

Н.А.Власенко, Н.Л.Кузьминская, А.А.Максименко

## **ЮНЕСКО о многоязычии и один из путей его эффективного использования**

**Введение.** Генеральная Ассамблея Организации Объединенных Наций провозгласила 2008 год Международным годом языков. Координация мероприятий возложена на ЮНЕСКО, как организацию, которая с момента своего учреждения (1945 год) уделяла огромное внимание сохранению культуры и языков. В послании Генерального директора ЮНЕСКО Коитиро Мацууры в связи с проведением Международного года языков говорится: «Следует повсюду продвигать идею о том, что «языки – это важно!», будь то с помощью инициатив в области образования и киберпространства и в интеллектуальной среде; будь то посредством проектов, направленных на сохранение языков, находящихся под угрозой исчезновения, или на поощрение языков в качестве средства социальной интеграции; будь то в целях изучения связи между языками и экономикой, между языками и знаниями коренных народов или между языками и творчеством.»

Количество языков, на которых разговаривают в современном мире, оценивается по-разному в промежутке между 6 000 и 7 000 языков. Сохранение этого разнообразия одна из главных задач ЮНЕСКО, потому что как минимум половина из этих языков могут исчезнуть в ближайшие десятилетия. Менее четверти этих языков представлены в школьном образовании, еще намного меньше - в киберпространстве. Недавно прошедшая 34-й сессия Генеральной конференции ЮНЕСКО снова обратилась к проблеме многоязычия в киберпространстве. Дискуссия по этому вопросу вызвала большой интерес у государств-членов, важность решения этой проблемы была поддержана всеми участниками Генеральной конференции. «В государствах-членах был достигнут определенный прогресс в том, что касается осознания важности "Рекомендации о развитии и использовании многоязычия и всеобщем доступе к киберпространству" для развития общества знаний. В государствах-членах на различных уровнях были приняты

меры административного, политического и законодательного характера, направленные на содействие применению положений этой Рекомендации» [1].

### **1. Позиция и инициативы ЮНЕСКО в отношении многоязычия.**

Культурное и языковое разнообразие понимается как часть общечеловеческого наследия, которое является источником обмена и возникновения новых идей, источником развития творчества. Культурное и языковое разнообразие также необходимо для человечества, как биологическое многообразие для живой природы. Поэтому его необходимо сохранять в интересах нынешнего и будущего поколений и рассматривать в качестве одного из основных прав человека.

Сознавая свою миссию, ЮНЕСКО на 31-й сессии Генеральной конференции принимает Всеобщую декларацию ЮНЕСКО о культурном разнообразии (Париж, 2 ноября, 2001 г.). В октябре 2003 года на 32-й сессии Генеральной конференции ЮНЕСКО принимает два стандартоустанавливающих документа: «Рекомендации о развитии и использовании многоязычия и всеобщем доступе к киберпространству» [2] и «Хартия о сохранении цифрового наследия» [3]. В этих документах особое внимание уделено потенциалу информационных и коммуникационных технологий (ИКТ) с точки зрения развития общества знаний, их роли в обеспечении свободы выражения мнений, возможности получения доступа к образованию и информации, в обеспечении языкового и культурного разнообразия. В преамбуле к Всеобщей декларации подчеркивается, что процесс глобализации, стимулируемый быстрым развитием новых информационных и коммуникационных технологий, хотя и представляет вызов для культурного разнообразия, вместе с тем создает условия для нового диалога между культурами и цивилизациями.

Позиция ЮНЕСКО состоит в том, что в подлинно многоязычном мире не могут и далее доминировать всего несколько «глобальных» языков, а должны использоваться также менее распространенные языки. При всей важности установочных документов по вопросам политики и программ на национальном уровне, ключевое значение имеет учет

местных потребностей, поскольку создание плюралистических и инклюзивных обществ знаний – это процесс, развивающийся, в основном, на низовом уровне.

## **2. Проблемы обеспечения многоязычия и доступа к информации**

Глобализация открыла новую эру интерактивного общения между народами, государствами и частными лицами, но она также вызывает определенную обеспокоенность, возможно, ведет к потере своеобразия своей культуры у того, кто сталкивается с другой и либо полностью растворяется в ней (при ассимиляции), либо оказывается “лишенным культуры” (в случае маргинализации). Международные структуры и государственные власти должны это учитывать, особенно, когда речь идет о доступе к знаниям, о распространении новых информационных и коммуникационных технологий и о развитии многоязычия в киберпространстве.

Решение этих проблем не может быть простым, поскольку даже если информационные технологии смогут существенно улучшить свободное распространение знаний, они же могут привести к еще большему неравенству между теми, кто имеет доступ ко всему информационному изобилию (информационно богатыми) и теми, кто этого лишен (информационно бедными). Аналогичным образом, доминирование группы языков может ограничить средства выражения и привести в итоге к унификации культуры.

Руководствуясь этими соображениями, ЮНЕСКО приняла решение стимулировать государства-члены к принятию рекомендаций о поддержке использования многоязычия и об обеспечении всеобщего доступа к киберпространству.

ЮНЕСКО предлагает в Рекомендациях большое количество конкретных мер. Например, таких: государства-члены должны определять и поддерживать национальную политику, стимулирующую изучение языков в киберпространстве, включая родные языки. Они должны также поддерживать усилия по предоставлению бесплатных общедоступных услуг по автоматическому переводу, а также по использованию интеллектуальных лингвистических систем, в частности, таких, которые способны выполнять многоязычный

поиск информации. В Рекомендациях также предлагается общественному и частному секторам, а также гражданскому обществу на местном, региональном и международном уровнях, работать в направлении уменьшения языковых барьеров и оказания помощи во взаимодействии людей посредством Интернета (32-я сессия Генеральной конференции, Париж, октябрь 2003г.,) [2].

Государства-члены и международные организации должны рассматривать предоставление доступа к Интернету как услугу, имеющую общественную значимость. Граждане должны иметь возможность доступа к общественным и правительственным документам, включая документы, содержащие всю информацию, которая необходима для жизни в современном демократическом обществе. И отдельные государства, и международные организации должны оказывать людям помощь в овладении навыками работы на компьютере и обеспечивать доступ к информации и услугам в Интернете. Они должны также стимулировать производство местного информационного содержания (контента). Более четырех миллионов европейцев в настоящее время работают в секторе создания и наполнения контента. Задачи создания контента в свою очередь дали толчок к развитию творчества и радикально изменили сами пути доступа к знаниям, средства получения знаний и послужили базисом для экономического роста, профессионального, социального и культурного развития граждан, их творческих и инновационных способностей. Сохранение многоязычия в настоящее время является приоритетной и стратегической задачей.

Наиболее интересным сетевым проектом, охватывающим 253 языка (практически все языки, представленные в Интернете) является стартовавший в 2001 году проект Википедия (англ. *Wikipedia*), получивший свое название от используемой технологии *Wiki* - технологии взаимодействия пользователя с веб-сайтом, предназначенной для коллективной разработки, хранения и структуризации информации (в основном гипертекста). Википедия - многоязычная общедоступная свободно распространяемая энциклопедия, публикуемая в Интернете. Создается эта энциклопедия коллективным

трудом добровольных авторов. Практически каждый пользователь, имеющий доступ к Интернету, может пополнить ее информацией, являющейся достоверной.

Это говорит о том, что технологически вопрос обеспечения языкового разнообразия в киберпространстве в настоящее время решен. Многоязычный материал Интернета представляет собой бесценный образовательный ресурс, который нужно умело использовать. Изучая многоязычные ресурсы, пользователи не только углубляют знания иностранных языков, но и лучше понимают особенности родного языка. Одной из наиболее творческих задач, является задача исследования многоязычных текстов.

### **3. Исследование проблем, связанных с обработкой многоязычных текстов**

Впервые с проблемой обработки многоязычных текстов авторы столкнулись при попытке проанализировать тексты учебных курсов на заданную тематику, представленные в Интернете на разных языках. Необходимо было проанализировать частоту встречаемости терминов в учебном тексте и их распределение в рамках определенного курса. Хотя критерий частотности, т.е. частоты встречаемости речевых (языковых) единиц, получает в настоящее время большое распространение в филологии, лингвистике, социологии, психологии и т.п., нельзя сказать, что ему отводится то место, которое он на самом деле заслуживает. Применение этого критерия в мультилингвистических исследованиях может дать весьма неожиданные результаты и стать для исследователя языка источником информации, определяющей специфику ассоциативного мышления носителей того или иного языка.

Авторы нашли в сети целый ряд программ, составляющих частотные словари, которые удовлетворительно работали с текстами на русском и английском языках, но были проблемы с анализом текстов на украинском языке, а особенно – на французском языке, где в изобилии представлены диакритические знаки (различные надстрочные, подстрочные знаки, применяемые для изменения или уточнения значения других знаков, например, Ââ Çç Èè) и лигатуры (знаки, образованные путем соединения двух и более графем, как например, Еёе). Различную роль в разных языках играет и небуквенный

орфографический знак апостроф. В зависимости от языка и выполняемой роли апостроф причисляют к знакам препинания, диакритическим знакам и иным категориям. Но даже внутри одного языка роль апострофа бывает различной. Например, в английском языке апостроф может

- служить орфографическим выражением притяжательного падежа для отличия от аналогичных форм множественного числа: *teachers* «учителя», *teacher's* «принадлежащий учителю», *teachers'* «принадлежащий учителям»;
- обозначать пропуск гласных (*I'm* вместо полного варианта *I am*), последовательностей звуков (*he'll*) и др.

Во французском языке апостроф чаще всего используется для пропуска гласных *a* и *e* в артиклях перед словами, начинающимися с гласных и *h*. Например, слова *l'année* = *la* + *année*, *l'an* = *le* + *an*, переводящиеся на русский одинаково «год», но одно из них является словом женского рода (*année*), а другое – мужского (*an*) и употребляются они в разных словосочетаниях и согласно правилам полная форма не употребляется. Если, например, мы исследуем тексты на французском языке с целью определить количество конструкций с пропуском гласных, нам следует рассматривать конструкции типа *l'année* и *l'an* как одно слово, если же мы исследуем соотношение в языке слов с определенным артиклем, нам при составлении частотного следует разбивать эту конструкцию на *l'* и *année* и др.

Другой пример. Иногда при составлении частотных словарей нам необходимо рассматривать число как отдельное слово и учитывать частоту его употребления в словаре (например, при анализе текстов на историческую тематику с указанием дат), а для других задач нам лучше числа опускать.

Если перед нами стоит задача исследовать творчество Л.Н.Толстого с точки зрения особенностей французской лексики, употребленной им в романе «Война и мир», нам следует создавать частотный словарь только французской лексики, если же мы исследуем особенности русской лексики, то при составлении частотного словаря опускаем

французские слова. Если же в целом исследуем этот роман, то необходимо строить как словарь французской лексики, так и русской.

Вышеизложенное говорит о невозможности создания универсальной программы составления частотных словарей, одинаково успешно работающей со всеми языками и для различных классов лингвистических задач. Выход мы увидели в создании простой интерактивной программы, предоставляющей возможность самому пользователю настраивать ее на различные языки и различные задачи контент-анализа и **позволяющей!!!** максимально возможно собирать статистику, как по анализируемым текстам, так и по созданным словарям. Была создана первая версия такой программы – TextAnalyzer\*, работоспособность которой была проверена на текстах на четырех языках и различных классах лингвистических задач.

#### **4. Описание программы TextAnalyzer**

Программа написана на объектно-ориентированном языке программирования C#, работающем в среде исполнения .NET Framework 2.0. В инсталляционный пакет программы TextAnalyzer входит .NET Framework 2.0, поэтому, в случае **ее!!!** отсутствия, не требуется дополнительная установка. Среда .NET Framework 2.0 работает под управлением Windows 2000/XP/2003/Vista.

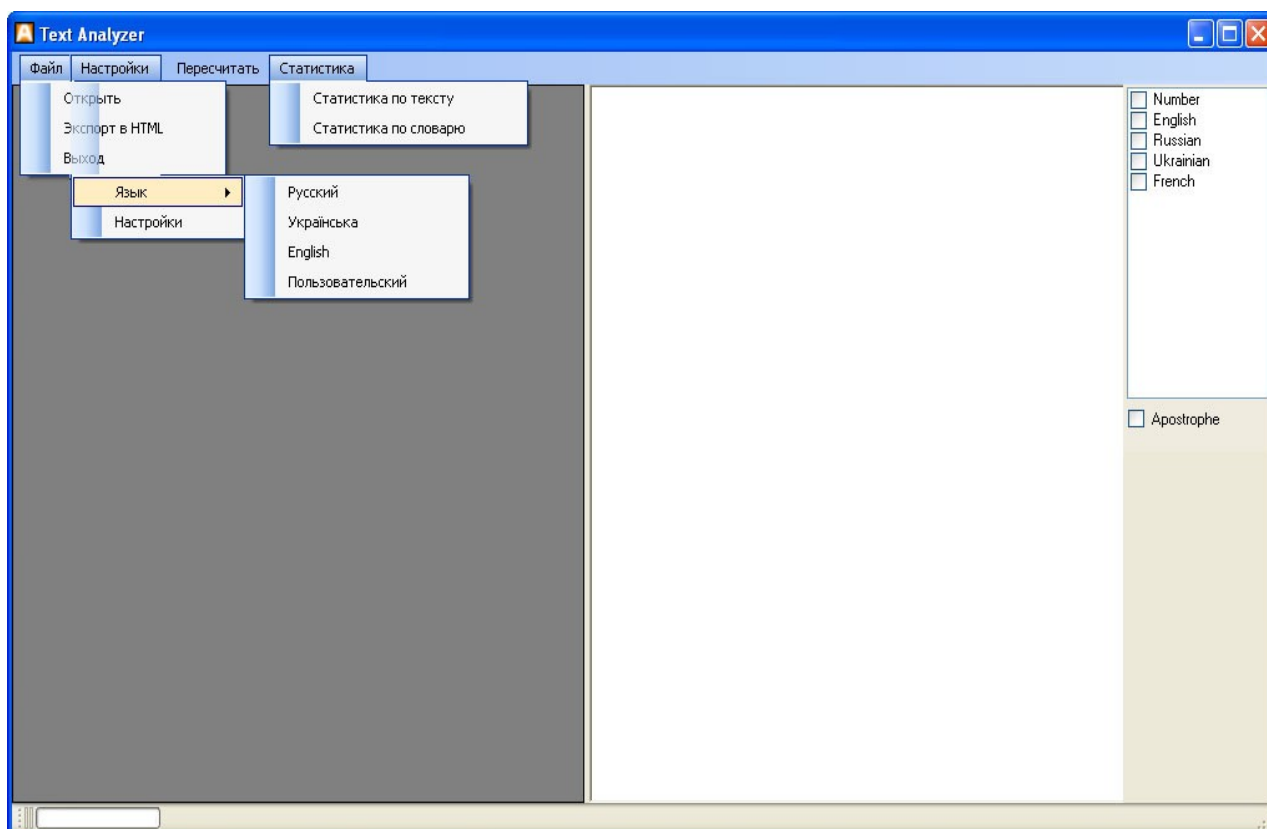
Минимальные системные требования: частота процессора 800 MHz, операционная система Windows 2000/XP, объем оперативной памяти 256 Mb RAM.

Программа TextAnalyzer составляет частотный словарь для любого текстового файла, сохраненного в формате txt. Перед работой с программой можно выбрать язык интерфейса (Настройки/Язык): русский, украинский, английский, пользовательский (изменяется в дополнительном к программе файле, по умолчанию используется английский язык).

---

\* Программа написана Кузьминским В.Н.

Окно и все меню программы имеют вид:



Перед загрузкой файла нужно отметить язык или несколько языков, лексика которых встречается в анализируемом тексте, в случае необходимости включать в словарь числа – отметить Number, а в случае необходимости рассматривать языковые конструкции с апострофом как одно слово - отметить Apostrophe (двойное нажатие левой клавишей мыши на нужный квадратик).

После загрузки текстового документа (**Файл/Открыть**) программа составляет частотный словарь. Первоначально слова в словаре представлены в порядке убывания частоты встречаемости в тексте.

Тексты для анализа можно загружать непосредственно в окно, используя операцию Копировать/Вставить.

Частотный словарь представлен в виде таблицы, состоящей из 4 столбцов: «Слово», «Длина» (количество букв в слове), «Частота» (частота появления слова в тексте), «Относительная частота» (относительная частота появления слова в тексте, равна



отношению частоты к общему количеству слов в тексте). В нижней части экрана приведено количество уникальных и общее количество слов в тексте.

Словарь можно легко сортировать простым нажатием на поле названия столбца. Так, при нажатии «Слово», словарь автоматически сортирует слова по алфавиту (от первой буквы алфавита до последней). При повторном нажатии этого поля слова в словаре сортируются в обратном порядке (от последней буквы алфавита до первой). При нажатии «Длина» идет сортировка по длине слова – от минимальной до максимальной и наоборот и т.д.

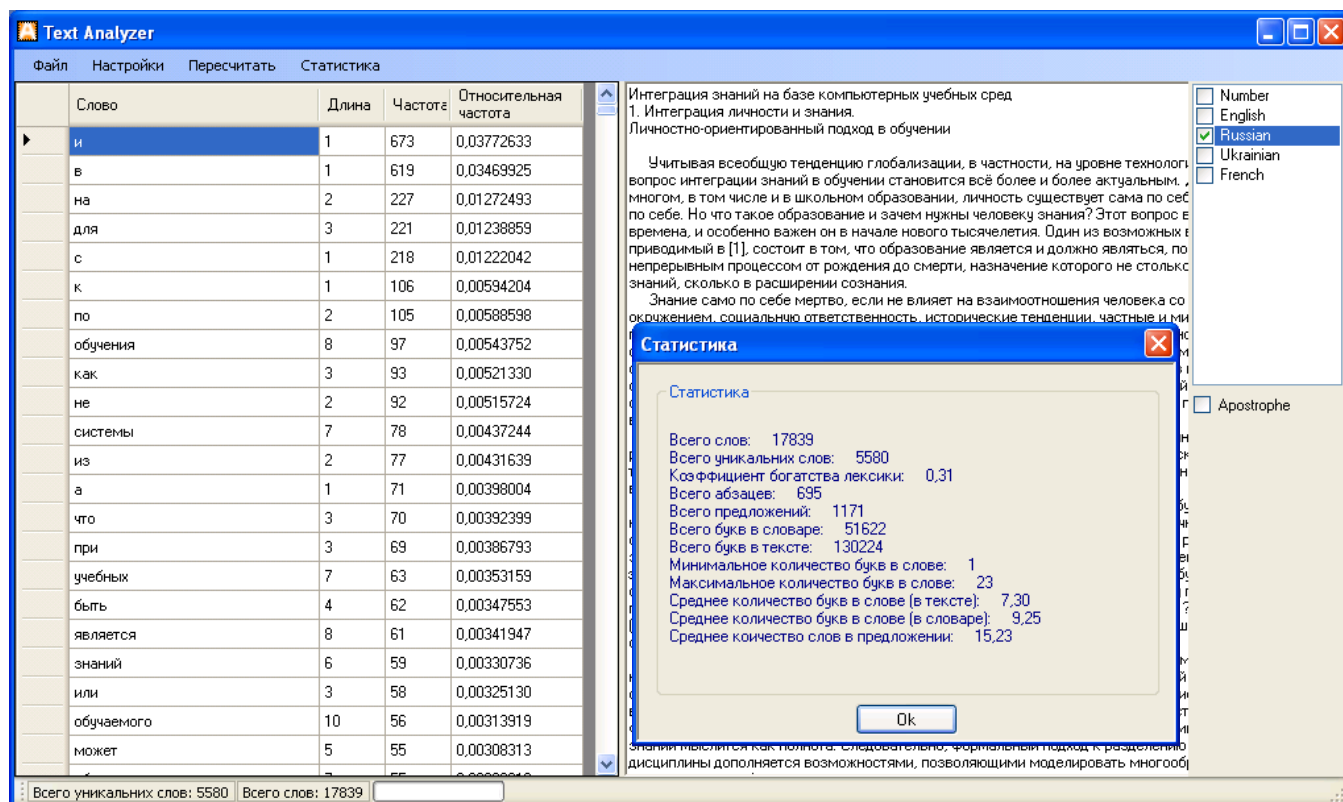
Программа также позволяет составлять частотный словарь после редактирования текста непосредственно в окне программы. Для этого после редактирования текста следует нажать пункт меню **Пересчитать**.

Пункт меню *Статистика* включает в себя два подпункта: *Статистика по тексту* и *Статистика по словарю*. В статистике по тексту представлена информация: всего слов, всего уникальных слов, всего абзацев, всего предложений, всего букв в словаре, всего букв в тексте, минимальное количество букв в слове, максимальное количество букв в слове, среднее количество букв в слове (в тексте), среднее количество букв в слове (в словаре). Статистика по словарю включает длину слова, его частоту в словаре и тексте, а также относительную частоту в словаре и тексте.

На экране приведены фрагменты частотного словаря, составленного на материале докладов, представленных на международную конференцию «Новые информационные технологии в образовании для всех», проводимую в Международном научно-учебном центре информационных технологий и систем в 2006 и 2007 годах и статистика по тексту.

Частотный словарь можно сохранить в формате HTML-документа. Этот формат позволяет его просматривать в виде таблицы, что достаточно удобно для дальнейшего использования.

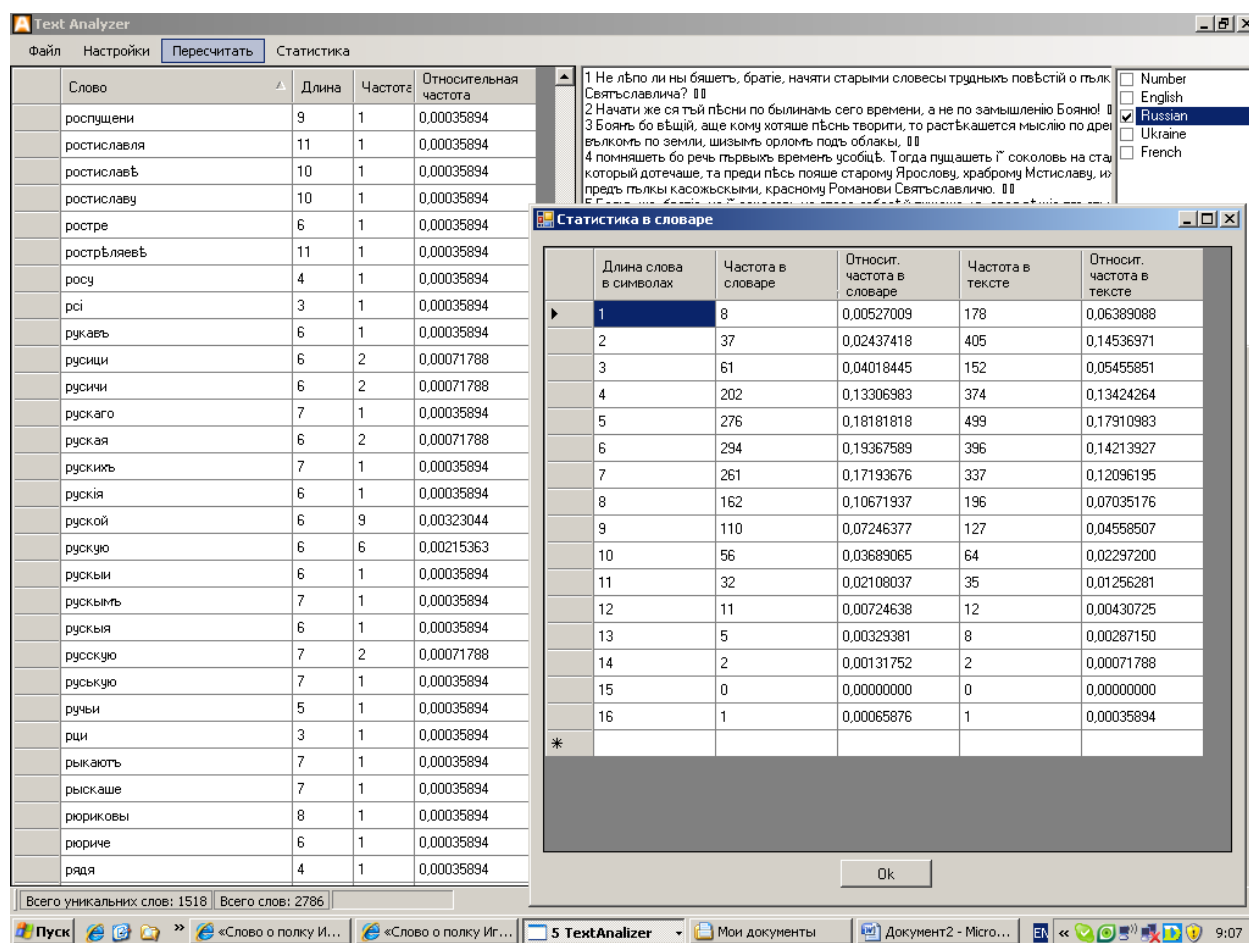
В меню **Настройки/Настройки** возможно проводить настройки фильтрации (добавление новых языков и их алфавитов) и настройки HTML-таблицы (настройка дизайна таблицы при сохранении частотного словаря).



## 5. Пример использования программы TextAnalyzer для анализа переводов «Слова о полку Игореве»

Проверить работу программы было решено с анализа памятника древнерусской литературы «Слова о полку Игореве» и его переводов на современный русский, украинский и английский языки – по два перевода на каждый язык. Тексты «Слова...» и его переводов брались из параллельного корпуса переводов этого произведения (<http://nevmenandr.net/slovo/>). Это электронный филологический инструмент для сравнения переводов памятника древнерусской литературы. Каждый текст в этом корпусе разбит на 218 фрагментов («звеньев») в соответствии с членением «Слова...», предложенным Р.О.Якобсоном. Поскольку в корпусе используется лишь один перевод «Слова...» на украинский язык В.Шевчука, решено было использовать еще один ритмический перевод на украинский язык Л. Е. Махновца (<http://litopys.org.ua/index.html>).

При построении частотного словаря древнерусского текста было решено не добавлять в *Настройке Древнерусский язык* и вносить туда его алфавит, а добавить в *Настройке Русский* несколько символов из древнерусских текстов (Ѣ, Ѣ̃, і, ї ) простым копированием из оригинала, в результате чего все символы корректно отображались как в тексте, так и в словаре. На экране представлен фрагмент частотного словаря древнерусского текста «Слова...», упорядоченный по алфавиту и статистика по словарю.



Проанализировав частотный словарь древнерусского текста летописи, сразу отмечается неодинаковое написание одних и тех же слов в тексте летописи. На экране мы специально остановились на букве «р», где видны разные варианты написания слов *русици* и *русичи* (оба с частотой 2); *рускаго*(1), *рускихъ*(1), *рускія*(1), *руской*(9), *рускую*(6), *русьи*(1), *русьмъ*(1), *русья*(1), *руську*(2) *руську*(1). Другой яркий пример - *Владимірь* и *Володимиръ*. Таких примеров в словаре десятки. Это говорит об отсутствии определенных норм произношения и написания древнерусских слов, что нашло потом свое отражение в формировании русского, белорусского и украинского языков. При чтении текста летописи на это просто можно не обратить внимание. Такие же словари были построены и для переводов «Слова...». Сводные статистические данные представлены в Таблице 1.

Таблица 1

| Источник<br>Статистика                                                | <i>Древне-<br/>рус</i> | <i>Рус.<br/>Лихачев</i> | <i>Рус.<br/>Творогов</i> | <i>Укр.<br/>Шевчук</i> | <i>Укр.<br/>Махновец</i> | <i>Анг.<br/>Набоков</i> | <i>Анг.<br/>Хани и<br/>Даль</i> |
|-----------------------------------------------------------------------|------------------------|-------------------------|--------------------------|------------------------|--------------------------|-------------------------|---------------------------------|
| <b><i>Всего слов</i></b>                                              | 2786                   | 2845                    | 2840                     | 2629                   | 2833                     | 4148                    | 3982                            |
| <b><i>Всего<br/>уникальных<br/>слов</i></b>                           | 1518                   | 1420                    | 1436                     | 1555                   | 1437                     | 1196                    | 1091                            |
| <b><i>Коэффициент<br/>богатства<br/>лексики</i></b>                   | 0,54                   | 0,50                    | 0,51                     | 0,59                   | 0,51                     | 0,29                    | 0,27                            |
| <b><i>Максимальное<br/>количество<br/>букв в слове</i></b>            | 16                     | 16                      | 16                       | 14                     | 15                       | 13                      | 13                              |
| <b><i>Среднее<br/>количество<br/>букв в слове<br/>(в тексте)</i></b>  | 5,10                   | 5,13                    | 5,12                     | 5,17                   | 5,07                     | 4,38                    | 4,41                            |
| <b><i>Среднее<br/>количество<br/>букв в слове<br/>(в словаре)</i></b> | 6,25                   | 6,37                    | 6,39                     | 6,28                   | 6,26                     | 5,78                    | 5,75                            |
| <b><i>Среднее<br/>количество<br/>слов в<br/>предложении</i></b>       | 13,46                  | 14,01                   | 13,21                    | 9,36                   | 13,96                    | 19,94                   | 17,26                           |

Можно анализировать как отдельные словари, так и несколько словарей сразу на предмет совпадения лексики. При анализе всех словарей отмечено, что наиболее частотными являются слова *Игорь* и *земля* (наиболее часто встречаемое сочетание

Русская земля) – ключевые слова произведения, но удивляет не одинаковая частота этих слов в переводах. Так, в переводе В.Шевчука словосочетание «землю Руськую» переводится «рідну землю»(7), у В.Набокова «землю Половецкую» переводится «the Kuman land», и т.п. При анализе словаря Набокова было отмечено (среди многих прочих) отсутствие слова соответствующего «Пирогощей». В анализ фрагмента 213 из параллельного корпуса, где это слово встречается, включаем еще и перевод Магнуса на английский язык. Результаты параллельного сравнения приведены в Таблице 2.

Таблица 2

| Источник                     | Текст                                                                        |
|------------------------------|------------------------------------------------------------------------------|
| Древнерусский текст          | Игорь ѣдетъ по Боричеву къ святѣй Богородици Пирогощей.                      |
| Перевод Д. С. Лихачёва       | Игорь едет по Боричеву ко святой богородице Пирогощей.                       |
| Перевод О. В. Творогова      | Игорь едет по Боричеву к святой Богородице Пирогощей.                        |
| Перевод В. Шевчука           | Вже Ігор їде Боричевим, радий Вклонитись чудотворній Пирогощій               |
| Перевод Л. Е. Махновца       | Ігор їде по Боричевім до святої богородиці Пирогощої.                        |
| Перевод В.Набокова           | Igor rides up the Borichev [slope] to the Blessed Virgin of the Tower;       |
| Перевод В. Хани и<br>Э. Даля | Igor rides up the Borichev slope to the Holy Mother of God Pirogoshchaia     |
| Перевод Магнуса              | Ígorí repairs up [the hill] Borířčev to the Holy Mother of God at Pirogóřšč. |

Как видно из этого и других фрагментов, переводчики испытывают трудности с транслитерацией русских букв *ш, щ, ч, ц*. Для передачи особенностей русской речи Магнус даже использует диакритические знаки, не существующие в английском языке.

Другие особенности языков можно легко проследить на графиках, представленных на рисунках 1 и 2, построенных с использованием данных из *Статистика по словарю* и отражающие зависимости относительной частоты слова от его длины в тексте и словаре.

Полученные результаты из Таблицы 1 и рисунков можно сравнить данными исследований, проводимыми в [1], где анализировались русские и английские научно-технические тексты объемом более 400Мб. В этих исследованиях определялась средняя длина русского и английского слова в заголовках документов и рефератах по Биологии по тесту и словарю. В заголовках этот показатель для русского языка -7,48 по тексту и 10,8 по словарю; для английского языка – 6,21 и 10,3 соответственно. Для рефератов эти показатели - 6,54 и 9,91 для русского и 5,56 и 10,01 для английского. В наших экспериментах с текстами докладов по информационным технологиям на русском языке

этот показатель был 7,30 и 9,25, для украинского 6,31 и 9,13, а для английского 5,63 и 9,11. Эксперименты с художественными текстами на этих языках показали, что слова научных текстов в среднем длиннее слов художественной литературы. Самыми короткими являются английские слова, затем украинские и, наконец, русские.

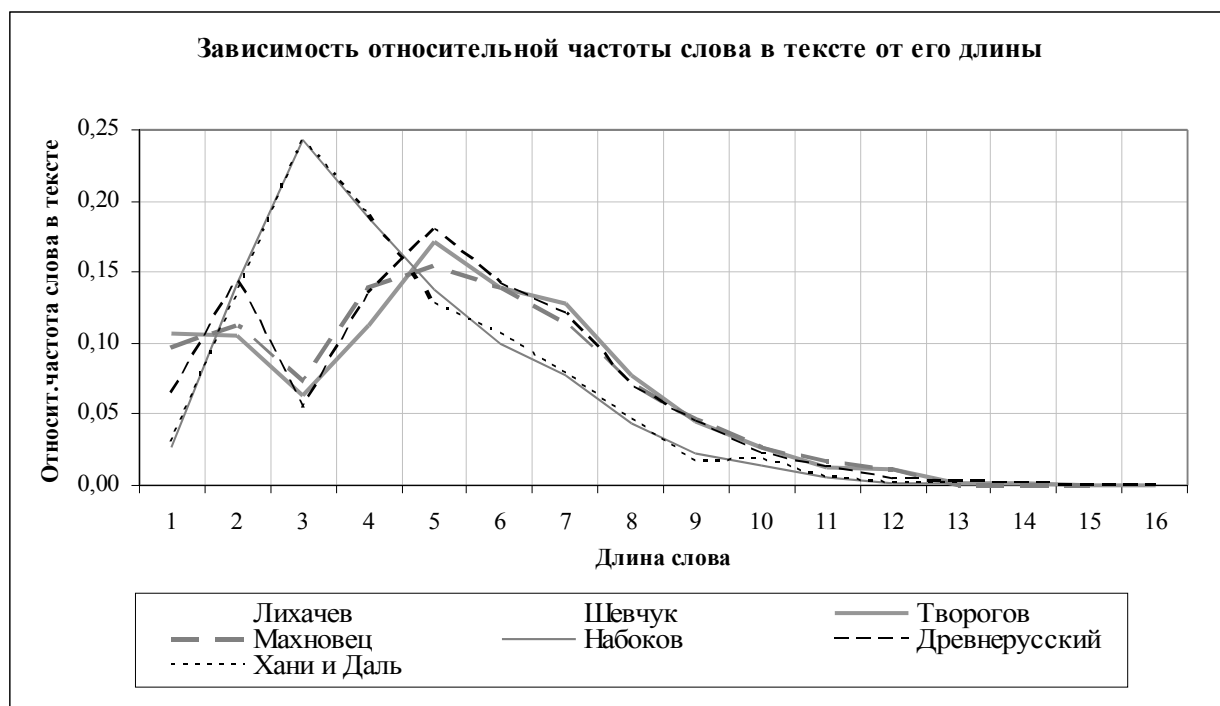


Рис.1

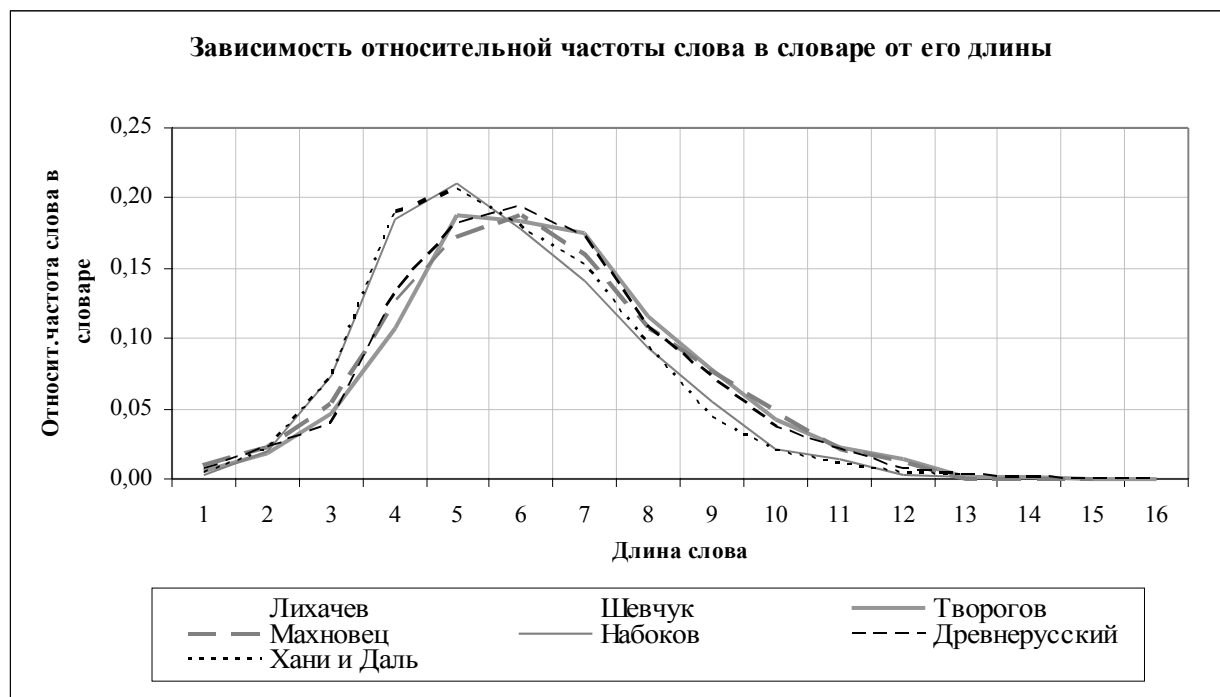


Рис.2.

## 6. Мультилингвистическое исследование паремий

Мир ассоциаций каждого отдельно взятого человека индивидуален. Тем не менее, направление лингвистической ассоциации носителей одного языка и культуры чаще всего совпадает и определяет особенность национального характера. Специфические для того или иного языка ассоциации следует прежде всего искать в области паремий.

Термин «паремия» (гр. *παροιμία* — притча, афоризм) сравнительно недавно вошел в лингвистическую терминологию для обозначения класса номинативных единиц языка, имеющих вид законченного суждения, которое легко укладывается в сознании. Чаще всего этот термин употребляется как объединяющий для пословиц, поговорок, крылатых выражений, афоризмов и т. п. Исследуя и сравнивая подсознательные ассоциации, заключенные в паремиях на различных языках, можно проникнуть в глубину коллективного сознания носителей этих языков, потому что именно ассоциативные ряды определяют тип языкового мышления. Изучение любого языка должно начинаться в первую очередь с изучения культурных традиций, обычаев, религиозных и нравственных представлений народа – носителя этого языка. Всю эту информацию мы можем легко отыскать в паремиях. Кроме того, в них с одной стороны отображены особенности (в том числе и грамматические) того или иного языка, а с другой – в паремиях можно легко найти контакты между носителями как близких, так и дальних языков (из других языковых семей и групп).

Попытаемся это сделать на материале коллекций паремий, собранных в украинском, русском и французском языках, посвященным дням недели. Неделя была выбрана не случайно. Существует мнение, что наши далекие предки имели шестидневную неделю.

Этот факт подтверждается археологически. В [4] описывается ритуальный сосуд, найденный в Ромашках на Роси, где выделяется гирлянда из шести квадратов, соответствующая древнерусской неделе. Последний день недели всегда был выходным и посвящался солнцу или другим божествам. Не потому ли среди чисто славянских

названий дней недели с прозрачной этимологией, мы встречаем неславянское слово суббота, этимологию которого возводят к иудейскому слову "саббат" - день покоя, в который запрещено работать и в который человек должен восстанавливать свои силы, размышлять о собственном месте в окружающем мире. Следующее название понедельник, может означать, что этот день следует за днем, в который ничего не делают (в нашем случае за субботой).

Систематизированные сборники славянских пословиц появляются только в XIX веке, хотя в древних летописях мы встречаем множество паремий, а «Моление Даниила Заточника» XIII ст. [9] можно уверенно рассматривать как первый древнерусский сборник паремий. Систематизированные французские сборники пословиц появляются уже в XVI веке, что позволяет говорить о том, что французские паремии сохранились лучше, чем славянские, где большая часть пословиц и поговорок была утрачена.

Некоторые статистические данные, полученные с использованием программы **TextAnalyzer** приведены в Таблице 3.

Таблица 3

| Дни          |      | Количество паремий | Общее количество слов в паремиях | Миним-е количество слов в паремии | Максим-е количество слов в паремии | Среднее количество слов в паремии |
|--------------|------|--------------------|----------------------------------|-----------------------------------|------------------------------------|-----------------------------------|
| Поне-дельник | Укр. | 26                 | 256                              | 3                                 | 25                                 | 9,846                             |
|              | Рус. | 35                 | 491                              | 3                                 | 37                                 | 14,029                            |
|              | Фр.  | 13                 | 117                              | 2                                 | 15                                 | 9                                 |
| Втор-ник     | Укр. | 7                  | 52                               | 3                                 | 22                                 | 7,429                             |
|              | Рус. | 8                  | 66                               | 2                                 | 15                                 | 8,25                              |
|              | Фр.  | 14                 | 138                              | 5                                 | 20                                 | 9,857                             |
| Сред-а       | Укр. | 57                 | 428                              | 3                                 | 17                                 | 7,509                             |
|              | Рус. | 23                 | 209                              | 3                                 | 24                                 | 9,087                             |
|              | Фр.  | 15                 | 192                              | 3                                 | 19                                 | 12,8                              |
| Чет-верг     | Укр. | 34                 | 248                              | 4                                 | 13                                 | 7,294                             |
|              | Рус. | 21                 | 177                              | 4                                 | 16                                 | 8,429                             |
|              | Фр.  | 12                 | 109                              | 5                                 | 19                                 | 9,083                             |
| Пят-ница     | Укр. | 39                 | 344                              | 4                                 | 41                                 | 8,821                             |
|              | Рус. | 43                 | 354                              | 4                                 | 23                                 | 8,233                             |
|              | Фр.  | 46                 | 518                              | 7                                 | 38                                 | 11,261                            |
| Суббо-та     | Укр. | 24                 | 146                              | 3                                 | 9                                  | 6,083                             |
|              | Рус. | 28                 | 195                              | 3                                 | 14                                 | 6,964                             |
|              | Фр.  | 36                 | 445                              | 4                                 | 34                                 | 12,361                            |



|                      |      |     |      |   |    |        |
|----------------------|------|-----|------|---|----|--------|
| сентябрь-воскресенье | Укр. | 48  | 424  | 3 | 32 | 8,833  |
|                      | Рус. | 16  | 161  | 4 | 16 | 10,063 |
|                      | Фр.  | 41  | 499  | 2 | 29 | 12,171 |
| Всего                | Укр. | 235 | 1898 | 3 | 41 | 8,077  |
|                      | Рус. | 174 | 1653 | 2 | 37 | 9,5    |
|                      | Фр.  | 177 | 2018 | 2 | 38 | 11,401 |

Сразу же отметим, что данные в Таблице 3 отражают общую устойчивую тенденцию, но не могут претендовать на роль абсолютной истины, поскольку (как отмечалось ранее) большая часть паремий была утрачена, не успев зафиксироваться в письменных источниках. Как видим, наибольшее количество удалось собрать украинских паремий, среди которых оказалось наибольшее минимальное количество слов в паремии – 3 и наибольшее максимальное количество слов в паремии – 41, но при всем при этом среднее количество слов в паремии – наименьшее (8,077). Во французском же языке средняя длина паремии оказалась наибольшей – 11,401 слова (за счет артиклей). Таблица показывает, что украинский язык оказался наиболее лаконичным в плане выражения. Во французском языке наибольшее количество паремий относится к *пятнице* (как наиболее неблагоприятному дню), а также к *субботе* и *воскресенью*. В украинском языке наибольшее количество паремий посвящено *среде*, *пятнице* и *воскресенью*, в русском же языке – *понедельнику*, *пятнице* и *субботе*. Интересно отметить, что во французском языке больше паремий посвящено благоприятным, счастливым дням, а в славянских – наоборот. Интересно отметить, что в славянских языках, чем легче и благоприятней считается день, тем меньше ему посвящено паремий. Самый яркий пример – *вторник*. Удивляет тот факт, что в русском языке сравнительно мало пословиц, посвященных *воскресенью*. Может этот факт также можно отнести к тому, что седьмой день недели у славян появился позже, но только в русском языке он получил название отличное от слова *неделя* (хотя слово *неделя* в значении *воскресенье* встречается в русских паремиях - *Неделя в неделю жди четверга*). Как видим даже сухие цифры могут многое рассказать об общих тенденциях распределения паремий по выбранной тематике в разных языках.

Проанализировав наиболее частотные слова коллекций можно многое сказать о представлении разных народов о неделе. Так, среди украинских паремий кроме названий дней недели и служебных слов в выбранном фрагменте наиболее частотными являются слова: *буде*(28), *тепер*(23), *робота*(12), *можна*(11)\*, *тиждень*(11), *день*(10), *сім*(10), *трапиться*(10); среди русских - *день*(12), *будет*(11), *нельзя*(11), *работу*(8), *великий*(8), *неделю*(7), *веселье*(7), *прясть*(6)\*\*, *семь*(6); среди французских – *soleil*-солнце(27), *saint*-святой(23), *femme*-женщина(20), *lune*-луна(10), *amour*-любовь(10), *faire*-делать(9), *tombe*-приходится(9). Первое, на что сразу же обращается внимание – это частота употребления слова *soleil*, причем 25 употреблений этого слова выпадает на субботу. При всем обилии французских пословиц со словом *солнце*, они несколько однотипны. Вот только некоторые из них:

*Il n'y a pas de **samedi** sans **soleil** comme il n'y a pas de fille sans amour.*

*Il n'y a pas de **samedi** sans **soleil** ni de femme sans conseil.*

*Jamais de veuve sans conseil ni de **samedi** sans **soleil**.*

*Il n'y a femme sans amour, **samedi** sans **soleil**, dimanche sans plaisir ni vieillard sans douleur.*

*Pas de **samedi** sans **soleil**, pas de jeune veuve sans deuil, pas de jeune fille sans amour, pas de femme enceinte sans douleur.*

Главная идея этих пословиц – *как нет субботы без солнца, так нет* (а дальше идут варианты *девушки без любви; женщины или вдовы без совета или без скорби; беременной женщины или старика без боли*). То есть, во французском языке декларируется, что субботы без солнца не бывает. В словарях находим объяснение этому феномену – в субботу потому бывает солнце (согласно легенде), чтобы успела высохнуть воскресная рубашка маленького Иисуса[10]. Но легенда не может породить такое количество паремий. К слову сказать, во французском языке не так много пословиц, посвященных солнцу (около 100) и тот факт, что четверть из них связывает слово *солнце* и *субботу* заставляет задуматься. Ведь предки французов, наверное, тоже имели шестидневную неделю и тоже в последний день поклонялись Солнцу, чем могло быть вызвано такое количество пословиц..

---

\* В украинских паремиях употребляется чаще выражение *не можна*.

\*\* В русских паремиях употребляется выражение *не прясть*.

Обращает на себя внимание тот факт, что во французской коллекции очень много паремий содержат слова *femme*(22)<sup>\*\*\*</sup>, *fille*-девушка(12), *veuve*-вдова(4). Напротив, в славянских языках, таких слов намного меньше. В украинской коллекции: *дівчина*(6), *баба*(2), *вдова*(1); в русской - *баба*(4), *женщина*(2), *девушка*(1). Зато слово *робота* превалирует в украинской коллекции с абсолютной частотой (учетом всех словоформ) – 27, в русской и французской – с абсолютной частотой – 15. Имея такие словари, легко делать сравнения типа, какие числительные, названия животных и растений, предметов быта и т.п. наиболее используются в том или ином языке. У французов более радостные ассоциации, связанные с днями недели, нежели у русских и украинцев.

Кроме всего прочего, можно сделать интересные выводы относительно лексики и грамматического строя паремий. В экспериментальных исследованиях[11], посвященных частоте словоформы французского глагола, что базировались на опросе носителей французского языка, были определены наиболее употребляемые современными французами грамматические категории времени. Это – *présent de l'indicatif*, *passé composé*, *futur simple*. Эта же тенденция прослеживается и при анализе французских паремий, а ведь между современным языком и языком пословиц дистанция в не одну сотню лет. Что же касается 12 наиболее часто употребляемых глаголов в современном французском языке, выделенных из словаря Ж.Бодо (*être, avoir, faire, pouvoir, devoir, aller, savoir, falloir, vouloir, prendre, donner, venir*)[12], то в коллекции из 177 пословиц французского языка (2018 слов), посвященных дням недели, употребляется 11 из них. Приведем частоту употребления этих глаголов: *être*(21), *avoir*(37), *faire*(14), *pouvoir*(2), *devoir*(0), *aller*(7), *savoir*(1), *falloir*(7), *vouloir*(3), *prendre*(1), *donner*(2), *venir*(3). Это говорит о том, что как сами глаголы, так и грамматические категории времени, во французском языке наименее подвержены изменениям во времени.

**Заключение.** Мы описали всего лишь несколько возможных применений программы TextAnalyzer. Другими применениями, где можно получить хорошие результаты – это использование этой программы в помощь разработчикам многоязычных

---

<sup>\*\*\*</sup> Здесь и далее указывается абсолютная частота слова.

узкопрофессиональных словарей (нужно построить частотные словари больших массивов текстов на заданную тематику на разных языках и провести соответствие в частотных словарях). Эта программа также может оказать помощь изучающим иностранные языки. Построив частотный словарь иностранного текста, можно сразу соотнести словарный запас автора со своими возможностями.

Первые опыты с этой программой выявили и ее слабые стороны. Так, хорошо бы было кроме словаря словоформ иметь еще и словарь основ для того, чтобы реально оценить лексическое богатство того или иного языка вообще и конкретного автора, в частности. В настоящее время готовится вторая версия этой программы с развернутой статистикой и ее графическим представлением.

Тем не менее даже те небольшие исследования, которые проводились с помощью описываемой программы говорят о том, что исследование многоязычных текстов может дать неожиданные результаты. Ведь не даром всё больше учёных приходит к тому выводу, что при отсутствии археологических, антропологических, документальных и других доказательств некоторые ответы на вопросы можно найти в самом языке, а сознательно культивируемое многоязычие может предоставить возможность языкам найти свое место в эпоху глобализации.