többszavas kifejezések egy altípusának tekinthetők (vö. Sag et al. (2002)). A többszavas kifejezések más típusaival való több szempont alapján történő összevetésből kiderült, hogy a félig kompozicionális szerkezetek **lexikális és szemantikai szinten idioszinkratikusak**, továbbá **szintaktikailag produktívak** és **szemantikailag részekre bonthatók**. Ezek a jellemzők alapvető hatással bírnak az értekezés későbbi fejezeteiben leírt elméleti és gyakorlati vizsgálódásokra.

### 5.2. A félig kompozicionális szerkezetek státusa

Az értekezés a félig kompozicionális szerkezeteket olyan produktív szerkezetekkel és idiómákkal vetette össze, melyek szintaktikai felépítése szintén a puszta köznév + ige sémát követi. Ennek érdekében létrehoztunk egy tesztegységtestet, amely elkülöníti egymástól a puszta köznév + ige szerkezetek különféle típusait. Fény derült arra is, hogy **a főnév + ige szerkezetek kontinuum**a a **produktivitás** és a **kompozicionalitás** mutatóival jellemezhető. A teszteredmények alapján a főnév + ige szerkezetek három nagy csoportra oszlanak – produktív szerkezetek, félig kompozicionális szerkezetek és idiómák. A félig kompozicionális szerkezeteket a produktív szerkezetektől és az idiómáktól a **variativitás** és **az ige elhagyhatósága** különbözteti meg, ugyanakkor a félig kompozicionális szerkezetek és az idiómák – szemben a produktív szerkezetekkel – a szemantikai idioszinkráziát mutató többszavas kifejezések közé sorolhatók. A félig kompozicionális szerkezetek további alcsoportokra bonthatók. A skálán elhelyezkedő



csoportok között azonban **nem húzható éles határ**, mert nem a vagy-vagy dichotómia, hanem az is-is vagy a sem-sem játszik szerepet a szerkezeteknek az adott csoportba való tartozásában, azaz a fokozatosság fogalmának segítségével lehet meghatározni egy adott szerkezet helyét a skálán. A teszteredmények segítségével kimutathatóvá vált az is, hogy a korábban egységként kezelt idiómák csoportja (Kiefer, 1990 1991; Komlósy, 1992) komplex képet mutat: bizonyos szerkezetekben erősebb az idiomatikus jelleg, mint más, hagyományosan idiómának tartott szerkezetekben.

### 5.3. A félig kompozicionális szerkezetek igei párjai

A félig kompozicionális szerkezetek legtöbbször megfelelhető egy (gyakran a főnévi komponenssel azonos tövű) igei variáns. Az értekezésben magyar és angol félig kompozicionális szerkezeteket állítottunk szembe egyrészt igei párjaikkal, másrészt a másik nyelvű megfelelőjükkel. Az eredmények azt mutatják, hogy a szerkezetnek és igei párjának használata nem minden esetben esik egybe **vonzatkeretbeli** (például *ajándékot ad vkinek* és *ajándékoz vkinek vmit*) vagy **akcióminőség- és aspektusbeli** eltéréseknek köszönhetően (például a *forgalomba hoz* szerkezet inchoatív akcióminőséget hordoz, ellentétben a *forgalmaz* igei variánssal). A produktív szerkezetekkel szemben (Kiefer, 2007) a félig kompozicionális szerkezetek **nem rendelkeznek inherens progresszív aspektussal**. Fény derült arra is, hogy a nyelvek közti különbségek nagyrészt a vonzatkeretbeli különbségek számlájára írandók, továbbá – a magyarral ellentétben – az angolban a szerkezetek rendelkezhetnek progresszív aspektussal. Végül bemutattuk, hogy a félig kompozicionális szerkezetek megítélése és **elfogadhatósága is fokozatosság kérdése**, hasonlóan a főnév + ige szerkezetek skáláján elfoglalt helyükhöz. A számítógépes nyelvészet terén az információ-visszakeresés hasznosíthatja a fejezet eredményeit, a nyelvek közti különbségek elemzése pedig a gépi fordításban juthat fontos szerephez.

### 5.4. A félig kompozicionális szerkezetek szintaxisa

A félig kompozicionális szerkezetek szintaxisát generatív és függőségi keretben tárgyaltuk, kiindulópontul Hale és Keyser (2002), Newson et al. (2006), Larson (1988), Chomsky (1995) és Alonso Ramos (1998) elemzései szolgáltak. Az elemzésekben különös figyelmet szenteltünk a

vonzatkeret-alternáció és a szerkezet és igei párja közti levezethetőség kérdéseinek. Az elméleti keretek és a különböző elemzések megegyeznek abban, hogy **a szerkezet vonzatai mély szinten a főnévhez tartoznak**, míg egyes vonzatok **az igehez kerülnek a felszínen**. Az argumentumtranszfer pontos szabályai (azaz melyik vonzat kerül át az igehez) azonban szerkezetfüggőek. Noha a félig kompozicionális szerkezetek frazális tulajdonságokkal is rendelkeznek, a szemantika szintjén egy egységként viselkednek. Ezzel összhangban bemutattunk egy olyan alternatív elemzést, amely a félig kompozicionális szerkezetet **komplex predikátumként** kezeli, ezáltal kiküszöböli az argumentumtranszfer kérdését. Utóbbi elemzés fontos következményekkel bír a számítógépes nyelvészeti, különösen információkinyerési alkalmazásokban.

### 5.5. A félig kompozicionális szerkezetek szemantikája

A félig kompozicionális szerkezetek szemantikáját az értelem–szöveg modell (l. például Mel'čuk et al. (1995)) keretében elemeztük. A főnév és ige közti szintaktikai-szemantikai relációt le lehet írni lexikai függvények segítségével. Elemzésünkhöz a magyar adatbázisban szereplő négy leggyakoribb ige (*ad*, *vesz*, *hoz* és *tesz*) egyikét tartalmazó félig kompozicionális szerkezeteket választottuk ki. A vizsgálatok alapján megállapítható, hogy **szemantikai összefüggések találhatóak a főnév, az ige és az alkalmazott lexikai függvény között**, például a beszédaktusokat jelölő főnevek a magyar nyelvben nagyon gyakran az *ad*, *tesz* és *hoz* igeik valamelyikével együtt alkotnak félig kompozicionális szerkezetet. A fejezetben lexikai szemantikai szempontból is vizsgálta a félig kompozicionális szerkezeteket: egyrésztől léteznek **szinonim** szerkezetek (*bajba jut – bajba kerül*), másrésztől bizonyos félig kompozicionális szerkezetek **konverzív** párokat alkotnak, azaz ugyanazt a szituációt írják le, de más résztvevő szemszögéből (*bérbe ad – bérbe vesz*). Ezek az eredmények a jelentés-egyértelműsítésben és a gépi fordításban is hasznosíthatók.

### 5.6. A félig kompozicionális szerkezetek lexikai reprezentációja

Az értekezésben bemutattuk a félig kompozicionális szerkezetek szótári reprezentációjának négy lehetséges módját, és megvitattuk azok előnyeit és hátrányait mind elméleti, mind gyakorlati megfontolások alapján. A félig kompozicionális szerkezetek fejének megállapításához

kapcsolódó nehézségek bemutatását és az elektronikus és papíralapú szótárak összevetését követően a négy szótári reprezentációt a *köt* és *hoz* ige és főnévi komponenseiket tartalmazó miniszócikkek segítségével illusztráltuk. A módszerek közül mind elméleti, mind gyakorlati szempontokból az került ki győztesen, amely **az ige és a főnév szócikkében is szerepelteti a szerkezetet**, noha a jelenlegi szótárak többsége az ige (Pusztai, 2003; Bárczi és Ország, 1959 1962) vagy a főnév (kollokációs szótárak, értelmező-kombinatorikus szótárak (Mel'čuk és Žolkovskij, 1984; Mel'čuk és mtsai, 1984 1999)) szócikkében szerepelteti a szerkezeteket. A fejezet végén bemutattuk, hogy a **Magyar WordNet** adatbázisban a félig kompozicionális szerkezetek **különálló lexikai egységként** jelennek meg. A félig kompozicionális szerkezetek szótári reprezentációja jelentős befolyással bír az információkinyerésben, jelentés-egyértelműsítésben és gépi fordításban.

## 6. Számítógépes nyelvészeti eredmények

Az értekezés második részében bemutattuk, hogyan kezeli a számítógépes nyelvészet több részterülete a félig kompozicionális szerkezeteket. Külön hangsúlyt fektettünk az elméleti eredmények gyakorlati alkalmazására. A következőkben az értekezés számítógépes nyelvészeti eredményeit tekintjük át röviden.

### 6.1. A félig kompozicionális szerkezetek automatikus felismerése

A félig kompozicionális szerkezetek automatikus felismerése megvalósulhat statisztikai, szabályalapú és hibrid módszerek segítségével is. Megvitatottuk a különböző módszerek előnyeit és hátrányait, és figyelmet fordítottunk arra is, hogy az elemzési láncban mikor és hogyan érdemes azonosítani a többszavas kifejezéseket, konkrétan a félig kompozicionális szerkezeteket. Bemutattuk saját **szabályalapú**, illetve **gépi tanulási technikákon** alapuló felismerőnket, amely angol nyelvű szövegekben képes azonosítani a félig kompozicionális szerkezeteket, szemben más alkalmazásokkal, amelyek ige + tárgy párokat osztályoznak aszerint, hogy idiomatikusak-e vagy sem (Bannard, 2007; Cook et al., 2007; Tan et al., 2006; Tu és Roth, 2011). Eredményeink azt igazolják, hogy **sekély morfológiai elemzéssel** (lemmatizálás, szó-

tövesítés, szófaji egyértelműsítés és toldalékok felismerése) már versenyképes eredményeket lehet elérni a félig kompozicionális szerkezetek automatikus felismerése terén, ugyanakkor a **szintaktikai információk** integrálása tovább javítja a rendszer teljesítményét. Egy **doménre és témakörre szabott funkcióige-listával** pedig bármely doménon sikeres eredményeket lehet elkönyvelni, különösen szintaktikai jellemzők felhasználásával. A félig kompozicionális szerkezetek felismerése ennél fogva leghatékonyabban a szintaktikai elemzést követően, egy utófeldolgozó lépésben valósulhat meg, annak végeredményét pedig jól tudják hasznosítani a magasabb rendű alkalmazások, például az információkinyerés és a gépi fordítás.

## 6.2. Félig kompozicionális szerkezetek a jelentés-egyértelműsítésben

A félig kompozicionális szerkezetek jelentés-egyértelműsítéséhez mind a főnévnek, mind az igének egy jól definiált jelentéshalmazát szükséges megadni. Lexikográfiai megfontolások alapján érdemes megkülönböztetni az igei komponens **funkciói** és **fői** jelentéseit. A számítógépes nyelvészeti alkalmazások sorrendjét tekintve amellet érveltünk, hogy a jelentés-egyértelműsítő algoritmusok jobban teljesítenek akkor, ha már a jelentés-egyértelműsítés előtt ismertek a félig kompozicionális szerkezetek, azaz a félig kompozicionális szerkezetek felismerése meg kell hogy előzze a jelentés-egyértelműsítést. Az automatikus jelentés-egyértelműsítést segíthetik a **morfoszintaktikai információk**, továbbá a **jelentések számának csökkenése**, illetve a **pontosabb jelentésdefiníciók** megkönnyíthetik mind a kézi, mind a gépi jelentés-egyértelműsítést.

## 6.3. Félig kompozicionális szerkezetek az információkinyerésben és -vissza-keresésben

Az információkinyerés három részterületén mutattuk be a félig kompozicionális szerkezetek kezelését: szemantikuskeret-illesztés, szemantikai szerepek azonosítása és modalitás azonosítása. A félig kompozicionális szerkezetet tartalmazó szemantikai keretek esetében javaslatot tettünk az eredetinél (Farkas et al., 2004) nagyobb lefedettséget nyújtó, ugyanakkor tömörebb reprezentációra, amely **lexikai függvényeket** és **wordnetbeli szinonimahalmazokat** is magában foglal. A félig kompozicionális szerkezeteket **komplex predikátumként** kezelő szintaktikai elemzés

pozitív hatásait a **szemantikai szerepek és a modalitás gépi azonosítása** területén mutattuk meg. A félig kompozicionális szerkezeteket és igei párjaikat tartalmazó **kétnyelvű listánk** pedig javítani képes a (nyelvközi) **információ-visszakereső** rendszerek fedését, illetve a magyar és angol wordnetek bővítésében is fontos szerephez juthat.

#### 6.4. Félig kompozicionális szerkezetek a gépi fordításban

A teljes kompozicionalitás hiánya nehézségeket okoz a többszavas kifejezések, konkrétan a félig kompozicionális szerkezetek fordításában (Dura és Gawrońska, 2005). **Lexikai függvények** segítségével ugyanakkor lehetséges formalizálni a félig kompozicionális szerkezetek főnévi és igei komponensének viszonyát, továbbá **az ige és főnév szemantikai típusa között összefüggések** figyelhetők meg, melyek hasznosíthatók a gépi fordításban és a szótárépítésben. Minthogy a gépi fordítási módszerek nagyméretű adatbázisokkal dolgoznak, amelyekben szerepelniük kell a többszavas kifejezéseknek is, a félig kompozicionális szerkezeteket és igei variánsaikat tartalmazó **kétnyelvű adatbázisunk** könnyen beépíthető a rendszerbe. A félig kompozicionális szerkezetek egyetlen lexikai egységként való reprezentációja szintén hasznosnak bizonyulhat a gépi fordításban. Végül kiemeltük, hogy a statisztikai gépi fordító alkalmazások a **SzegedParalellFX** párhuzamos korpuszból is profitálhatnak, amelyben angolul és magyarul is annotálva vannak a félig kompozicionális szerkezetek.

### 7. A kontrasztív vizsgálatok eredményei

Az értekezésben különös figyelmet fordítottunk arra is, hogy a magyarra elért eredményeinket összevessük a más nyelvekre leírt eredményekkel. Alapvetően ugyanaz a **tesztesyüttes** alkalmazható a magyar főnév + ige szerkezetekre, mint az angolra, ami azt jelenti, hogy az angol nyelv ilyen szerkezeteire ugyanolyan skála jellemző, mint a magyarra. A különféle nyelvű félig kompozicionális szerkezetek **alapvető szintaktikai jellemzőikben** megegyeznek, továbbá a vizsgált nyelvek mindegyikében kimutathatók **szemantikai összefüggések a főnév és az ige szemantikai típusa között** (például a beszédaktust jelölő főnevek többnyire az orosz *davat* 'ad' és a magyar *ad* igével társulnak). Eltérést jelenthet ugyanakkor, hogy a szerkezetek idegen

nyelvű megfelelője időnként **más formájú vonzatokkal vagy más aspektussal/akcióminőség-gel** rendelkezik.

A számítógépes nyelvészeti eredményeket tekintve a félig kompozicionális szerkezetek **automatikus azonosítására** kidolgozott módszer **csak pontosításokkal emelhető át** más nyelvekre, hiszen a magyarban és más agglutináló nyelvekben ugyanaz a szerkezet számos formában megjelenhet, ami kevésbé jellemző a szegényebb morfológiájú nyelvekre (l. angol). Ugyanakkor a **jelentés-egyértelműsítés** és **információkinyerés** terén **nem mutatkozik nagy eltérés** az angol és magyar nyelvre való alkalmazhatóság között. A **gépi fordítás** területén pedig kifejezetten a **nyelvek közti hasonlóságok és különbségek** rendszeres feltérképezése vezethet eredményre.

## 8. Összefoglalás és jövőbeli tervek

Az értekezésben félig kompozicionális szerkezeteket elemeztünk elméleti és számítógépes nyelvészeti szempontok alapján. Az értekezés fő célja az volt, hogy az elért elméleti eredményeket a lehető legnagyobb mértékben alkalmazzuk a gyakorlatban, konkrétan a számítógépes nyelvészetben. Az elméleti eredmények és a számítógépes nyelvészeti területek kapcsolatát a következőkben foglalhatjuk össze:

- minthogy a félig kompozicionális szerkezetek a produktív szerkezetek és idiómák között helyezkednek el, automatikus azonosításuk nem történhet pusztán szintaktikai alapokon;
- az igei tőből, pontosabban a szerkezet igei párjával azonos tőből származó főnevek azonosítása megkönnyíti a félig kompozicionális szerkezetek automatikus azonosítását;
- az angol és magyar félig kompozicionális szerkezetek, valamint igei párjaik összevetéséből származó nyelvek közti különbségek a gépi fordításban hasznosíthatók;
- a félig kompozicionális szerkezeteket és igei párjaikat tartalmazó kétnyelvű lista növelheti a gépi fordítórendszerek és információ-visszakereső rendszerek hatékonyságát;
- a félig kompozicionális szerkezetek komplex predikátumot feltételező szintaktikai elemzése hasznosítható a szemantikus szerepek és a modalitás gépi azonosításában;

	azonosítás	WSD	IE	IR	MT
státus	•				
igei párok	•			•	•
szintaxis			•		
szemantika		•			•
lexikai reprezentáció		•	•		•

1. táblázat. Az értekezés témakörei közti kapcsolatok

- a félig kompozicionális szerkezetek két komponense közti szemantikai viszonyokat lexikai függvények segítségével lehet formalizálni, ami a gépi fordítórendszerek szótári adatbázisának építése során játszhat fontos szerepet;
- a főnév és az ige szemantikai típusa közti korrelációk a jelentés-egyértelműsítésre vannak hatással;
- a félig kompozicionális szerkezetek lexikai reprezentációjára javasolt módszerek integrálhatók a szemantikuskeret-illesztő, jelenés-egyértelműsítő és gépi fordító alkalmazásokba.

Az értekezés elméleti nyelvészeti és számítógépes nyelvészeti témaköreinek összefonódását szemlélteti az 1. táblázat.

A félig kompozicionális szerkezetek elméleti szempontú vizsgálata arra is rámutatott, hogy az eltérő nyelvi szinteken (nevezetesen a szintaxis és a szemantika szintjén) nem feltétlenül esnek egybe a nyelvi egységek: a *döntést hoz* kifejezés szintaktikailag részekre bontható (rá lehet kérdezni a főnévi komponensre, illetve szerepelhet önálló válaszként), azonban szemantikailag ugyanazt a jelentést hordozza, mint a szintaktikai részekre nem bontható *dönt*. A **szemantikai** és **szintaktikai** aktáns fogalmának elkülönítése az értelem-szöveg elméletben (Mel'čuk, 2004a; Mel'čuk, 2004b) is azt jelzi, hogy a szintaktikai és szemantikai egységek nem feltétlenül feleltethetők meg egymásnak egy az egyben, azaz nem feltételezhetünk homomorfizmust a szintaxis és szemantika között. A **félkompozicionalitás** fent megfogalmazott definíciója, továbbá **a szintaktikai és szemantikai egységek megkülönböztetése** maga után vonja, hogy a félig kompozicionális szerkezeteknek kettős természetük van: bizonyos vonatkozásokban egységként, máskor pedig két részből álló elemként viselkednek. Amennyiben viszont a szerkezetek

elméleti státusa kettősséget mutat, a dolgozat központi céljának megfelelően ennek a gyakorlatban is tükröződnie kell. A félig kompozicionális szerkezetek számítógépes nyelvészeti kezelési módjainak bemutatásakor ezért arra törekedtünk, hogy ahol lehetséges volt, mindkét – az egységen vagy részeken alapuló – megoldást ismertessük.

E kettősség különféle módokon érhető tetten az elméleti és számítógépes nyelvészeti területeken. A szerkezetek egy igei és egy főnévi komponensből állnak, ami a következőkre van hatással:

- a két rész viszonya határozza meg a szerkezet félproduktivitását és félkompozicionalitását;
- a főnév módosíthatósága arra utal, hogy a szerkezetnek van belső szintaktikai szerkezete;
- az ige és a főnév között szemantikai korrelációk figyelhetők meg;
- más félig kompozicionális szerkezetekkel konverzíva viszonyban állhatnak (az igei komponensek jelentését figyelembe véve);
- a lexikonban megjelenhetnek a főnév, az ige vagy mindkettő szócikkében;
- automatikus azonosításuk a főnév + ige párok megtalálásán alapszik;
- a jelentés-egyértelműsítésben lehetséges az igeinek speciális funkcióigei jelentést tulajdonítani;
- a transzferalapú gépi fordításban szabályok segítségével feleltetik meg egymásnak a nyelvenként eltérő szintaktikai szerkezettel rendelkező félig kompozicionális szerkezeteket.

A következő pontokban a szerkezet egysége tükröződik:

- rendelkezhetnek hasonló jelentésű igei párral;
- a szerkezet és igei párjának vonzatkerete eltérhet egymástól;
- a szerkezet és igei párja között lehetnek különbségek aspektus és akcióminőség terén;
- komplex predikátumoknak tekinthetők;



- más félig kompozicionális szerkezetek szinonimái lehetnek (egységként tekintve őket azonos a jelentésük);
- a lexikonban egy jelentéses egységként is megjelenhetnek;
- a jelentés-egyértelműsítésben egy egységként is lehet őket szemlélni;
- a szemantikuskeret-illesztésben, szemantikus szerepek gépi azonosításában és a modalitás gépi azonosításában egy szintaktikai egységként (komplex predikátum) érdemes őket kezelni;
- az igei párjukkal való variativitást ki lehet aknázni az információ-visszakeresésben;
- a statisztikai gépi fordítás egy egységként tekint rájuk.

A fentiek összefoglalása a 2. táblázatban látható.

A jövőben mind az elméleti, mind a számítógépes nyelvészeti kutatásokat szeretnénk tovább folytatni. Például a félig kompozicionális szerkezetek adatbázisát lexikai függvények segítségével tudjuk formalizálni. A főnév és ige közti lexikai szemantikai relációk vizsgálatát az értekezésben elemzett négy igt tartalmazókon kívül más szerkezetekre is ki lehet terjeszteni. A funkcióigék különféle jelentéseit érdemes lenne feltérképezni különféle félig kompozicionális szerkezetekben, ezáltal segítve a lexikológiai munkálatokat, továbbá olyan szótárak és lexikonok létrehozása is kívánatos, ahol a félig kompozicionális szerkezetek az ige és a főnév szócikében is előfordulnak. Ami a számítógépes nyelvészeti kutatásokat illeti, a félig kompozicionális szerkezetek azonosítását célzó algoritmusok tovább fejleszthetők, illetőleg más nyelvekre adaptálhatók. E cél eléréséhez további nyelvi erőforrások – annotált korpuszok és adatbázisok – létrehozása szükséges. A jelentés-egyértelműsítés, információkinyerés és -visszakeresés, illetve a gépi fordítás mellett más számítógépes nyelvészeti területeken is hasznos lenne kidolgozni a félig kompozicionális szerkezetek hatékony kezelési módjait. Reményeink szerint ezeket a feladatokat hamarosan elvégezzük a Szegedi Tudományegyetem Mesterséges Intelligencia Kutatócsoportjában.

	részek	egység
<b>Státus</b>		
félproduktivitás	•	
félkompozicionalitás	•	
<b>Igei pár</b>		
vonzatkeretbeli eltérések		•
akcióminőség/aspektusbeli eltérések		•
<b>Szintaxis</b>		
a főnév módosíthatósága	•	
komplex predikátum		•
<b>Szemantika</b>		
összefüggések az ige és főnév között	•	
szinonímia		•
konverzió	•	
<b>Lexikai reprezentáció</b>		
ige szócikke	•	
főnév szócikke	•	
önálló szócikk		•
ige és főnév szócikke	•	
<b>Automatikus azonosítás</b>	•	
<b>Jelentés-egyértelműsítés</b>		
egy jelentés lexikai egységként		•
funkcióigei jelentés	•	
<b>Információkinyerés</b>		
szemantikuseret-illesztés		•
szemantikai szerepek azonosítása		•
modalitás azonosítása		•
<b>Információ-visszakeresés</b>		•
<b>Gépi fordítás</b>		
transzferalapú módszerek	•	
statisztikai módszerek		•

2. táblázat. A félig kompozicionális szerkezetek kettős természete és az értekezés témakörei

## Hivatkozások

- Alonso Ramos, Margarita. 1998. *Etude sémantico-syntaxique des constructions à verbe support*. Doktori értekezés, Université de Montréal, Montreal, Kanada.
- Bannard, Colin. 2007. A measure of syntactic flexibility for automatically identifying multiword expressions in corpora. In *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions*, MWE '07, pp. 1–8, Morristown, NJ, USA. Association for Computational Linguistics.
- Bárczi, Géza; Ország, László (szerk.). 1959–1962. *A magyar nyelv értelmező szótára*. Akadémiai Kiadó, Budapest.
- Chomsky, Noam. 1995. *The Minimalist Program*. Current studies in linguistics. MIT Press, Cambridge, MA.
- Cook, Paul; Fazly, Afsaneh; Stevenson, Suzanne. 2007. Pulling their weight: exploiting syntactic forms for the automatic identification of idiomatic expressions in context. In *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions*, MWE '07, pp. 41–48, Morristown, NJ, USA. Association for Computational Linguistics.
- Dobos, Csilla. 1991. *Leíró kifejezések az orosz jogi szaknyelvben*. Doktori értekezés, Debreceni Egyetem, Debrecen.
- Dobos, Csilla. 2001. *A funkcióigés szerkezetek vizsgálata (különös tekintettel az orosz jogi szaknyelvre)*. Doktori értekezés, Debreceni Egyetem, Debrecen.
- Dura, Elżbieta; Gawrońska, Barbara. 2005. Towards Automatic Translation of Support Verbs Constructions: the Case of Polish *robić/zrobić* and Swedish *göra*. In *Proceedings of the 2nd Language & Technology Conference*, pp. 450–454, Poznań, Lengyelország. Wydawnictwo Poznańskie Sp. z o.o.
- Farkas, Richárd; Konczer, Kinga; Szarvas, György. 2004. Szemantikuseret illesztés és az IE-rendszer automatikus kiértékelése. In Alexin, Zoltán; Csendes, Dóra (szerk.), *II. Magyar Számítógépes Nyelvészeti Konferencia*, pp. 49–53, Szeged. Szegedi Tudományegyetem.
- Grétsy, László; Kemény, Gábor. 1996. *Nyelvművelő kéziszótár*. Auktor Könyvkiadó, Budapest.
- Hale, Kenneth; Keyser, Samuel J. 2002. *Prolegomenon to a Theory of Argument Structure*. MIT Press, Cambridge.
- Hale, Bob. 1997. Grundlagen section 64. *Proceedings of the Aristotelian Society*, XCVII:243–261.
- Heltai, Pál; Gósy, Mária. 2005. A terpeszkedő szerkezetek hatása a feldolgozásra. *Magyar Nyelvőr*, 129:470–487.
- Keszler, Borbála. 1992. A mai magyar nyelv szófaji rendszere. In Kozocsa, Sándor Géza; Laczkó, Krisztina (szerk.), *Emlékkönyv Rácz Endre hetvenedik születésnapjára*, pp. 131–139, Budapest.

- Kiefer, Ferenc. 1990–1991. Noun Incorporation in Hungarian. *Acta Linguistica Hungarica*, 40(1–2):149–177.
- Kiefer, Ferenc. 2007. *Jelentélemélet*. Corvina, Budapest.
- Komlósy, András. 1992. Régensek és vonzatok. In Kiefer, Ferenc (szerk.), *Strukturális magyar nyelvtan. I. Mondattan*, pp. 299–527, Budapest. Akadémiai Kiadó.
- Langer, Stefan. 2004. A Linguistic Test Battery for Support Verb Constructions. *Linguisticae Investigationes*, 27(2):171–184.
- Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry*, 19:335–391.
- Mel’čuk, Igor; Clas, André; Polguère, Alain. 1995. *Introduction à lexicologie explicative et combinatoire*. Duculot, Louvain-la-Neuve, Franciaország.
- Mel’čuk, Igor; mtsai. 1984–1999. *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexico-sémantiques I–IV*. Presses de l’Université de Montréal, Montreal, Kanada.
- Mel’čuk, Igor; Žolkovskij, Aleksander. 1984. *Explanatory Combinatorial Dictionary of Modern Russian*. Wiener Slawistischer Almanach, Bécs, Ausztria.
- Mel’čuk, Igor. 2004a. Actants in semantics and syntax I: Actants in semantics. *Linguistics*, 42(1):1–66.
- Mel’čuk, Igor. 2004b. Actants in semantics and syntax II: Actants in syntax. *Linguistics*, 42(2):247–291.
- Meyers, Adam; Reeves, Ruth; Macleod, Catherine. 2004. NP-External Arguments: A Study of Argument Sharing in English. In Tanaka, Takaaki; Villavicencio, Aline; Bond, Francis; Korhonen, Anna (szerk.), *Second ACL Workshop on Multiword Expressions: Integrating Processing*, pp. 96–103, Barcelona. Association for Computational Linguistics.
- Newson, Mark; Hordós, Marianna; Pap, Dániel; Szécsényi, Krisztina; Tóth, Gabriella; Vincze, Veronika. 2006. *Basic English Syntax With Exercises*. Bölcsész Konzorcium, Budapest.
- Pusztai, Ferenc (szerk.). 2003. *Magyar értelmező kéziszótár*. Akadémiai Kiadó, Budapest.
- Sag, Ivan A.; Baldwin, Timothy; Bond, Francis; Copestake, Ann; Flickinger, Dan. 2002. Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, pp. 1–15, Mexikóváros, Mexikó.
- Sziklai, Lászlóné. 1986. Terpeszkednek vagy körülírnak? *Magyar Nyelvőr*, 110:268–273.
- Tan, Yee Fan; Kan, Min-Yen; Cui, Hang. 2006. Extending corpus-based identification of light verb constructions using a supervised learning framework. In *Proceedings of the EACL Workshop on Multi-Word Expressions in a Multilingual Contexts*, pp. 49–56, Trento, Olaszország. Association for Computational Linguistics.
- Tu, Yuancheng; Roth, Dan. 2011. Learning English Light Verb Constructions: Contextual or Statistical. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, pp. 31–39, Portland, Oregon, USA. Association for Computational Linguistics.

- Vincze, Veronika; Felvégi, Zsuzsanna; R. Tóth, Krisztina. 2010. Félig kompozicionális szerkezetek a SzegedParalell angol–magyar párhuzamos korpuszban. In Tanács, Attila; Vincze, Veronika (szerk.), *MSzNy 2010 – VII. Magyar Számítógépes Nyelvészeti Konferencia*, pp. 91–101, Szeged. Szegedi Tudományegyetem.
- Vincze, Veronika; Nagy T., István; Berend, Gábor. 2011. Multiword expressions and named entities in the Wiki50 corpus. In *Proceedings of RANLP 2011*, Hissar, Bulgária.
- Vincze, Veronika; Csirik, János. 2010. Hungarian corpus of light verb constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 1110–1118, Peking, Kína. Coling 2010 Organizing Committee.

## Az értekezés témájához kapcsolódó publikációk

1. Vincze Veronika 2007. A félig kompozicionális szerkezetek gépi fordításainak lehetőségéről. In: Váradi Tamás (szerk.): *I. Alkalmazott Nyelvészeti Doktorandusz Konferencia*. Budapest, MTA Nyelvtudományi Intézet, pp. 207–218.
2. Vincze Veronika 2008. A puszta köznév + ige komplexumok státusáról. In: Sinkovics Balázs (szerk.): *LingDok 7. Nyelvész-doktoranduszok dolgozatai*. Szeged, Szegedi Tudományegyetem, pp. 279–297.
3. Vincze Veronika, Szarvas György, Almási Attila, Szauter Dóra, Ormándi Róbert, Farkas Richárd, Hatvani Csaba, Csirik János 2008. Hungarian word-sense disambiguated corpus. In: *Proceedings of 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Marrakech, Marokkó.
4. Vincze Veronika 2009. On the Machine Translatability of Semi-Compositional Constructions. In: Váradi Tamás (szerk.): *Válogatás az I. Alkalmazott Nyelvészeti Doktorandusz Konferencia előadásaiból – Selected Papers from the 1st Applied Linguistics PhD Conference*. Budapest, MTA Nyelvtudományi Intézet, pp. 166–178.
5. Vincze Veronika 2009. Angol–magyar főnév + ige szerkezetek és igei párjaik. In: Váradi Tamás (szerk.): *II. Alkalmazott Nyelvészeti Doktorandusz Konferencia*. Budapest, MTA Nyelvtudományi Intézet, pp. 112–122.
6. Vincze Veronika 2009. Félig kompozicionális szerkezetek a Szeged Korpuszban. In: Tanács Attila, Szauter Dóra, Vincze Veronika (szerk.): *VI. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, Szegedi Tudományegyetem, pp. 390–393.
7. Vincze Veronika 2009. *Előadást tart* vs. *előad*: főnév + ige szerkezetek igei variánsai. In: Sinkovics Balázs (szerk.): *LingDok 8. Nyelvész-doktoranduszok dolgozatai*. Szeged, Szegedi Tudományegyetem, pp. 265–278.
8. Vincze Veronika 2009. Főnév + ige szerkezetek a szótárban. In: Váradi Tamás (szerk.): *III. Alkalmazott Nyelvészeti Doktorandusz Konferencia*. Budapest, MTA Nyelvtudományi

- Intézet, pp. 180–188.
9. Vincze Veronika 2010. Félig kompozicionális főnév + ige szerkezetek a számítógépes nyelvészetben. In: Gecső Tamás, Sárdi Csilla (szerk.): *Új módszerek az alkalmazott nyelvészeti kutatásban*. Budapest, Tinta Könyvkiadó, pp. 327–332.
  10. Vincze Veronika, Csirik János 2010. Hungarian corpus of light verb constructions. In: *Proceedings of COLING 2010*, Peking, Kína, pp. 1110–1118.
  11. Vincze Veronika, Felvégi Zsuzsanna, R. Tóth Krisztina 2010. Félig kompozicionális szerkezetek a SzegedParalell angol–magyar párhuzamos korpuszban. In: Tanács Attila, Vincze Veronika (szerk.): *MSzNy 2010 – VII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, Szegedi Tudományegyetem, pp. 91–101.
  12. Vincze Veronika 2011. Mi fán terem a főnév + ige szerkezet?. In: Gécseg Zsuzsanna (szerk.): *LingDok 10. Nyelvész-doktoranduszok dolgozatai*. Szeged, Szegedi Tudományegyetem, pp. 225–243.
  13. Vincze Veronika, Nagy T. István, Berend Gábor 2011. Detecting noun compounds and light verb constructions: a contrastive study. In: *Proceedings of the ACL Workshop on Multiword Expressions: from Parsing and Generation to the Real World*. Portland, Oregon, USA, Association for Computational Linguistics.
  14. Nagy T. István, Vincze Veronika, Berend Gábor 2011. Domain-dependent identification of multiword expressions. In: *Proceedings of Recent Advances in Natural Language Processing (RANLP 2011)*. Hissar, Bulgária.
  15. Vincze Veronika, Nagy T. István, Berend Gábor 2011. Multiword expressions and named entities in the Wiki50 corpus. In: *Proceedings of Recent Advances in Natural Language Processing (RANLP 2011)*. Hissar, Bulgária.

	ALKNYELVDOK1 2007	LINGDOK7 2008	LREC 2008	ALKNYELVDOK2 2009	LINGDOK8 2009	MSZNY 2009	ALKNYELVDOK3 2009	APPLINGPHD 2009	ALKNYELV 2010	COLING 2010	MSZNY 2010	LINGDOK10 2011	MWE 2011	RANLP 2011a	RANLP 2011b
Korpuszok						•				•	•			•	
Státus		•													
Igei párok				•	•										
Szintaxis												•			
Szemantika	•							•							
Lexikai reprezentáció							•								
Automatikus azonosítás									•	•			•		•
Jelentés-egyértelműsítés			•						•						
Információkinyerés és -visszakeresés									•						
Gépi fordítás	•							•	•						

3. táblázat. A szerző publikációinak és az értekezés témáinak kapcsolata

## Az értekezés témájához kapcsolódó konferencia-előadások

1. Funkcióigés szerkezetek vizsgálata lexikai függvények segítségével. Nyelvészdoktoranduszok 9. Országos Konferenciája, Szeged, 2005. november 18.
2. A funkcióigés szerkezetekről. Kari Doktori Konferencia, Szegedi Tudományegyetem, Bölcsészettudományi Kar, 2006. június 21.
3. A főnév + ige szerkezetek státusáról. Nyelvészdoktoranduszok 10. Országos Konferenciája, Szeged, 2006. december 1.
4. A félig kompozicionális szerkezetek gépi fordításainak lehetőségeiről. I. Alkalmazott Nyelvészeti Doktorandusz Konferencia, Budapest, 2007. február 2.
5. Félig kompozicionális szerkezetek és kollokációs láncok. Kari Doktori Konferencia, Szegedi Tudományegyetem, Bölcsészettudományi Kar, Szeged, 2007. június 20.
6. *Előadást tart vs. előad*: főnév + ige szerkezetek igei variánsai. Nyelvészdoktoranduszok



11. Országos Konferenciája, Szeged, 2007. december 6.
7. Angol–magyar főnév + ige szerkezetek és igei párjaik. II. Alkalmazott Nyelvészeti Doktorandusz Konferencia, Budapest, 2008. február 8.
8. (Szarvas Györggyel, Almási Attilával, Szauter Dórával, Ormándi Róberttel, Farkas Richárdal, Hatvani Csabával és Csirik Jánossal) Hungarian word-sense disambiguated corpus. 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Marokkó, 2008. május 30.
9. A főnév + ige szerkezetek lexikai reprezentációjáról. Nyelvészdoktoranduszok 12. Országos Konferenciája, Szeged, 2008. december 3.
10. Főnév + ige szerkezetek a szótárban. III. Alkalmazott Nyelvészeti Doktorandusz Konferencia, Budapest, 2009. február 6.
11. Félig kompozicionális főnév + ige szerkezetek a számítógépes nyelvészetben. Új módszerek az alkalmazott nyelvészeti kutatásban, Budapest, 2009. október 27.
12. Mi fán terem a főnév + ige szerkezet? Nyelvészdoktoranduszok 13. Országos Konferenciája, Szeged, 2009. november 25.
13. Félig kompozicionális főnév + ige szerkezetek a Szeged Korpuszban. VI. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, 2009. december 3.
14. (Csirik Jánossal) Hungarian corpus of light verb constructions. COLING 2010, Peking, Kína, 2010. augusztus 24.
15. (Felvégi Zsuzsannával és R. Tóth Krisztinával) Félig kompozicionális szerkezetek a SzegedParalell angol–magyar párhuzamos korpuszban. VII. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, 2010. december 2.
16. (Nagy T. Istvánnal és Berend Gáborral) Detecting noun compounds and light verb constructions: a contrastive study. ACL Workshop on Multiword Expressions: from Parsing and Generation to the Real World, Portland, Oregon, USA, 2011. június 23.

17. (Nagy T. Istvánnal és Berend Gáborral) Domain-dependent identification of multiword expressions. Recent Advances in Natural Language Processing (RANLP 2011). Hissar, Bulgária, 2011. szeptember 12.
18. (Nagy T. Istvánnal és Berend Gáborral) Multiword expressions and named entities in the Wiki50 corpus. Recent Advances in Natural Language Processing (RANLP 2011). Hissar, Bulgária, 2011. szeptember 12.