

## **Simon Eszter**

tudományos főmunkatárs  
Magyar Tudományos Akadémia  
Nyelvtudományi Intézet  
Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály  
Nyelvtechnológiai Kutatócsoport

A nyelvtechnológia számos ágával foglalkozik, kutatási területei közé tartozik a tulajdonnév-felismerés, a morfológiai elemzés, a korpuszépítés és -annotáció, a történeti korpuszok fejlesztése, a digitális bölcsészet és az uráli nyelvek számítógépes nyelvészeti támogatása. Több nagyszabású számítógépes nyelvészeti projekt és kutatócsoport résztvevője és koordinátora, mint a Magyar Generatív Történeti Szintaxis, vagy az Urali Nyelvek Szintaxisa. Egyik létrehozója az Ómagyar Korpusznak és az e-magyar.hu digitális nyelvfeldolgozó rendszernek. Több nyelvtechnológiai kurzus oktatója.

## **Magyar nyelvű történeti korpuszok**

A nyelvi kulturális örökség elérhetővé tételében kulcsfontosságú szerep jut a nyelvtechnológiának, melynek módszereivel a kutatók egységes, következetes, nyelvi információval ellátott adatbázisokhoz juthatnak. A nyelvtörténészek és nyelvtechnológusok egyik legfontosabb együttműködési terepe a történeti korpuszok építése, melyek kiváló alapanyagot szolgáltatnak az elméleti és történeti nyelvészeti kutatásoknak. Az elmúlt évtizedekben számos történeti korpuszt fejlesztettek – elsősorban indoeurópai nyelvekre, de a magyarra is készült néhány. Időrendi sorrendben haladva ezek a következők. Az Ómagyar Korpusz tartalmazza az összes ómagyar korból fennmaradt szövegelemet és néhány középmagyar kori bibliafordítást is. A Történeti Magánéleti Korpusz az ó- és középmagyar kor magánéleti nyelvi regiszteréhez közelebb álló műfajokat tartalmazza: 1772 előtti magánlevelekből és peres eljárások jegyzőkönyveiből épül fel. A Magyar Történeti Szövegtár pedig 1772-től, vagyis az újmagyar kor kezdetétől egészen a 20. század végéig tartalmaz szövegeket. Előadásomban ezeket a korpuszokat és a hozzájuk tartozó lekérdezőfelületeket fogom ismertetni, és néhány példán keresztül azt is illusztrálom, hogy milyen kutatási kérdésekre hogyan tudunk választ kapni ezeknek az adatbázisoknak a segítségével.

**kulcsszavak:** történeti korpuszok, nyelvtechnológia, történeti nyelvészet, korpuszok