

A szótárkészítés támogatása párhuzamos korpuszokon végzett szóillesztéssel

Héja Enikő

MTA Nyelvtudományi Intézet

Az előadásban egy kétnyelvű szótárak készítésének gépi támogatására irányuló módszert ismertetünk. A javasolt megközelítés alapja a párhuzamos korpuszokon végzett automatikus szóillesztés. A korpusz-vezérelt módszertan, ezen belül különösen a párhuzamos korpusz használata több okból is hasznosnak bizonyult a lexikográfia számára. Ezek közül a legfontosabb, hogy – megfelelő méretű reprezentatív korpusz használatával – a javasolt megközelítés garantálja, hogy a legrelevánsabb fordítások fognak szerepelni a szótárban. További előnyt jelent, hogy ha több fordítás is elérhető, az automatikusan generált fordítási valószínűségek alapján ezeket sorrendezhetjük, így biztosítva, hogy a leggyakrabban használt fordítási ekvivalensek elöl szerepeljenek a szótári bejegyzésben. Ezen felül, mivel az összes korpuszbeli példamondat könnyedén hozzáférhető, a poliszém jelentések közül nagy mennyiségű természetes adat alapján választhatjuk ki a legmegfelelőbbet. A két fenti tulajdonság különösen alkalmassá teszi az általunk javasolt módszert aktív szótárak előállítására.