

## Zárójelentés a K 72983 számú OTKA-projektről

Jelentésünket az éves időközi szakmai beszámolókra alapozva, azok lényegi részeit kiemelve és röviden kommentálva állítjuk össze: kutatási tevékenységünk és eredményeink bemutatását évekre lebontva, kronologikusan haladva végezzük el.

Előljáróban néhány általános megjegyzést teszünk.

1. Az összes időszaki beszámolónk értékelése egyöntetűen pozitív volt. Egyetlen észrevétel fogalmazott meg fenntartást egy olyan treebank létrehozásával kapcsolatban, amely nem szerepelt az eredeti tervek között. Amint azt a beszámoló megfelelő részében kifejtjük, úgy látjuk, hogy a végeredmény igazolja a döntésünket (vagyis a plusz vállalásunkat): a sikeresen kifejlesztett treebank folyamatosan hozzájárult a projekt célkitűzéseinek a minél komplexebb megvalósításához, valamint az elért eredmények gyakorlati hasznosíthatóságához, ráadásul komoly elismerést váltott ki hazai és nemzetközi szakmai körökben.
2. A projekt nemzetközi beágyazottsága és elismertsége az eredeti terveinket, várakozásainkat messze felülmúlta. A *ParGram* (*Parallel Grammar* 'Párhuzamos Nyelvtan') elnevezésű immáron húsz éves, egyre szélesedő nemzetközi együttműködésbe a mi HunGram projektünk szervesen illeszkedett bele. Az OTKA támogatásnak köszönhetően majdnem mindig részt tudtunk venni az évi egy vagy két ParGram műhelykonferencián, azokon rendszeresen beszámoltunk a projektünk aktuális eredményeiről, kihívásairól és terveiről, nagyon hasznos szakmai együttműködések alakítottunk ki (konzultációk, kölcsönös „segítségnyújtás” és közös tudományos közlemények formájában), és nem kis megtiszteltetésként 2011-ben is és 2013-ban is mi adhattunk otthont a soron következő műhelytalálkozónak.
3. Alább azt is megmutatjuk, hogy részben szintén az OTKA támogatásnak köszönhetően további két jelentős nemzetközi rendezvényt is szerveztünk: egy műhelykonferenciát 2012-ben és az éves LFG konferenciát az idén. Mindkettőnek, csakúgy mint a két ParGram-es rendezvényünknek kiemelkedően jó volt a visszhangja. Következésképpen tudományszervezői tevékenységünk is jóval felülmúlta minden előzetes tervünket, várakozásunkat.
4. Idézetek két időszaki beszámoló véleményezéséből. „A pályázat szervezeten és eredményesen folyik. Szoros szakmai együttműködés van a résztvevők között. Az eredményeket rangos fórumokon folyamatosan publikálják.” (2009) „A részjelentésben ismertett eredmények egyértelműen bizonyítják, hogy a munkálatok a tervezettek szerint haladnak és igen gazdag, publikációkban és nemzetközi tudományos konferenciákon tartott előadásokban megmutatkozó eredményeket hoztak.” (2011) A projekt lezárultakor úgy véljük, hogy joggal megállapítható, hogy az iménti két idézet összességében és együttesen a projekt teljes futamidejére vonatkoztatható. Az eredeti tervben szereplő vállalásokat tekintélyes mértékben meghaladó számú és minőségű fórumon ismertettük tudományos eredményeinket. Az ötéves ciklus alatt 41 konferenciaelőadást tartottunk, és 42 olyan tudományos közleményünk jelent meg, amelyben az OTKÁ-ra való hivatkozás szerepel.
5. Külön örömünkre szolgál az, hogy egy-egy résztéma erejéig az Angol Nyelvészeti Tanszéken dolgozó kollégák közül többen is bekapcsolódtak a munkába, illetve az, hogy rátermett PhD-s hallgatók is érdemben hozzájárultak a kutatási tevékenységünk eredményességéhez.

## 2009

elméleti kutatás	implementáció
eseménynominalizáció	igen
elliptikus főnévi csoportok	igen
relációs főnevek a magyarban és az angolban	nem
birtokviszony a főnévi csoportban	igen
a melléknévi, határozószói és névutói csoportok alapszerkezete	igen
az ige utáni tartomány alapszerkezete	igen
koreferenciaviszonyok kódolása a magyarban	nem

<b>további implementációs tevékenység</b>	az implementációs platform (XLE) installálása, használatba vétele
	szimbólum-manipulációs eljárás vs. konnekciónizmus az implementációban: elvi és gyakorlati kérdések
	a lexikai többértelműség implementációs kezelésének lehetőségei és korlátai

## 2010

elméleti kutatás	implementáció
magyar igekötős igék és angol, német megfelelőik elméleti problémái	igen
a magyar relációs főnevek argumentumszerkezete	igen
a birtokviszonynak egy új LFG-s kezelése (közönséges főnevek, képzetlen relációs főnevek, és esemény jelentésű deverbális főnevek esetében)	nem
a magyar és az angol nyelv diskurzusfunkcióinak LFG-beli reprezentációs problematikája	igen
a kauzalitásnak az igei vonzatkeret meghatározásában játszott szerepe	nem

<b>további implementációs tevékenység</b>	morfológiai elemző tesztelése, installálása
-------------------------------------------	---------------------------------------------

## 2011

elméleti kutatás	implementáció
igemódosítók, igekötők, fókuszált kifejezések viselkedése	igen
a magyar morfológiailag összetett reflexív elemek nyelvtana	igen
a koreferenciaviszony PP-típusú összetevőkön belüli jelölési módjai	igen
diskurzusjelölők a mondat bal periferiáján	nem

<b>további implementációs tevékenység</b>	a magyar egyszerű mondat bal periferiájának implementációja: kvantorok, mondathatározók és topikok
	1,5 millió szavas tree-bank létrehozásának előkészítő munkálatai, rendszerterv
	névelemek LFG alapú kezelésének implementációs vonatkozásai
	a hálózatos (konnekcionista) nyelvmodellezés hatása a lexikonról kialakult elképzeléseinkre

<b>tudományszervezés</b>	<i>ParGram/ParSem Meeting</i> <a href="http://hungram.unideb.hu/ParGram_ParSem_2011/">http://hungram.unideb.hu/ParGram_ParSem_2011/</a>
	<i>Workshop on Spatial and Temporal Relations in LFG</i> <a href="http://hungram.unideb.hu/LFG_Workshop_2011/">http://hungram.unideb.hu/LFG_Workshop_2011/</a>

## 2012

elméleti kutatás	implementáció
igekötős igék	igen
a mondat bal perifériája	igen

<b>további implementációs tevékenység</b>	az összetett mondatok és a nem véges tagmondatok implementációs nyelvtanának alapvetése
	a HunGram2 kidolgozása és tesztelése a HG-1 Treebank számára
	a treebank fejlesztése és teljes kiépítése (a kiinduló korpusz összeállítása, az adatbázis (modell) megtervezése, implementálása, valamint a felhasználói interfész kialakítása, programozási és tesztelési feladatok, a nyelvtan különböző változatainak futtatása szuperszámítógépen)
	hálózatos (konnekcionista) nyelvmodellezésre irányuló vizsgálatok, számítógépes kísérletek

<b>tudományszervezés</b>	<i>Workshop on Argument Structure</i> <a href="http://was.unideb.hu/">http://was.unideb.hu/</a>
--------------------------	----------------------------------------------------------------------------------------------------

## 2013

elméleti kutatás	implementáció
létigés mondatok a magyarban	igen
argumentum-szerkezet, argumentum vs. adjunktum	igen
az igekötős igék és a szóképzés	igen
az egyszerű mondat szerkezete – átfogó, összegző vizsgálatok	igen

<b>„független” implementációs tevékenység</b>	a HunGram1 továbbfejlesztése különös tekintettel a ParGramBank együttműködésben való részvételre
	részvétel egy 50 mondatos ParGram-es parallel treebank (ParGramBank) kifejlesztésében

<b>tudományszervezés</b>	<i>18th Lexical-Functional Grammar Conference</i> <a href="http://lfg13.unideb.hu/">http://lfg13.unideb.hu/</a>
	<i>ParGram/ParSem Meeting</i> <a href="http://hungram.unideb.hu/ParGram_ParSem_2013/">http://hungram.unideb.hu/ParGram_ParSem_2013/</a>

- **Megjegyzések a HG-1 Treebankról**

Kutatómunkánkhoz kapcsolódóan létrehoztunk egy 1,5 millió szót (több, mint 280 000 magyar mondatot) tartalmazó, morfológiai és mondattani annotációval ellátott szövegtörzset (HG-1 Treebank). A treebank közvetlenül támaszkodik a pályázatban vállalt nyelvészeti kutatómunkára és annak a – szintén felvállalt – számítógépes implementációjára. Ehhez a projektemhez a szükséges erőforrásokat részben a Debreceni Egyetem TÁMOP-4.2.1/B-09/1/KONV-2010-0007 pályázata biztosította.

A mondatok elemzését az NIIF szuperszámítógépes szolgáltatására támaszkodva végeztük el az általunk erre a célra kifejlesztett HunGram2 XLE nyelvtan révén. Nyersanyagként a Magyar Webkorpuszt használtuk, mely magyar weblapokról gyűjtött, elemzés nélküli szövegeket tartalmaz. Az elemzőnk által előállított kimenetet feldolgozva a mondattani fák

tárolását a Tiger-XML dokumentumleíró-nyelv segítségével oldottuk meg. A projekt website-ja (<http://corpus.hungram.unideb.hu/>) tartalmaz egy on-line lekérdezési felületet, melynek segítségével minden felhasználó szabadon kereshet szóra vagy lemmára, valamint lehetséges a keresés szűrése morfológiai jegyekre és a keresett szót tartalmazó összetevőre (szűrés beállítása űrlap segítségével). Az online felületen megtekinthetők a keresési feltételeknek megfelelő mondatok, valamint elérhető a találati listából kiválasztott mondatelemzés ágrajza. A treebankhez tartozó honlapon ugyancsak publikáltuk a treebankról szóló alapvető információkat, valamint egy részletes leírást, útmutatót az elemzésekben található alaktani és mondattani jellemzők értelmezéséhez.

A treebank legfőbb célja és értéke a munkacsoport által kifejlesztett magyar LFG nyelvtan kézzelfoghatóvá és felhasználhatóvá tétele, valamint alkalmazható a nyelvoktatás, nyelvtanulás területén, a korpusz alapú megoldások összes előnyével: motiváló autentikus élőnyelvi szövegekkel dolgozhatunk olyan módon, hogy a tanulás nyelvi felfedezéssé válik. A treebank felhasználható elméleti nyelvészeti kutatásokban, melyre közvetlen példa saját nyelvitanirási projektünk is, amelyhez a treebank folyamatos tesztelési lehetőséget és visszajelzést biztosított. Az pedig különleges hozadéka volt a treebankes „alprojektünknek”, hogy egy alternatív implementációs nyelvtan (a HunGram2) kifejlesztésére készítetett bennünket. Mivel az volt a célunk, hogy a teljesen automatizált folyamat eredményeként a lehető legkevesebb elemzészváltozatot kapjuk meg (csupán az egy vagy (maximum) két elemzést kapó mondatok kerültek be a treebankbe), arra törekedtünk, hogy a nyelvtanunk szabályai minél egyszerűbbek és egyértelműbbek legyenek, vagyis igyekeztünk kiiktatni az elméletileg fennálló, és egy „mélyebb” implementációs nyelvtanban modellezendő két- vagy többértelműséget (például a birtokviszonyt tartalmazó, határozott névelős magyar főnévi csoportokhoz eredendően két fajta szerkezet tartozik, mi viszont olyan szabályt írtunk, amely ezt a kiszámítható kétértelműséget nem vezeti be az elemzésbe). A HunGram1 és a HunGram2 fejlesztése során nagyon sok hasznos, előre mutató nyelvitanirási tapasztalatot szereztünk: bizonyos (jól működő) szabályokat, szabályrészleteket átemeltünk az egyik nyelvtanváltozathoz a másikba, és viszont. Ezeket a tapasztalatokat és eredményeket a későbbi nyelvtanfejlesztések során mindenképpen hasznosítani tudjuk és kívánjuk.

- ***Megjegyzések a ParGramBank-ről***

Ez a párhuzamos treebank nemzetközi (és ParGram-en kívüli) viszonylatban is nívóvannak, rendkívüli vállalkozásnak számít egyrészt az elemzésbe bevont nyelvek tekintélyes száma, másrészt a lehető legnagyobb mértékű egységesítésre, párhuzamosításra való törekvés miatt. Örömmel vettünk részt ebben a fejlesztésben is, hiszen ez tovább növelte a mi projektünk és nyelvtanunk nemzetközi beágyazódását és kompatibilitását is. Az 50 mondat(típus) megfelelő elemzéséhez szükséges nyelvtanváltozat kifejlesztése (a HunGram2 párhuzamos fejlesztése mellett) meglehetősen sok időt és energiát igényelt tőlünk, ezért kevesebb idő és energia jutott a projekt vége felé az összetett mondatokra és a nem véges tagmondatokra (l. a *Záró megjegyzések* 2. pontját). Mindent mérlegelve a *ParGramBank*-ben való együttműködést fontosabbnak ítéltük meg.

### **Záró megjegyzések**

1. A munkatervvel összhangban napirenden tartottuk azt a kérdést, hogy lehetséges-e és érdemes-e nyelvtani szerkezetek elsajátítását kutatnunk konnekcionalista eszközökkel. A nyelvtan gépi tanulással elsajátítása (*grammar induction*) a ParGram együttműködésben nem vált fókuszterületté. Ennek egyik oka az, hogy a megfelelő mennyiségű tanítókörpusz nem áll rendelkezésre: kis méretű (néhány száz ill. ezer mondatos) korpuszok erre a célra nem alkalmasak. A mi munkacsoportunk a ParGram együttműködés keretén belül a

manuálisan implementált nyelvtanra alapozva egy jelentős méretű (1,5 millió szavas) korpuszt állítottunk össze, ami a nyelvtan továbbfejlesztésének függvényében tovább növelhető olyan módon, hogy a gépi tanuláshoz elégséges méretű legyen. Ez esetleg egy későbbi pályázat egyik eleme lehet.

2. Az egyszerű, ragozott igét tartalmazó mondatok elemzésére kifejlesztett nyelvtanunk megfelelő hatásokkal működik. Az összetett mondatok, illetve az igeves szerkezet esetében a nyelvtan még gyakran túl sok (részben helyes, részben helytelen) elemzést ad. Ennek a megszorításához – a jelenségek bonyolultsága miatt – további fejlesztő munkára lesz majd szükség.
3. A pályázatunkban szintén vállalt monográfia-kézirattal kapcsolatban a kutatás előrehaladása során számos elképzelés körvonalazódott. Többször is változott a tartalmi és szerkezeti koncepció, és legújabban a nyelv. A jelenlegi (kiérleltnek és véglegesnek tekinthető) terünk a következő. Tematikusan összeraktuk (de még nem egységesítettük, nem szerkesztettük egybe) a releváns tudományos közleményeinket és prezentációinkat. Három nagy fejezetben, esettanulmány-sorozatban gondolkodunk (a szokásos bevezető és záró fejezeten túlmenően). Az elsőben lennének a kifejezetten elméleti témák (implementációs vetület nélkül), például a relációs főnevek kezelése. A másodikban azok, amelyeket elméleti és implementációs szempontból egyaránt körüljártunk, például az igekötős igék kezelése. A harmadikba kerülnének a kifejezetten számítógépes, implementációs témák, pl. a hálózatos (konnekcionista) nyelvmodellezésre irányuló vizsgálatok, számítógépes kísérletek. Tekintettel arra, hogy a releváns publikációink túlnyomó többsége angol nyelven íródott, illetve arra, hogy kezdettől fogva megkülönböztetett figyelmet fordítottunk a nemzetközi beágyazottságra, eredetileg angol nyelvűnek terveztük a monográfiát. Később azonban (és nem kis mértékben a treebank-es alprojekt sikere után) úgy ítéltük meg, hogy magyarországi viszonylatban sokkal fontosabb és hasznosabb lenne, ha magyar nyelven jelentetnénk meg ezt a kötetet (és majd később kapna egy angol változatot). Ehhez viszont még le kell fordítani magyarra a cikkek zömét.
4. A projektben folytatott munka eredményességét és sikerességét az is jelzi, hogy a három meghatározó résztvevője egyéni kutatói ösztöndíjakat nyert el, amelyek keretében tovább folytathatják kutatómunkájukat (Laczkó Tibor: Hajdú Péter Vendégkutatói Ösztöndíj az MTA Nyelvtudományi Intézetébe, Rákosi György és Tóth Ágoston: Magyary Zoltán Posztdoktori Ösztöndíj).

Debrecen, 2013. augusztus 30.

Dr. Laczkó Tibor  
hab. egyetemi docens, tanszékvezető,  
a K 72983 számú OTKA-projekt vezetője  
Debreceni Egyetem  
Angol Nyelvészeti Tanszék