

Konstituensfák automatikus átalakítása függőségi fákká vagy kézi annotáció?

Simkó Katalin Ilona¹, Vincze Veronika^{1,2}, Szántó Zsolt¹, Farkas Richárd¹

¹Szegedi Tudományegyetem, Informatikai Tanszékcsoport,
Szeged, Árpád tér 2.

kata.simko@gmail.com, {szantozs,rfarkas}@inf.u-szeged.hu

²MTA-SZTE Mesterséges Intelligencia Kutatócsoport,
Szeged, Tisza Lajos körút 103.
vinczev@inf.u-szeged.hu

Kivonat A magyar azon ritka nyelvek egyike, ahol rendelkezésre áll kézzel annotált konstituens és függőségi annotáció is ugyanazon a szövegállományon, a Szeged (Dependencia) Treebanken. Ez lehetővé teszi, hogy megvizsgáljuk a szabályalapú automatikus átalakítás minőségét, valamint hogy összehasonlítsuk az etalon treebankeken tanított konstituens és függőségi elemzők kimenetét a konvertált mondatokon tanított elemzők kimenetével. Eredményeink szerint bár a különböző módszerek szám szerint hasonló eredményeket érnek el, különböző hibákat ejtenek.

Kulcsszavak: szintaxis, konstituens, dependencia, statisztikai szintaktikai elemzés

1. Bevezetés

Manapság a statisztikai szintaktikai elemzésben legelterjedtebb megközelítések konstituens vagy függőségi nyelvtanon alapulnak. Konstituensalapú treebank számos nyelvre létezik, míg a függőségi treebankeket legtöbb esetben automatikusan, nyelvészeti szabályokon alapuló konverterekkel hozzák létre: például az SPMRL-2013 Shared Task [1] versenyen használt kilenc nyelv közül csak a baszkra [2] és a magyarra [3,4] létezik kézzel annotált függőségi treebank. A konvertált függőségi treebankek minősége ennek ellenére kevésbé vizsgált.

A magyar azon ritka nyelvek egyike, ahol rendelkezésre áll kézzel annotált konstituens és függőségi annotáció is ugyanazon a szövegállományon, a Szeged (Dependencia) Treebanken [3,4]. Ez lehetővé teszi, hogy megvizsgáljuk a szabályalapú automatikus átalakítás minőségét, valamint hogy összehasonlítsuk az etalon treebankeken tanított konstituens és függőségi elemzők kimenetét a konvertált mondatokon tanított elemzők kimenetével.

Az automatikus konverzióknak az elemzésre gyakorolt hatását is megvizsgáljuk. Köztudott, hogy az angol esetén egy konstituenselemző kimenetét automatikusan függőségi elemzésre konvertálva hasonló eredmények érhetőek el, mint ha függőségi elemzőt tanítanánk konstituensből automatikusan konvertált fákon. Ennek egy lehetséges magyarázata, hogy az angol mint kötött szórendű nyelv

esetén a konstituenselemzők jobban működnek a függőségi elemzőknél. Megvizsgáljuk, hogy ez a megállapítás igaz-e a magyar esetén is, ami egy tipikus szabad szórendű nyelv.

Cikkünkben három pár függőségi elemzést hasonlítottunk össze, hogy megvizsgáljuk a konvertált fák használhatóságát. Először magának a konvertálásnak a hibáit vizsgáljuk a konvertált fák és a kézzel annotált etalon függőségi elemzés összehasonlításával. Másodszor összehasonlítjuk a konvertált fákon tanított függőségi elemző kimenetét az etalon fákon tanított függőségi elemző kimenetével. Harmadszor megmutatjuk, hogy az angolhoz hasonlóan a magyar esetén is igaz az, hogy a konstituens treebanken tanított elemző kimenetét függőségi elemzésre konvertálva hasonló eredményeket érhetünk el, mint a függőségi elemzőt az automatikusan konvertált fákon tanítva, bár a tipikus hibák eltérőek a két módszer esetén.

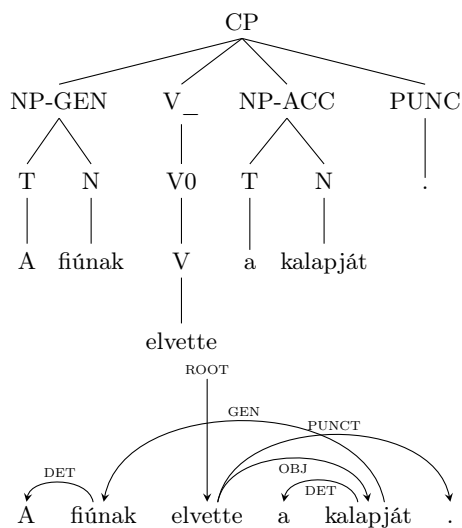
2. Magyar szintaxis a Szeged Treebankben

A magyar nyelv szintaktikai elemzése nagyban eltér az angolétól, főként a nyelv morfológiai gazdagsága és szabad szórendje miatt. Külön nehézséget okoznak a magyar esetén a távoli függőségek, például egyes birtokos szerkezetek (*A fiúnak vette el a kalapját.*), és azok a mondatok, amelyekben nem jelenik meg ige (*A kalap piros.*).

A Szeged Treebank [3] egy 82 000 mondatból álló, konstituensnyelvtan szerint kézzel annotált treebank magyarra. A frázisstruktúra mellett az igék argumentumainak nyelvtani szerepe és a szavakhoz tartozó morfológiai információ is annotálva van benne. Hat különböző doménből (üzleti rövidhírek, újságcikkek, irodalmi szövegek, fogalmazások és informatikai szövegek) származó szövegeket tartalmaz, ezek közül cikkünkben az üzleti rövidhírek alkorpuszt használtuk.

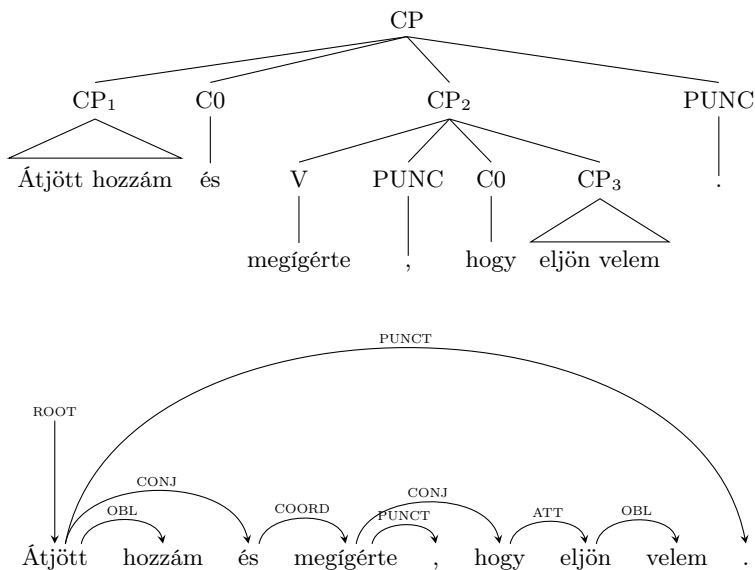
A Szeged Dependencia Treebank [4] kézi dependenciaannotációt tartalmaz ugyanezekre a szövegekre. Bizonyos nyelvtani összefüggések (például távoli függőségek) csak ebben a treebankben vannak jelölve, a konstituensben nem. A függőségi treebankben a birtokos a birtokhoz van kötve, míg a konstituenselemzés nem tartalmazza ezt az információt. A két szerkezet a 1. ábrán látható.

További különbség a két treebank között az összetett mondatok reprezentációja, ahogy a 2. ábrán látható. A függőségi treebankben az alá- és mellérendelések kezelése nagyon hasonló. Az egyik tagmondat feje (alárendelés esetén az alárendelt tagmondaté, mellérendelés esetén a második tagmondaté) a másik tagmondat fejéhez (alárendelés esetén a főmondatéhoz, mellérendelés esetén az első tagmondatéhoz) kapcsolódik, és csak a közöttük lévő reláció különbözteti meg a kétféle szerkezetet. A 2. ábrán látható mondat függőségi elemzésében a három tagmondat feje (*átjött*, *megígérte* és *eljön*) egymáshoz vannak kapcsolva a kötőszavakon keresztül ATT relációval alárendelés és COORD relációval mellérendelés esetén. A konstituens treebank eltérő szerkezetet rendel az alá- és mellérendelésekhez: alárendelés esetén az alárendelt tagmondat része a főmondatnak: a CP₃ a CP₂-n belül található a 2. ábrán. A mellérendelt tagmondatok



1. ábra: Távoli függőség kezelése konstituens- és függőségi elemzésben.

egy szinten vannak a struktúrában: az ábrán CP₁ és CP₂ mellérendelt tagmondatok.



2. ábra: A mellérendelés kezelése konstituens és függőségi elemzésben.

A kézzel annotált treebankek hasonlóságai megfelelővé teszik őket az automatikus függőségi átalakítással kapcsolatos hipotéziseink megvizsgálásához. Különbségeiket a konstituens és függőségi nyelvtanok alapvető eltérései okozzák.

3. Konstituensfák átalakítása függőségi fákká

Ebben a részben bemutatjuk a konstituensfák függőségi fákká alakításához alkalmazott módszerünket, valamint az átalakítás közben felmerült legtipikusabb hibákat.

3.1. Átalakítási szabályok

A konstituensfák függőségi fává alakításához egy szabályalapú rendszert használtunk. A virtuális csomópontokat tartalmazó mondatokat kihagytuk a vizsgálatból, mivel ezek a konstituens treebankben nincsenek külön jelölve, továbbá függőségi nyelvtanbeli kezelésük is problémás [5,6]. Így 7372 mondattal és 162960 tokennel dolgoztunk.

Első lépésben meghatároztuk a tagmondat (CP) fejét és a CP-k közötti kapcsolatokat az összetett mondatokban. A CP feje általában egy finit ige, ha a CP nem tartalmaz finit igét, akkor a fej egy főnévi igenév vagy határozói igenév, ha egyik sem található a CP-ben, akkor a fej egy névszói összetevő. A CP fejek közötti kapcsolatok alkotják a függőségi struktúra alapját: a főmondat feje ROOT relációval kapcsolódik egy absztrakt kiinduló csomóponthoz, a mellérendelt tagmondatok fejei COORD, az alárendelt tagmondatok fejei ATT relációval kapcsolódnak a főmondat fejéhez, esetleg a CP-k között lévő kötőszón keresztül, CONJ relációval.

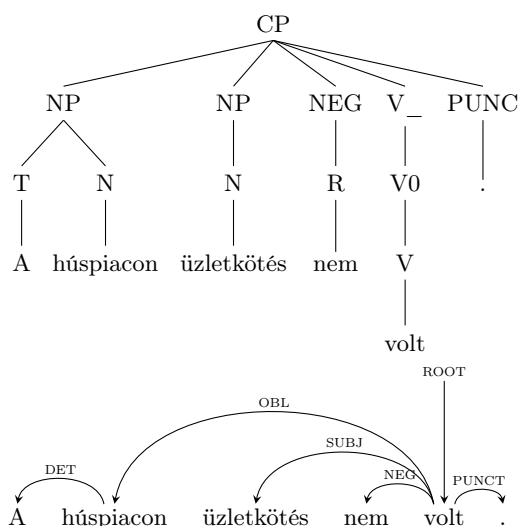
A Szeged Treebankben az igék, főnévi igenevek és határozói igenevek össze vannak kapcsolva az argumentumaikkal, azok nyelvtani szerepét is jelölve. Ezt az információt felhasználva állapítottuk meg a megfelelő függőségi relációt az igei kifejezések és argumentumaik között. A fő nyelvtani szerepek, azaz az alany, tárgy és részeshatározó, saját függőségi címkével rendelkeznek, míg az egyéb főnévi vonzatok egy összevont (OBL) relációt kapnak. Ezután az argumentumok módosítóit a fejhez vagy más módosítókhöz kapcsoltuk a frázisstruktúrájuknak és morfológiai kódjuknak megfelelően.

A távoli függőségek, mint a birtokos és birtok között lévő kapcsolat, nincsenek jelölve a konstituens treebankben. Ezekben az esetekben a morfológiai információt használtuk fel a megfelelő függőségi viszony megteremtéséhez.

A 3. ábrán egy mondat konstituensnyelvtanból függőségi nyelvtan szerinti átalakítása látható.

3.2. Hibaelemzés

A konstituens treebanket automatikusan függőségi fákká alakítottuk a honlapunkon leírt szabályoknak megfelelően (<http://rgai.inf.u-szeged.hu/SzegedTreebank>). A kiértékeléshez a labeled attachment score (LAS) és unlabeled attachment score



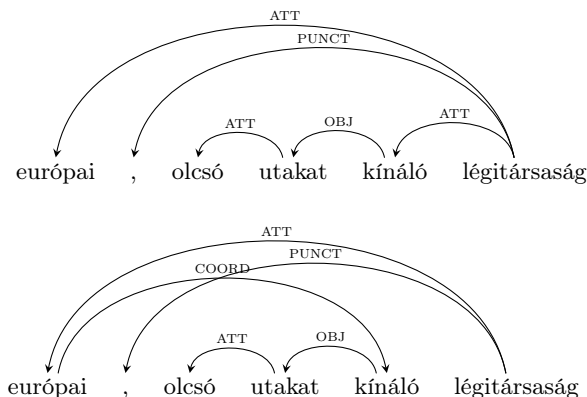
3. ábra: *A húspiacon üzletkötés nem volt* mondat konstituens annotációjának dependenciára alakítása.

(ULA) metrikákat alkalmaztunk, az írásjelek figyelembevétele nélkül. Az átalakítás pontossága 96,51 (ULA) és 93,85 (LAS). Az átalakítás hibáit az üzleti rövidhírek alkorpuszból véletlenszerűen kiválasztott 200 mondat kategorizációjával vizsgáltuk, a leggyakoribb hibák a 1. táblázat, *konvHiba* oszlopában láthatóak.

1. táblázat. Hibatípusok. *konvHiba*: konstituensfák függőségi fákká alakítása során vétett hibák. *etalonTrain*: a Bohnet parser etalon fákon tanított kimenetének hibái. *silverTrain*: a Bohnet parser silver standard fákon tanított kimenetének hibái. *BerkKonv*: etalon fákon tanított Berkeley parser kimenetének automatikus átalakítása során vétett hibák. *KonvDep*: függőségi címkék nélküli, silver standard fákon tanított Bohnet parser kimenetének hibái.

Hibatípus	konvHiba		etalonTrain		silverTrain		BerkKonv		KonvDep	
	#	%	#	%	#	%	#	%	#	%
Mellérendelés	26	13,00	39	13,22	59	14,82	55	16,37	64	19,57
Több módosító	26	13,00	30	10,17	49	12,31	52	15,48	47	14,37
Determináns	7	3,50	28	9,49	25	6,28	31	9,23	31	9,48
Kötőszó/határozószó kötés	33	16,50	23	7,80	45	11,31	39	11,61	42	12,84
Ige argumentuma	10	5,00	27	9,15	34	8,54	59	17,56	44	13,46
Alá- vagy mellérendelés	7	3,50	9	3,05	12	3,02	–	–	–	–
Birtokos	9	4,50	14	4,75	16	4,02	28	8,33	22	6,73
Rossz gyökerelem	14	7,00	17	5,76	23	5,78	35	10,42	27	8,26
Egymást követő főnevek	4	2,00	11	3,73	14	3,52	13	3,87	15	4,59
Többszavas NE	8	4,00	25	8,47	33	8,29	8	2,38	19	5,81
Rossz MOD címke	25	12,50	26	8,81	34	8,54	–	–	–	–
Egyéb rossz címke	17	8,50	33	11,19	30	7,54	–	–	–	–
Egyéb	14	7,00	13	4,41	24	6,03	16	4,76	16	4,89
Összesen	200	100	295	100	398	100	336	100	327	100

A leggyakoribb hibaforrás, ha egy frázisban egynél több módosító is volt, mint a 4. ábra mutatja. A következő ábrák mindegyikén bal oldalon, illetve felül látható az etalon elemzés, jobb oldalon, illetve alul pedig a hibás.



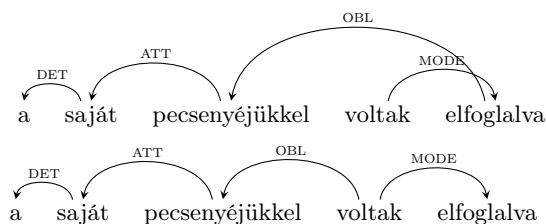
4. ábra: Több módosító miatti hiba.

Mellérendelési hibák akkor fordultak elő, amikor egy több tagból álló mellérendelés tagjai rosszul lettek összekötve. Másrészt a kötőszavak és néhány határozószó kapcsolása is problémás volt. Az 5. ábrán az *is* kötőszó az etalon elemzésben az igéhez van kötve, míg az átalakított változatban a főnévhez.



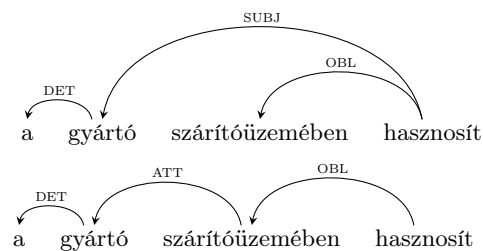
5. ábra: Kötőszó kapcsolásának hibája.

Bizonyos nyelvtani relációkat a konstituens treebank nem jelölt (például a számnevek és determinánsok egyszerűen csak az NP részei külön címkézés nélkül, mint *[NP az öt [ADJP fekete] kutya]*), de a dependencia reprezentációban ezekhez is szükséges volt szülőt és címkét rendelni. Ez nem minden esetben volt teljesen egyértelmű: például a *[NP nem [ADJP megfelelő] módszerek]* kifejezés konvertált reprezentációjában a tagadószó a melléknév helyett a főnévhez van kötve. A determináns hibák esetén a determináns rossz főnévhez lett kötve olyan NP-kben, ahol a fejet egy másik főnév módosítja. A több igei összetevőt is tartalmazó CP-k esetén (egy finit ige és egy főnévi vagy határozói igenév) az argumentumok néha rossz igei összetevőhöz kapcsolódtak, mint a 6. ábrán látható esetben.



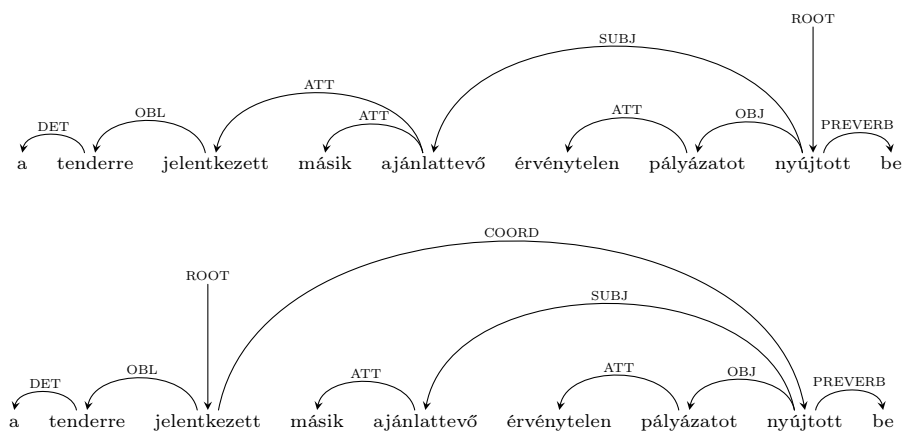
6. ábra: Ige argumentuma rossz helyre kapcsolva.

Mivel a konstituensannotációból ez hiányzik, így a birtokos megtalálásában is előfordultak hibák, mint a 7. ábra esetén.



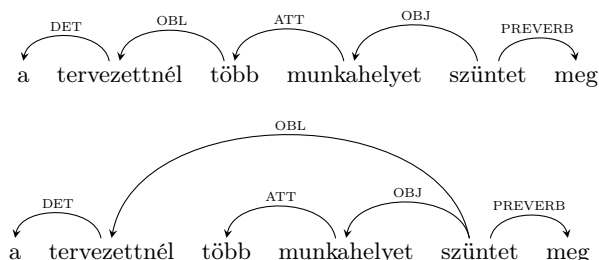
7. ábra: A birtokos kapcsolásának hibája.

Több igei összetevőt tartalmazó CP-kben nem mindig a megfelelő gyökérem lett kiválasztva, mint a 8. ábrán.



8. ábra: Rossz gyökérem.

Néhány esetben egymást követő, de különálló NP-k egy egységként lettek kezelve, mintha az egyik főnév a másikat módosítaná, mint a 9. ábrán.



9. ábra: Egymást követő főnevek kapcsolásának hibája.

A többszavas névelemek is okoztak átalakítási problémákat, mint a 10. ábrán látható.



10. ábra: Többszavas NE hiba.

Bizonyos esetekben a konstituens és dependencia treebankben előforduló annotációs hibák is okoztak eltéréseket az etalon és az átalakított fák között. Erre tipikus példa a rossz MODE címke hiba. A treebank a magyar határozószavakat tér- és időbeliség, valamint irányhármasság figyelembevételével megkülönbözteti, így hat külön címkével írja le ezeket a relációkat, a további határozószavak pedig egy összevont MODE relációval szerepelnek. Mivel ez a megkülönböztetés szemantikai jellegű, és gyakran hibásan lett annotálva a konstituens treebankben, ezek a hibák később a függőségi treebank annotációjában javítva lettek, így az átalakítás során hibákat okoztak, mint a 11. ábrán.



11. ábra: Rossz MODE címke.

Más hibák annyira ritkák voltak (például egy determináns hibásan az igéhez lett kötve), hogy egy kategóriába soroltuk őket, ezek a 1 táblázat „egyéb hiba” sorában láthatóak.

4. Tanítás etalon és silver standard fákon

Kísérleteztünk a Bohnet függőségi elemző [7] kézzel annotált (gold standard) és átalakított (silver standard) fákon tanításával. A Bohnet parser egy széles körben használt gráfalapú parser¹, amely perceptronra épülő online tanítást alkalmaz.

A korpuszunkból 5892 mondatot (130211 token) használtunk tanító adatbázisként, a megmaradt 1480 mondatot (32749 token) tesztelésre. A kiértékelés itt is a LAS és ULA metrikák alapján történt. Az eredmények a 2. táblázat *etalonTrain* és *silverTrain* sorában látható.

Látszik, hogy jobb eredmények érhetőek el etalon adatokon tanítva, 1,6% (ULA) és 3,16% (LAS) különbséggel. A számszerű különbségeken túl itt is összehasonlítottuk a hibákat: az átalakítási hibák vizsgálatánál is használt, véletlenszerűen kiválasztott mondatokat felhasználva kézi hibaelemzést végeztünk az etalon annotációval összehasonlítva a két kapott kimenetet, l. 1. táblázat, *etalonTrain* és *silverTrain* sorok.

Bizonyos jelenségek mindkét esetben okoztak hibákat a parsernek. A mellérendelés és a több módosító például a leggyakoribb hibatípusok között van mindkét esetben, bár a hibák számát nézve észrevehetjük, hogy az etalonon tanítás mindkét esetben csökkenti a hibák számát. Emellett a kötőszó és határozószó szülő csomópontjának megtalálásában is sokat segít az etalon adat. Ennek oka valószínűleg az, hogy a konstituens treebank nem jelöli ezeket a nyelvtani kapcsolatokat, így a silver standard treebank ilyen szempontból nagyon zajos. Így összességében megállapíthatjuk, hogy bizonyos nyelvtani kapcsolatok (például a kötőszavak és határozószavak kapcsolása) kézi ellenőrzést igényel akkor is, ha automatikusan konvertálunk konstituens fákból függőségi fákat.

5. Tanítás előtt vagy tanítás után konvertáljunk?

Az angol nyelv esetén ismert, hogy egy konstituens parser kimenetét dependenciává alakítva hasonló ULA eredmények érhetőek el, mint egy átalakított fákon tanított függőségi parser használatával [8,9]. Ennek egyik lehetséges magyarázata, hogy mivel az angol kötött szórendű nyelv, így a konstituenselemzőktől jobb eredményre számíthatunk. Cikkünkben megvizsgáljuk, hogy igaz-e ez az állítás a magyar esetén, ami a morfológiailag gazdag, szabad szórendű nyelvek prototípusa.

Kísérleteinkhez a Berkeleyparser [10] nyelvtanok szorzatára épülő eljárását [11] alkalmaztuk. Ennek segítségével eltérő beállítások mellett ugyanazon tanító halmazon 8 különböző nyelvtant tanítottunk, és kiértékeléskor az egyes fák

¹ A Bohnet parsernek más parserekkel való összevető elemzéséről l. [5]. Az ott leírtak alapján a Bohnet parser érte el a legjobb eredményeket a treebanken, így jelen kísérleteinkben is ezt a parsert alkalmaztuk.

2. táblázat. A kísérletek eredménye. Konverzió: konstituensfák függőségi fákká alakítása. etalonTrain: a Bohnet parser etalon fákon tanítva. silverTrain: a Bohnet parser silver standard fákon tanítva. BerkeleyKonv: etalon fákon tanított Berkeley parser kimenetének automatikus dependenciává alakítása. KonvDep: függőségi címkék nélküli, silver standard fákon tanított Bohnet parser.

Kísérlet	LAS	ULA
Konverzió	93.85	96.51
etalonTrain	93.48	95.17
silverTrain	90.32	93.57
BerkeleyKonv	–	92.78
KonvDep	–	93.23

valószínűségét ezen nyelvtanok által adott valószínűségek szorzataként kaptuk meg. A parserek kimenetét ezután automatikusan átalakítottuk a 3. részben leírt szabályok használatával (*BerkKonv*). Emellett a silver standard függőségi treebanken a Bohnet parsert tanítottunk (*KonvDep*). Mivel a konstituensparserünk kimenetében nincsenek nyelvtani szerepek, így a megfelelő összehasonlíthatóság céljából a Bohnet parsert címkézetlen dependencia fákon tanítottuk (ez okozza a különbséget az 1. táblázat *BerkKonv* és *KonvDep* oszlopa között).

Láthatjuk, hogy a két módszer által elért eredmények hasonlóak, 92,78 és 93,23 ULA. Ez azt jelenti, hogy magyar nyelvre ugyanúgy igaz a megállapítás, mint angolra. Így – meglepő módon – azt mondhatjuk, hogy a tanítás utáni konverzió módszerével elérhető jó eredmény nem függ össze a kötött szórenddel.

Az előzőekhez hasonló kézi hibaelemzés alapján megállapítható, hogy ebben az esetben is előfordulnak hasonlóan gyakori hibatípusok, mint például a mellérendelés, a kötőszavak, módosítók és determinánsok kapcsolása. Másrészt a konstituensfákon tanítás egyéb hibákat is okozott. Egyrészt a birtokos szerkezetekben a birtokos ritkábban van helyesen a birtokhoz kötve, valószínűleg, mivel a konstituens treebankben nincs jelölve ez a kapcsolat, így a parser nem tudta megtanulni. Másrészt a több igei kifejezést tartalmazó tagmondatokban az argumentumok megfelelő helyre kapcsolása is több hibát okozott, főként határozói igenevek és főnévi igenevek esetén. A 6. ábrán látható, hogy a helyes elemzésben a *pecsenyéjükkel* főnév a határozói igenévhez van kötve, míg a másokban a főigéhez. Harmadrészt a CP fejének megtalálása is nehezebben ment ebben az esetben. A [9] cikkben a szerzők azt találták, hogy a CP fej megtalálása jobb eredménnyel történik tanítást megelőző konverzió esetén, így ez érdekes nyelvspecifikus különbségnek tűnik az angol és a magyar között. Emellett a konstituensfákon tanítás jó hatással van a többszavas NE-k felismerésére. Ezek alapján megállapíthatjuk, hogy bár számszerűen hasonló eredmények érhetőek el a két módszerrel, ezek mégis különböznek egymástól a hibatípusok tekintetében.

6. Összegzés

Cikkünkben különböző módszerekkel nyert magyar nyelvű függőségi elemzéseket hasonlítottuk össze. Megállapítottuk, hogy bár a különböző módszerek számszerűleg hasonló eredményeket érnek el, különböző hibákat ejtenek. Másrészt bizonyos nyelvtani összefüggések (például mellérendelés, több módosító vagy a kötőszavak és határozószavak kötése) általában nehezek a függőségi elemző számára.

A konstituenskorpuszunk konvertálásával egy silver standard függőségi treebanken is tudtunk kísérleteket végezni. Az eredményeink azt mutatják, hogy jobb eredményeket érhetünk el az etalon annotáció használatával, ezért kívánatos a kézi függőségi annotáció. Másrészt ennek hiányában a konstituensparser kimenetének dependenciává konvertálásával vagy konstituensből átalakított dependencián tanítással az angolhoz hasonlóan jó eredmények érhetőek el, bár a hibák típusa itt is különbözik [12].

A jövőben szeretnénk megvizsgálni, hogy milyen további előnyei vannak a magyar konstituens és függőségi reprezentációjának, valamint mindkét elemző esetében szeretnénk „uptraining” kísérleteket folytatni.

Köszönetnyilvánítás

A jelen kutatás a futurICT.hu nevű, TÁMOP-4.2.2.C-11/1/KONV-2012-0013 azonosítószámú projekt keretében az Európai Unió támogatásával és az Európai Szociális Alap társfinanszírozásával valósult meg.

Hivatkozások

1. Seddah, D., Tsarfaty, R., Kübler, S., Candito, M., Choi, J.D., Farkas, R., Foster, J., Goenaga, I., Gojenola Galleitebitia, K., Goldberg, Y., Green, S., Habash, N., Kuhlmann, M., Maier, W., Marton, Y., Nivre, J., Przepiórkowski, A., Roth, R., Seeker, W., Versley, Y., Vincze, V., Woliński, M., Wróblewska, A.: Overview of the SPMRL 2013 shared task: A cross-framework evaluation of parsing morphologically rich languages. In: Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages, Seattle, Washington, USA, Association for Computational Linguistics (2013) 146–182
2. Aduriz, I., Aranzabe, M.J., Arriola, J.M., Atutxa, A., Diaz de Ilarraza, A., Garmendia, A., Oronoz, M.: Construction of a Basque dependency treebank. In: Proceedings of the 2nd Workshop on Treebanks and Linguistic Theories (TLT), Växjö, Sweden (2003) 201–204
3. Csendes, D., Csirik, J., Gyimóthy, T., Kocsor, A.: The Szeged TreeBank. In Matousek, V., Mautner, P., Pavelka, T., eds.: Proceedings of the 8th International Conference on Text, Speech and Dialogue, TSD 2005. Lecture Notes in Computer Science, Berlin / Heidelberg, Springer (2005) 123–132
4. Vincze, V., Szauter, D., Almási, A., Móra, Gy., Alexin, Z., Csirik, J.: Hungarian Dependency Treebank. In: Proceedings of LREC 2010, Valletta, Malta, ELRA (2010)

5. Farkas, R., Vincze, V., Schmid, H.: Dependency Parsing of Hungarian: Baseline Results and Challenges. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, Association for Computational Linguistics (2012) 55–65
6. Seeker, W., Farkas, R., Bohnet, B., Schmid, H., Kuhn, J.: Data-driven dependency parsing with empty heads. In: Proceedings of COLING 2012: Posters, Mumbai, India, The COLING 2012 Organizing Committee (2012) 1081–1090
7. Bohnet, B.: Top accuracy and fast dependency parsing is not a contradiction. In: Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). (2010) 89–97
8. Petrov, S., Chang, P.C., Ringgaard, M., Alshawi, H.: Uptraining for accurate deterministic question parsing. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, Association for Computational Linguistics (2010) 705–713
9. Farkas, R., Bohnet, B.: Stacking of dependency and phrase structure parsers. In: Proceedings of COLING 2012, Mumbai, India, The COLING 2012 Organizing Committee (2012) 849–866
10. Petrov, S., Barrett, L., Thibaux, R., Klein, D.: Learning accurate, compact, and interpretable tree annotation. In: Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics. (2006) 433–440
11. Petrov, S.: Products of random latent variable grammars. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, California, Association for Computational Linguistics (2010) 19–27
12. Simkó, K.I., Vincze, V., Szántó, Zs., Farkas, R.: An Empirical Evaluation of Automatic Conversion from Constituency to Dependency in Hungarian. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland, Dublin City University and Association for Computational Linguistics (2014) 1392–1401