Инструкция по выполнению семантического анализа значения предлогов на базе текстов Национального корпуса русского языка

Программные средства. Средства поиска и представления данных НКРЯ. Текстовый процессор MS Word. Электронная таблица MS Excel.

Ход работы.

А) Исследование частотных характеристик предлогов

- 1. Произвести по поиск в основном корпусе НКРЯ по исследуемому предлогу.
- 2. Зафиксировать число документов, число вхождений и объём корпуса.
- 3. Используя функцию «поискать в других корпусах» выполнить поиск в газетном, устном и поэтическом корпусах
- 4. Результаты оформить в таблице следующего вида

Корпус	объём	документы	вхождения	IPM	Вхождений на
НКРЯ	корпуса				документ

5. Для подсчета IPM и параметра «вхождений число вхождений на документ таблицу перенести в MS Excel

) 🖬 🤊 - (- E	+				чер	ез - в	корпусах	Нкря.xlsx - Microsoft Exc	el			
	Главная	Встав	ка Разметк	а страни	цы Формулы	Да	анные Рецензирование	В	ид Ра	зработчик				
	Å	Calibri	- 11	• A /		≫⁄	Перенос текста			Числовой	·	S €		I,
В	тавить	Ж К	<u>ч</u> • 🖽 •]	🏷 - <u>A</u>		i	🛛 📴 Объединить и помести	ИТЬ В	центре 👻	∰ ~ % 000 % ∻	3 Ус форма	ловное тирование ч	Форматироват как таблицу *	ъ Ст яча
Буф	ер обмена 🗔		Шрифт		Fa		Выравнивание		Gi.	Число	ā 📃		Стили	
	E2 • fx =D2/B2*1000			2*1000000										
	А		В		С		D			E		F		G
1	Корпус		Объём корг	iyca	документы		вхождения		IPM		Вхож,	дений на д	окумент	
2	Основной НК	РЯ	283	431966	3	33503	22	4145		790,	82		6,69	
3	Газетный НКР	РЯ	228	521421	10	01136	15	2312		666,	51		1,51	
4	Устный НКРЯ		12	113491		1946		6804		561,	69		3,50	
5	1оэтический НКРЯ 10967173 2683		2683	3436		313,	30		1,28					
6														
7														

6. В окно «вставить функцию» вставить формулу, как показано на рис. 1

Рис. 1. Таблица «Частотные характеристики предлога «через» в текстах корпусов НКРЯ»

- 7. IPM рассчитывается по формуле: число вхождений (в таблице ячейка D2) / объём корпуса (в таблице ячейка B2) * 1000000. Значение число вхождений на документ подсчитывается делением числа вхождений на число документов. Для ячеек выбрать числовой формат с двумя знаками после запятой.
- Б) Количественный анализ значений предлогов

Произвести поиск в корпусе НКРЯ по исследуемому предлогу

Задать настройки представления результатов как показано на рис 2

Упорядочити	5.	x
случайно		
Формат выд KWIC (key v	ачи: vord in context) 💌	
Показывать:		
5	слов в контексте KWIC-выдачи	
100	документов на странице	
10	примеров в документе	
50	примеров на странице	
Применить	Восстановить исходные значения	

Рис 2. Настройки представления результатов поиска

Результаты на странице выделить с помощью мыши и перенести в файл MS Word, где они будут иметь вид сложной таблицы с большим числом вставных таблиц и ссылок, как показано на рис. 3.

Ссылки	Рассылки	Рецензирование	Вид	Разра	ботчик	Конструктор	Макет			
12 • • × ² Aa•	A A ●		· ₽ 20 - E	¶ -	AaBbCc 1 Обычни	Dc AaBbCcDo ый 1 Без инте	АаВЬС Заголово	АаВЬСс Заголово	АаВ _{Название}	АаВbСс. А Подзагол С
фт	G	Абзац		G.				Стил	и	
	1 + 1 + 2 + 1	3 4 5 .	「一樓」	17.1	1 1 8 1 1	· 9 · 1 · 10 ·	1 + 11 + 1 + 12	1 I I I I I I I	• 14 • - 1 • 15 •	· · · 16 · · · 17
9										
+										
	то∙наш∙д	руг•выздоровее	т".∙И°че	рез	годон	выздоров	ел. (Замеч	іу попутн	[∞] 00	, α α
	ещё∙с	лов•не•понимае	т,∙но⁰че	pes	звук∙го молити	олоса∙може венную∞ <u>←</u>	ет∙воспри - <u>→</u> ·¤	нять		
	и.смыс	ла∙тварного∙ми	ра,∙а⁰че	резо	него-	-·к·созерца	анию вечн	юсти:∙веч	ной∞ <u></u> .	<u>→</u> ·¤
	собой•чел	товеческую∙дуп перелив	цу∙и, _{°че} аясь	резо	край, и	сполняет,	наполняе	ет∙собойч	I [∞] ←→	×a
д	о∙краёв∙и	·снова·перелива	ается° <u>че</u>	рез	край.∘	<u>°←…→</u> ∙¤				
	и-дадим	·Ему·просто·лю	бить°че	резо	нас,∙де	йствовать	через∙нас	:,∙быть∞्	- <u>→</u> ·¤	
С	игнал∙от∙і	перемещения∙р управл	учки ения	резо	блокза	апаздыван	ия∙подава	лсяна∙м	онитор∞	<u> →</u>
α			ण्पूल	рез	мериді запазд	ианы,⊤де∙р ывания∞्	оазность∙у <u>→</u> ¤	/ГЛОВ∙		
	Траег	ктория, проходя	ящая°че	pes	любук полнуі	о́начальну ю∞ <u>←…→</u> ≀	ю•точку,∘ а	представ	ляет	
α			प्पe	pe3'	дефект ток∞ <u>←</u>	ную∙изоля •… <u>→</u> ∙¤	нцию∙мож	ет∙прохо	дить	
веј	іичиной∙я напр	нвляется∙возвра яжение,∙измере	тное _{че}	pes	30-с-по	осле•отклю	чения•из(оляции∞⊹	<u>→</u> •¤	
		Отфильтрован	нные ^о че	резо	гидроі	циклон·и·д	жутовую∙	ткань∙стс	ки [∞] <u>←</u>	<u>.→</u> ¤
	Повторн	юполитурунан	юсят ^о че	резо	2∙3· <u>N</u>	ин∙после∙	нанесения	<u>a∞←…→</u>	a	
возм	ожно, тој	тько∙50%•перев осуществля	одов _{оче} нется	pe3°	офици	альные ка	налы•(11).	···~=	¥α	
								H		

Рис. 3 Результаты поиска в НКРЯ после перенесения их в MS WORD

В таком представлении данные не поддаются автоматической нумерации, практически невозможно добавить необходимые для ввода дополнительных данных строки и столбцы. В связи с этим необходимо преобразовать таблицу в более удобный формат. Для этого следует выделить таблицу. Выделять таблицу следует с помощью мыши, так как функция «выделить таблицу» Word не работает. Затем следует использовать функцию «преобразовать в текст», задав при этом следующие параметры (рис. 4)

Преобразование в текст <mark>?</mark> 🗙								
Разделитель								
🔘 знак <u>а</u> бзаца	🔿 знак <u>а</u> бзаца							
. 💽 знак <u>т</u> абуляци	Энак табуляции							
🔘 точ <u>к</u> а с запято	й							
О дру <u>г</u> ой: ,	Одру <u>г</u> ой: ,							
🔽 вложенные табл	ицы							
ОК	Отмена							

Рис. 4 Параметры преобразования таблицы в текст

Полученный результат будет иметь следующий вид (рис. 5)

К 🔏 Ч талык ж. ж. Алт 💇 т 🛆 т 📑 🚍 🚍 📜 🖓 т 🖽 т Обычный 11 Без инте... Заголово... Заголово... Название Подза G Абзац 5 Стили Шрифт 1 · 2 · 1 · 1 · 1 · 1 · 1 · 1 · 2 · 1 · 3 · 1 · 4 · 1 · 5 · 1 · 6 · 1 · 7 · 1 · 8 · 1 · 9 · 1 · 10 · 1 · 11 · 1 · 12 · 1 · 13 · 1 · 14 · 1 · 15 · 1 · 16 ¶ то∙наш∙друг∙выздоровеет". И.¶ → черезчодон выздоровел. (Замечу попутно∞ ... →. ¶ ещё·слов·не·понимает, но·¶ → °через°звуктолоса·может воспринять молитвенную[∞] ... → ¶ и смысла тварного мира, а · собой человеческую душу и, переливаясь ¶ → °через°край, исполняет, наполняет собой и[∞] ← ... → ¶ до краёв∙и снова переливается ¶ → °через°край. ∞<u>←...→</u> ¶ и∙дадим•Ему•просто•любить•¶ → °через°нас, действовать через нас, быть ∞ — ... → .¶ Сигнал от перемещения ручки управления → °через °блок запаздывания подавался на монитор °° ←... → ¶

Рис. 5 Вид данных после преобразования таблицы в текст

Далее следует выполнить несколько контекстных замен.

Знак абзаца – знак табуляции на пробел

Знак табуляции на пробел

Неразрывный пробел на удаление

Знак гиперссылки на удаление (для этого выделить гиперссылку и открыть функцию «заменить»)

Два пробела на один

Два знак абзаца на один

Результат – текст в формате Word разбитый по абзацам, каждый из которых представляет собой контекст использования предлога. Рекомендуется с помощью контекстной замены задать цвет шрифта исследуемого предлога (рис. 6), что несколько упростит дальнейшую работу.



Рис. 6. Промежуточный результат преобразования исходных данных.

Примечание. Если необходимо обработать несколько страниц результатов поиска в НКРЯ, то следующую страницу рекомендуется загрузить в тот же файл, после завершения данного этапа. С ней производятся описанные выше операции, затем тексты соединяются в один.

Текст выделить и через меню «вставить таблицу» преобразовать его в таблицу с начала с одним столбцом, а за тем добавить ещё два — слева и справа (рис. 7). Столбцы озаглавить

«№», «Контекст НКРЯ (Основной корпус)», «Значение». Открыв буфер обмена, в него следует загрузить наименования значений (рис 7).

Буфер обмена (5 из 24) 🔹 🔻 🗙	L	3 + 1 + 3	2 * 1 * 1 * 1 🏛 🔹 1 * 1 * 1		<mark> </mark>
鴅 Вставить все	F		ſ		
🔀 Очистить все	-		Ne¤	Контекст∙НКРЯ∙(основной∙корпус)¤	Значение
высерите вставляемый объект:	-		1.→¤	то наш друг выздоровеет". И через год он выздоровел. «	Темпоратив
iegi upaseonoriusm	2		2.→¤	ещё·слов·не·понимает,·но· через ·звук·толосаможет·воспринять· молитвенную·¤	a a
🗐 Деменсив	м -		3.→¤	и смысла тварного мира, а через него · — к созерцанию вечности: вечно <u>й</u> ⊂	a a
🗐 Медиатив	- - 4		4.→¤	собой человеческую душу и, переливаясь через край, исполняет, ∙ наполняет собой и о	α
	LO.		5.→¤	до краёв∙и снова переливается • <mark>через</mark> • край∘¤	α
🗐 Транзитив	9		6.→¤	и дадим Ему просто:любить через ∙нас, действовать через ∙нас, быть ¤	a 1
👰 Темпоратив	<u>-</u>		7.→¤	Сигналот-перемещения ручки управления через блок- запаздывания подавался на монитор а	a 1
			8. →¤	Через•меридианы, где разность углов запаздывания ©	α
	00 - -		9.→¤	Траекторня, проходящая <mark>через</mark> :любую начальную точку, представляет полную [™]	α
	σ		10. •¤	Через дефектную изоляцию может проходить ток ©	α a
	10 + 1		11.*¤	величиной является возвратное напряжение, чзмеренное через 30 с•после•отключения изоляции ∞	a r
	-		12.•¤	Отфильтрованные-через-гидроциклон-и-джутовую-ткань-стоки-¤	α
-	11		13.*¤	Повторно политуру наносят через 2 3 мин после нанесения · П	α
Параметры 🔻	2		14.•¤	возможно, только-50% переводов осуществляется • через официальные каналы (11). · □	α

Рис. 7 Вставка наименования значения контекста

Установив в соответствующую ячейку курсор и щелкнув левой клавишей на нужном термине в буфере обмена, вставляем термин в ячейку. После того, как таблица заполнена, произвести подсчет значений. Для этого используется функция «выделение при чтении». Выделить столбец «значение». В поисковое окно вкладки найти вставить наименование значения. Система выделит искомые термины и сообщит число выделений (рис 8).

	№ Контекст НКРЯ (основной корпус)						
-	Найти и заменить	×∣ од∙он∙выздоровел.⊷¤	Темпоратив				
~	Найти Заменить Перейти	сголоса может воспринять	Медиатив¤				
-		1					
m		ок.созерцанию вечности:	Медиатив¤				
-							
4		ваясь· через ·край, исполняет, ·	Транзитив¤				
-	Приложение Word выделило 9 элементов. Чтобы снять выделение, выберите "Выделение при						
un l	чтении", а затем — "Снять выделение".	·край¤	Транзитив¤				
- 11	<< Меньще Выделение при чтении ▼ Найти в ▼ Найти далее Закрыть	ас, · действовать · через · нас, ·	Медиатив¤				
۔ و	Параметры поиска						
- 11	Направление: Вперед 💌	вления•через•блок•	Медиатив¤				
	Учит <u>ы</u> вать регистр Учитывать пр <u>е</u> фикс	α					
- 11	Полько слово целиком Учитывать суффикс	взапаздывания ©	Транзитив¤				
- ∞	Подстановочные знаки	/ю•начальную•точку,∙	Транзитив¤				
- 11	Бее словоформы						
თ		роходить ток∙¤	Транзитив¤				
- 11		ряжение, измеренное через	Темпоратив				
9	Найти						
- 11	Пормат т Специальный т Снять форматирование	рни∙джутовую•ткань•сток蕤	Транзитив¤				
÷.		3 мин после нанесения □	Темпоратив				
- 11		уществляется через	Медиатив¤				
2	официальные каналы (11)¤						

Рис. 8 Определение числа значений «медиатив»

Результаты подсчетов оформляются следующим образом

Основной корпус

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.

через

Найдено 224 145 вхождений.

Отобрано 204 контекста

Значение	Число контекстов	% от общего числа		
		контекстов		
темпоратив	103	51,0		
транзитив	61	30,4		
медиатив	35	17,6		
дименсив	4	2,0		
фразеологизм	1	0,5		

Сводная таблица значения предлога «через» по корпусам, % от числа отобранных

Корпус НКРЯ	темпоратив	транзитив	медиатив	дименсив	фразеологизм
Основной	51,0	30,4	17,6	2,0	0,5
Газетный	49,0	16,5	35,2	0	0,4
Устный	48,0	27,5	23,2	0,5	0,9
Поэтически	23,2	70,4	5,5	0	1,2
й					