

# **A nyelvtudomány technológiai alapú evolúciója**

**(MTA, Budapest, 2015. november 10.)**

*Nagy Annamária Lilla*

*E-mail: annamarialillanagy@gmail.com*

A Magyar Tudományos Akadémia Nyelvtudományi Intézete a Magyar Tudomány Ünnepe keretein belül „A nyelvtudomány technológiai alapú evolúciója” címmel szervezte meg előadásorozatát 2015. november 10-én. A konferenciának a Tudományos Akadémia Nagyterme adott otthont. Az Akadémia 1997 óta számon tartja a Magyar Tudomány Ünnepe, az országgyűlés azonban csak 2003-ban nyilvánította november 3-át a Magyar Tudomány Ünnepevé, mivel 1825-ben Széchenyi István ezen a napon ajánlotta fel birtokainak éves jövedelmét a Tudós Társaság megalapítására.<sup>1</sup>

2015-ben a Tudományos Akadémia novemberi rendezvénysorozatának előadásait minden tudományterületen „A tudomány evolúciója: a valós és virtuális világok” tematika köré csoportosították, így került előtérbe a nyelvészetek területén is a nyelvtudomány és a számítógépes világ kapcsolata.

Az előadásorozatot Kenesei István, a Nyelvtudományi Intézet igazgatója nyitotta meg, aki felhívta a közönség figyelmét arra, hogy a nyelvtudomány sokáig semmilyen technikát nem igényelt, az utóbbi időben azonban a tudomány észrevehetően megváltoztatta a nyelvet. Az előadók egytől egyig az Intézet fiatal kutatói voltak, akik nyelvészeti kutatásukhoz a technológiát hívták segítségül.

Az első előadó Pintér Lilla volt, szakterülete a kísérletes nyelvészet és elsősorban azt vizsgálja, különbözőképpen értelmezik-e a felnőttek és a gyerekek a mondatokat. Tesztmondatok segítségével mondat-kép megfeleltetési kísérleteket végzett 60 főn (óvodás korú, első, második és harmadik osztályos gyerekeken, valamint felnőtteken). A kísérlethez az SR Research Experiment Builder programot használta. Az alanyoknak a mondat és a kép viszonyát kellett háromfokú skálán (szomorú arc, szájhúzó arc, mosolygó arc) osztályozni a következő három mondatcsoportba: „csak” partikulás mondatok, szerkezeti fókuszos mondatok, semleges szórendű mondatok fókusz nélkül. Az eredmények korcsoportonként eltérőek lettek, így beigazolódni látszik, hogy a mondatok értelmezése valóban függ az életkortól.

---

<sup>1</sup> <http://parameter.sk/rovat/technika-tudomany/2009/10/25/kozeleg-magyar-tudomany-unnepe>

A második előadó Káldi Tamás volt, aki a szemmozgáskövetés világába vezette be a hallgatóságot. Kísérletét az infra-illuminátor elnevezésű gép segítségével végezte, amellyel a pupilla és a szaruhártya tükröződését lehet felmérni és követni. Káldi az „és” és „vagy” mondatok közötti különbséget vizsgálta, amelyből arra a következtetésre jutott, hogy az „és” mondatok feldolgozása egyértelműbb és gyorsabb a „vagy” mondatokénál. A kísérlet legfőképp azért érdekes a nyelvészek és elsősorban a fordításkutatók számára, mert a szem mozgása az agy figyelmének levetülése, így információkat kapunk arról, mi megy végbe az agyban.

A harmadik előadó Gráczy Tekla Etelka, a fonetika kutatója. Előadásában előbb a fonetika kutatásának történetét ismertette különböző eszközök bemutatásával, amelyeket a hangok felismerésére fejlesztettek ki. Az idő előrehaladtával ezek az eszközök egyre fejlettebbé váltak, különösképp a 20. században, amikor is egyre több kutatást folytattak a beszédtechnológia fejlesztésének érdekében, többek között a beszélőfelismerés, a beszédfelismerés és a spontánbeszéd-elemzés területein. A mai modern kutatásokat elősegítik továbbá a számítógépes adatbázisok és a külső zajtól mentes megfigyelőtermek is.

Sipos Mária és Kornai András közös előadásban informálták a közönséget a veszélyeztetett nyelvekről. Az előadást Sipos Mária kezdte, aki ismertette, minek alapján sorolják be a nyelveket a veszélyeztetettségük szempontjából egy úgynevezett vitalitási skálába. Ezután néhány kutató munkáját ismertette (Bloomfield, Ladefoged és Dorian), majd beszélt a veszélyeztetett nyelvek és a politikai mozgalmak kapcsolatáról, valamint arról, hogy az idő teltével egyre érzékenyebb témaként kezelik a jelenséget. A nyelvészekre azonban kétségtelenül óriási felelősség hárul. Kornai András egy másik besorolási módot mutatott be, ahol 4 fokozata van a veszélyeztetettségnek: viruló, élő, örökségi, halva született. A nyelvtechnológia abban segít, hogy az élő, de nem viruló, valamint a kihaltfélben lévő nyelvek digitális túlélővé válnak általa. Az életerős nyelvek esetében felhívta a figyelmet arra, hogy számtalan technikai újítás elérhetővé válik számukra, mint például a számítógépes szintaktikai elemzések, a beszédfelismerés vagy a gépi fordítás.

Ezután Makrai Márton következett, akinek szakterülete a nyelvmodellezés, amelynek célja nem más, mint a kérdésmegválaszolás, a fordítás és a modellezés. Rövid előadásában ismertette a szómodellezést, amelynek során több száz dimenziós térben helyezük el a szavakat hasonlóság, rokonértelműség, asszociáció és analógia alapján. Bemutatta még az úgynevezett háromszögletes módszert is, amelynek során 3 nyelv lexikai kapcsolata mutatható ki. Makrai saját kísérletében adott szavak, valamint azok fordításának leképezését vizsgálta. A vizsgálatban kimutatta, hogy a fordítások esetében a szóhoz legközelebb eső leképezés volt a szó célnyelvi megfelelője.

Simon László tíz éve a nagyszótáron dolgozik, az állat- és növénynevekért felelős. A szótár létrehozása során gyakorlatilag a szócikkek adminisztrációja folyik lektorok és szaklektorok bevonásával az XMetal programban. A jelenlegi célkitűzés, hogy minden változtatást adott helyen tudjanak jelölni, ami elősegítené a hatékonyabb, gyorsabb munkát. Mesélt továbbá arról a feldolgozási munkáról is, amellyel munkája során találkozik. Ezekhez szövegfelismerő

– OCR – programokat használnak, azonban a papíralapú, kézirásos anyagok digitalizálásához elengedhetetlen az emberi munka. Az előadás címe *Illúzió és valóság* volt, amelyet Simon előadása zárásakor elegánsan azzal magyarázott, hogy az illúzió az, hogy szoftverekkel minden megoldhatóvá válik, a valóság azonban az, hogy ez nem igaz.

Az utolsó előadást Sass Bálint tartotta a korpuszlekérdezésről. A prezentáció során elmagyarázta, mi a haszna az ilyen programoknak, és bemutatta a legfontosabb magyar nyelvű korpuszokat, amelyeknél a számítógép alapú korpuszlekérdeezést kialakították. Ezek közül a legújabb az MNSZ (Magyar Nemzeti Szövegtár 2), amely 2014-ben jelent meg és 587 millió szót tartalmaz, és többek között lehetőség nyílik vele képzőszűrésre, gyakorisági listák megtekintésére és kollokációs vizsgálatokra is.

A konferenciát Kenesei István zárta, kiemelve, hogy a technológia idővel minden tudomány, így a nyelvtudomány területén is egyre nagyobb teret kap majd, és rámutatott arra, hogy érdemes figyelmünket erre a területre összpontosítani, hiszen a tudományok jövője ebben rejlik. Külön kiemelném még, hogy az előadások teljes ideje alatt jelyelvi tolmácsot biztosítottak az azt igénylő néző számára, ami jól érzékelteti az Akadémia rendezvényeinek magas színvonalát. A Magyar Tudomány Ünnepe rendezvénysorozat minden év novemberében megrendezik érdekes, modern témákkal.