

Seminarium

w Pracowni Historii Języka Polskiego XVII i XVIII w. IJP PAN
Warszawa, 21 listopada 2013

Korpusy tekstów dawnych i gwarowych

Cel spotkania

- Prezentacja projektów realizowanych i planowanych.
- Wymiana doświadczeń między zespołami realizującymi poszczególne projekty.
- Wypracowanie zasad, które będą umożliwiały (a przynajmniej nie będą uniemożliwiały) w przyszłości korzystania z korpusów (zwłaszcza polskojęzycznych) za pomocą jednego zestawu narzędzi.

Proponowany program seminarium

- 11:00-11:15 – powitanie uczestników, zagajenie, omówienie programu, sprawy organizacyjne.
- 11:15-14:15 – prezentacja projektów (ok. 0,5 h na każdy projekt):
 - korpus łaciny średniowiecznej w Polsce (IJP PAN);
 - korpus polszczyzny XVI wieku (IBL PAN);
 - korpus tekstów polskich z XVII i XVIII w. (IJP PAN);
 - korpus staropolski (IJP PAN);
 - korpus polszczyzny XIX-wiecznej (IJP UW);
 - korpus polskich tekstów gwarowych (IJP UW).

Proponowany program seminarium

- 14:15-15:00 obiad (w sali konferencyjnej);
- 15:00-17:00 dyskusja i prezentacje uzupełniające.

Zagadnienia do dyskusji (1)

- Czy jest szansa, by w przyszłości korpusy historyczne i gwarowe mogły stać się podkorpusami NKJP?
- Jakie warunki musiałyby zostać spełnione?
- Czy poszczególne zespoły są zainteresowane realizacją tak wyznaczonego celu?

Zagadnienia do dyskusji (2)

- Zakres i sposób znakowania tekstów.
- Czy możliwe jest uzgodnienie sposobu i zakresu znakowania strukturalnego i językowego?
- Inne rodzaje znakowania – ich zakres i metody wprowadzania (morfosyntaktyczne, leksykalno-semantyczne).

Zagadnienia do dyskusji (3)

- Problemy związane z wyszukiwarką.
- Jak radzić sobie z wariantywnością zapisów form tej samej jednostki leksykalnej (wariantywnością ortograficzną lub – w wypadku tekstów gwarowych – fonetyczną)?
 - prezentacja rozwiązania zastosowanego przez Aleksandrę Wieczorek w jej pracy doktorskiej.
- Czy nowy Poliqarp może stać się narzędziem obsługującym wszystkie (polskojęzyczne) korpusy?
- Możliwości zastosowania innych narzędzi, np. wyszukiwarki Holmes.

Zagadnienia do dyskusji (4)

- Problemy związane z lematyzacją (sprowadzaniem form fleksyjnych do postaci hasłowej).
- Czy nowy Morfeusz może stać się narzędziem uniwersalnym?
- Jakie warunki muszą spełniać słowniki dołączane do Morfeusza?
- Możliwości zastosowania innych narzędzi, np. wyszukiwarki Holmes.