

Корпус метафорических и неметафорических употреблений русских глаголов с разметкой глагольных аргументов

Представленный корпус доступен для просмотра в графическом виде (в формате визуализации инструмента BRAT Rapid annotation tool) по ссылке:

http://web-corpora.net/tools/brat/#/YBadryzlova/Diploma/public_html/

Текущая версия корпуса содержит разметку для **10 переходных глаголов**: бомбардировать, очертить, пилить, причесать, распылять, разбавлять, выкроить, взорвать, взвесить, зажигать. Для каждого глагола размечено порядка **100-120 употреблений** с контекстами, с примерно **равным соотношением метафорических и неметафорических употреблений**.

Содержащиеся в корпусе глагольные употребления и их контексты получены методом случайной выборки из **корпуса ruTenTen[2011]** (<https://the.sketchengine.co.uk>).

Каждая строка в размеченных файлах содержит одно употребление рассматриваемого глагола в контексте.

Тэг <Focus> в каждой строке обозначает **ключевой глагол**.

Тэг <MET> или <NONMET> в начале каждой строки обозначает, что глагольное употребление в данном контексте является **метафорическим либо неметафорическим** соответственно.

Во всех метафорических употреблениях тэг <Focus> сопровождается тэгом <Sense>, в примечании к которому содержится **непрямое значение**, наблюдаемое у глагола в данном контексте. Для просмотра примечания наведите курсор на тэг <Sense>.

Аргументная разметка состоит из двух типов тэгов:

- **Тэги, обозначающие семантический класс глагольных аргументов:**
 - 1) <Concr> - конкретная сущность (за исключением человека);
 - 2) <Abstr> - абстрактная сущность;
 - 3) <Pers> - человек.

- **Тэги, обозначающие тип глагольного аргумента, т.е. его семантическую роль:**
 - 4) <Arg0> – субъект (актор) действия, обозначенного глаголом;
 - 5) <Arg1> – объект действия, выраженный прямым дополнением в прямой диатезе предиката;
 - 6) <INSTR> - инструмент;
 - 7) <LOC> - место;
 - 8) <SOURCE> - источник;
 - 9) <MEDIUM> - средство;
 - 10) <BENEF> - бенефициант;
 - 11) <CONT> - контейнер;
 - 12) <WITH> - участник совместного с актором действия;
 - 13) <TEMP> - время;
 - 14) <MANNER> - образ действия;
 - 15) <MEASURE> - мера;
 - 16) <QUANT> - количество;

- 17) <GOAL> - цель;
- 18) <COND> - условие, выполнение которого требуется для совершения действия;
- 19) <CIRCUM> - условия, обстоятельства, в которых совершается действие;
- 20) <CONJ_АСТ> - действие, происходящее одновременно с основным.

Тэги 4-12 являются **парными** – они всегда употребляются с одним из обозначений семантического класса (тэги 1-3). Тэги 13-20 являются **непарными** и употребляются без обозначения семантического класса.

Семантические роли аргументов метафорических употреблений глагола рассматриваются как результат **метафорического переноса** с ролей аргументов его неметафорических употреблений. Таким образом, например, следующие два употребления (неметафорическое и метафорическое) будут иметь одинаковую структуру аргументов:

	<Arg0>	<Focus>	<Arg1>	<GOAL>	<SOURCE>
<NONMET>	Она	выкраивает	пояс	для юбки	из ткани.
<MET>	Они	выкраивают	деньги	на отпуск	из семейного бюджета.

Помимо вышперечисленных тэгов, разметка некоторых глаголов содержит **дополнительные тэги**:

<Set_Expr> - устойчивое сочетание;

<Co-occur> - лексические единицы, употребление которых регулярно наблюдается вблизи рассматриваемого глагола.

Корпус доступен для **скачивания в виде отдельных файлов** по ссылке:

<http://web-corpora.net/~badryzlova/>

Для каждого глагола имеется два файла: *name.txt* и *name.app*. В первом содержится размечаемый текст, во втором – его разметка. Помимо этого, в корпус входит два конфигурационных файла: *annotation.conf* и *visual.conf*. Подробнее о структуре и функциях файлов, размеченных при помощи инструмента BRAT rapid annotation tool см. :

<http://brat.nlplab.org/standoff.html> - brat standoff format;

<http://brat.nlplab.org/configuration.html> - configuration files.