

ЛИНГВИСТИЧНИ КОМПОНЕНТИ НА ДИГИТАЛНА БИБЛИОТЕКА НА БЪЛГАРСКИЯ ПЕСЕНЕН ФОЛКЛОР

д-р Константин Рангочев

НЕОБХОДИМОСТ ОТ РЕЧНИЦИ

- Основна съставляваща на лингвистичните изследвания на българския фолклор е анализът на неговата лексика.
- Счита се, че данните в един честотен речник са достатъчно достоверни, ако той съдържа минимум 20 000 лексикални единици.
- Колко и какви лексеми съдържа?
- Има ли и каква е доминацията на едни групи лексеми и съответно липса на други?
- Какви парадигматични групи има/няма във лексиката българската фолклорна песенност?
- До днес този лингвистичен анализ не е направен.
- Основна проблем – честотните речници се създават и използват от и за лингвисти – специалисти по динамиката на развитието на българския език.

Типове речници

- **Честотен речник.** Започват да се създават в края на XIX в. Основната им функция е да решат проблемите пред чуждоезиковото обучение – показват кои са най-често употребимите думи в един език и с тях да започва обучението по чужд език
- **Речник – конкорданс.** Създават се още през Средновековието за книгите на Библията. Това са речници, които показват лексемата в нейния контекст – т. е. показват лексемата заедно с предходната и следходната лексема на изучаваната.

Честотен речник

- Честотния речник представлява списък от лексеми, подредени по азбучен ред, които се съдържат в определени текстове. На всяка дума се дава морфологична характеристика (това не е задължително) и число, което показва колко пъти думата се среща в съответните текстове. Честотния речник съдържа рангов списък на най-често срещаните думи. В него думите са подредени по намаляване на честотата им.

Речник - конкорданс

- Това са речници, които показват лексемата с/в нейния контекст - показват предходната една (или повече) лексема и следходната (или повече) лексема на изучаваната. Пример: „*Вино **пият**, петдесет юнака*” – подчертаната лексема е изучаваната, а лексемите в курсив са нейния контекст. Естествено, при песните това може да бъде (и най-често е) речник-конкорданс на стиховете им (някои библейски конкорданси са направени така), а при наративните текстове (описания на обреди и пр.) – изреченията, в които те се съдържат (от точка до точка ...).

Честотните речници по света ...

- 1898 г. – публикуван е първият честотен речник е на немския език на Кединг (290 текста, 11 000 000 лексеми).
- 1911 г. – честотен речник на Елдрич на английския език
- 1920 г. – честотен речник на Кенингстън испанския език

И у нас

- 1968 г. Дипломна работа на В. Радованова “Репрезентативен фреквентен речник на текст с дължина 500 000 вокабулеми, СУ “Св. Кл. Охридски”, непубликуван.
- 1987 г. Николова, Цв. Честотен речник на българската разговорна реч. София. – 100 000 лексеми.
- 1994 г. Рангочев, К. Структурни особености на епическия текст (по материал от българския юнашки епос), Канд. дис., СУ “Св. Кл. Охридски”, Приложения I – X (Честотен речник на българския юнашки епос. 40 062 лексеми - непубликуван.
- 2001 г. Електронен архив на българските диалекти – <http://www.bultreebank.org/veda/indexbg.html>

Проблемите ...

- Радованова – стандартизирана извадка: използват се текстове от различни стилове с цел елиминиране на стилового разнообразие; липсват на антропоними и топоними.
- Николова – липсват антропоними и топоними, но тяхната фреквентност е включена в общия брой на съществителните имена.
- Рангочев – относително малък обем на извадката; липсва лингвистично описание на лексемите
- ЕАБД –направен е само върху корпуса “Веда Словена”, т. 1 и т. 2. Налице е смесване на метаданните за “Веда Словена”: предговора на Ст. Веркович, заглавията на песните, поставени от него - със самия корпус от фолклорни текстове.

И ИЗВОДИ ДО ТУК ...

- Лексиката на българските фолклорни песни и проза като цяло е неизучена.
- Всички речници, които са направени до сега с едно изключение (Рангочев 1994) са направени така, че да бъдат използвани изключително от специалисти с интереси в областта на развитието на българския език, а не от широк кръг специалисти - етнологзи, фолклористи, културолози и дори лингвисти на текста.
- Езикът на българския фолклор е неизследван

Сравнението I

Радованова 1968

- и – 21 800
- на – 20346
- 1. съм – 18 343**
- се – 12 235
- да – 11 278
- в – 10 244
- от – 8 136
- за – 6 685
- си – 6 305
- с – 5 926
- не – 4 189
- че – 3 154
- ще – 2 745
- по – 3 120
- това – 2 226
- а – 2 223
- 1. той – 1797**
- но – 1746
- при - 1 521
- които – 1 382
- го – 1341
- до – 1 317
- 1. аз – 1 310**
- 2. един – 1 178**
- като – 1 051

Николова 1987

- 1. съм – 4 041**
- и – 3764
- да – 3 148
- 1. аз – 2 433**
- 2. той – 2 288**
- не – 1 956
- се – 1 928
- този – 1 701
- на – 1 669
- ти – 1 249
- ще – 1 183
- 1. един – 1 131**
- в – 1 099
- си – 1065
- 1. казвам – 1 045**
- тя – 1044
- 1. викам – 1 031**
- те – 1014
- какъв – 938
- за – 913
- че – 874
- с – 809
- 1. имам – 768**
- така – 742
- от – 731

Сравнението II

Николова 1987

1. съм – 4 041
1. и – 3764
2. да – 3 148
3. аз – 2 433
1. той – 2 288
1. не – 1 956
2. се – 1 928
3. този – 1 701
4. на – 1 669
5. ти – 1 249
6. ще – 1 183
1. един – 1 131
1. в – 1 099
2. си – 1065
1. казвам – 1 045
1. тя – 1044
1. викам – 1 031
1. те – 1014
2. какъв – 938
3. за – 913
4. че – 874
5. с – 809
1. имам – 768
1. така – 742
2. от – 731

Рангочев 1994

1. съм – 1342
1. да – 1 247
2. си – 548
1. Марко – 1 036
1. се – 828
2. на – 801
3. и – 796
4. па – 657
5. у – 582
6. я – 553
7. та – 526
8. не – 412
1. юнак – 396
1. го – 338
2. му – 320
3. че – 318
4. а – 286
1. кон – 276
1. от – 272
2. ми – 233
1. ти – 225
1. що – 222
2. по – 218
1. добър – 201
1. три – 201

Речници - конкорданс

- Това са речници, които показват лексемата с/в нейния контекст - показват предходната една (или повече) лексема и следходната (или повече) лексема на изучаваната. Пример: „*Вино **пият**, петдесет юнака*” – подчертаната лексема е изучаваната, а лексемите в курсив са нейния контекст. Естествено, при песните това може да бъде и е речник-конкорданс на стиховете им, а при наративните текстове (описания на обреди и пр.) – изреченията, в които те се съдържат (от точка до точка ...). Създаването и използването на речници-конкорданси на текстовете от ДББФП би дало чудесна възможност на фолклористи и етнологзи да изучат редица проблемни области като: наличие/липса на формули във фолклорната песенност и епика, структурата на фолклорния текст и пр.

Процедури: честотни речници

- Честотни речници на цялата БД: азбучен и рангов списък
- Честотни речници на антропонимите
- Честотни речници на топонимите
- Честотни речници на отделни функционални групи, обособени чрез метаданните : по тематика, по функция, по информатор, по населено място, по дата на запис ...
- NB! – минималният обем на корпуса от текстове трябва да бъде 20 000 лексеми, т. е. приблизително 80 /100 песни.

Процедури: речници - конкорданс

- Речници - конкорданс на антропонимите
- Речници - конкорданс на топонимите
- Речници - конкорданс на отделни функционални групи, обособени чрез метаданните : по тематика, по функция, по информатор, по населено място, по дата на запис ...

И накрая ...

- Търсения в текстовата база данни и в гъвкави корпуси от текстове
 - По отделна дума
 - По 2 и повече думи – търсене на словесни формули във фолклорната лексика: „*Вино пият*“, „*Седнал Марко*“.
 - Търсене по корен на дума за изследване на фолклорното словообразуване: ‘*пия*’ – „*пи**“ (*пия, пих, пиха ...*)
 - ...