

Методичка по пополнению данных RussNet для инкорпорирования в YARN

Для процедуры корректировки потребуются следующие ресурсы:

- подготовленная семантическая группа данных RussNet;
- корпус текстов кафедры математической лингвистики в менеджере Bonito;
- электронная версия толкового словаря MAC <http://feb-web.ru/feb/mas/mas-abc/01.htm>;
- браузер WordNet 2.1;
- Викисловарь <https://ru.wiktionary.org/wiki/> (вспомогательные данные)
- Русский ассоциативный словарь (РАС) <http://thesaurus.ru/dict/dict.php>;
- Национальный корпус русского языка (НКРЯ) <http://www.ruscorpora.ru/>;
- Sketch Engine <http://sketchengine.co.uk> (30 дней можно использовать бесплатно).

Процесс пополнения данных RussNet для инкорпорирования в YARN

Имеющиеся данные RussNet представляют ядро лексико-семантической области, привязанной к некоторой части речи, в сочетании с некоторыми элементами периферии. Они включают распределение по частотам значений частотных лексем из ядра области.

Для пополнения данных необходимо выполнить следующие этапы

- (1) сформировать синсеты ядра (базовых понятий) для имеющихся данных;
- (2) пополнить синсеты имеющимися синонимами из Викисловаря, включая информацию о семантически связанных лексемах (гипонимы, гиперонимы, антонимы);
- (3) скорректировать словарное определение синсетов, приводя его к родовидовой форме;
- (4) найти соотносимые синсеты в WordNet 2.1;
- (5) пополнить область синсетами других частей речи, которые выражают понятия описываемой семантической области.

Отредактированные данные следует заносить в таблицу из файла «Шаблон для данных RussNet». При заполнении таблицы файл следует переименовать в соответствии с названием семантической области.

Ниже приводятся рекомендации для этапов пополнения RussNet.

1. Формирование синсетов для базовых понятий семантической области

(а) В структуре имеющихся данных RussNet следует найти группу значений *ipm*, которые превышают для существительных 80, для глаголов 60, для прилагательных 40, для наречий 50. Если встречаются показатели, значительно превышающие эти цифры, то это указывает на объединение реальных значений в одно. Надо понять, какие значения объединены (см. пример 1 ниже).

Самое частотное слово (ядро группы) дает представление о смысловой неоднородности группы (= подгруппах) (см. пример 2 ниже).

Имеющиеся данные помогают структурировать значения: большая *ipm* частота значения доминанты синсета указывает на более высокое положение в родовидовой иерархии (за исключением абстрактных «вершин» связывания иерархии) (см. пример 1 ниже).

Помимо ядерных значений, которые можно выделить на основании порогового значения, приведенного выше, есть предъядерные значения – содержательные

противопоставления в семантической группе значений. Частоты *ipm* таких значений варьируются в интервале $0,4 \times (F_{\max} \pm 0,2 \times F_{\max})$. Интервалы значений *ipm* для основных частей речи: существительные (48–16), глаголы (36–12), прилагательные (24–8), наречия (30–10). Не следует забывать, что в подсчетах частот обычно есть ошибка экстраполяции $\pm 15\%$, что несколько размывает соответствующие границы интервалов $\approx (55–13,5), (41,5–10), (20,5–7), (34,5–8,5)$. Значения *ipm*, меньшие минимальной границы интервала частот, показывают, что это периферийные значения – «листы» семантической иерархии, пересекающиеся значения из двух и более областей. Слова, которые встречаются с частотой менее 0,1 *ipm*, считаются потенциальными и добавляются в словарь по усмотрению составителя: основанием является естественность слова для языкового сознания и участие в семантических корреляциях (противопоставлению по семантическим признакам). Если слово не встречается в корпусе КМЛ, необходимо дополнительно проверить его по НКРЯ. Если его частота превышает пороговое значение (для стандартного объема НКРЯ в 256 млн словоупотреблений это будет 26 вхождений), то оно считается реальным.

(b) Выделенные базовые значения семантической области необходимо поместить в графу 2 (Синсеты) таблицы с уточнением номера значения по частотному распределению (см. примеры 1 и 2 ниже). Значения слов, имеющие малые доли ($<5\%$) в распределении контекстов по значениям, могут упорядочиваться произвольно (например, в соответствии со словарной нумерацией). Значения, представленные в виде оттенков в МАС, в RussNet перечисляются как самостоятельные при условии их реальной (а не потенциальной) частотности в корпусе КМЛ.

Часть речи следует привести в графе 3 (ЧР) в сокращенном виде
N (существительные или эквивалентные им словосочетания),
V (глаголы или эквивалентные им словосочетания),
A (прилагательные или эквивалентные им словосочетания),
B (наречия или эквивалентные им словосочетания).

Значение *ipm* доминанты поместить в графу 6 (Частота) без указания единицы измерения (*ipm*).

2. Пополнение синсетов семантической области

(a) Полученные синсеты необходимо дополнить используя материалы Викисловаря (в него, как правило, внесены синонимы из МАС) и стилистическими вариантами, которые можно найти в РАС. Также полезную информацию содержит английский синсет WordNet 2.1, состав которого поможет дополнить элементы синсета (важно! не вставлять варианты перевода без рассмотрения семантически связанных синсетов).

В синсет следует также добавить и словосочетания, которые можно увидеть в синонимичных толкованиях, ассоциациях РАС и контекстных употреблениях при сортировке контекстов по правому или левому контексту (Simple Sort в Bonito). Стилистически окрашенные значения следует приводить с надстрочным индексом:

книжн – значение, указывающее на стилистически повышенный статус текста (в словарях может быть отмечено по-разному: *книжн*, *научн*, *поэт* и проч.);

разг – значение, характерное для стилистически сниженных текстов (в словарях обозначено *разг*, *прост*, *вульг* и проч.)

Для слов, которые добавляются в синсет, необходимо уточнить номер значения по частотности. В минимальном случае надо проверить толкование (например, в

Викисловаре), если оно одно, то надо приписать единицу в виде подстрочного индекса: *размышлять*₁. Если значений несколько, то необходимо ввести слово в менеджер Bonito, отсортировать контексты по правому (левому) контексту и оценить долю контекстов, в которых данное слово можно заменить на доминанту. Для таких элементов, как словосочетания обычно не характерна многозначность. Поэтому единица в качестве номера значения подразумевается.

Разделителем для элементов синсета является знак |. Элементы синсета необходимо приводить в алфавитном порядке за исключением доминанты и значений предъядра. Например: думать₃ | размышлять₁ | делать умозаключение^{книж} | мыслить₃ | рассуждать₂ | соображать₃ | строить умозаключение^{книж}

Начало синсета – доминанта и элементы предъядра выступают в качестве идентификаторов-ссылок при оформлении родовых синсетов или других синсетов, семантически связанных с данным. Если удобнее скопировать синсет целиком, то можно и так. Не следует забывать, что синсеты могут дополняться как в рамках работы по интеграции, так и впоследствии при подборе синсетов YARN. Ключом к идентификации синсета на человеческом уровне является доминанта (слово с порядковым номером значения), это необходимо для контроля структуры дополняемых в YARN данных RussNet. Во внутренней структуре YARN идентификатором будет числовое значение (как в других WordNet словарях).

(b) Приведенные в Викисловаре гипонимы, антонимы и проч. необходимо вставлять в таблицу для последующей проверки в графу 2 «Синсет» с указанием в графе 8 «Отношение» стандартного обозначения отношения (см. ниже перечень отношений) и в графе 9 ссылки на данный синсет в виде доминанты или полного синсета.

Некоторые семантические отношения симметричны:

- антонимия;
- конверсия (похожие на антонимы слова, которые позволяют создавать предложения-перифразы с перестановкой субъекта и объекта: *давать–получать; превосходить–уступать*). Достаточно указывать это отношение один раз: от описываемого синсета к тому, которое будет рассмотрен ниже. Обратное отношение будет автоматически дополняться в YARN.

Основная группа семантических отношений несимметричны. Для некоторых из них обратное отношение имеет определенный смысл: это, в первую очередь, родовидовое отношение: от родового к видовому оно обозначается Нурупум, от видового к родовому – Нурегпум. Если приводятся согипонимы от родового термина, то желательно их группировать по «фасетам» – признакам, положенным в основание противопоставления. Например, для родового *человек*₁... один фасет – характеристика возраста *старик, мужчина, юноша, мальчик*; другой – обозначений в плане отношений приязни/ неприязни: *друг, приятель, знакомый, неприятель, враг* и т.п.

Отношение часть-целое имеет содержательную интерпретацию от целого к части (Меропум), реже от части к целому (Holonum).

Некоторые семантические отношения имеют истинное (осмысленное) значение только в одностороннем порядке, а обратное отношение будет справедливым только под отрицанием. В описаниях структуры WordNet такие отношения объединялись под рубрикой Entailment (отношения лексического вывода). Это отношения – тропонимии (Тропонум): разновидности способов выполнения действий, состояний и процессов, например *маршировать–идти, заикаться–говорить* и проч.;

- каузации (Cause) – лексического обозначения действия/бездействия с объектом/субъектом, которое приводит с некоторой вероятностью к определенному состоянию этого объекта/субъекта. Например: *убить–умереть, искать–найти, пропустить–пройти, высушить–сухой* и проч.;
- пресуппозиции (Presupposition) – обозначение необходимой для данного значения предпосылки: *промахнуться–целиться, шорох–трение, лязг–удар*;
- действия-части (Subevent), сходного с меронимией отношения при обозначении признаков: *храпеть–спать, уйти/прийти–вернуться, держать–нести*.

В рамках YARN для таких пар будут установлены обратные отношения, помеченные как формально сгенерированные.

Перечисленные отношения связывают между собой синсеты.

Для связи значений разных частей речи в концепции WordNet словарей используется префикс XPOS. Обычно таким образом передается значение транспозиции – выражение значения, характерного для одной части речи, при помощи другой: *любить–любовь, шуметь–шум–шумный, дерево–деревянный, смелый–смело* и проч. Поскольку во всех этих случаях нарушается принятый тест на взаимную замену слов в контексте, отношение называется довольно сложно XPOS_Near_Synonymy. В интерфейсе WordNet 2.1 используется в обобщенное описание Related_to:

force (v) Sense 1 coerce, hale, squeeze, pressure, force -- (to cause to do through pressure or necessity, by physical, moral or intellectual means : "She forced him to take a job in the city"; "He squeezed her for information")

RELATED TO->(noun) force#5

=> force -- (a powerful effect or influence; "the force of his eloquence easily persuaded them")

RELATED TO->(noun) force#2

=> power, force -- (one possessing or exercising power or influence or authority; "the mysterious presence of an evil power"; "may the force be with you"; "the forces of evil")

Для таких языков, как русский, словообразовательные связи задают множество семантически маркированных отношений, поэтому будет использован специальный префикс DER с уточнением характера связи: Synonym, Hyponym, Antonym, Troponym, Role или Involved. Отношение Role указывает на образование существительных, обозначающих участников действия через отсылку к глаголу (например: *расточник* “Рабочий, который растачивает” по значению *расточить1* “Тех. Расширить, углубить какое-л. отверстие на чем-л., вытачивая”), а Involved – образование глагольных значений через обозначение участника при помощи существительного (например: *рассупонить* – “Развязать супонь у хомута при распряжке лошадей”). Эти два отношения могут быть уточнены в плане характеристики роли актанта: Agent, Subject, Instrument, Location, Object. В первом случае это отношение DER_Role_Agent, во втором – DER_Involved_Object.

Название отношения	Определено на	Обратное отношение	Характеристика обратного отношения
Симметричные отношения			
Antonym	синсет<→синсет	Antonym	true
Converse terms	синсет<→синсет	Converse terms	true
Несимметричные отношения			
Hypernym	синсет→синсет	Hyponym	true
Holonym	синсет→синсет	Meronym	true/false
Отношения лексического вывода			
Troponym	синсет→синсет	Trop_hyponym	false
Cause	синсет→синсет	Is_caused_by	false

Presupposition	синсет→синсет	Is_presupposed	false
Subevent	синсет→синсет	Complex_event	false
Словообразовательные отношения			
XPOS_Near_Synonym	слово-значение<→> слово-значение	Related_to	true
DER_Antonym	слово-значение<→> слово-значение	Related_to	true
DER_Hyponym	слово-значение<→> слово-значение	Related_to	true
DER_Troponym	слово-значение<→> слово-значение	Related_to	true
DER_Role_Agent	слово-значение<→> слово-значение	Related_to	true
DER_Role_Subject	слово-значение<→> слово-значение	Related_to	true
DER_Role_Object	слово-значение<→> слово-значение	Related_to	true
DER_Role_Instrument	слово-значение<→> слово-значение	Related_to	true
DER_Role_Location	слово-значение<→> слово-значение	Related_to	true
DER_Involved_Agent	слово-значение<→> слово-значение	Related_to	true
DER_Involved_Subject	слово-значение<→> слово-значение	Related_to	true
DER_Involved_Object	слово-значение<→> слово-значение	Related_to	true
DER_Involved_Instrument	слово-значение<→> слово-значение	Related_to	true
DER_Involved_Location	слово-значение<→> слово-значение	Related_to	true

Если между парой синсетов можно установить несколько типов семантических отношений, то выбираются такие, которые расположены выше в приведенной таблице. Например, глаголы открыть-закрыть являются антонимами, поскольку позволяют нейтрализовать предыдущее действие, эти глаголы одновременно можно рассматривать как пресуппозицию друг друга: можно закрыть только то, что открыто и наоборот. Из двух вариантов выбираем антонимию.

Аналогично при конкуренции словообразовательных связей и семантических отношений выбираем семантическое отношение.

3. Корректировка определения синсета

Довольно часто словарная дефиниция в МАС и Викисловаре удовлетворяет требованию родовидовой структуры: надо указать родовой термин и видовое отличие. Требуется лишь небольшая корректировка.

(а) В определение МАС регулярно входит информация о синонимах. Например в МАС: *размышлять* “Предаваться мыслям о чем-л., думать”. Эти синонимы будут внесены в синсет, поэтому их надо исключить из определения.

(б) Разные элементы синсета по-разному представлены в словаре даже в том случае, когда их эквивалентность подчеркнута наличием синонимичных отсылок. Например в МАС ядерный глагол *думать* определяется так: “о ком-чем, над чем и без доп. Размышлять, предаваться раздумью” (ср. с определением для *размышлять* выше). Из

нескольких определений необходимо выбрать то, которое лучше всего передает родовидовую структуру определения. В данном случае основная проблема состоит в том, что глагол описывает «вершину» семантической группы мышления. В таком случае структура английского определения может подсказать «ключ» к передаче того, что может быть родовым термином: “use or exercise the mind or one's power of reason in order to make inferences, decisions, or arrive at a solution or judgments”. Само определение выглядит довольно громоздко, однако характеристика мышления как вида человеческой деятельности (=контролируемых и целенаправленных действий человека) весьма полезна. То есть начало определения “использовать разум для” довольно очевидно. Видовое отличие глагольных значений описывается в синтаксической модели сочетаемости (*о ком-чем, над чем*). Тогда необходимо указать на объектную направленность (в широком смысле: предмет, явление, отношения между ними) процесса, т.е. его результативность (постижение фактов, формулировка мнения и т.п.). Собственно результат мышления четко представлен в самом частотном употреблении глагола *думать* (см. пример 2 ниже) – это приобретение мнения (модальной характеристики фактов), которое основывается на логическом выводе из доступных аргументов. Как видно из пример 2, последовательная логичность не является абсолютно необходимой – возможна догадка (предварительное суждение), однако чисто внешне логичность не отрицается. Тогда обобщенное обозначение объекта как “использовать разум для логичного построения мнения о проблеме” вполне подходит.

Как правило, таких сложностей при формулировке словарных дефиниций не возникает. В Викисловаре определения обычно скорректированы в силу новизны ресурса, поэтому предпочтительны.

(с) Родовидовой характер определения позволяет выявить пропущенные узлы родовидовой иерархии. Ср. *ясень* “высокое дерево сем. маслиновых, с перистыми листьями и плотной древесиной, идущей на различные изделия”, *береза* “лиственное дерево с белой корой”, *клен* “лиственное дерево с широкими резными листьями”. Родовым термином для всех этих согипонимов будет “лиственное дерево”. Видовое отличие – внешний характерный признак, а использование древесины для изделий, вероятнее всего, будет показана в словаре через метонимический перенос – материал, используемый при создании артефактов.

4. Указание соотносимых синсетов в WordNet 2.1

Для того, чтобы сделать «привязку» к общей системе понятий будем использовать версию WordNet 2.1. Почему используем эту версию, а не последнюю, потому что ее можно установить на компьютер и использовать локально. Второй аргумент менее очевиден, но в текущие версии 3.0 и далее было внесено довольно много изменений, которые затрудняют работу с wordnet словарем. В некоторых работах по автоматическому использованию wordnet указывается, что «удобнее» использовать версии 2.1 (?!).

Как видно из приводимых выше примеров в методичке, структура WordNet в отдельных случаях позволяет прийти к более оптимальному структурированию иерархии и формулировке значений. Это не означает, что абсолютно все подструктуры словаря в одинаковой степени могут использоваться без перестройки. Общность некоторых схем значения может указывать на общее происхождение наших языков, но в принципе определенная «неконгруэнтность» некоторых семантических групп, например, частей тела человека, когда в русском *палец* объединяет название пальцев рук и ног, несмотря на их

непохожесть, разные объемы понятий эмоциональных состояний человека, терминов родства и проч.

(а) В случае полного или практически полного совпадения объема лексического значения (как в пример с глаголами мышления) соответствующий синсет с дефиницией необходимо поместить в графу 7 «Английский синсет».

(b) В RussNet могут встретиться конкретные или стилистически окрашенные значения, для которых нет соответствий в WordNet. Например разговорное употребление *шкура* для обозначения кожи человека, которая дает базу для широкого круга сочетаний *быть/побывать/посидеть в нашей шкуре, испытать на собственной шкуре* и проч., которые приводятся в неполном объеме как фразеологизмы, но по частоте использования и обширности синтаксического окружения подходят под регулярное значение.

В таком случае графа 7 остается пустой, а связь с более абстрактным или нейтральным употреблением показана через связь родового термина *кожа*₁ | *кожный покров* “Наружный покров тела человека или животного” *skin, tegument, cutis* -- (*a natural protective body covering and site of the sense of touch; "your skin is the largest organ of your body"*).

(с) Сложной проблемой является способ отображения неполного соответствия. Известный пример *тоска*, одновременное обозначение тяжелого гнетущего чувства (т.е. эмоции в трактовке WordNet) и скуки, которая представлена в словаре как разновидность желания *longing, yearning, hungriness* -- (*prolonged unfulfilled desire or need*), что противоречит стандартным коллокациям типа *безысходная/смертная тоска*. В таком случае придется приводить ближайший по значению родовой синсет, указав перед синсетом тип семантического отношения *Hyponym* [*anger, choler, ire* -- (*a strong emotion; a feeling that is oriented toward some real or supposed grievance*)].

5. Пополнение семантической области синсетами других частей речи

Данные RussNet за редкими исключениями выполнены в рамках значений одной части речи, как правило, мотивирующей другие слова для данной семантической области.

(а) Если представлены данные для группы глаголов, направления расширения – абстрактные существительные, иерархическая структура которых повторяет глагольную структуру: базовую группу составляют однокоренные слова;

- дериваты-прилагательные и наречия;
- необходимо проверить все возможные приставочные и суффиксальные дериваты глаголов, соотносимые по значению.

(b) Если представлены данные по прилагательным, направления расширения:

- абстрактные существительные;
- наречия-дериваты прилагательных;
- приставочные и суффиксальные дериваты прилагательных.

(с) Если представлены данные по существительным, направления расширения:

- дериваты-прилагательные;
- дериваты-глаголы;
- предложно-падежные конструкции в наречном значении;
- приставочные и суффиксальные дериваты существительных.

Пример 1. Слова, обозначающие эмоциональные состояния

В группе значений эмоциональных глаголов, подготовленных в рамках студенческой работы, есть частоты ipm: 130, 101, 79,8, 76,8. Первые два значения являются слишком большими, что подтверждается тем, что в работе базовое значение толкового словаря МАС не отделялось от оттеночных, которые имеют разные семантические типы:

«3. *также с неопр.* Чувствовать склонность, интерес, влечение, тяготение к чему-л. — *Я люблю искусство...*

|| Испытывать удовольствие от созерцания, ощущения чего-л. *Люблю тебя, Петра творенье...*

|| Иметь пристрастие к чему-л., отдавать предпочтение чему-л. *Любить острую пищу*

|| с *придаточным дополнительным.* Быть довольным чем-л., чувствовать удовольствие, удовлетворение от чего-л. [Алмазов] *любил, чтобы кто-нибудь сыграл на гармонии или на гитаре...*»

Эти значения представлены как самостоятельные в Викисловаре (значения 4–7).

Среди 8 значений, приведенных в нем, 3 значения более-менее точно соответствуют значениям, приведенным для глаголов в WordNet 2.1:

1. (43) love -- (have a great affection or liking for; "I love French food"; "She loves her boss and works hard for him")
2. (26) love, enjoy -- (get pleasure from; "I love cooking")
3. (13) love -- (be enamored or in love with; "She loves her husband deeply")
4. roll in the hay, love, make out, make love, sleep with, get laid, have sex, know, do it, be intimate, have intercourse, have it away, have it off, screw, fuck, jazz, eff, hump, lie with, bed, have a go at it, bang, get it on, bonk -- (have sexual intercourse with; "This student sleeps with everyone in her dorm"; "Adam knew Eve"; "Were you ever intimate with this man?").

Очевидно, что 4-ое значение WordNet также присутствует у русского глагола *любить* (например в контексте корпуса: *Любите и предохраняйтесь!* и реакция *спать* в РАС), однако в силу непонятных обстоятельств не приведено ни в МАС, ни в Викисловаре.

Чтобы правильно структурировать систему значений и понять, относятся ли они к ядру или периферии эмоциональных глаголов, необходимо ввести в Бонито запрос на контексты глагола *любить*, отсортировать 2 позиции правого контекста, просмотреть совокупность, выбирая часто повторяющиеся сочетания со словами правого контекста и ряды перечисления, соединенные союзом *и*. Эта совокупность сочетаний покажет базовые парадигмы (противопоставления или градации) в семантической группе. Аналогично можно отсортировать два слова левого контекста и повторить процедуру.

Наиболее частотные слова как раз показывают общую схему, в рамках которой можно представить все остальные глагольные значения. Для глагола *любить* можно привести следующие: *любить и ненавидеть, любить и дружить, любить и ревновать...*

Довольно часто (41 ipm) встречаются конструкции с инфинитивом: *любят вспоминать/ выпить/ выразиться / говорить...* В этих контекстах русское *любить* эквивалентно английскому *like*. Интересно, что значения *like* частично похожи на *love*, но не объединяются в один синсет:

1. (78) wish, care, like -- (prefer or wish to do something; "Do you care to try this dish?"; "Would you like to come along to the movies?")
2. (73) like -- (find enjoyable or agreeable; "I like jogging"; "She likes to read Russian novels")
3. (19) like -- (be fond of; "I like my nephews")
4. (2) like -- (feel about or towards; consider, evaluate, or regard; "How did you like the President's speech last night?")
5. (1) like -- (want to have; "I'd like a beer now!")

В русском языке глаголы *любить* и *нравиться* имеют разные синтаксические схемы валентностей, и поэтому по стратегии МАС не могут объединяться в один синсет, хотя они и используются для синонимической замены в текстах

НРАВИТЬСЯ, -влюсь, -вишься; несов., кому (сов. понравиться).

1. Производить на кого-л. хорошее, приятное впечатление, вызывать расположение к себе. *Мне особенно нравилось выражение ее лица: так оно было просто и кротно.* Тургенев, Свидание. *Егорка взглянул на деда Фишку приветливо. Старик, видимо, нравился ему.* Марков, Строговы. *Городок начал ей даже нравиться, особенно когда пришла зима и завалила его снегом.* Паустовский, Снег. || также без доп. Вызывать интерес, влечение у лиц другого пола. [Сусанна:] *А что, дядя, я могу нравиться?* А. Островский, Красавец-мужчина. *Она просыпалась каждое утро с единственной мыслью: «нравиться!».* --- *Ей каждый день нужно было очаровывать, пленять, сводить с ума.* Чехов, Ариадна.

2. Соответствовать чьему-л. вкусу, быть приятным кому-л. [Марье Ильиничне] *вовсе не нравились ни похвалы старине, ни порицания новейших обычаев.* Пушкин, Арап Петра Великого. *Бабушке нравилось, что я терпеливо отношусь к боли.* М. Горький, В людях. || безл., с неопр. Любить делать что-л., получать удовольствие от чего-л. *Нравилось ей В пестром смешении звуков знакомых Птиц различать, узнавать насекомых.* Н. Некрасов, Саша. *Было видно, что ему нравится шагать налегке в это солнечное утро.* Тендряков, Под лежац камень...

Пример 2. Слова, обозначающие мышление

Самым частотным глаголом в группе слов, обозначающих мышление, является думать. В словарной статье МАС приведено несколько значений, указывающих на неоднородность структуры данной группы (с опорой на синтаксический метод реализации связей).

ДУМАТЬ, -аю, -аешь; несов.

1. о ком-чем, над чем и без доп. Размышлять, предаваться раздумью. *Думать над задачей.* || о ком-чем. Иметь какое-л. мнение. *Дурно думать о ком-л.*

2. Полагать, считать, предполагать *Скажи, братец, долго ли граф пробудет у нас?* [Альбер:] *Думаю, сестра, что это будет зависеть от тебя.* Пушкин, Сцены из рыцарских времен. || (сов. подумать) на кого. Разг. Подозревать, считать виновным в чем-л. *Он нисколько не виноват, а ты на него думаешь,* --- забормотал Алька. Гайдар, Военная тайна.

3. с неопр. Намереваться, собираться делать что-л. [Лариса:] *Вы когда же думаете ехать в деревню?* А. Островский, Беспреданница.

4. о ком-чем. Заботиться, беспокоиться. *Думать только о себе.*

Из набора этих значений в Викисловаре представлены следующие:

1. осуществлять мыслительную деятельность ♦ Вот я теперь думаю, думаю, думаю. Из сетевого дневника юноши, 2003 г. (цитата из Национального корпуса русского языка, см. Список литературы)
2. мысленно сосредоточиваться на чём-либо; обрабатывать информацию в уме, размышлять ♦ Я повернулся на правый бок, стал думать о том, как мне провести день, и незаметно заснул. Зоя Масленникова, «Жизнь отца Александра Меня», 1992 г. (цитата из Национального корпуса русского языка, см. Список литературы)
3. иметь мнение, полагать ♦ Но я думаю, что команды, достойной этих задач, пока нет. «Свободы — сколько возможно. Государства — сколько необходимо», 2 января 2003 г. // «Новая газета» (цитата из Национального корпуса русского языка, см. Список литературы)
4. разг. предполагать, замысливать, намереваться ♦ Поначалу я думала выкупить у Литфонда нашу дачу. Елена Митько, «Донцова переезжает на ПМЖ в Шервудский», 2002.10.10 г. // «Вечерняя Москва» (цитата из Национального корпуса русского языка, см. Список литературы)

По синонимии с глаголом *размышлять* можно отождествить 1 значение МАС и 2 значение Викисловаря. Приведенные синонимы помогают понять способ употребления

глагола (и показывают состав синсета). Однако порядок следования значений не совпадает с частотностью употребления значения в текстах.

1. мыслить
2. осмысливать, обдумывать, размышлять
3. полагать, считать, находить, рассматривать, смотреть, усматривать, видеть
4. предполагать, задумывать, замышлять, намереваться, собираться, сов. вознамериться, намерен, планировать, запланировать

Для сравнения: близкий английский эквивалент think имеет следующую структуру:

1. (277) think, believe, consider, conceive -- (judge or regard; look upon; judge; "I think he is very smart"; "I believe her to be very smart"; "I think that he is her boyfriend"; "The racist conceives such people to be inferior")
2. (190) think, opine, suppose, imagine, reckon, guess -- (expect, believe, or suppose; "I imagine she earned a lot of money with her new novel"; "I thought to find her in a bad state"; "he didn't think to find her in the kitchen"; "I guess she is angry at me for standing her up")
3. (111) think, cogitate, cerebration -- (use or exercise the mind or one's power of reason in order to make inferences, decisions, or arrive at a solution or judgments; "I've been thinking all day and getting nowhere")
4. (8) remember, retrieve, recall, call back, call up, recollect, think -- (recall knowledge from memory; have a recollection; "I can't remember saying any such thing"; "I can't think what her last name was"; "can you remember her phone number?"; "Do you remember that he once loved you?"; "call up memories")
5. (4) think -- (imagine or visualize; "Just think--you could be rich one day!"; "Think what a scene it must have been!")
6. (3) think -- (focus one's attention on a certain state; "Think big"; "think thin")
7. (2) intend, mean, think -- (have in mind as a purpose; "I mean no harm"; "I only meant to help you"; "She didn't think to harm me"; "We thought to return early that night")
8. think -- (decide by pondering, reasoning, or reflecting; "Can you think what to do next?")
9. think -- (ponder; reflect on, or reason about; "Think the matter through"; "Think how hard life in Russia must be these days")
10. think -- (dispose the mind in a certain way; "Do you really think so?")
11. think -- (have or formulate in the mind; "think good thoughts")
12. think -- (be capable of conscious thought; "Man is the only creature that thinks")
13. think -- (bring into a given condition by mental preoccupation; "She thought herself into a state of panic over the final exam")

Очевидно, что первые 2 значения для *think* в WordNet соответствуют 3 значению из МАС, что отражается в повышенной частоте этого значения 136 ipm, которая в два раза больше частоты ядерного значения (60 ipm). Синтаксически эти употребления выглядят идентично, как предикаты мнения, после которых приводится квалифицируемый предикат: *Не думаю, что в Тюмени особенно хорошие гостиницы... Я думаю, что в будущем мы получим доказательства того, что известные нам как "реформаторы" люди – на самом деле подставные фигуры, не ведавшие, что творят.* Первый контекст амбивалентен (то ли предположение из общих соображений, то ли догадка), второй – более точно указывает на предположение, поскольку относит нас в план неопределенного будущего. Необходимость разделения значений поддерживается также соотношением с существительными данной семантической области: *предположение* 1. предварительное суждение о чем-л.; догадка. Надо помнить, что выбор того или иного значения при автоматическом анализе будет неоднозначным (см. пример выше). Однако, временная характеристика квалифицируемого предиката может служить отправной точкой для принятия решения.

Следовательно, первые 2 значения по частотной схеме будут хорошо соотносимы с английскими синсетами.

Обратите внимание, что объединение значений глаголов совершенного и несовершенного вида довольно часто не имеет под собой лексико-статистического

основания. В частности, глагол *размышлять* (специализированное обозначение собственно процесса мышления) в МАС в первую очередь трактуется как видовая пара для глагола *размыслить* 1. Разг. Подумать как следует о чем-л., всесторонне взвесить в мыслях. Это употребление вообще не представлено в корпусе КМЛ, в НКРЯ представлено 48 употреблений (на 265 млн), т.е. частота 0,19 ipm (явная периферия семантической области). В то время как глагол *размышлять* имеет частоту 21 ipm и входит в предъядро слов, содержательно и однозначно передающих значение данной области.

СООБРАЖАТЬ, -аю, -аешь; несов.

1. перех. или *с придаточным дополнительным*. Постигать умом, осознавать окружающее с необходимой остротой; понимать. *После чая Аркадий Иванович устроил игру в фанты, но дети устали, наелись и плохо соображали, что нужно делать.* А. Н. Толстой, *Детство Никиты*. — *От грохота стрельбы у меня так разбаливается голова, что я ничего не соображаю целый день потом.* Степанов, *Порт-Артур*. || *в чем и без доп.* Разг. Быть сведущим в чем-л.; смыслить. *Он, как видно, кое-что соображал в литейном деле.* В. Беляев, *Старая крепость*. — *Ефим Петрович, --- ты, слышал я, мастер насчет техники?* — *Кое-что соображаю.* Бабаевский, *Кавалер Золотой Звезды*.

2. перех. (устар.) или *с придаточным дополнительным*. Обдумывать, взвешивать, размышлять, чтобы найти нужный ответ, решение и т. п. *[Алексей Александрович] сел к столу и --- стал соображать маршрут путешествия.* Л. Толстой, *Анна Каренина*. *Сегодня, очнувшись после многих дней беспамятства, я долго соображал, где я.* Гаршин, *Художники*. || *без доп.* Думать, рассуждать. — *Конечно, скверно, что я про портрет проговорился, — соображал князь про себя.* Достоевский, *Идиот*. *Расходясь по домам, обыватели соображали:* — *Видать, что и взаправду будут перемены.* М. Горький, *Городок Окуров*.

3. Разг. Обладать острой сообразительностью, находчивостью. — *Парень-то соображает, толковый, видно, парень.* Крымов, *Инженер*.

4. (сов. сообразить) перех. Устар. Приводить в нужное соотношение, связь с чем-л.; сопоставлять. *Мне приснился сон ---, в котором до сих пор вижу нечто пророческое, когда соображаю с ним странные обстоятельства моей жизни.* Пушкин, *Капитанская дочка*. *После сильного припадка падучей болезни она обыкновенно некоторое время не могла соображать свои мысли.* Достоевский, *Униженные и оскорбленные*.