Когниция, коммуникация, дискурс. – 2016 – № 13. – С. 23–32. http://sites.google.com/site/cognitiondiscourse/

УДК 811.161.1

ЭЛЕКТРОННЫЕ МАССИВЫ ИНФОРМАЦИИ И КОНЦЕПТОЛОГИЧЕСКИЙ АНАЛИЗ

П.Н. Донец (Харьков, Украина)

П.Н. Донец. Электронные массивы информации и концептологический анализ. Статья посвящена возможностям, открывающимся для концептологического анализа при использовании метода поиска по ключевому слову и по маске в больших массивах информации. К последним относятся: а) электронные версии обычных словарей, б) онлайн-словари и в) архивы средств массовой информации. Поиск в электронных словарях позволяет установить этимологию прямого наименования концепта, его сигнификат (ядро), парадигматические связи и соответствующее семейство слов. Поиск в онлайн-словарях представляет интерес, прежде всего, с точки зрения синтагматических связей концепта, в частности т.н. «кооккурентностей». Полезным в этом отношении может оказаться также поиск в архивах СМИ, но основная его ценность видится в возможности установления референциала концепта (совокупности отражаемых им референтов). Потенциал метода поиска по ключевому слову и по маске в больших массивах информации иллюстрируется на примере немецкого концепта GRENZE (ГРАНИЦА). Основной вывод исследования состоит в том, что применение этого метода кратно облегчает и убыстряет поиск и обработку языкового материала в целях концептологического анализа.

Ключевые слова: концепт, концептологический анализ, онлайн-словарь, поиск по ключевому слову и по маске, референциал концепта, электронный словарь, электронный массив информации.

П.М. Донець. Електронні масиви інформації та концептологічний аналіз. Стаття присвячена можливостям, що відкриваються для концептологічного аналізу при використанні методу пошуку за ключовим словом чи маскою у великих масивах інформації. З-поміж останніх виділяють такі: а) електронні версії звичайних словників, б) онлайн-словники та в) архіви засобів масової інформації. Пошук в електронних словниках дає змогу встановити етимологію прямого найменування концепту, його сигніфікат (ядро), парадигматичні зв'язки і відповідне сімейство слів. Пошук в онлайн-словниках викликає інтерес, передусім, з погляду синтагматичних зв'язків концепту, зокрема т.з. «кооккурентностей». Корисним при цьому може виявитися також пошук в архівах ЗМІ, але основна його цінність полягає у можливості встановлення референціалу концепту (сукупності відображених ним референтів). Потенціал методу пошуку за ключовим словом і маскою у великих масивах інформації проілюстровано на прикладі німецького концепту GRENZE (МЕЖА, КОРДОН). Основний висновок дослідження полягає в тому, що застосування цього методу кратно полегшує і прискорює пошук та обробку мовного матеріалу з метою концептологічного аналізу.

Ключові слова: електронний масив інформації, електронний словник, концепт, концептологічний аналіз, онлайн-словник, пошук за ключовим словом і маскою, референціал концепту.

P.N. Donec. Large arrays of information and the conceptological analysis. The paper concerns the means for using the keyword and mask search in large arrays of information. The latter include the electronic versions of conventional dictionaries, online dictionaries and media archives. The search in electronic dictionaries makes it possible to set the etymology of direct nomination of the concept, its significate (core), paradigmatic relations and the corresponding word family. The search in online dictionaries is of interest primarily in terms of syntagmatic relations of the concept, in particular the so-called «cooccurrences». The media archives search mainly contributes to identifying the concept referential, i.e. the set of referents represented by this concept. The purpose of the study is to illustrate possible ways of keyword and mask search in large arrays of information on the example of the German concept GRENZE (BORDER, BORDERLINE, FRONTIER). The study shows that the use of this method considerably simplifies and speeds up the search and processing of language material and its conceptological analysis.

Key words: concept, concept referential, conceptological analysis, electronic arrays of information, electronic dictionary, online dictionary, search by keyword and by mask.

1. Введение

Появление электронных словарей и других банков данных, охватывающих большие массивы дискурса (например, многолетних архивов различных СМИ) и позволивших поиск по ключевому слову или по маске, произвело настоящую революцию в методике лингвистических исследований, в разы облегчив отбор и анализ исследуемого языкового материала, что касается и весьма актуального ныне концептологического анализа.

Эти возможности иллюстрируются в настоящей статье на примере немецкого концепта GRENZE и на материале его манифестаций в таких массивах информации, как: 1) электронные версии традиционных словарей (в частности, этимологического, толкового и синонимического словарей Duden, идеографического словаря Dornseiff, двуязычных словарей в оболочке Lingvo), 2) онлайн-словари (в частности, Лейпцигский онлайн-словарь и Маннгеймский онлайн-словарь OWID) и 3) большие массивы дискурса (в частности, электронный архив еженедельника «Die Zeit» за 1994—2009 гг.).

2. Процедура концептологического анализа

Современный концептологический анализ отличается комплексностью и включает в себя обычно несколько этапов:

- 1) этимологический,
- 2) семасиологический,
- 3) ономасиологический,
- 4) психолингвистический и логико-когнитивный (среди прочего, определение прототипа соответствующего класса, а также построение фрейма),
 - 5) дискурсивно-дистрибутивный.

Что касается электронных словарей, то они могут принести пользу, прежде всего, на этапах 1–3.

3. Результаты анализа и их обсуждение

3.1. Концептологический анализ с помощью электронных версий традиционных словарей

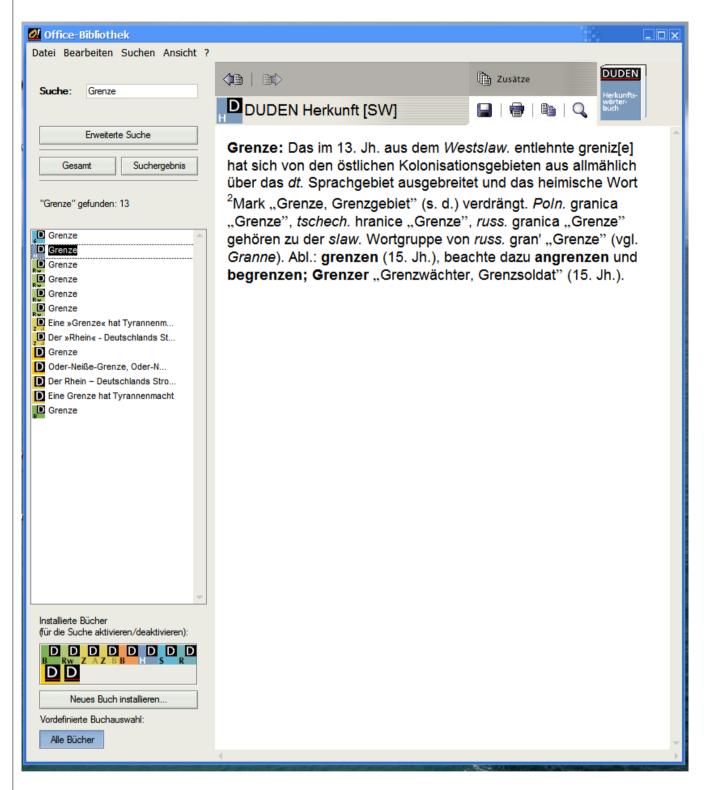
Электронные словари можно разделить на два типа: закрытые (аналоги «бумажных») и открытые – онлайн-словари, словник, дефиниции, переводы и прочие данные в которых постоянно подвергаются изменениям.

Электронные версии бумажных словарей ничем не отличаются от своих оригиналов и могут быть одно- и двуязычными, толковыми и переводными, этимологическими и фразеологическими, орфографическими и синонимическими и т.д.

Применительно к немецкому языку стандартным является семейство словарей «Duden», которое представлено и в электронном виде, причем поиск по ключевому слову и по маске можно осуществлять сразу во всех представленных в нем словарях.

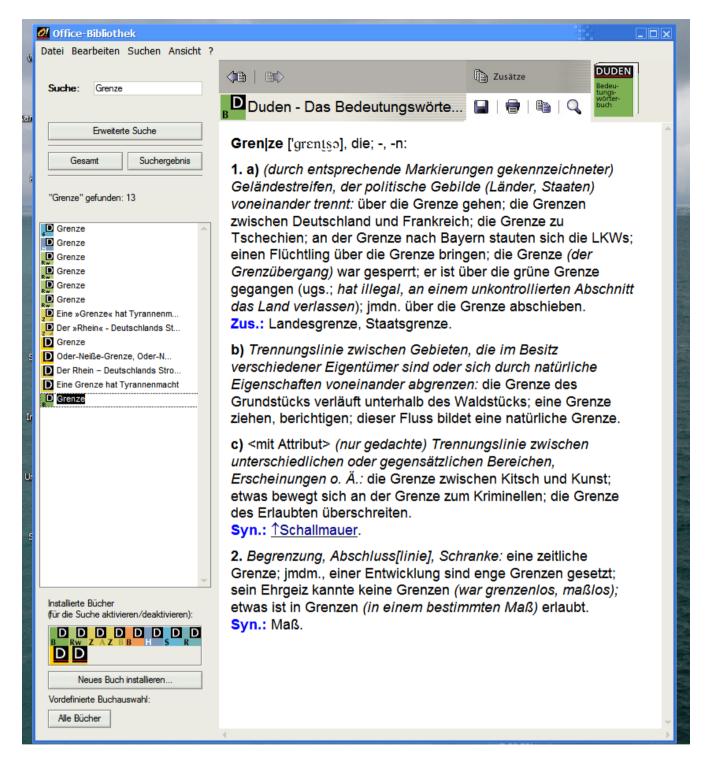
Как было отмечено выше, концептологический анализ целесообразно начинать с этимологического этапа, посколько это дает возможность определить «зародыш» соответствующего концепта и динамику его развития.

Этимологический словарь Duden выдает следующую историю слова Grenze:



Наиболее вероятной причиной вытеснения исконно германского слова *Mark* заимствованным из западно-славянских наречий *granica* представляется то, что именно зона контакта со славянским населением, скорее всего, воспринималась немцами как «настоящая» граница — дело в том, что рубежи Священной Римской империи на севере и юге были, большей частью, естественными (моря и горы), а на западе — достаточно размытыми, в связи с претензией этого государственного образования на универсализм, с его мультиэтническим характером, с одной стороны, и культурным родством этих этносов — с другой.

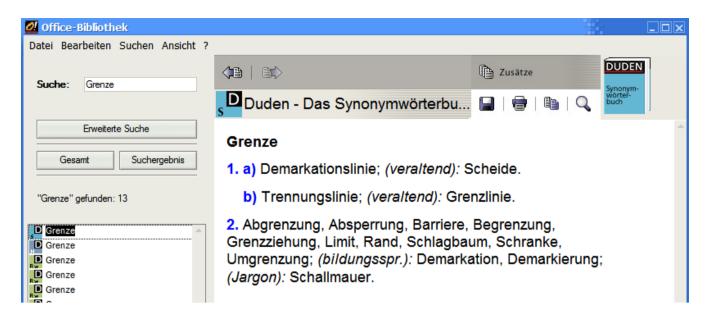
Определение сигнификата (ядра) концепта удобно начинать с анализа дефиниции соответствующего слова в толковом словаре – в нашем случае, «Duden – Das Bedeutungswörterbuch»:



Как показали данные нашего исследования (см. [Донец 2005; Донец 2005; Donec 2014]), это толкование вряд ли можно признать удачным, но отправной точкой для анализа оно вполне может послужить («линия», «линия, разделяющая две сущности»; «линия, означающая окончание чего-либо»; «линия барьера для чего-либо» и т.д.).

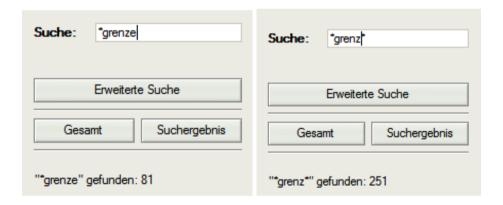
Здесь мы также можем обнаружить отдельные «семемы» Grenze, которые в онтологическом смысле соответствуют крупным подтипам отражаемого феномена: «барьер», «окончание», «мера» и т.д.

Решению этой же задачи способен послужить и анализ синонимов соответствующего слова, что означает переход к ономасиологическому уровню анализа:



На роль таких «семем» могут, в частности, претендовать Barriere, Limit и Rand.

В электронных словарях Duden существует также возможность поиска по т.н. маске, с помошью астериска (звездочки).



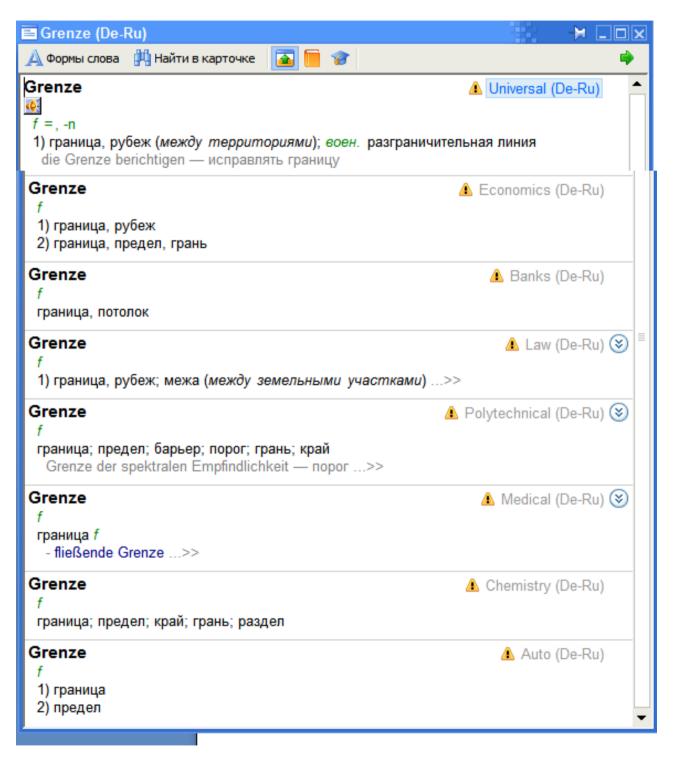
Введя стимул ***grenze**, мы получим список сложных слов с компонентом *-grenze*, охватывающий 81 лексему. Он может, в частности, оказаться полезным при построении типологии границ.

Стимул *grenz* дает нам возможность получить практически все семейство слов вокруг корня -grenz- (251 единица), которое может обрабатываться под самыми различными углами зрения. Если говорить о концепт-анализе, то полученный список способен стать отправной точкой для определения прототипа концепта (в понимании Э. Рош), то есть наиболее частотного «типичного» представителя класса. Результаты анализа указанного лексического семейства говорят о том, что примерно 2/3 слов с корнем -grenz- имеют отношение к семеме «государственная граница» — другими словами, именно эта семема концепта является прототипической. Эти результаты в целом подтверждаются данными дискурсивно-дистрибутивного анализа, а также результатами проведенного нами психолингвистического эксперимента.

Устройство электронной версии словаря Duden предусматривает также вариант т.н. «полнотекстового поиска» (Volltextsuche) слова, позволяющего найти не только заглавные лексемы (леммы), но и все остальные случаи его употребления. В словаре имеется три рубрики, в которых может встретиться искомая лексема: 1) словник как таковой (список толкуемых слов), 2) дефиниции (толкование) различных слов, где фигурирует, в нашем случае, Grenze и 3) иллюстрации его употребления в тех или иных, как правило, литературных произведениях или текстах публицистического жанра. В последних, правда, искомое слово может оказаться случайно, не имея прямого отношения к собственно лемме,

а вот вторые способны представлять интерес с точки зрения концепт-анализа, в частности, в плане установления сетки его идеографических связей (см. [Донец 2009]).

Определенную пользу способен принести также поиск в таких двуязычных (переводных) словарях, как всем известная серия словаря Lingvo, который, как известно, состоит из набора «оцифрованных» универсальных и специализированных обычных словарей. Этот поиск может, среди прочего, показать, какие субконцепты превалируют в различных подъязыках:



Одним из самых трудных для лексикографического поиска типом словарей являются идеографические (тезаурусные) словари, каковым для немецкого языка является словарь Дорнзайфа, поскольку они устроены по иному принципу, и слова в них систематизируются не по алфавиту, а по понятийным полям. Электронная версия словаря способна облегчить решение этой проблемы. Отдельного электронного издания словаря Дорнзайфа нам обнаружить не удалось, однако он присутствует в качестве составной части известного

онлайн-словаря Лейпцигского университета [http://wortschatz.uni-leipzig.de/cgi-bin/wort_www?Wort=Grenze&site=1&cs=1&x=0&y=0], см. нижнюю часть соответствующей статьи (**Dornseiff-Bedeutungsgruppen**):

Links zu anderen Wörtern:

- Synonym von: **Grenzlinie**
- Grundform: Grenze
- ist ein(e) Begrenzung
- Teilwort von: Grenze zwischen, über die Grenze, nahe der Grenze, An der Grenze, bis zur Grenze, innerdeutsche Grenze, an der Grenze von, obere Grenze, deutsch-deutsche Grenze, untere Grenze, die Grenze ziehen, Hart an der Grenze, oberste Grenze, äußerste Grenze, An die Grenze, Grüne Grenze, Schmuggel über die Grenze, Grenze schließen, Überschreitung der Grenze, Eine Grenze hat Tyrannenmacht, eine Grenze festlegen, Steine an der Grenze, sich der Grenze nähern, Berg an der Grenze, frei Grenze
 - Synonym von: Demarkationslinie, Trennungslinie, Scheidelinie
 - Form(en): Grenzen, Grenze
- Unterbegriffe: Ende, Rand, Limit, Obergrenze, Altersgrenze, Extrem, Landesgrenze, Gema rkung, Armutsgrenze, Strafraumgrenze, Stadtgrenze, Schmerzgrenze, Seegrenze, Scheide, Höc hstgrenze, Untergrenze, Staatsgrenze, Schneefallgrenze, Kapazitätsgrenze, Beitragsbemessun gsgrenze, Verwaltungsgrenze, Einkommensgrenze, Leistungsgrenze, Belastungsgrenze, Lohnuntergrenze, Millionengrenze, Grundstücksgrenze, Gemeindegrenze, Westgrenze, Tole ranzgrenze, Südgrenze, Nordgrenze, Promillegrenze, Ostgrenze, Freigrenze, Kreisgrenze, Mil liardengrenze, Kantonsgrenze, Versicherungspflichtgrenze, Zuverdienstgrenze, Landgrenze, Waldgrenze, Baumgrenze, Schuldengrenze, Außengrenze, Gemarkungsgrenze, Gehaltsobe r-grenze, Oder-Neiße-Grenze, Defizitgrenze, Ortsgrenze, Kostenobergrenze

Dornseiff-Bedeutungsgruppen:

- 1.16 Landbezirk: Begrenzung, Grenze
- 3.17 Außen: Fassade, Grenze, Oberfläche, Peripherie, Rand, Schliff, Straßenseite, Überzug, Vorderseite
- 3.22 Umgrenzung, Rand: Baulinie, Demarkationslinie, Gemarkung, Grenze, Grenzfluss, Grenzlinie, Grenzsperre, Grenzstein, Grenzwall, Landesgrenze, Limes, Rubikon, Scheidelinie, Staatsgrenze, Trennungslinie, Zonengrenze
- 4.34 Trennen: Absperrung, Fahne, Flagge, Grenze, Scheide, Scheidelinie, Symbol, Wand, Wasserscheide

С точки зрения концептологического анализа эти подразделы не представляют значительного интереса — уж слишком малочисленный и разнородный материал они охватывают. Единственное, на что здесь стоит, пожалуй, обратить внимание, так это на появление в списках фреймовых элементов (Fahne, Flagge, Symbol), а также индивидуальных представителей класса (Limes, Rubikon, Zonengrenze).

3.2. Концептологический анализ с помощью электронных онлайн-словарей

Аналогичный упрек можно высказать и в адрес самого онлайн-словаря (см. верхнюю часть статьи) — что, прежде всего, касается раздела Teilwort von: («...является частью...») и особенно Unterbegriffe : («видовые понятия»), в котором смешаны типы (Ende, Rand, Limit, Extrem) и подтипы, количество которых, в принципе, с трудом поддается обозрению.

Кроме того, в статье отсутствует раздел «таксон верхнего уровня», которым в нашем случае, очевидно, является *Linie*.

Гораздо более ценным в отношении концепт-анализа представляется OWID — еще один электронный онлайн-словарь корпусного типа, разрабатываемый в Институте немецкого языка Маннгейма, что в особенности касается раздела т. н. «кооккурентностей» [http://corpora.ids-mannheim.de/ccdb/?preload=http://corpora.ids-

mannheim.de/ccdb/db/4772/4772656e7a65/t4772656e7a6500.html?src=elex]:

2 770202 verschiedene Kookkurrenzpartner gespeichert		
Grenze	als Kookkurrenzpartner suchen	
C alphabetisch nach Kohäsion nach Wortart		
☐ Suchmuster (regexp) ve	erwenden	

Wort	Synsemantika	
Wort	ohne	mit
überschreiten	<u>32475</u>	<u>32475</u>
nahe ®	<u>29917</u>	<u>29917</u>
nahen	<u>23962</u>	23962
innerdeutsch	<u>23832</u>	23832
überschritten	<u>22958</u>	22958
jenseits ®	<u>21906</u>	<u>21906</u>
entlang ®	<u>14850</u>	14850
grünen	<u>10976</u>	10976
magisch	9031	9031
Belastbarkeit ®	8882	8882
polnisch	8389	8389
grün	<u>7471</u>	<u>7471</u>
tschechisch	<u>6874</u>	<u>6874</u>
Kilometer ®	6628	6628
überqueren	<u>6300</u>	6300
Albanien ®	<u>5648</u>	<u>5648</u>
unweit ®	<u>5611</u>	<u>5611</u>
Kosovo ®	<u>5149</u>	<u>5149</u>
zwischen ®	<u>4876</u>	4876
ungarisch	<u>4838</u>	4838
entfernt ®	<u>4720</u>	4720
illegal ®	<u>4645</u>	<u>4645</u>
schmuggeln ®	<u>4616</u>	<u>4616</u>
ober	<u>4197</u>	4197
Afghanistan ®	<u>4155</u>	<u>4155</u>

entfernen	<u>4071</u>	<u>4071</u>
verlaufen	<u>4036</u>	4036
untere ®	3802	3802
überquert	3439	3439
Öffnung	3404	3404
beiderseits ®	3241	3241
albanisch	3102	3102
gelegen	3088	3088
Mazedonien ®	3055	3055
diesseits ®	3002	3002
nördlich	<u>2956</u>	2956
angelangt ®	2798	2798
mazedonisch	<u>2735</u>	<u>2735</u>
Ungarn ®	<u>2718</u>	<u>2718</u>
Flüchtling	<u>2682</u>	<u>2682</u>
Polen ®	<u>2660</u>	<u>2660</u>
syrisch	<u>2652</u>	<u>2652</u>
Pole	<u>2649</u>	2649
anlangen	<u>2614</u>	<u>2614</u>
Pakistan ®	<u>2614</u>	<u>2614</u>
Nähe	<u>2607</u>	<u>2607</u>
obere ®	<u>2591</u>	<u>2591</u>
Pol	<u>2543</u>	<u>2543</u>
holländisch	<u>2516</u>	<u>2516</u>
Tschechien ®	2485	2485
Leistungsfähigkeit ®	2470	2470

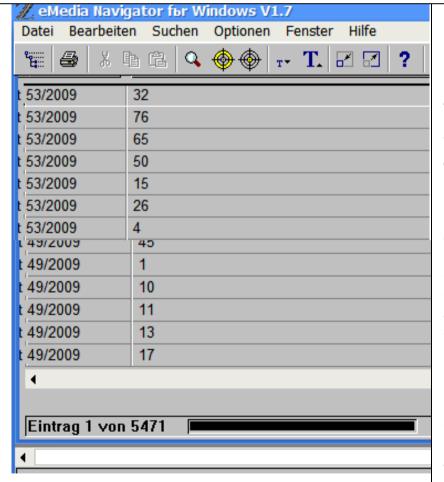
К каким же выводам можно придти на основе изучения первых по частотности 50 «попаданий»? Во-первых, они подтверждают высказанный ранее тезис о прототипичности территориальной (чаще всего, государственной) границы для этого концепта — к иным «семемам» («лимит», «предел») относятся разве что Belastbarkeit и Leistungsfähigkeit.

Значительное число коокурентностей связано с пространственным положением самой границы или тех или иных субъектов/объектов по отношению к ней: *nah*, *unweit*, *entfernt*, *verlaufen*, *gelegen*, *anlangen*, *Nähe* и т.д.

Многие «попадания» можно отнести к фрейму концепта (под которым мы понимаем его иерархические и коллатеральные связи с другими концептами в лингвокультурном сознании носителя соответствующих языка и культуры). Так, самое частотное действие, связанное с «границей», — это ее пересечение (überschreiten, überschritten, überqueren, überquert). Встречаются также нелегальная транспортировка через границу отдельных предметов (schmuggeln) или ее нелегальное пересечение отдельными лицами (Flüchtling) Чаще всего в немецкоязычном дискурсе обнаруживается упоминание границ ближайших соседей ФРГ (innerdeutsch, polnisch, tschechisch, ungarisch, holländisch), либо стран, где в последнее время имели место пограничные конфликты (Albanien, Kosovo, Afghanistan, Mazedonien, Pakistan) и т.д.

3.3. Концептологический анализ с помощью электронных версий архивов СМИ

Приведенные данные онлайн-словарей были получены в результате автоматической компьютерной обработки больших массивов дискурса. Лингвистические исследования, однако, часто требуют более точных результатов, которые можно получить только путем гораздо более тонкого «ручного» анализа. Такую возможность дает поиск по ключевому слову в электронных архивах различных органов СМИ, предлагаемых, в том числе, в форме CD- или DVD-ROM. Ниже (слева) приводятся результаты поиска по стимулу «Grenze» в электронном архиве еженедельника «Die Zeit» за 1994—2009 гг.



AUSGABE 53/09, Chancen

Gottes Wort im Oderbruch

Die Menschen ziehen weg aus den Dörfern Ostbrandenburgs. Der Berliner Vikar Robert Parr ging den entgegengesetzten Weg. Autor: Scholter, Judith.

Damit nicht auch noch Gott diese Gegend verlässt. Robert Parr nach Golzow gekommen. «Das sieht grausam aus», dachte er, als er zum ersten Mal hier ausstieg, außer dem verlassenen Bahnhäuschen weit und breit nichts, woran er sich orientieren konnte außer seinem an Glauben. Parr. 29, kommt eigentlich aus Berlin. Er ist Vikar, evangelischer Pfarrer in Ausbildung. Seit gut einem Jahr Oderbruch, lebt er im Kilometer von der polnischen Grenze entfernt, in einem Dorf mit 954 Einwohnern.

Как видим, общее число «попаданий» составило 5741 — объема языкового материала (полученного буквально за секунду), более чем достаточно для описания различных граней концепта GRENZE. С помощью одного клика компьютерной «мыши» можно развернуть

каждое из них в полный текст статьи с указанием года и номера издания, названия статьи и ее автора, страницы (см. колонку справа, где представлен фрагмент одной из таких статей). Искомое слово будет при этом выделено красным цветом.

В аспекте концептологического анализа этот материал интересен, прежде всего, тем, что позволяет с относительно небольшими затратами времени и усилий определить *референциал* концепта, то есть классы референтов, отражаемых им в общественном дискурсе.

4. Итоги и перспективы исследования

Подводя итоги вышеизложенного и оценивая дальнейшие **перспективы** исследований, отметим, что потенциал, открываемый компьютерными методами обработки лексикографического материала и больших массивов дискурса, еще не до конца осознан в лингвистике. Однако даже первые опыты использования таковых дают основания утверждать, что они способны кардинально облегчить решение, по крайней мере, части задач, стоящих перед языкознанием и другими смежными направлениями науки.

ЛИТЕРАТУРА

- 1. Донец П.Н. Концепт «GRENZE» в немецком языке: семасиологический угол зрения / П.Н. Донец // Известия Волгоград. гос. пед. ун-та. Серия «Филологические науки». 2005. № 3 (12). С. 18–22.
- 2. Донец П.Н. Концепт «GRENZE» в немецком языке: ономасиологический угол зрения / П.Н. Донец // Вестник Харьков. нац. ун-та имени В.Н. Каразина. 2006. № 726. С. 3–5.
- 3. Донец П.Н. Электронные словари и концептологический анализ зрения / П.Н. Донец // Вестник Моск. гор. пед. ун-та. Серия «Филология. Теория языка. Языковое образование». -2009. -№ 1(3). -С. 101-105.
- 4. Донец П.Н. Метод электронно-лексикографического поиска и концептология / П.Н. Донец // Вестник Харьков. нац. ун-та имени В.Н. Каразина. 2012. № 1023. С. 6–9.
- 5. Donec P. Die GRENZE: eine konzeptanalytische Skizze der Limologie / P. Donec. Würzburg : Königshausen & Neumann, 2014.

REFERENCES

- Donec, P.N. (2005). Koncept «GRENZE» v nemeckom jazyke: semasiologicheskij ugol zrenija [The concept «GRENZE» in German: a semasiological point of view]. *Izvestija Volgogr. gos. ped. un-ta. Volgograd State Pedag. Univ. Messenger*, *3* (12), 18–22 (in Russian)
- Donec, P.N. (2006). Koncept «GRENZE» v nemeckom jazyke: onomasiologicheskij ugol zrenija [The concept «GRENZE» in German: an onomasiological point of view]. *Visnyk Kharkiv. nats. un-tu im. V.N. Karazina. V.N. Karazin Kharkiv Uni. Messenger*, 726, 3–5.
- Donec, P.N. (2009). Elektronnyje slovari i konceptologicheskij analiz [Electronic dictionaries and the conceptological analysis]. *Vestn. Mosk. gor. ped. un-ta. Moscow State Ped. Univ. Messenger*, 1(3), 101–105.
- Donec, P.N. (2012). Metod elektronno-leksikograficheskogo poiska i konceptologija [The method of electronic-lexicographic search and conceptology]. *Visnyk Kharkiv. nats. un-tu im. V.N. Karazina. V.N. Karazin Kharkiv Uni. Messenger, 1023*, 6–9.
- Donec, P. (2014). *Die GRENZE: eine konzeptanalytische Skizze der Limologie*. Würzburg: Königshausen & Neumann.

Павел Николаевич Донец — доктор филологических наук, доцент, профессор кафедры немецкой филологии и перевода Харьковского национального университета имени В.Н. Каразина (пл. Свободы, 4, Харьков, 61022, Украина); e-mail: p_donec@list.ru

Pavel Donets – PdD, Full prosessor, German Language and Translation Dept., V.N. Karazin Kharkiv University (Svobody Sq., 4, Kharkiv, 61022, Ukraine); e-mail: p_donec@list.ru