

# Beszámoló a XIII. Magyar Számítógépes Nyelvészeti Konferenciáról

2017. január 26–27-én került sor az ez évi Magyar Számítógépes Nyelvészeti Konferenciára (MSZNY) a Szegedi Tudományegyetem József Attila Tanulmányi és Információs Központjában. A sorban immár tizenharmadik konferencia célja a nyelv- és beszédtechnológia különböző területein végzett, illetve folyamatban lévő legújabb kutatások eredményeinek bemutatása, hallgatói projektek és ipari alkalmazások ismertetése volt, például a *morfológiai és szintaktikai elemzés*, a *korpusznyelvészet*, a *szemantikai elemzés*, a *jelentés-egyértelműsítés*, a *név-elem-felismerés*, az *információkinyerés*, a *gépi fordítás*, a *beszédfelismerés és -szintézis* terén.

A korábbiaktól eltérően nem absztrakttal, hanem teljes cikkel lehetett jelentkezni, ami huszonhat előadást, öt poszter- és négy laptopos bemutatót eredményezett. A korábbiakhoz hasonlóan egyes témák mind az előadások, mind a laptopos bemutatók között is szerepeltek, ezáltal is tér és lehetőség nyílt a kutatások széles körű megismertetésére. A konferenciakötet elérhető a konferencia honlapján (<http://rgai.inf.u-szeged.hu/project/mszny2017/files/kotet.pdf>). A szervezőbizottság tagjai: Csirik János, Farkas Richárd, Simon Eszter és Vincze Veronika.

A beszámolóban minden szekcióból kiemelek egy-egy előadást, amelyek jól mutatják a hazai nyelv- és beszédtechnológiai kutatások színvonalát.

A konferencia programbizottsága új hagyományt teremtett: a cikkek bírálatai és az előadások alapján kiemelte a legjobb munkát, amely eredményeivel leginkább hozzájárult a magyarországi nyelv- és beszédtechnológiai kutatásokhoz. Az MTA Nyelvtudományi Intézete által biztosított díjat Tihanyi László és Oravecz Csaba *First experiments and results in English–Hungarian neural machine translation* című cikke nyerte el. A szerzők neurális modellek alkalmazásával végeztek angol–magyar gépi fordítási kísérleteket.

A konferencia plenáris előadását Labádi Gergely, a Szegedi Tudományegyetem Magyar Irodalmi Tanszékének docense tartotta *Filológia az információs társadalomban* címmel. Előadásában kiemelte a számítógépes nyelvészet szerepének fontosságát a digitális bölcsészetben, hiszen „*bármilyen eredményeket érnek el ma az irodalmárok a digitális bölcsészet területén, azt mind a számítógépes nyelvészetenek, nyelvészeknek köszönhetik*”. Rávilágított, hogy bár a filológiai munka alapműveletei – a hozzáadás, az elvétel, az ellenőrzés és a szerkesztés – mind a számítógép segítségével elvégezhető feladatok, a számítógépek elterjedése a humán tudományokban eddig nem változtatta meg alapvetően a textuális kultúrát, a szövegekkel történő munka gyakorlatát. A digitális bölcsészet történetének fontos fejezeteit felvillantva mutatta meg, hogyan vált egyre fontosabbá a digitalizáció és a számítógéppel való feldolgozhatóság az irodalomtudományban és általánosabban a filológiában. Az előadás tanulsága, hogy a számítógépes nyelvészet eredményeire támaszkodó digitális bölcsészet területein rengeteg

lehetőség van ezeknek az eredményeknek kamatoztatására, tehát fontos a két tudományterület együttműködése, nyitottsága egymás felé.

Az *információkinyerés* szekcióban hangzott el Novák Attila és Siklósi Borbála előadása, *A Dologfelismerő*, egy olyan eszközről, amely a szavakhoz diszkrét szemantikai jegyeket rendelve segíti a folytonos modellben ábrázolt jelentések értelmezését, és ugyanakkor a reprezentált tudást hozzáférhetővé teszi gépi tanuló vagy keresőalgoritmusok számára is. Ehhez létrehoztak egy angol és egy magyar szóbeágyazási modellt, a szóbeágyazási modellben szereplő szóvektorok klaszterezésével pedig egy reprezentációt, amelyben az egyes klaszterekbe tartozó szavak valamilyen szempont alapján hasonló jelentésű szavaknak tekinthetők. Megmutatták, hogy a különböző nyelvekre létrehozott szóbeágyazási modellek által definiált szemantikai terek leképezhetők egymásba, ami lehetővé teszi, hogy az angol erőforrásokból létrehozott szemantikai kategóriacímkekhez rendelt vektorokat az angol térből leképezzék a magyar nyelvű szóbeágyazási modell terébe. Ezáltal a módszer olyan nyelvekre is alkalmazható, amelyekre nem áll rendelkezésre megfelelő lexikai erőforrás. Az eszköz jól címkézi a kézzel készített erőforrásokban nem szereplő ritka szavakat is, például a neveket, a rövidítéseket, a szlenget és általában a nem sztenderd szóalakokat.

A programbizottság külön szekciót szentelt az e-magyar elnevezésű digitális nyelvfeldolgozó rendszer bemutatásának. Az *e-magyar.hu* projekt egységes magyar előfeldolgozó láncot fejlesztett ki a magyar nyelvtechnológiai műhelyek együttműködésével. Az *e-magyar digitális nyelvfeldolgozó rendszer* című nyitó előadás átfogó ismertetést nyújtott a rendszeréről, amelynek célja az eddig előállított eszközök továbbfejlesztése, egységesítése és koherens technológiai láncba szervezése. A projekt a korábban különböző műhelyek által létrehozott eszközök és erőforrások interoperábilissá tételét célozta meg, ahol az eszközök modulárisan egymásra épülnek, tehát önállóan is működnek, de elemzési láncba is szervezhetők. A kutatóknak fontos volt a nyílt rendszer és a széles körű elérhetőség, ezért az infrastruktúra egésze és minden eszköze külön-külön is elérhető és felhasználható kutatásfejlesztési célra. Emellett létrejött az *e-magyar.hu* webszolgáltatás is, amely alkalmas rövidebb szövegek azonnali elemzésére, és olyan felhasználók számára is elérhető, akik kevésbé járatosak az informatikában. A rendszer az írott szövegek feldolgozása mellett a beszédfeldolgozással is megbirkózik, hála a beszédatadabázisnak és a beszédelemző moduloknak. Az e-magyar projektben részt vevő kutatók: Váradi Tamás, Simon Eszter, Sass Bálint, Geröcs Mátyás, Mittelholcz Iván, Novák Attila, Indig Balázs, Prószéky Gábor, Farkas Richárd és Vincze Veronika.

A *beszédtechnológia* szekció egyik előadása az *Élő labdarúgó-közvetítések gépi feliratozása* volt. Ennek a kutatásnak a célja az élő televíziós műsorok akadálymentesítése a siket- és nagyothalló nézők számára. A gépi feliratozás lépést tud tartani az emberi beszéd tempójával, ezért a végeredmény a veszteségmentes feliratozás, cserébe viszont meg kell küzdenie olyan nehézségekkel, mint az idegen nyelvű személynevek felismerése, amelyek gyakran előfordulnak a sportközvetítésekben. A szerzők az előadás során különböző szótárbővítési eljárásokat vetettek össze hatékonyságuk alapján, amelyek segítségével egy már betanított felismerő modell szótárát a kívánt feladathoz lehet adaptálni. A kísérletekben kimutatták, hogy a személynevek pontos ragozással történő feliratozásához komplex, környezetfüggő modellezésre van szükség, de ha az esetragok felismerése nem elvárás, akkor az egyszerűbb, környezetfüggetlen megoldások jól megfelelnek a célnak. A cikk szerzői: Tarján Balázs, Szabó Lili, Balog András, Halmos Dávid, Fegyő Tibor és Mihajlik Péter.

A *szentimentelemzés* szekcióból Pólya Tibor *A kognitív disszonancia narratív markereinek azonosítása termékleírásokban* című előadását emelem ki. A bemutatott kutatás célja, hogy az autóleírásokban megállapítsa a termékleírást végző személy kognitív disszonanciájának

mértékét. A szerző autókkal kapcsolatos tapasztalatok megosztását, cseréjét lehetővé tévő honlapról származó autóleírásokat elemezett a szöveg kompozicionális kategóriáinak elemzésére kidolgozott automatikus eljárással. Az eredmények azt mutatják, azonosíthatók azok a narratív markerek, amelyek alapján megbízhatóan megállapítható, hogy a leírás magas vagy alacsony mértékű kognitív disszonancia állapotában készült-e.

A *többnyelvűség* szekciót Simon Eszter *Négy hatás alatt álló nyelv – Korpuszépítés kis uráli nyelvekre* című előadása nyitotta. Az ismertetett projekt annotált korpuszt épít négy oroszországi kisebbségi uráli nyelvre (udmurt, tundrai nyenyec, színjai és szurguti hanti). Fő célkitűzése az, hogy a korpusz építése sztenderd eszközökkel történjen, és a végeredmény strukturált adatbázis legyen. A projekt három elméleti célja a fent említett nyelvek sajátos mondattani tulajdonságainak a leírása, a nyelvek szintaktikai változásainak vizsgálata, valamint a nyelvekre gyakorolt orosz hatás kimutatása. A korpuszépítés során a kutatók törekedtek arra, hogy szabadon elérhető eszközöket használjanak. Az előálló szöveges és feldolgozó erőforrást szabadon hozzáférhetővé tették a projekt honlapján (<http://www.nytud.hu/oszt/elmnyelv/urali/adatbazisok.html>).

A konferencia előadásai és a konferenciakötet cikkei alapján elmondható, hogy a XIII. Magyar Számítógépes Nyelvészeti Konferencia résztvevőinek sikerült megmutatniuk, milyen sokszínű és értékes kutatások zajlanak a hazai nyelv- és beszédtechnológia különböző területein.

Vadász Noémi

FERCSIK ERZSÉBET, RAÁTZ JUDIT

KERESZTNEVEK

240 oldal 1990 Ft

Ajánljuk a könyvet mindazoknak, akik érdeklődnek az egyes nevek eredete, jelentése iránt, valamint szívesen megismerkednének a hozzájuk kapcsolódó szó-lásokkal, népszokásokkal. Az egyes nevek mellett található statisztikai adatok a gyermeket váró szülők számára is fontos információt jelenthetnek.

Megvásárolható a kiadóban:

TINTA KÖNYVKIADÓ

1116 Budapest, Kiskőrös utca 10.;

tel.: (1) 371-0501; fax: (1) 371-0502

E-mail: [info@tintakiado.hu](mailto:info@tintakiado.hu);

honlap: [www.tintakiado.hu](http://www.tintakiado.hu)