

Dömötör Adrienne–Novák Attila:

*A korpuszépítő, a felhasználó – és a történeti morfológia útvesztői.  
1772 előtti magánéleti szövegek morfológiailag elemzett adatbázisa*

Egy hónappal ezelőtt zárult le a „Morfológiailag elemzett nyelvtörténeti korpusz a magánéleti nyelvhasználat köréből” című négy éves projektum. (OTKA 81189. sz.) A munkálat arra vállalkozott, hogy terjedelmes adatbázist épít az élőnyelvet leginkább megközelítő ó- és középmagyar kori szövegekből: magánlevelekből és peres eljárások jegyzőkönyveiből. Az előadás bemutatja a korpuszépítés menetét, kitérve a választott módszer előnyeire és nehézségeire. A munkafázisok közül részletesebben foglalkozik az úgynevezett szövegnormalizálás menetével, valamint a morfológiai elemzés gépi és kézi munkálataival. Példák szemléltetik a szakmai kérdések típusait, amelyekkel a korpuszépítő szembesül, amikor nyelvtörténeti szövegeket kell kezelnie, és szó esik arról, hogyan befolyásolják a felhasználói érdekek a korpuszépítő szempontjait. Az előadók ezután adatokkal is jellemezve bemutatják a megépült korpuszt és a keresőfelületet.