

Светла Коева

Институт за български език „Проф. Л. Андрейчин“  
при Българската академия на науките  
София, България

### КОРПУСНАТА ЛИНГВИСТИКА В СПИСАНИЕ *БЪЛГАРСКИ ЕЗИК*

В световен мащаб корпусната лингвистика има сериозни постижения и традиционно приложение основно при създаването на различни типове речници и обучението по чужд език. Компютърната лингвистика също използва големи колекции от анотирани и неанотирани текстове, за да извлича статистически закономерности за употребата на дадено лингвистично явление. Компютърната лингвистика обаче може да бъде в помощ и при ефективното използване на различен тип корпусни данни за нуждите на лексикографията, при формулирането на правописни и правоговорни правила, различен тип граматика и т.н. За съжаление невсички граматика и лексикографи в България осъзнават необходимостта от приложението на компютърната лингвистика за усъвършенстване на работата им от гледна точка на прецизност, пълнота и бързина (още по-малко те си сътрудничат ефективно компютърни лингвисти).

Целта на този брой на списание „Български език“ е да се популяризира корпусната лингвистика, да се отразят постиженията в българската корпусна лингвистика и да се насочи вниманието на изследователите към широките възможности на компютърната лингвистика, включително и за обработка на корпусни данни. В книжката са включени различен тип статии – уводната статия показва ефективното приложение на компютърната лингвистика, за да се подпомогне и оптимизира лексикографската работа (Рандъл и Килгариф) – в програмата Sketch Engine има и български анотирани корпуси, както и българска граматика на скици за употребата на думите, съставена от авторката на предговора. Представен е Българският национален корпус – неговата структура, метаданни, анотация, системата за търсене в корпуса, както и някои примери за приложението му (Коева, Благоева и Колковска). По-тясно специализирани, но поради това особено ценни за изследователите, са Корпусът с устна българска реч (Тишева, Джонова) и Корпусът от студентска научна реч (Алексова, Ласкова, Велкова). Паралелните корпуси – корпуси от преводни документи на два или повече езика, които са съотнесени по изречения, се използват в редица области като машинен превод, търсене на информация в документи на различни езици и т.н. Представени са Корпусът RuN- Euro, многоезиков корпус, който включва и българския език (Грьон, Ро Хауге), и Българо-английският-X+ паралелен корпус (Коева, Стоянова, Декова). Приложението на паралелните корпуси е илюстрирано на основата на балканските езици (Търпоманова), а на Българския национален корпус – в по-обща (Благоева, Колковска) и по-конкретни изследвания (Несторова), както и в бележките за езикова култура (Благоева; Търпоманова). Включена е и статия, посветена на създаването на диахронни корпуси (Димитрова).

В списанието е включена рецензия на *Българо-полската граматика* – съпоставителните изследвания не са възможни без съответния паралелен емпиричен материал (Лясковски), както и на книгата *Българският семантично анотиран корпус* (Бъркалова, Георгиева). Завършваме с хроника за *Третия конгрес по корпусна лингвистика*, проведен през пролетта на 2011 година.

Трета книжка на списанието представя някои от основните постижения на българската корпусна лингвистика при създаването, анотирането и използването на

различни видове корпуси. Едновременно с това се показват основните насоки в бъдещото развитие на корпусната лингвистика: съставянето на все по-големи по обем корпуси, отразяващи максимално достоверно състоянието на даден език през определен етап от неговото развитие; все по-детайлното аотиране на корпусите с подходяща лингвистична информация; все по-широкото използване на компютърната лингвистика за обработка на лингвистичните корпуси и приложение на получените резултати в различни изследователски сфери: лексикография, граматични изследвания, компютърна обработка на естествения език.

✉ Светла Коева  
*svetla@dcl.bas.bg*

Публикувано: 30 септември 2011