

Светла Коева, Ивелина Стоянова, Росица Декова
Институт за български език „Проф. Л. Андрейчин“
при Българската академия на науките
София, България

БЪЛГАРО-АНГЛИЙСКИ-Х+ ПАРАЛЕЛЕН КОРПУС

(Резюме)

След кратък преглед на съществуващите паралелни корпуси, включващи български език, статията се спира обстойно на структурата на Българско-английския-Х паралелен корпус, общите принципи при съставяне и структуриране на паралелните корпуси, както и метаданните, използвани при описанието на паралелните корпуси. Подробно са представени отделните субкорпуси – Художествена литература, Публицистика, Законодателство на Европейския съюз, Субтитри, Здравеопазване (административни текстове на Европейската агенция по лекарствата). Описани са различните начини за тяхното съставяне, дадени са детайлни данни за езици и брой думи на всеки от субкорпусите, разпределението на текстовете по тематична област, жанр и период на съставяне. Специално внимание е обърнато на аотирането на корпуса на различни равнища на лингвистична анотация: морфологично, морфосинтактично, синтактично и семантично, както и на възможностите, които произтичат от това аотиране. Поради специфичността на създаване на този вид (сравнително нов за България) езикови ресурси, сме се спрели и на позицията на паралелните корпуси по отношение на Закона за авторското право и сродните му права.

Ключови думи: паралелни корпуси, корпусна лингвистика, анотация

✉ Светла Коева
svetla@dcl.bas.bg
✉ Ivelina Stoyanova
iva@dcl.bas.bg
✉ Rositsa Dekova
rosdek@dcl.bas.bg