

Babarczy Anna

tudományos munkatárs
Magyar Tudományos Akadémia
Nyelvtudományi Intézet
Pszicho-, Neuro- és Szociolingvisztikai osztály
Pszicho- és Neurolingvisztikai Kutatócsoport

A Budapesti Műszaki és Gazdaságtudományi Egyetem Kognitív Tudományi Tanszékének egyetemi docense. Elismert kutatója és oktatója a gyermeknyelvnek és a pragmatikának, melyekhez korpusznyelvészeti eszközöket is használ. Érdeklődési körébe tartozik a kísérleti pragmatika, az absztrakció pszicholingvisztikája, és a szó szerinti, valamint a nem szó szerinti jelentések automatikus gépi azonosítása. Több nagyszabású projekt vezetője és résztvevője volt ezekben a témákban, illetve egyetemi oktatóként számos kurzust tartott ezekkel kapcsolatban.

Gyermeknyelvi korpuszok és erőforrások

Az előadás két gyereknyelvi korpuszt és a hozzájuk tartozó nyelvtechnológiai eszközöket mutatja be. Az egyik, a CHILDES (Child Language Data Exchange System, <https://childes.talkbank.org>) egy nemzetközi vállalkozás Brian MacWhinney, a Carnegie Mellon University pszicholingvista professzora szervezésében. A CHILDES kutatási célokra szabadon hozzáférhető; egy folyamatosan bővülő, néhol hanganyaggal vagy videóval is összekötött gyermeknyelvi spontánbeszéd és történetmesélés korpuszból, továbbá a korpusz elemzését, valamint újabb anyagok átírását segítő számítógépes eszközökből áll. Az angol és néhány más nyelvű anyag egyértelműsített morfológiai elemzéssel együtt érhető el, a magyar anyag jelenleg elemzetlen. A rendszer kínál előre kidolgozott elemzési paradigmákat, és lehetőséget ad saját annotációs rendszerek kidolgozására és alkalmazására.

A másik bemutatott korpusz a MONYEEK (Magyar Óvodai Nyelvi Korpusz), Mátyus Kinga doktori disszertációjának anyaga, melyet az én témavezetésemmel hozott létre, és amely az MTA Nyelvtudományi Intézetének honlapjáról érhető el (<http://www.nytud.hu/oszt/korpusz/monyek.html>). A MONYEEK óvodás gyerekekkel szigorú forgatókönyv szerint folytatott interjú és képleírási feladat átírt anyagából áll. Az átírás a CHILDES szabályai szerint történt; a szöveg automatikus morfológiaelemzését a HuMor végezte, melynek kimenetét a PurePos egyértelműsítő rendszer gyermeknyelvre adaptált változata egyértelműsítette utólagos kézi ellenőrzéssel.

kulcsszavak: gyereknyelv, CHILDES, kötetlen annotáció, computer-aided annotáció, kép-hang-szöveg szinkronizáció