

Szakmai záró beszámoló

Mikes-szótár

K 81337

Célkitűzések

A projekt célkitűzése egyfelől a hazai írói szótáriródalom hiánypótló művének elkészítése, másfelől a digitális szótárkészítés speciális módszerének kialakítása volt. A vállalkozással az első magyar elektronikus írói szótár közel jutott a befejezéshez, s az elért eredményei jóval tovább mutatnak a kutatási tervben vállalt feladatoknál.

A kutatás lépései

1. A munka első szakaszában a szövegvariánsokkal együtt digitálisan rögzítettük Mikes Kelemen összes műveinek kritikai kiadását. Az életmű terjedelme mintegy 6000 osztott oldal, zömében prózai szöveg. A főszöveget beszkeneltük, a karakterfelismertetés tesztelése során speciális beállításokat alakítottunk ki a XVIII. századi szöveg sajátosságaihoz. A MEK-kel kötött együttműködési megállapodás lehetőséget teremtett a gyors ütemű szkennelésre és a szövegfelismertetés elvégzésére. A korrektúra során nemcsak a szoftveres szövegfelismertető hibáit javítottuk, hanem a kéziratok és a mikrofilm alapján korrigáltuk a kritikai kiadás nyomdahibáit is. Elkészítettük Mikes Kelemen összes műveinek betűhív átiratát elektronikus formában. A hibák javításával egyidejűleg előkészítettük a szövegeket szótári és adatbázisban történő feldolgozásra: a teljes szöveganyagba tartalmi jelölőelemeket, ún. XML tageket illesztettünk. Külön jelöltük többek között a versbetéteket, a különféle címfokozatokat, az idegen nyelvű szövegrészleteket, a fordításrészleteket, az idézeteket, a margináliákat és a rövidítéseket, azaz minden szótári feldolgozáshoz szükséges egységet. A műcímekhez kialakítottuk a rövidítésjegyzéket. A digitálisan rögzített anyag lehetővé tette, hogy különféle szövegstatistikai vizsgálatokat végezzünk.

2. Létrehoztuk a Folio-adatbázist, amely új szókeresési lehetőségeket nyújt a szövegek feldolgozásához, az XML-formátumban történt rögzítésnek, valamint a különféle szókeresési módzatoknak köszönhetően.

3. Az elektronikusan rögzített anyagból szoftveres támogatással elkészítettük a konkordancialistát, amely tartalmazza az összes szóalakot a hozzájuk tartozó szóelőfordulásokkal, mennyiségi mutatókkal, valamint minimális, gép generálta szöveggörnyezettel és a forrásmegjelöléssel együtt. A konkordancalista önmagában is többféle kutatás forrásanyagául szolgált a projekt tartama alatt.

4. Kidolgoztuk a szótári struktúrát a konkordancialista anyagának feldolgozásához. Ennek megfelelően kialakítottuk a szócikkstruktúrát, amelybe grammatikai szempontokat figyelembe véve, a mai címszónak megfelelően rendszereztük a példamondatokat. A történeti anyag rendszerszerű feldolgozását megnehezítette a rendkívüli terjedelem: a feldolgozandó szöveganyag nyomtatásban mintegy 60 ezer oldalt tesz ki, több mint 1,5 millió példamondattal. A szóalak, szóelőfordulás megfelelő címszóhoz illesztése jelentős szellemi befektetést jelent, mivel minden egyes szót a maga kontextusában kellett értelmezni, minden egyes példamondot a megfelelő címszóhoz, azon belül alakváltozathoz és a megfelelő szóalakhoz kellett sorolni. A gépi automatizmusok ebben a munkafázisban csak minimális segítséget nyújtottak. A régies írásmódú alakok kézi feldolgozásának nehézsége, ezen belül különösen a tulajdonnevek, idegen szavak változatos formáinak egységes felvétele jelentősen lassította e feladat elvégzését. Ennek oka, hogy a gépi konkordancialista a két szóköz közti egységet tudta csak szónak tekinteni, a régies írásmód miatti egybe-, illetve különírási alternációkat egyesével, kézzel kellett feldolgozni.

6. A konkordncialista feldolgozásával párhuzamosan kidolgoztuk a szerkesztési elveket, amelyeket szabályzatban rögzítettünk, és főbb elemeiben a szótár internetes felületén megjelenítettünk. A problémás esetek tisztázása után kialakítottuk a végleges szócikkstruktúrát. Grammatikai, lexikológiai és felhasználói szempontokat egyaránt figyelembe vettünk. Elvégeztük a címszavak más szótárakkal történő összevetését, meghatároztuk a tulajdonnevek, idegen szavak, török szavak felvételének rendszerét. Létrehoztunk egy újfajta, ún. referencia-címszó típust.

7. A szócikkek minőségének és az egységes színvonal biztosítása érdekében szótári munkában nagy tapasztalattal rendelkező szakemberrel lektoráltattuk az elkészült anyagot.

8. Újabb kutatási feladatokat is meghatároztunk, amelyek a tervezéskor még nem, de a projekt előrehaladtával már körvonalazódtak, és szükségesnek bizonyultak. Azáltal, hogy a teljes szöveganyag részint számítógépes módszerrel, részint szóról szóra történő kézi feldolgozással történt, eddig meg nem válaszolt kérdések tisztázására nyílt lehetőség: a jegyzetapparátus adatainak a ponosítására, különféle textológiai, lexikológiai, grammatikai, filológiai vizsgálatokra.

9. Az informatikai munkálatok első lépéseként megterveztük a szövegkorpusz adatszerkezetét, majd ezt követően félautomatikus módszerrel a rögzített szövegkorpusz hibáit javítottuk. A teljes, strukturált konkordancialista elkészítése kizárólag informatikai eszközökkel történt. Jelentős mennyiségű szócikk elkészítését követően az internetes megjelenítéshez megterveztük a szótár teljes körű online megjelenítését, és létrehoztuk a szótári portált, amelyhez speciálisan e munkához

kialakított szótári keresőrendszert dolgoztunk ki, s biztosítottuk a hosszútávú karbantartási, javítási, fejlesztési-bővítési munkálatok elvégzését is. Az 1000 feletti előfordulású szavak (közel 150) előrendezéséhez a pályázat utolsó évében megbízási szerződés keretében az eredeti munkatervben nem szereplő két külső munkatársat vontunk be.

Elért eredmények

1. Mikes Kelemen összes műveinek betűhív átirata elektronikus formában elérhetővé vált a MEK-en (<http://mek.oszk.hu/09000/09000/>).

2. Az elkészült Mikes-korpusz a középmagyar kor jelenlegi legnagyobb digitális szövegtörzs. A több szempontú keresési lehetőségeket nyújtó adatbázis fontos forrássá vált különféle történeti, irodalmi, grammatikai, lexikológiai kutatások számára. A konkordanciaalista a szótárkészítés alapjaként a teljes életműben lehetővé teszi a szóalakonkénti kereshetőséget, emellett nyelv- és irodalomtörténeti kutatásokhoz is példanyagot biztosít.

3. A Mikes-szótár jelenleg a legnagyobb terjedelmű, írói életművet feldolgozó magyar szótár, s az első elektronikus teljes írói szótár. Nemcsak egy-egy szónak a kontextuális vonatkozásait láttatja, hanem egy kevésbé feldolgozott korszak meghatározó szövegtörzsét is nyújtja. Mikes összes szava megtalálhatóvá válik a mai címszavak szerint rendezve. A szócikkstruktúra a későbbiekben további szerkezeti bővítésekre ad lehetőséget, különös tekintettel a jelentésre és a frazeológiára. A szócikkek XML struktúrája révén címszótípusokra és különféle szó megjelenési módok szerinti régies formákra is kereshetünk. A *b, c, cs, d, dz, dzs, e, é, f, g, gy, h, i, í, j, k, o, ó, ö, ő, t, ty* betűk teljes feldolgozottsága mellett az abc összes betűjéből készült már el szócikk, a jelenlegi állomány kb. 12 500 kész szócikk. Más szótárakkal történő összevetés eredményeképpen az online felületen listázhatjuk Mikes eddig nem szótárzott szavait.

4. Kialakítottuk a szótár internetes arcát, és létrehoztuk a több szempontú keresést lehetővé tevő lekérdező-felületet. Ezen egyaránt kereshetünk külön címszó, szóalak, szóelőfordulás szerint a teljes életműből, és egy, két vagy több műből. Külön listázhatók a címszóállományon belül a tulajdonnevek, a saját szóalkotások és a török szavak. Mindezek a keresési lehetőségek kombinálhatók is. Utalások révén jelöltük a szótáron belüli és kívüli hivatkozásokat. Elkészítettük a felhasználók számára szükséges részletes dokumentációt. Az eddig feldolgozott anyagot közzétettük az interneten (<http://mikesszotar.iti.mta.hu>), s azt folyamatosan bővítjük.

5. A munka már a projekt kezdeti szakaszában országos szakmai érdeklődést keltett. A szótár irodalomtörténeti, grammatikatörténeti vizsgálatok alapjává vált, amelyek korszerű, szemléletváltó eredményeket hoztak az alapkutatásokban. Az új módszernek köszönhetően, amellyel történeti szövegeket informatikai eszközökkel dolgozunk fel, a projekt helyet kapott az egyetemi oktatásban is. A kutatás utolsó három évében előadásokat tartottunk az ELTE Magyar nyelv és irodalom szak Korpusznyelvészet kurzusán, a BME Méréstechnika és Információs Rendszerek Tanszékével együttműködésben önálló labor témát írtunk ki a hallgatók számára. Konzultációkat tartottunk, egyetemi hallgatók (ELTE, BME) szakmai gyakorlatát irányítottuk.

6. Informatikai fejlesztések eredményeképpen elkészült az automatikus, kereshető, strukturálható konkordancialista; szógyakorisági mutatók készültek; megvalósult a teljes Mikes-korpusz Folio-adatbázisa, és az internetes szótári portál a történeti szóalakokhoz illeszkedő keresési módokkal.

7. Nyári szakmai gyakorlata során a BME egy hallgatója olyan eljárást dolgozott ki, amellyel a Mikes-korpuszból azok a visszatérő mondat-, illetve szövegrészletek listázhatók, amelyek nem feltétlenül teljes szó szerinti egyezések, hanem variánsként szoros hasonlóságot mutatva konkrét szókeresés nélkül is kigyűjthetők. Ezzel megteremtettük a későbbiekben elvégzendő textológiai-filológiai kutatások alapját, amely eddig megválaszolatlan kérdések tisztázására adhat lehetőséget.

8. A szótár munkálatai lehetőséget adtak arra, hogy kidolgozzuk az elektronikus írói szótár speciális műfajának a módszertanát. A kötet formában meglévő művek digitalizálásától kezdve a korszerű XML feldolgozásán át, a lexikográfiai munkákig az összes munkafázisra kialakított eszközkészlettel rendelkezünk, melyek más hasonló munkák alapjául szolgálhatnak a jövőben.

9. A szótár készítésén túlmutató, a projekt kezdeti szakaszában előre nem tervezhető eredményeink jelentős mértékben gazdagították az alapkutatásokat: grammatikatörténeti, irodalomtörténeti vonatkozásban. Számolni kell a szótár további, ma még rejtett és előre nem látható alkalmazási lehetőségeivel, elsősorban a nyelv- és irodalomtudományban, azon belül különösen a grammatikatörténet, szókészlettan, textológiai területén.

10. Szakmai együttműködéseket alakítottunk ki hazai és külföldi kutatóinézetekkel, amelynek során megismertettük munkánkat más szakemberekkel. Összevethető nemzetközi kutatások

munkafolyamataiba pillantottunk be, és tapasztalatokat szereztünk a további fejlesztések megtervezéséhez.

11. Folyamatosan számot adtunk kutatásaink eredményeiről konferenciákon, tanulmánykötetben, szakfolyóiratokban. A közzétett publikációk száma: 21. Ebből könyv: 2, könyvfejezet: 4, folyóiratcikk: 6, konferenciatanulmány: 7, ismertetés, egyéb 2.