

Представление устойчивых лексических сочетаний в компьютерном тезаурусе RussNet

Азарова Ирина Владимировна

Синопальникова Анна Алексеевна

Кафедра математической лингвистики СПбГУ

Смрж Павел

Технологический университет Брно (Чехия)



Компьютерный тезаурус RussNet

Является моделью человеческого лексикона:

- в тезаурус включены смыслы, которые имеют словесное выражение в русском языке
- элементарной единицей является синонимический ряд (СИНСЕТ)
- между синсетами установлены семантические отношения
- семантическое отношение род-вид организует синсеты в набор деревьев

Пример словарной статьи RussNet

Синсет, состоящий из одного слова «дом»

VisDic

Files Dictionary Help

Russian WordNet 0.2.a

дом

[n] дом:1

View Tree RevTree Edit Nouns Adjectives Verbs XML

POS: n ID: **RUS-1234560004**

Synonyms: дом:1

Definition: Здание, строение, предназначенное для жилья, размещения различных учреждений и предприятий.

Usage: Передним фасадом дом обращен к востоку в сторону Вознесенского проспекта.

-->> [has_hyponym] [n] постройка:1

<<< [synonym] [n] здание:1

<<< [has_hyponym] [n] вилла:1

<<< [has_hyponym] [n] вокзал:1

<<< [has_hyponym] [n] дворец:1

<<< [has_hyponym] [n] магазин:1

<<< [has_hyponym] [n] особняк:1

<<< [has_hyponym] [n] охотничий домик:1

<<< [has_hyponym] [n] ресторан:1

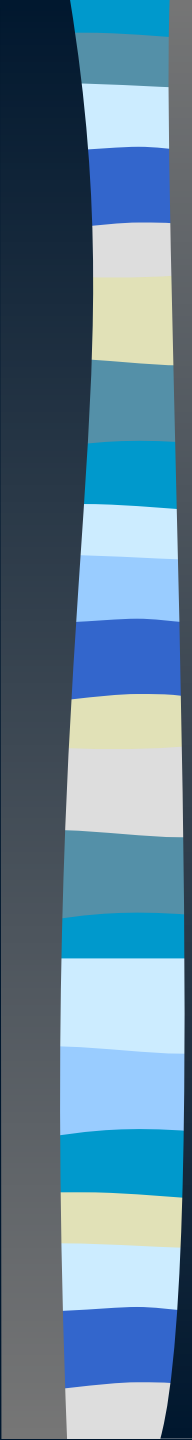
<<< [has_hyponym] [n] станция скорой помощи:1

<<< [mero_part] [n] подъезд:1

<<< [mero_part] [n] этаж:1

<<< [mero_part] [n] помещение:1

Number of entries: 4



Корпус современных текстов Бокренок 3

Объем: 21 млн. словоупотреблений

Временной интервал: 1985 –2004

Типы текстов:

40% – газетные тексты

30% – научно-популярные тексты

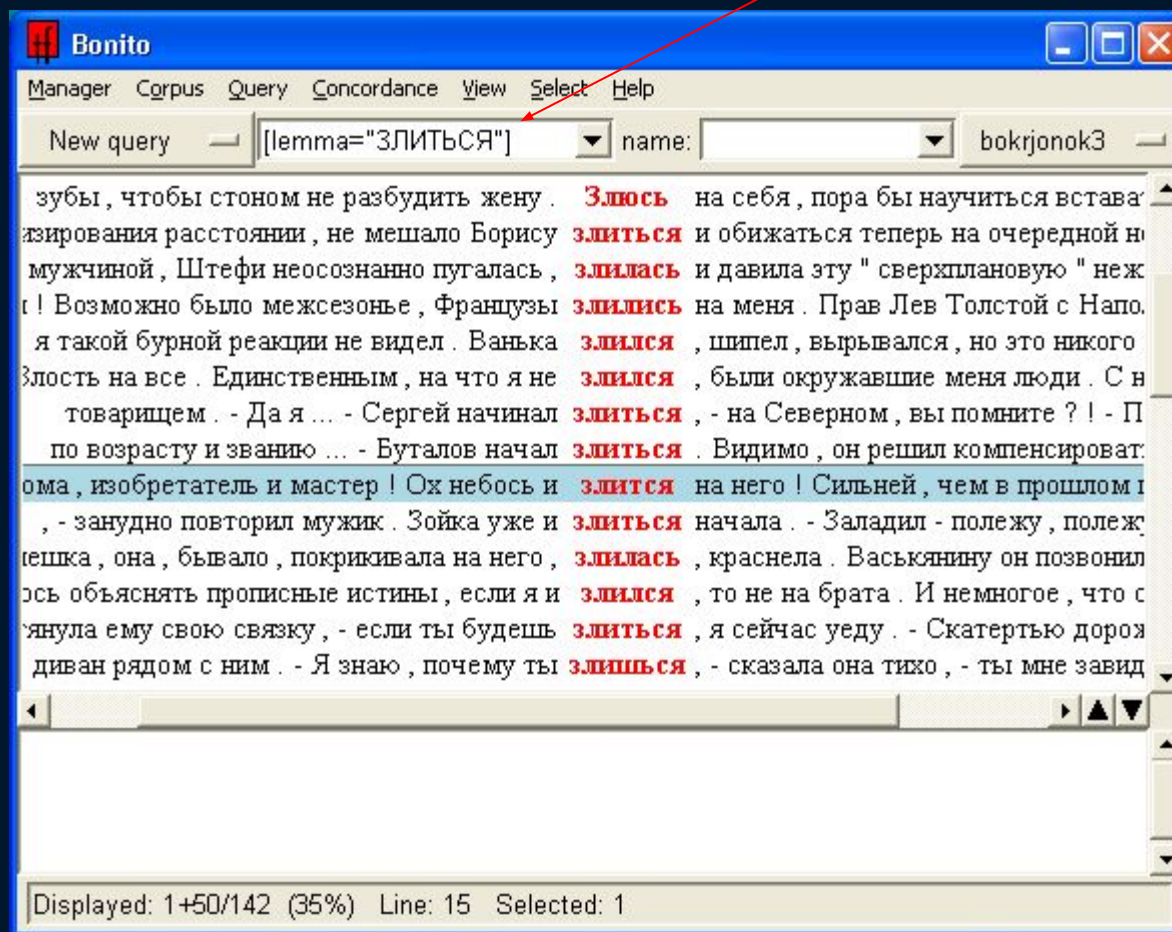
20% – фрагменты художественных текстов

10% – тексты законов, договоров...

- баланс стереотипности, информативности и экспрессивности текстов
- отсутствие больших фрагментов текста

Отбор контекстов в программе Бонито

Запрос на поиск контекстов слова



The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manager", "Corpus", "Query", "Concordance", "View", "Select", and "Help". The "Query" field contains "[lemma="ЗЛИТЬСЯ"]" and the "name" field contains "bokrjonok3". The main window displays a list of text excerpts with the word "ЗЛИТЬСЯ" highlighted in red. A red arrow points from the "Select" menu item to the search query field. The status bar at the bottom indicates "Displayed: 1+50/142 (35%) Line: 15 Selected: 1".

зубы, чтобы стоном не разбудить жену . **Злось** на себя , пора бы научиться вставка
изирования расстояния , не мешало Борису **злиться** и обижаться теперь на очередной н
мужчиной , Штефи неосознанно пугалась , **злилась** и давила эту " сверхплановую " неж
! Возможно было межсезонье , Французы **злились** на меня . Прав Лев Толстой с Напо.
я такой бурной реакции не видел . Ванька **злился** , шипел , вырывался , но это никого
Злость на все . Единственным , на что я не **злился** , были окружавшие меня люди . С н
товарищем . - Да я ... - Сергей начинал **злиться** , - на Северном , вы помните ? ! - П
по возрасту и званию ... - Буталов начал **злиться** . Видимо , он решил компенсироват
ома , изобретатель и мастер ! Ох небось и **злится** на него ! Сильней , чем в прошлом
 , - занудно повторил мужик . Зойка уже и **злиться** начала . - Заладил - полежу , полеж
пешка , она , бывало , покрикивала на него , **злилась** , краснела . Васьянину он позвонил
ось объяснять прописные истины , если я и **злился** , то не на брата . И немного , что с
янула ему свою связку , - если ты будешь **злиться** , я сейчас уеду . - Скатертью дорож
диван рядом с ним . - Я знаю , почему ты **злишься** , - сказала она тихо , - ты мне завид

Displayed: 1+50/142 (35%) Line: 15 Selected: 1



Контекстные маркеры значений лексем в текстах

Контекстные маркеры можно задавать в виде

- **рамок валентности – устойчивых семантических и морфо-синтаксических признаков контекстов;**
- **сочетаний слов, имеющих собственное связанное значение и форму выражения.**

Схема описания активной валентной рамки

$V_n \langle N, \text{ROLE}, \text{SEM}, \text{SYN}, f_o \rangle \dots$

N – порядковый номер (при объективном порядке слов)

ROLE семантическая роль в пропозиции;

значения: СУБЪЕКТ, ОБЪЕКТ1, ОБЪЕКТ2...

SEM семантический тип слов, заполняющих позицию;

значения: ЧЕЛОВЕК, ПРЕДМЕТ, АРТЕФАКТ...

SYN частотная форма морфо-синтаксического

заполнения позиции; *значения:* acc, dat, abl,

«на» + acc ...

F_o 1 – обязательная (> 66-85%),

0 – факультативная (> 15-33%)

Валентная рамка RussNet

Направиться (358 контекстов - 100%)

“двинуться в каком-л направлении”

Рамка валентностей (2 обязательные позиции)

(1) *субъект движения*

N_1 [лицо / группа лиц] СУБЪЕКТ

(2) *направления движения*

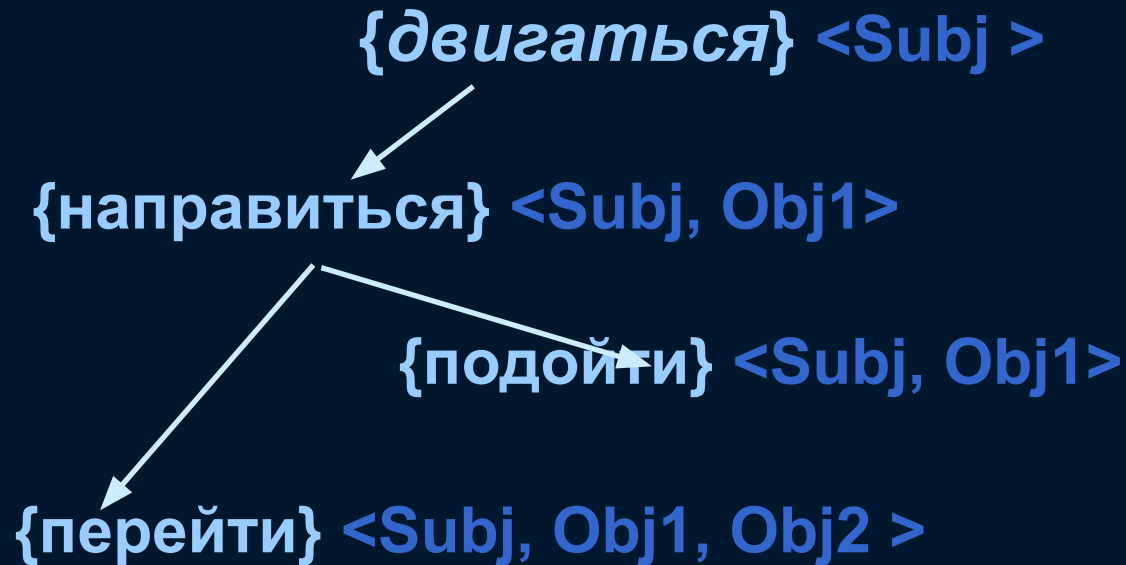
"к + N_3 " (44%) [локализация/ лицо] направление движения?

"в + N_4 " (27%) [локализация] ОБЪЕКТ1

"в стороону + N_4 " (4%)...

(к дивану, к другу, к спуску, к нему...; в комнату, в деревню, в угол гостиной)

Семантические роли валентных позиций в RussNet



- семантические деревья RussNet обладают конечным набором семантических ролей
- типы объектов в дереве устойчивы

Пассивная валентная рамка RussNet

$_V$ (SYN) <ROLE, SEM>...

SYN морфо-синтаксическая форма слова, имеющего пассивную валентность

в лицо "В + N₄"

ROLE семантическая роль в пропозиции;

SEM семантический тип слов, присоединяющих валентность

{*сказать*}, {*знать*}, {*плюнуть*}

Пассивная валентная рамка RussNet (2)

The image shows two windows of the Bonito software interface. The background window displays a search result for the query 'bokrjonok3'. The text snippet includes the words 'странно', 'он не знал', 'объект', 'в лицо', and 'всякий раз'. The foreground window shows a detailed view of the search results, including a filter 'P filter' and navigation options like 'First', 'Last', 'From: -5', 'Positions', 'from', '< KWIC', 'To: -1', and 'Positions'. The text in the foreground window is a full paragraph with many words highlighted in red, including 'странно', 'он не знал', 'объект', 'в лицо', 'и всякий раз', 'когда в офис входил', 'предприниматель', 'знавший', 'объект', 'в лицо', 'К убийству майор готовился тщательно', 'и клиента здесь еще не знают', 'в лицо', 'сохраняйте инкогнито', 'Если', 'которого противник не знает', 'в лицо', '- особенно эф - фективен', 'Слившийся', 'или он просто не знал этих людей', 'в лицо', 'или заметил их слишком поздно', 'среди нас, но мы их знаем только', 'в лицо', '39. Чингиз Айтматов - матерый', 'Профессионалов высокого полета знают', 'в лицо', 'Пусть это будет знатная ткачиха', 'людей, которых вы знаете хотя бы', 'в лицо', 'Кстати, понятия имеют право', 'людей, которых вы знаете хотя бы', 'в лицо', 'Кстати, понятия имеют право', '... Тысячи людей знали Иисуса', 'в лицо', 'он проповедовал в Храме, творил', 'родителей, которых и не знают', 'в лицо', ', точно также, как те своих', 'старомодная преподавательница знает', 'в лицо', 'Какие - нибудь депутаты? Которых', 'деле офицер ЦРУ. Маша знала его', 'в лицо', 'когда - то он преподавал в разведшколе', 'происшествий и знали друг друга', 'в лицо', 'Но сейчас никаких признаков'.

Displayed: 1+50/455 (10)

Displayed: all/41 Line: 1



Словосочетания в wordnet-словарях

Принстонский WordNet

give up

Sense 1

forfeit, give up, throw overboard, render, waive, forgo

=> abandon -- ("We abandoned the old car in the empty parking lot")

Sense 2

drop out, give up, throw in, throw in the towel, quit, leave, admit defeat, chuck up the sponge -- (give up in the face of defeat of lacking hope)

Sense 3

spare, give up, part with, dispense with

=> give -- (transfer possession of something concrete or abstract to somebody; "I gave her my money"; "can you give me lessons?")

Устойчивые словосочетания (1)

Контактность существительных

The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manager", "Corpus", "Query", "Concordance", "View", "Select", and "Help". Below the menu bar, there are input fields for "P filter", "name:" (containing "bokrjonok3"), and "From:" (set to 1) and "To:" (set to 5). The search criteria are set to "from < KWIC". The main window displays a concordance table with two columns: the original text and the KWIC (Key Word In Context) results. The KWIC results for the word "время" are highlighted in red. A red arrow points from the word "время" in the KWIC column to the status bar at the bottom, which reads "Displayed: 1+50/219 (22%) Line: 1".

Original Text	KWIC
русской Лиги по борьбе за права человека , во	время весеннего призыва 1995 года 30
туры Чаще труп обнаруживается в холодное	время года на открытой местности . Г.
, поэтому эти отравления бывают в жаркое	время года заболевание связано с поп:
Так что я в общем - то " агитирую за смену	времен года " -- они ведь и без моей аг
заемого потомства , адаптироваться к смене	времен года или мигрировать в дальние
лись в памяти эти две встречи ! Какое было	время года - помню ! Какая погода ст
узлом столицы , его назначением и образом .	Время реализации Генерального плана
льшом поме - щении или на улице . В теплое	время года собака способна учуять не
тех или иных продуктов в магазине в данное	время года . Мы ведем разговор лишь
я тепла . На процесс дыхания влияют также	время дня и время года . В состоянии i
процесс дыхания влияют также время дня и	время года . В состоянии покоя кошка
молочных продуктах , полученных в разное	время года , неодинаково . Так , в лет
бесчастные случаи могут произойти в любое	время года , в любом возрасте животн
ценивает длину шерсти , учитывая при этом	время года и степень линьки . У длин:

Воспроизводимость

Устойчивые словосочетания (2)

Контактность может нарушаться при перечислении.

The screenshot shows the Bonito software interface with a concordance search for the word "палец". The search parameters are "P filter" (empty), "name:" (empty), and "bokrjonok3". The search results are displayed in a table format with columns for the text, the word being searched, and the context. A red arrow points to the word "большой" in the text, which is highlighted in red. The text is as follows:

Text	Word	Context
бутылки в кулак , ногтем большого	палеца	(специально один ноготь оставил
вытянул большой и указательный	пальцы	. Остальные один за другим сделали
под грязным ногтем большого	палеца	поролоновый шарик и передвигать
Сергеевич показал большой волосатый	палец	. - Завтра меня весь день не
волны и показал мне большой	палец	- это у него морской сигнал
материнской линии . Большой	палец	одного из шоферов был перевязан
Вот так . - Я сунул большой	палец	в рот и стал громко чмокать
ладонь --- передней . Большой	палец	" несет ответственность " за
1984 - м с гангреной большого	палеца	на правой ноге в сопровождении
удовольствием все слопал . И большой	палец	мне показал . - Как вообще они
посмотрела на свой натертый большой	палец	, распухший , в крови , и потеряла
посмотрела на свой натертый большой	палец	, распухший , в крови , и потеряла
на пакете отпечаток большого	палеца	, который впоследствии стал
фигурировал отпечаток большого	палеца	, обнаруженный на присланном

Displayed: 81+50/137 (36%) Line: 81

Устойчивые словосочетания (3)

Глагольные словосочетания

Контактность нарушается наречиями, частицами и проч.

понятно , что все к тому идет . Я **имею** в виду не мое назначение ,
- скандал ? - Смотря что вы **имеете** в виду ? Если битье посуды
полной независимости . - Вы **имеете** в виду : разработать систему
перед федеральным центром . Я **имею** в виду , что у нас заключено
плохая кредитная история . Я **имею** в виду кредит доверия у населения
А он обаятельный человек . Я **имею** в виду в разговоре . - У вас
любого человека в животное / Я **имею** здесь в виду негативную характеристику
новые доселе нам неизвестные . Я **имею** в виду Кувейт , с которым мы
организации - спецэкспортеры . Не **имею** в виду посредников , которые
тельности и опирается система BTC . **Имею** в виду закон о военно - техническом
В другом случае он заявил , **имея** в виду расовую " теорию " :
мы будем говорить о Сталине , **имея** в виду , что все , касающееся
не осталось никаких страхов (**имею** , естественно , в виду страхи
----- *Примечание : Л . Бородина **имеет** в виду Андрея из Львова , чье

Displayed: 41+50/944 (5%) Line: 41

Устойчивые словосочетания (4)

Глагольные словосочетания

В менее очевидных случаях контактность нарушается другими валентными позициями.

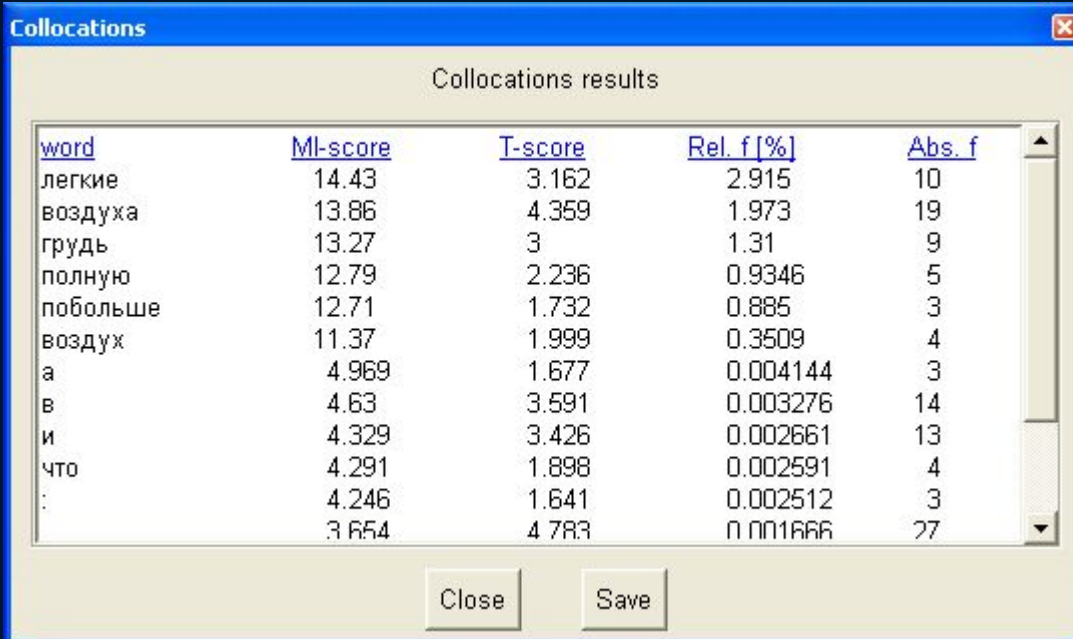
The screenshot shows the Bonito software interface with a concordance search for the word "воздуха". The search results are displayed in a table format, showing the context of the word in a text snippet and the word itself in red. The search parameters are: P filter (empty), name: bokrjonok3, From: -5, To: 5, Positions: from <<MIC, To: 5, Positions. The search results are as follows:

Context	Word	Context
абсолютно ничего . Пришлось	набрать	в грудь побольше воздуха и выдать
попытке выглянул в коридор ,	набрал	в легкие воздуха , а тут идут
растерялся - схватил микрофон и ,	набрав	в легкие побольше воздуха ,
Юля брезгливо поморщилась .	Набрав	воздуха я что есть силы дунул
? - Приду , расскажу . Жди .	Набрав	воздуха в легкие , промчался
пожар . Вперед , к звездам ! Он	набрал	полную грудь воздуха и вернулся
... Тоска зеленая . Длинный	набрал	воздуха и медленно выдохнул
представляем . -- Ой ли - Лукой ли	набрал	в легкие воздуха , чтобы продолжить
гримасу , словно разминая щеки ,	набрал	воздуха и произнес ... что
ныряльщик перед погружением ,	набрал	полную грудь воздуха и показал
же следует выбираться так :	набрать	в грудь воздуха и нырнуть поглубже
разрешила ей что - то . Саня	набрала	полную грудь воздуха , выдохнула
перебивай ... - Она уловила , что он	набрал	воздух для ответа . - Я хочу
Она мучительно , со всхлипом ,	набрала	полную грудь воздуха . И вдруг

Displayed: all/26 Line: 1

Устойчивые словосочетания (5)

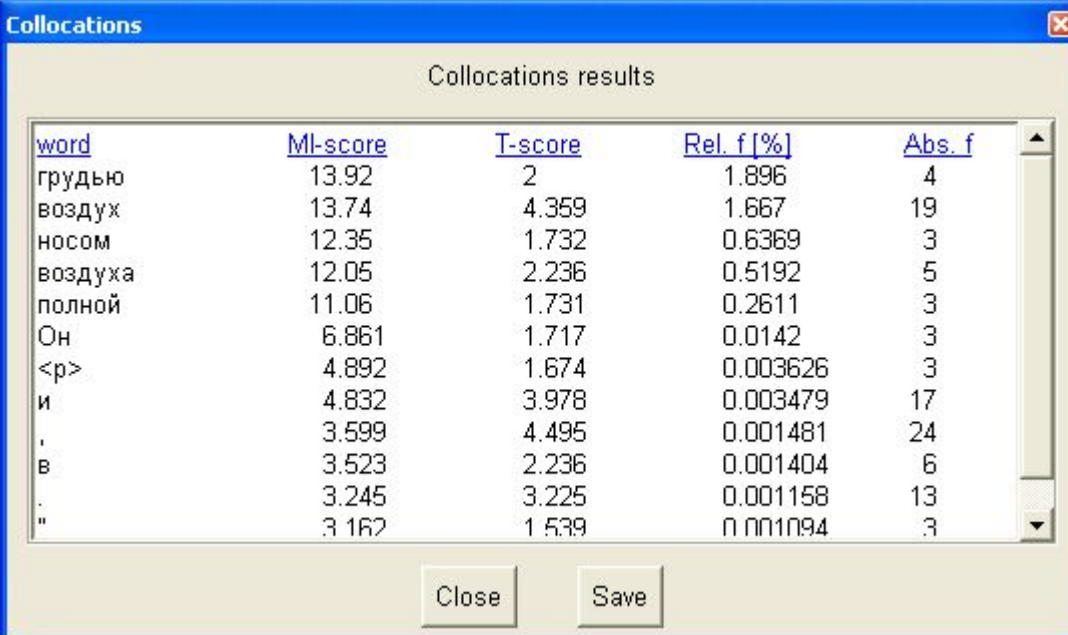
Статистическая оценка «устойчивости» сочетания *набрать воздух(а)* в диапазоне ± 5 слов. Коэффициент «неслучайности» MI показывает, насколько «информативно» данное сочетание слов. Доля контекстов с данным значением 3%, общее число вхождений глагола 664 (31 ipm)



<u>word</u>	<u>MI-score</u>	<u>T-score</u>	<u>Rel. f [%]</u>	<u>Abs. f</u>
легкие	14.43	3.162	2.915	10
воздуха	13.86	4.359	1.973	19
грудь	13.27	3	1.31	9
полную	12.79	2.236	0.9346	5
побольше	12.71	1.732	0.885	3
воздух	11.37	1.999	0.3509	4
а	4.969	1.677	0.004144	3
в	4.63	3.591	0.003276	14
и	4.329	3.426	0.002661	13
что	4.291	1.898	0.002591	4
:	4.246	1.641	0.002512	3
	3.654	4.783	0.001666	27

Устойчивые словосочетания (6)

Статистическая оценка сочетаний
вдохнуть воздух в диапазоне ± 5 слов
(частотность в корпусе – 5 ipm, 101 вхождение)



The screenshot shows a window titled 'Collocations' with a subtitle 'Collocations results'. It contains a table with five columns: 'word', 'MI-score', 'T-score', 'Rel. f [%]', and 'Abs. f'. The table lists various words and their associated scores and frequencies. At the bottom of the window, there are 'Close' and 'Save' buttons.

<u>word</u>	<u>MI-score</u>	<u>T-score</u>	<u>Rel. f [%]</u>	<u>Abs. f</u>
грудью	13.92	2	1.896	4
воздух	13.74	4.359	1.667	19
носом	12.35	1.732	0.6369	3
воздуха	12.05	2.236	0.5192	5
полной	11.06	1.731	0.2611	3
Он	6.861	1.717	0.0142	3
<p>	4.892	1.674	0.003626	3
и	4.832	3.978	0.003479	17
,	3.599	4.495	0.001481	24
в	3.523	2.236	0.001404	6
.	3.245	3.225	0.001158	13
"	3.162	1.539	0.001094	3



Контактная информация

RussNet

<http://www.phil.pu.ru/depts/12/RN/>