## Каталог лингвистических ресурсов

### Программы анализа и лингвистической обработки текстов

1. [Свежий взгляд/Fresh Eye](http://www.kirsanov.com/fresheye/) (версия 1.21, 1995) (Дмитрий Кирсанов)

<http://www.kirsanov.com/fresheye/>

DOS-утилита, реализующая стилистическую проверку русскоязычных текстов. Программа отыскивает в тексте места, где фонетически и морфологически схожие слова расположены в непосредственной близости, что порождает так называемую *паронимию* или "*нечаянную тавтологию*". Программа распространяется без каких-либо ограничений вместе с исходным текстом на C. Есть версия для OS/2.

2) [Link Grammar Parser](http://bobo.link.cs.cmu.edu/link/)

Link Grammar Parser – это синтаксический парсер английского языка. Работает со словарем, включающем около 60000 словарных форм. Реализован на C для Unix. Есть также версия для [Windows API32](http://bobo.link.cs.cmu.edu/link/ftp-site/link-grammar/system-3.0/). Имеет консольный интерфейс.

Исходные предложения для разбора могут вводиться вручную с клавиатуры или задаваться в ASCII-файле для пакетной обработки. Программа распространяется бесплатно.

3) [Russian Morphological Dictionary](http://www.geocities.com/SiliconValley/Bit/1116/)

Программа для синтаксического и морфологического анализа русскоязычных текстов. Работает с входным ASCII-текстом. Используется морфологический словарь, включающий 120000 слов. Реализована на SWI-Prolog для Windows.

Программа распространяется бесплатно.

4) [Лингвоанализатор](http://www.rusf.ru/books/analysis/index.htm)

On-line версия программы математического анализа структуры текста. Целью анализа является определение близости любого из предлагаемых пользователем текстов к одному из авторских эталонов, определенных заранее. (*Авторский эталон* - это набор текстов данного автора, взятый из [ресурсов Русской Фантастики](http://www.rusf.ru/books/analysis/list.htm)). Программа анализирует входной текст и выдает имена трех писателей, которые могли бы быть его наиболее вероятными авторами. Кроме этого, программа находит три произведения каждого из авторов, которые наиболее близки данному тексту.

5) [Galaktika-ZOOM](http://zoom.galaktika.ru/)

Автоматизированная система поиска и аналитической обработки информации. Мощный инструмент анализа и обработки текста (Text Mining), позволяющий извлекать необходимые сведения из огромного объема данных. Это коммерческая система, имеющая клиентов в рекламе, органах управления и средствах массовой информации.

### Психолингвистические программы

6) [ПСИ-Офис версия 2.1](http://psy-two.narod.ru/)

Система психолингвистического анализа текстов, включающая 3 компоненты:

- программа поиска вложенных слов в тексте, т.е. слов, "спрятанных" внутри и на переходах между словами.

- программа поиска повторяющихся фрагментов текста при анализе "автоматического письма" (такие тексты пишутся с целью анализа текущих подсознательных процессов).

- программа синтеза подсознательного компонента текста.

Поддержка русского (словарь 690 тыс.слов) и английского (словарь 430 тыс.слов) языков. Возможность подключения дополнительных словарей.

7) [ВААЛ-2000](http://vaal2000.newmail.ru/vaal2000.htm)

Психолингвистическая экспертная система лексического и контент-анализа текстов. Основные задачи - прогноз эффекта неосознаваемого воздействия текста на массовую аудиторию, анализ текстов с точки зрения такого воздействия, генерация текста с заданным вектором воздействия, выявление личностно-психологических качеств автора текста. Система реализована в виде набора DLL-библиотек, которые подключаются к текстовому процессору Word для Windows. Успешно используется в ГосДуме, правительстве, крупных банках и PR-компаниях. Для бесплатной загрузки доступна старая DOS-версия системы образца 1995 года.

### Программы преобразования текстов

8) [Словогрыз](http://www.softwaremaniacs.org/TR/)

Словогрыз - это бесплатная утилита для Windows, которая предназначена для поиска и замены текста по заданным маскам. Программа умеет находить и заменять в тексте не только строго заданные слова и фразы, но и текстовые элементы, определяемые произвольными поисковыми шаблонами. Позволяет сохранять/загружать сценарии преобразований.

9) [ClearText Reader's Edition](http://www.gribuser.ru/freeware/cleartxt/index.shtml)

Freeware программа для переформатирования текстовых или html-файлов (Windows). Позволяет осуществлять как вычистку разметки, так и ее добавление. Понятный графический интерфейс, автораспознавание кодировки.

### Системы обработки естественного языка (NL-Processing)

10) [Natural Language Projects at ISI](http://www.isi.edu/natural-language/nlp-at-isi.html)

Описание группы проектов по вычислительной лингвистике и машинному переводу, выполняемых в Университете Южной Калифорнии.

[Webclopedia](http://www.isi.edu/natural-language/projects/webclopedia) - проект, связанный с созданием информационных агентов, способных генерировать ответы пользователю на основе анализа разнородных коллекций ресурсов, доступных в сети.

[ONTOSAURUS](http://mozart.isi.edu:8003/sensus/sensus_frame.html) - это онтологический тезаурус, включающий 90,000 концептуальных сущностей (словарная база взята из проекта [WordNet](http://www.cogsci.princeton.edu/~wn/)).

[SUMMARIST](http://www.isi.edu/natural-language/projects/SUMMARIST.html) - мультиязычная система реферирования текстовой информации.

[ReWrite](http://www.isi.edu/natural-language/projects/rewrite/) - система двуязычного перевода, основанная на статистическом анализе пар текстовых корпусов.

11) [Каталог программ по вычислительной лингвистике](http://www.sil.org/computing/)

Обширный каталог программ по вычислительной лингвистике, разработанных в рамках проекта SIL. Представленные в каталоге программы носят явно академический характер и требуют значительных усилий для изучения и применения. Не рекомендуются неискушенным в академической лингвистике пользователям. Все программы, а их более 60, являются бесплатными и доступны для online загрузки. Есть исходные коды.

12) [Каталог программ в Corpus Linguistics](http://www.ruf.rice.edu/~barlow/corpus.html#Software)

Интересная подборка доступного в сети лингвистического программного обеспечения, составленная уважаемым профессором Майклом Барлоу из Райского Университета в Техасе.

13) [TELRI Research Archive of Computational Tools and Resources (TRACTOR)](http://www.tractor.de/)

Архив лингвистических программ и ресурсов в рамках проекта TELRI. TELRI - это ассоциация исследователей, целью которой является выявление и продвижение лингвистических ресурсов и программ, связанных с обработкой текстовых и речевых данных для языков центральной и западной Европы.

14) [Словарь устаревших и диалектных слов](http://www.telegraf.ru/misc/day/dis.htm)

Небольшой словарик устаревших слов русского языка на сервере Центрального телеграфа.

15)[The Wordsmyth Educational Dictionary-Thesaurus (WEDT)](http://www.wordsmyth.net/)

Web-сайт, предлагающий on-line версию образовательного словаря-тезауруса английского языка. Содержит около 50 тысяч коренных слов. Удобные средства поиска. Пользователь может получить расклад интересующего слова по слогам, его произношение, список всех словоформ и толкование.

16) [British National Corpus (BNC)](http://info.ox.ac.uk/bnc)

Коллекция более чем 100 миллионов слов современного английского письменного и устного языка. BNC включает в себя более 4000 различных текстов, слова которых классифицированы и размечены с использованием стандарта [Text Encoding Initiative](http://info.ox.ac.uk/bnc/what/tei.html)(TEI). В режиме [on-line поиска](http://thetis.bl.uk/lookup.html) пользователь может найти примеры употребления как отдельных слов, так и выражений. Зарегистрированным подписчикам BNC предоставляются услуги специальной поисковой машины [SARA](http://info.ox.ac.uk/bnc/sara/index.html).