

Ember-ember, ember-gép, kommunikáció:
A HuComTech multimodális korpuszról

A HuComTech-korpusz az első, magyar nyelvű dialógusokon alapuló olyan korpusz, ami videó- és audiófelvételek sokszempontú annotációjára támaszkodva az eddigi legátfogóbb képet nyújtja a verbális és nemverbális kommunikáció általános és egyéni sajátosságairól. Célja, hogy hozzájáruljon a különböző helyzetekben megfigyelhető viselkedés jellemző mintázatainak a feltárásához, valamint e mintázatok implementálásához az ember-gép kommunikációban.

Az 50 órás, mintegy másfél millió adatot eredményező, kézi és gépi annotáció kiterjed a beszélők nemverbális, fizikai jegyeinek a leírására (tekintet, fej-, kéz-, testmozgás) éppúgy, mint e jegyek pragmatikai, funkcionális interpretálására (pl. társalgásszervezés, együttműködés, érzelmek). A verbális jegyek annotációja kiterjed a beszéd fonetikájára (a beszéddallam, intenzitás, tempó automatikus lejegyzésére), morfológiájára és szintaxisára. A közel félmillió szövegszó időbeli illesztése lehetővé teszi, hogy a szöveges megnyilvánulásokat akár szószinten is egyeztetni lehessen a nemverbális jegyekkel.

A korpusz adatai egy adatbázison keresztül is hozzáférhetővé válnak nyelvészek, kommunikáció-kutatók, pszichológusok, nyelvtechnológusok számára.

A HuComTech korpusz egy még nem végleges változata elérhető az alábbi címen:

https://clarin.nytud.hu/ds/imdi_browser/
under External Resources

<https://archive.mpi.nl/islandora/search/%2A%3A%2A?f%5B0%5D=cmd.Country%3A%22Hungary%22>