

# ЕНЦИКЛОПЕДИЧНИЯТ ТЕКСТ ПРЕЗ ПРИЗМАТА НА ЛИНГВОСТАТИСТИКАТА

Маргарита Димитрова

## 1. Обект и обем на изследването.

Обект на изследване е езикът на енциклопедичния текст, представен в изданието “Българска енциклопедия “А...Я”, БАН, БЕ, КК “Труд”, С. 1999 год. Наблюдаваният корпус се състои от 10 000 графични думи и е формиран чрез 10 случайни извадки по 1000 графични думи всяка<sup>1</sup>. Обемът от 1000 графични думи за първична извадка притежава характеристиките на свързан текст (вж. Якубайтис 1981:48-62). Изследваният корпус е достатъчен за изводи, основани на приложените в работата статистически методи.

Изследването е осъществено изцяло чрез високите технологии. Наблюдаваният корпус е в електронен вид<sup>2</sup>, а обработката му и извличането на данни са извършени чрез пакет от три програми, създаден за нуждите на Електронния архив при Института за български език - БАН. Използването на високите технологии позволи да се разшири кръгът на наблюдаваните явления. Освен частите на речта, характеризиращи текста в рамките на езиковата система, се изследват и явления, които характеризират енциклопедичния текст като речев продукт.

Изборът на енциклопедичен текст като обект на лингвостатистическо изследване е свързан и със състоянието на изследванията по такава методика върху български текстове. Досега са изследвани художествен текст (вж. Мутафчиев 1967; Мутафчиев 1972; Димитрова 1995), публицистичен текст (вж. Тодорова 1993; Тодорова, Панчовска 2001) и разговорна реч (вж. Николова 1987). В рамките на традиционните пет функционални стила неописани остават административно-деловите и научните и научнопопулярните текстове. Като представителен за последната група беше избран енциклопедичният текст.

В прагматичен план енциклопедията, както и всеки енциклопедичен текст, обслужва единен дискурс независимо от тематичното разнообразие. Този дискурс е различен от дискурсите, обслужвани от други текстове, например поезия, художествена проза, закон, вестникарска статия, научна статия в някоя частна област и т. н. От тази гледна точка в настоящото изследване текстът на енциклопедията се разглежда като хомогенен спрямо регистровата си принадлежност. Енциклопедичният текст е „най-популярният“ от специализираните текстове. Той е мостът между специалното знание и общата информираност. Същевременно енциклопедията е интегрален текст, докато научните и научнопопулярните текстове засягат дадена тясна област.

## 2. Минимална наблюдавана единица.

В изследването минимална единица за наблюдение е графичната дума. В повечето досегашни лингвостатистически изследвания върху българския език минимална единица за наблюдение е граматичната дума (Мутафчиев 1967; Мутафчиев 1972б; Тодорова, Панчовска 2001; Димитрова 1995), а в единични – графичната дума

---

<sup>1</sup> Тъй като всяка извадка завършва с цяло изречение, обемът им се движи от 994 до 1002 думи, а общият обем на изследвания корпус е точно 9988 думи.

<sup>2</sup> Общият обем на предоставения от КК „Труд“ на Електронния архив при Института за български език електронен текст е 5521 KB (приблизително 3116 стандартни печатни страници). Текстът съдържа около 890 000 графични думи, обединени в приблизително 20 500 статии.

(Николова 1987). Този подход не е нов и е приложен в лингвостатистически изследвания върху други езици (срв. ЧСРЯ; ЧРПЕ). В съвременната практика за работа с лингвистични бази данни и в корпусната лингвистика (вж. напр. Синклер 1992; VulTreeBank 2004; Коева 2004) се работи с графична дума. Този избор е свързан и с развитието на лингвистичната теория и с промяна в парадигмите на научното изследване: съвременните формални теории за езика в повечето случаи анализират поотделно частите на сложните форми на думите и на съставните думи като експлициращи различни езикови значения (срв. напр. Пенчев 1993).

### 3. Енциклопедичният текст – съчетание от много знакови системи.

Настоящото изследване е основано върху десетте части на речта според класическата българска граматика (вж. ГСБКЕ, т. II, 1983). В анализа водещ е езикът в неговата системност, който първично и пряко се реализира в устната реч. При писменото представяне на думите в изследвания корпус са използвани различни средства. Особено типични за енциклопедичния текст са многобройните езикови и графични съкращения. Те обаче не са обект на наблюдение и затова като словоформи на една и съща лексема бяха определени следните явления:

- пълният запис
- обичайното съкращение
- функционалното съкращение в определена статия

Напр.:

- (1) *Тундрова растителност, **аркт.** пустиня.*
- (2) *Субарктичен и **арктичен** климат с мор. влияние по крайбрежието (измерени са едни от на-'ниските темп. – ок. -70°C).*
- (3) *Адм. **ц.** - Готхоб.*
- (4) *ГРЕНЛАНДСКИ АНТИЦИКЛОН - област с вис. атм. налягане в ср. многогод. карти с **център** над Гренландия.*

В писмената комуникация съжителстват много знакови системи – както езикови, така и графични. Това е особено очевидно в енциклопедичен текст. Значителното наличие на чужди собствени имена в изследвания корпус наложи обособяването на определителния член като допълнителна част на речта. Членът присъства в транскрибирани на български собствени имена от различни езици (напр. “*Ла Скала*”, *Харун ар Рашид*, *Али Шейх*, *Лаго ди Комо*). Включването на такава част на речта като равноредна на традиционните десет отразява мултикултурната страна на енциклопедичния текст: в него чрез собствените имена присъстват своеобразни „цитирания” от езици с друг граматичен строеж.

Важна графична особеност на енциклопедичния текст представлява неговата интегралност по отношение на запис. Той се фиксира не чрез една, а чрез много графични системи. Това се оказва негова същностна характеристика сред останалите писмени текстове. В художествения текст (традиционен за лингвистичните изследвания) се използва азбучната знакова система. Повече графични системи се срещат в публицистичните и в административните текстове (букви в небуквена функция, цифри, абревиатури, различни знаци, най-често за градус и процент и др.). В научнопопулярните текстове и особено в научните текстове се срещат, освен изброените, и специфични за отделните науки знакови системи (букви от различни азбуки в небуквена функция, формули и други символи с различен характер). Единствено енциклопедичният текст обаче поради интегралния си характер съчетава всички графични системи, съществуващи в българските писмени текстове: азбучен запис, букви в небуквена функция, римски и арабски цифри, символи и други знаци от различните научни области, съкращения с различен характер и т. н.

Очертаната знакова интегралност на енциклопедичния текст доведе до отделянето на още една група графични „думи” – наречена „Знаци”, също равноредна на десетте части на речта. Тази група обхваща графемни редици, които не могат да бъдат определени като принадлежащи към някоя част на речта в българския език. Тук влизат:

- математически, химически и физически знаци, символи, формули и изрази
- означения на измерителни единици
- буквени абривиатури
- десетични дроби
- букви и препинателни знакове с графична функция, извън състава на дума и изречение

Графичните „думи” от групата „Знаци” са еднотипни по няколко свои характеристики. Основен техен общ белег е, че буквите не са използвани в типичната си функция на знакове за фонетични езикови явления. Често тези графични „думи” се съотнасят с повече от една дума в устната реч и представляват изрази или изречения, напр. *АН* - ‘академия на науките’;  $a=b$  - ‘*a* е равно на *b*’; *3,14* - ‘три цяло и четиринайсет’; *?* - ‘три четвърти’ и т. н. Не представляват изрази или изречения само буквените символи в различните точни науки, напр. *a, b, c, ?, σ*, а също и буквите и знаците с графична функция. Подобни явления се наблюдават при публицистичния и административния текст. Там обаче няма такова разнообразие от знаци и количеството им е статистически пренебрежимо. За енциклопедичния текст графичните „думи” от групата „Знаци” са типично явление както с разнообразието, така и с количеството си (по-подробен анализ на групата вж. у Димитрова 2006).

#### 4. Характеристика на енциклопедичния текст в рамките на езиковата система. Части на речта.

Обща представа за съотношението в енциклопедичния текст между количествата на десетте части на речта и приравнените към тях явления – определителния член и знаците дава следната хистограма.



Най-многобройни са съществителните, които заемат около 2/5 от единица текст. Прилагателните и предлозите са също много чести, с почти равни количества, което е естествено при голямата честота на съществителните. За разлика от другите типове текст (срв. напр. Николова 1987; Тодорова 1993; Димитрова 1995) в изследвания корпус числителните са четвъртата по количество група думи. Многобройността им може да се отнесе към спецификите на енциклопедичния текст.

Местоименията, глаголите, наречията и съюзите са отчетливо по-редки в сравнение с посочените по-горе четири части на речта. Като се има предвид обичайната за наречията функция на обстоятелствено пояснение, логично е честотата им да е свързана с честотата на глаголите. По-малкото количество на съюзите се определя от спецификата на изказа в енциклопедичния текст. В него се избягват сложни изречения и често се употребяват отглаголни съществителни (предимно с наставки *-не*, *-ие*) като начин за по-кратко и компактно изразяване чрез номинализация на вербалната фраза, срв.:

(5) **ПРИТРИВАНЕ, лепинговане** - **окончателно обработване** на мет. детайли чрез **отнемане** на метал с абразив или с повърхностно активни в-ва.

Много по-ниската честота на местоименията в сравнение с другите имена може да получи прагматично обяснение. В енциклопедичния текст се предпочита назоваване, а не реферирание. Най-често се среща, разбира се, клистиката *се*, която има само приглаголна употреба (*се* 93.<sup>3</sup>). От другите видове най-чести са относителните местоимения и местоименни наречия (*която* 13. *който* 7. *които* 6. *което* 5. *където* 3. *както* 1.) и кратките форми на личните и на притежателните местоимения (*му* 17. *си* 9. *й* 5. *тях* 5. *им* 3. *я* 3. *ги* 2. *го* 2.). Формите за 1 л. са най-редки (*нас* 1. *нашите* 1.) и се срещат в специфични единични случаи – в заглавия на творби на изкуството. Форми за 2. л. няма в изследвания корпус. Останалите местоимения принадлежат към различни групи и са с единични употреби (*всички* 3. *всеки* 1. *всяко* 1.; *него* 3. *нея* 1.; *себе* 1.; *те* 1.; *негов* 1. *негова* 1. *неговия* 1. *нейна* 1. *нейно* 1.; *кой* 1.; *тези* 1.; *няколко* 3.).

Най-рядката част на речта в енциклопедичния текст са частиците. Такива лексеми бяха установени в половината от корпусите с по 1-2 появи.

Междуметията като част на речта са нехарактерни за енциклопедичния текст. В изследвания корпус не попаднаха междуметия, но в целия текст те се срещат в единични случаи – при заглавия на произведения на изкуството.

Определителни членове като отделни лексеми се появиха само в два от корпусите и са част от чужди собствени имена.

Във всеки от корпусите е представена групата на знаците, което говори, че употребата на разнообразни и/или разнофункционални графични системи е постоянна особеност на енциклопедичния текст.

Следващата Таблица 1. съдържа основните статистически характеристики<sup>4</sup> (вж. напр. Янакиев 1977; Хамерл, Самбор 1990) на енциклопедичен текст при частите на речта и приравнените към тях явления – „Членове” и „Знаци”. Важна информация се съдържа в колоната „Доверителен интервал”. Тя показва (тук с вероятност за грешка от  $\pm 5\%$ ) границите, в които ще се движи количеството на съответния параметър, ако се проведат нови изследвания.

<sup>3</sup> Данните са за целия корпус.

<sup>4</sup> За автоматичното изчисляване на средното аритметично, доверителните интервали, стандартното отклонение и корелацията ми помогна ст. н. с. д-р Е. Стоименова от Института по математика и информатика при БАН, за което сърдечно ѝ благодаря.

Наблюдаваните в изследването явления са разделени в пет съдържателни групи според съотношението между средното аритметично  $\bar{x}$  при всяко явление (за основанията на това групиране вж. Стоименова 2000). По своя характер групите са съдържателни и отразяват езиковите особености на енциклопедичния текст. Наименованието на всяка група отразява мястото ѝ в йерархията и чрез него – значението ѝ за езиковото изграждане на енциклопедичния текст.

### Основни статистически характеристики на частите на речта и приравнените към тях явления

Таблица 1.

Част на речта	Средно аритметично $\bar{x}$	Доверителен интервал -95% +95%		Размах min max		Стандартно отклонение s	Коеф. на вариация v %
<i>Изключително чести</i>							
Съществителни	413,9	399,90	429,90	380	443	19,58	4,73
<i>Много чести</i>							
Прилагателни	172,7	161,61	183,78	150	199	15,49	8,97
Предлози	171,0	155,58	186,42	132	205	21,56	12,61
<i>Чести</i>							
Числителни	80,1	55,64	104,56	16	123	34,19	42,68
Глаголи	69,8	64,23	75,37	57	79	7,79	11,15
Съюзи	44,8	37,84	51,76	29	58	9,73	21,72
<i>Редки</i>							
Местоимения	20,8	14,50	27,10	9	37	8,80	42,33
Знаци	14,6	8,61	20,59	6	35	8,37	57,32
Наречия	9,6	6,62	12,58	4	19	4,17	43,42
<i>Случайни</i>							
Членове	0,8	-0,58	2,18	0	6	1,93	241,52
Частици	0,7	0,11	1,29	0	2	0,82	117,61
Междуметия	0	0	0	0	0	0	0

**Най-често явление** в енциклопедичния текст са съществителните. Този клас думи присъства в най-вече чрез отглаголни съществителни и чрез разнообразни собствени имена (на лица, места, институции и под.).

Прилагателните и предлозите образуват групата на **много честите думи**: те са около два пъти по-чести от следващите ги по честота групи, но са наполовина по-редки от съществителните. Тези две части на речта в синтактичен план функционално зависят непосредствено от съществителните. При текст, почти половината от който заемат съществителните, естествено е прилагателните и предлозите да са следващата по количество група.

В групата на **честите явления** попадат числителните, глаголите и съюзите. В енциклопедичния текст се използват широко безглаголни изречения, но дори и най-икономичният откъм думи текст обаче не може без глаголи и съюзи. Количествата на тези две части на речта се формират от статии, в които се съчетават дескрипция и наратив (вж. Дьо Богранд, Дреслер, Стоянова-Йовчева 1995:184-185) – основно свързаните с история. Броят на числителните, които също са в групата на честите явления, най-силно се колебае в отделните извадки (коефициентът на вариация v% е

малко над допустимата граница от 40%, вж. Носенко 1981). Най-много числителни не се употребяват в текстове от областта на природоматематическите и техническите науки, а в текстове на историческа тема и в персоналии. Следователно темата и типът дискурс са определящи за употребата на по-значителни количества от тази част на речта. За енциклопедичния текст е характерна фактологичността на информацията, постигана чрез съотнасяне с календара и измерителните системи (най-често СИ). Този тип фактологичност е свързан с хора и места, но не и с понятия

Към *редките* за енциклопедичния текст части на речта спадат местоименията и наречията. От приравнените към частите на речта явления в групата на *редките* попадат и знаците. Ниската честота на местоименията може да се обясни със спецификата на изказа в изследвания текст. В енциклопедичния текст нарацията като текстов сорт, при който най-често се корелира, не е обичайна. Типичната за този текст дескрипция е организирана езиково по специфичен начин, често с непълни изречения. Срв.:

(6) *ХЕРКУЛАН (Herculaneum) - ант. град в обл. Кампания, Юж. Италия, близо до Помпей. През 79 е частично разрушен и затрупан с вулк. пепел при изригването на Везувий. Открит през 1709. Археол. проучвания се провеждат системно от 1927.*

Липсата на експлицитни корелации в текста води и до намаляване на броя на местоименията и превръщането им в рядка категория за енциклопедичния текст.

Наречията са редки думи в енциклопедичния текст по две причини. От една страна, използват се по-малко глаголи и това се отразява върху употребата на наречия. От друга страна, обстоятелствените пояснения – обичайната синтактична реализация на наречията – не са задължителен изреченски член при голяма част от глаголните сказуеми. Фактът, че наречията са малко, говори, че за енциклопедичния текст употребата на този клас думи не е наложителна и в комуникативен план.

Знаците, макар и типични за енциклопедичния текст, закономерно попадат сред редките категории. Тази група обхваща доста разнообразни явления. Нейните елементи “дублират” (в широк смисъл) буквения запис на езиковия изказ. Групата на знаците се попълва най-вече от две явления: абривиатурите и знаците за нелинейни мерки. Срв. напр. големият принос само на една сравнително доста кратка статия:

(7) *ГРЕНЛАНДСКО МОРЕ (Greenland Sea) - периферно море, част от Сев. Ледовит ок., между Скандинавския п-в и о-вите Гренландия, Исландия, Ян Майен, Мечи, Шпицберген. Площ 1,195 млн. km<sup>2</sup>. Дълб. до 5537m. Соленост 33,0 - 34,0‰. Приливи до 4,4 m. Лов на тюлени; риболов. Гл. пр-ща: Лонгир, Баренцбург – о-в Шпицберген; Акюрейри, о-в Исландия.*

Частите на речта от петата група са определени като *случайни*. Тази характеристика за определителния член като самостоятелна графична “дума” в български писмен текст е безспорна. Сред случайните за енциклопедичния текст явления естествено попадат и двете части на речта с предимно оценъчна функция – частици и междуметия. Характерна за енциклопедичния текст е липсата на пряка оценъчност при представянето на информацията.

При групите на редките явления и на случайните явления коефициентът на вариация  $v\%$  надхвърля в различна степен граничната стойност. За редките явления може да се предположи, че изследвания върху по-голям корпус ще намалят  $v\%$ . Едва ли обаче трите редки групи ще променят мястото си в скалата. При случайните явления високите стойности на  $v\%$  доказват, че частиците и членовете (също както и междуметията) са нетипични части на речта в енциклопедичния текст.

**5. Характеристика на енциклопедичния текст като речев продукт (текст).  
Лексемно разнообразие. Изречения и параграфи.**

Използването на високите технологии за настоящето изследване позволи да се направят обстойни наблюдения, които са свързани с енциклопедичния текст като речев продукт. Подобни факти се разглеждат само в някои от досегашните лингвостатистически изследвания и само за някои от посочените по-долу характеристики (срв. Тодорова 1993; Димитрова 1995). Приведените данни дават пълна картина на статистическите особености на енциклопедичния текст.

**Лексемно разнообразие**

Таблица 2а.

Основни статистически характеристики							
Явление	Средно аритметично $\bar{x}$	Доверителен интервал -95% +95%		Размах		Стандартно отклонение $s$	Коеф. на вариация $v\%$
				min	max		
Графични думи (най-често словоформи)	998,8	996,9 3	1000,67	994	1002	2,62	2,62
Основни форми (най-често лексеми)	590,9	545,0 1	636,79	463	646	64,15	10,86
Различни графични думи (най-често словоформи)	624,0	591,2 5	653,74	545	670	45,76	7,33

Характеристиката „Лексемно разнообразие” се извежда чрез две съотношения: 1. между броя на графичните думи (най-често словоформи) и броя на лексемите; 2. между броя на графичните думи (най-често словоформи) и броя на различните, неповтарящите се графични думи (най-често словоформи). Тази характеристика разкрива степента на новост в информацията, предадена чрез текст от 1000 графични думи.

Данните показват, че количествата на различните графични думи (най-често словоформи) и на различните основни форми са доста близки. Това говори за голямото тематично разнообразие на енциклопедичния текст: в много случаи една дума се появява в една своя словоформа.

Най-голям дял от повтарящите се словоформи се пада на най-честите думи в корпуса. С честота до 10 появи включително това са основно предлози, съюзи и глагола *съм*, който е представен с две от формите си – *е* и *са*. Въпреки че съществителните като част на речта са около половината от думите в корпуса, отделните съществителни имат рядка употреба. Затова сред думите с честота до 10 употреби включително няма съществителни, а сред думите с честота до 5 употреби включително съществителните са единични.

## Количествени особености на изреченията

Таблица 2б.

Основни статистически характеристики							
Явление	Средно аритметично $\bar{x}$	Доверителен интервал -95% +95%		Размах		Стандартно отклонение $s$	Коеф. на вариация $v\%$
				min	max		
Изречения	94,4	83,96	104,84	63	110	14,59	15,46
Средна дължина на изречение в графични думи (най-често словоформи)	10,4	9,04	11,76	9	15	1,89	18,24
Максимален брой графични думи (най-често словоформи) в изречение	37,7	33,20	42,10	29	53	6,29	16,68

Средният и максималният брой на графичните думи (най-често словоформи) в изречение дават обща представа за синтактичния строеж на енциклопедичния текст. Малкият брой графични думи (най-често словоформи) говори за по-опростен синтаксис, а големият – за по-сложен синтактичен строеж.

Статиите в енциклопедичния текст са организирани езиково еднотипно. Броят на изреченията в отделните извадки от изследвания корпус е съотносим и не зависи както от количеството на статиите, попаднали в съответната извадка, така и от тяхното съдържание. Изреченията в енциклопедичния текст са къси (вж. последните два реда от Таблица 2б.). По този показател няма статистически големи разлики между отделните извадки (вж. колоните „Доверителен интервал” и „Размах” от Таблица 2б.). Няма съществени разлики между отделните извадки и при максималния брой графични думи в изречение. Следователно една от стабилните характеристики на енциклопедичния текст е наличието на кратки изречения.



## Количествени особености на параграфите

Таблица 2в.

Основни статистически характеристики							
Явление	Средно аритметично $\bar{x}$	Доверителен интервал -95% +95%		Размах min max		Стандартно отклонение s	Коеф. на вариация $v\%$
Параграфи	26,4	21,48	31,32	12	36	6,88	26,07
Средна дължина на параграф в графични думи (най-често словоформи)	40,9	29,39	52,41	27	83	16,09	39,35
Средна дължина на параграф в изречения	3,4	2,17	4,65	2	8	1,71	50,37
Максимален брой изречения в параграф	14,4	3,67	25,13	6	56	14,10	104,15

В енциклопедичния текст обикновено всяка заглавка представлява отделен параграф. Изключение правят статиите за държавите, които са оформени в няколко тематични параграфа. Няма свободен избор за смислово графично членение на текста на параграфи според новата микротема. Затова данните от реда „Параграфи” на Таблица 2в. представят количеството нова информация, зададена в енциклопедичния текст чрез по-голям брой статии, а не чрез по-голям брой микротеме, както е при другите типове текст.

В рамките на всеки параграф обемът на информацията и начинът на представянето ѝ са свързани с броя на графичните думи (най-често словоформи) и с броя на изреченията, в които тези графични думи (най-често словоформи) са организирани. Големината на параграфите, оценена в графични думи (най-често словоформи) и в изречения, зависи от комуникативни и прагматични условия. Няма статистически големи различия между тематично различните енциклопедични статии според количествата графични думи и изречения (вж. колоните „Доверителен интервал” и „Размах” в Таблица 2в). Това говори, че независимо от съдържанието енциклопедичният текст се организира езиково еднотипно. Тематичното разнообразие се проявява в разликите между отделните статии по максималния брой изречения в тях (вж. последния ред на Таблица 2в.). Статиите с нисък максимален брой изречения са най-често свързани с техниката. Статиите с висок максимален брой изречения са представящите на държави с по-богата история. По този показател разликите са значителни.

## Обобщения

Статистическият подход към лингвистичните явления позволява да се открият спецификите на отделните типове текст. Възможностите за автоматизация позволяват вниманието да се насочи и към явления, които в традиционните лингвистични изследвания са пренебрегвани.

Проведеното изследване на енциклопедичен текст помогна за установяването на много специфични черти, които досега бяха само интуитивно ясни. Енциклопедията съчетава много знакови системи, които съществуват в допълнителна дистрибуция, обусловена прагматично. Основна е кодифицираната писмена норма на българския език. Успоредно с нея и най-широко се в този тип текст се използват различни други знакови системи: от общо разпространените означавания на различни мерни единици до тяснонаучните записи на обекти и отношения (математически, химически, физични). Знаковете, синонимни на цели изрази при буквен запис, макар и с не особено голям абсолютен брой, са статистически значимо количество и определят специфичния облик на енциклопедичния текст. Друга характерна черта е проникването на езикови категории, които липсват в българския език: членът като самостоятелна дума.

Съдържателната същност на енциклопедичния текст води до специфика в употребата най-вече на две части на речта. Изключително чести са съществителните, които заемат 2/5 от текста. Числителните са четвъртата по количество част на речта и се нареждат преди глаголите. Това е характерна отлика на енциклопедичния текст в сравнение с другите изследвани български текстове (вж. Димитрова 2006).

Енциклопедичният текст представлява предизвикателство за изследвателя и от семиотична гледна точка. Освен съжителството на символен и буквен запис, при него се наблюдават разнообразни по тип и много на брой съкращения, които обаче не възпрепятстват разбирането. За българския енциклопедичен текст е характерно и съжителството на много азбуки: освен българската и няколко небългарски варианта на кирилица, няколко варианта на латиница. Този план остана извън кръга на интереси в настоящото изследване поради целта да се проучат количествата на частите на речта. Семиотичният план е включен само дотолкова, доколкото влияе на основната цел, и присъства чрез категориите „Знакове” и „Членове”.

Изследването на количествата на частите на речта и на другите езикови явления по същността си е съсредоточено върху езиковата форма. Чрез него могат да се очертаят типове и общи зависимости, които не зависят от тематичната насоченост на текста, а са обусловени от комуникативни и прагматични условия. Представените данни очертават такъв тип специфики на енциклопедичния текст.

## Цитирана литература

**ГСБКЕ 1983:** Граматика на съвременния български книжовен език, Изд. на БАН, София, 1985, т. I-III.

**Димитрова 1995:** М. Димитрова. Частите на речта и стиловата характеристика на текста (Лингвостатистическо изследване). – Проблеми на граматичната система на българския език. Българско езикознание, т. 1. АИ „Проф. М. Дринов”, София, 1995, с. 321-370.

**Димитрова 2006:** М. Димитрова. Езикови особености на текста на енциклопедия. (под печат).

**Дьо Богранд, Дреслер, Стоянова-Йовчева 1995:** Р. Дьо Богранд, В. Дреслер, Ст. Стоянова-Йовчева. Увод в текстовата лингвистика. УИ „Св. Кл. Охридски”, София, 1995

- Коева 2004:** Св. Коева. Съвременни езикови технологии (приложения и перспективи). – Закони на/за езика. Изд. „Хейзъл”, 2004, с. 111-156.
- Мутафчиев 1967:** Р. Мутафчиев. Статистически наблюдения върху стила на българската мемоарна литература от последното десетилетие на миналия век. – Изв. на Инст. за бълг. език, 1967, №14, с. 43-102.
- Мутафчиев 1972:** Р. Мутафчиев. Структура на лексиката в български художествен и публицистичен текст. Статистически наблюдения. – Изв. на Инст. за бълг. език, 1972, №21, с. 3-73.
- Николова 1987:** Цв. Николова. Честотен речник на българската разговорна реч. „Наука и изкуство”, София, 1987.
- Носенко 1981:** И. А. Носенко. Начала статистики для лингвистов. „Высшая школа”, Москва, 1981.
- Пенчев 1993:** Български синтаксис. Управление и свързване. Пловд. унив. издателство, Пловдив, 1993.
- Синклер 1992:** J. Sinclair. Corpus, Concordance, Collocation. Oxford University Press, First published 1991, Second impression 1992.
- Стоименова 2000:** Е. Стоименова. Измерителни качества на тестове. Изд. НБУ, София, 2000.
- Тодорова 1993:** Е. Тодорова. Езикът на българската публицистика (1944-1989). Лингвостатистическо изследване. Изд. „Евтимов-Илинда”, София, 1993.
- Тодорова, Панчовска 2001:** Е. Тодорова, Р. Панчовска. Честотен речник на българската публицистика (1944-1989). Второ допълнено издание. АИ „Проф. М. Дринов”, София, 2001.
- Хамерл, Самбор 1990:** R. Hamerl, J. Sambor. Statystika dla jezykoznawcow. Warszawa, 1990.
- ЧСРЯ:** Частотный словарь русского языка (ред. Л. Н. Засорина). Москва, 1977.
- ЧРПЕ:** I. Kurcz, A. Lewicki, J. Sambor, K. Szafran, J. Woronczak. Słownik frekwencyjny polszczyzny współczesnej (red. Zygmunt Saloni). Т. 1-2. PAN-Institut Jenzyka Polskiego, Krakow, 1990.
- Якубайтис 1981:** Т. А. Якубайтис. Части речи и типы текстов. „Зинатне”, Рига, 1981.
- Янакиев 1977:** М. Янакиев. Стилистиката и езиковото обучение. ДИ „Нар. просвета”, София, 1977.
- BulTreeBank 2004:** Р. Osenova, К. Simov. BulTreeBank project. – <http://www.BulTreeBank.org> (31.08.2004).