

## Перевод на основе базы прецедентных словосочетаний, или переводных фрагментов

Михаил Юрьевич Михеев  
ст.науч.сотр. НИВЦ МГУ,  
[miheev@srcc.msu.su](mailto:miheev@srcc.msu.su)

Изложен проект и даны выборочные иллюстрации проделанной в 2000-2001 г. работы по написанию синтаксических правил для русско-английского и англо-русского автоматического перевода ("Когнитивные технологии"), что было осуществлено силами одного лингвиста и 1,5 программистов. В результате на сегодня имеется в основном пословный перевод (его можно посмотреть в интернете). Затронут также вопрос о желаемых на завтра задачах таких систем – о составлении базы прецедентных параллельных словосочетаний.

В продолжение приблизительно полутора лет, с 2000 и до середины 2001, в сотрудничестве с А.В. Мисюревым (программная часть) мной были написаны лингвистические правила для того, что предполагалось как система англо-русского и русско-английского автоматического перевода ООО «Когнитивные [или вариант: «интеллектуальные»] технологии». Насколько известно, до сих пор система существует только в отладочном режиме и в полном объеме воплотить изначально заложенные в ней идеи не удалось. Но результаты того, что было реализовано, можно видеть в <http://cs.isa.ac.ru:10000>. Фактически это снова только «пословный» перевод (с элементами «сочетательного»), а хотелось замахнуться на что-то вроде перевода при помощи речевых шаблонов (больших чем слово). Надеюсь, тем не менее, что какому-нибудь будущему инженеру излагаемая здесь стратегия построения перевода может быть полезна. В начале были взяты Русская грамматика и какой-то учебник английского языка для вузов – и переписаны с учетом различий русского и английского синтаксиса, в списки правил (см. I и II). Правила, иллюстрации которых приводятся ниже, состоят из «объектов», или «образцов», на которых проверяются определенные свойства; приводимые примеры выделены курсивом, а после знака // следуют примечания и комментарии, иногда в жанре диалога.

### I. Правила для рус. синтаксического анализа

[Оглавление (жирным выделено иллюстрируемое ниже в тексте):]

**1. Подлежащее и Сказуемое** 2. **Согласованная группа** 2.3. Группа с Числительным 2.4. Причастный оборот 2.4.1. Обособленный Причастный\_об 2.4.2. Неособобленный Причастный\_об 2.4.3. Причастный оборот с Вкраплениями 2.4.4. Границы группы Причастия 2.4.5. Группа Причастия самостоятельного 3. **Группа Существительного** с Родит. падежом 4. Предложные группы и Глаголы с Предлогами 4.1. Сбор Предложной группы 4.2. Поиск "хозяина" для Предложной группы 5. Прямое дополнение (...) 8.

Сложные Предложения. Определение границ Фразы и Предложений внутри нее 8.0.  
Разграничение между Однородными членами и Сложным предложением 8.1. Деление простых предложений (...)

### 1. Сбор группы (Подлежащее + Сказуемое)

о) прототипическая (наиболее частая) форма: Сущ + Глагол\_личн

Уточнения:

а) Сущ: Им.п.; Гл\_личн: 3\_лицо; условия: (согласование\_число)

Сущ: Им.п.; Гл\_личн: 3\_лицо; Прош\_вр; условия: (согласование\_род\_число)

б) здесь вместо первого (Сущ) могут быть: Личн\_местоим Числит\_порядк  
Числит\_собирает Местоимен\_прилаг Имя\_собственное (все остальные условия те же, что и для а.)

Личн\_местоимен: *я ты он она они оно мы вы они*

Числит\_порядк: *перый(ая-ое-ые-ого...) второй(ая-ое-ые)... сто первый(...)*

Числит\_собирает: *двое трое четверо пятеро шестеро семеро восьмеро девятеро десятеро*

Местоимен\_прилаг: *это(-т-а-и) то(-т-а-е) ... сам(-а-о-и)...*

в) здесь на месте второго (Гл\_личн) могут быть: Прил\_полн Прил\_кратк Причастие.

Условия для этого: (согласование\_род\_число\_падеж); (порядок\_жесткий);

остальные условия те же, что для а. и/или для б.

вв) к условиям в. может добавляться и такие ограничения на второй компонент:

(падеж\_творительный) и (наличие\_глагола\_связки)

список возможных глаголов\_связок: *быть являться оказываться выдаваться*

*получаться...* [не закончен]. При этом Прилагательное может быть как в Им.п., так и в

Твор.п. и тоже согласовано с Сущ. по Роду и Числу, а также связано Глаголом\_вставкой

"был /-а /-о /-и; выдался / оказался / получился / является..."; причем Глагол-вставка в 3-м

Лице и согласован с Сущ. по Роду и Числу: // "Погода была / оказалась теплая"; "День

*выдался солнечный"*

г) ограничения на второй компонент: вместо Глагол\_личн может быть Соглас\_группа в

Род.п., состоящая из Параметрич\_прил и Сущ\_параметра, где последнее – одно из слов

списка: // *Иван (был) высокого роста / он не робкого десятка / это – вопрос второго*

*порядка / Она – женщина хрупкого телосложения*

список Сущ\_параметров: *"рост, вес, возраст, характер, нрав, глубина, высота, объем,*

*цена, телосложение (сложение), толщина, уровень;... десяток, порядок,..."* [список не

полон], гг) вместе с г. возможно также (наличие глагола\_связки) "быть оказываться"

д) ограничение на второй компонент: вместо Глагол\_личн может быть Сущ: Имен.п. и

выполняться свойство: (согласование\_число\_падеж): // "Ее мать врач"; "Она – живодер",

дд) вместе с условиями пункта д. возможно также (наличие глагола\_связки) "быть"

е) ограничения на второй компонент: вместо Глагол\_личн может быть Сущ: Твор.п. и

Глагол\_связка "служить являться выступать работать..." [список неполон] и должно

выполняться свойство: (согласование\_число): // *Он служит средством.../ Я работаю*

*волшебником / Он высок ростом / Она ростом с телеграфный столб*

ее) вообще если не находится полноценного Глагола на роль второго компонента

(Сказуемого), то берем (пытаемся) всё, что есть в Предложении, например,

Предложную\_группу // "Чай – с сахаром"

ж) первый компонент может быть не в Имен.п., а в Дат\_пад., а второй тогда –

Глагол\_личн. в 3\_лице Ед\_числа (а в Прош.времени еще и Ср\_рода) или Существительное

или же Предикат из специального списка // "ребенку (мне) жарко (плохо, охота)"

- з) оба компонента могут быть Инфинитивом (Неопред\_форм\_глагол) или (оба же) Глагол: 2\_лицо; ед\_число со стоящим между ними Тире // *"Поспешишь – делу навредишь"*
- зз) то есть когда и само Полежащее оказывается вырожденным (например, Инфинитивом), то на роль Сказуемого берем всё, что подвернется под руку: Сущ // *"Дозвониться – проблема"* Прилагат\_краткое (возможно с Глагол\_связкой) или Сравнит\_степень\_прилаг: // *"Промолчать – (было) разумно (разумнее)"* Наречие или вообще все что угодно (желательно с тире перед Сказуемым): // *"Сделать – быстро"*
- и) вырожденные формы предложения: односоставные: состоящие из одного Глагола: Глагол\_личн: 2\_лицо; Ед\_число // *"Работай по-путински"; "Не стой под стрелой"*, и Мн\_число // *"Поедете в мягком вагоне"; "Вытирайте ноги"*; Глагол\_личн: 3\_лицо (обычно Сред\_род в Прош\_врем): // *"Уезжают домой"; "Светает"; Вечерееет"; "Меня знобит"; "Потеплело"*; Глагол\_личн: 1\_лицо // *"Просыпаюсь"*
- ии) Инфинитив (когда он является вершиной и к нему нет Подлежащего): // *"Не курить"; "Не сорить"; "Не трогать руками"*
- к) вырожденные формы предложения: односоставные: состоящие из одного Имени (т.е. Существительного Прилагательного или Наречия) – когда уже Сказуемое отсутствует: // *"Ночь, улица, фонарь, аптека."; "Осень"; "Тепло"; "Желтое"*;
- л) когда Сказуемое – это Глагол с Отрицанием или Предикат "нет" ("нету"), то при нем Подлежащее (любое из пунктов а. б.) может быть в Род.п. (это своего рода трансформация Отрицания для самого русского языка): // *(Там (никого (не было)))*; *(отца (нет дома))*; *(сестры (я (не увидел)))*).

## 2. Согласованные группы

условия: если два Слова, одно из которых – это условное "Прилагательное" (подробнее см. перечисление ниже пунктов от а) до гг), а другое – условное "Существительное" (подробнее см. х-у), согласуются в Роде, Числе и Падеже, // [примеры: *"пионерский галстук"*, *"моя милиция"*, *"первый сентябрь"*, в отличие от *"первое сентября"*, последнее из которых должно собираться уже не как Согласованная группа, а как Обстоятельственный оборот, т.е. как специализированная Временная\_группа] **то вывод:** из этих двух Слов образуется единый узел с атрибутами (Род, Число, Падеж), взятыми у Существительного (или же: с максимально уточненными формами: то есть в том случае, когда они оказываются более точными у Прилагательного, их надо заимствовать, наоборот, у него: например. в сочетании "наш староста" то, что его Род = муж ("наш староста ответил"), определяется именно по Прилагательному, потому что у данного Существительного Род = Общий), и данный узел называется с этих пор Согласов\_Группа (или Группа\_соглас), // а соответствующее Прилагательное в ней (или Местоим. или Причаст. или Числит. – см. ниже) получает атрибут Согласов\_прил (Согласов\_местоим. Согласов\_числит. Согласов\_причаст.)

### область применения (более подробно):

- а) Прилагательное полное или краткое + Существительное (в любом порядке, включая и инвертированный  
// т.е. *"фиалка обыкновенная"* сюда тоже подойдет
- аа) Прилагательное Краткое (или, правильнее, Притяжательное\_Прил., когда оно будет в списке, а пока такого нет) + Существительное (*"бабушкино, платье; ленино наследство; декартовы координаты"* итп.);
- б) Местоимение притяжательное или вопросительное или относительное или неопределенное или отрицательное или определительное или указательное + Существительное // например:

"мои родственники; эти люди; ничья собственность" итп., – причем также как в прямом порядке слов, так и в обратном – "родители мои и твои"

**в)** Причастие + Существительное // причем в обе стороны: слева от Сущ. стоят, как правило, только одиночные Причастия "оторванные пуговицы", а справа – целый Причастный оборот: "пуговицы, оторванные в давке", т.е. группа, зависящая от заглавного слова Существительного. Но впрочем, и слева нужно разрешить целый причастный оборот – "красивое, останавливающее на себе взгляд лицо"); подробнее об этом см. в 1.2. Причастный оборот.

**г)** Порядковое Числительное + Сущ. // "первый космонавт" или в обратном порядке "номер первый"; **гг)** Количественное Числительное (180 или Числит\_прилагат.)

"один/одна/одно/одни" + Сущ. // "одного чудака" или "чудаку одному"

**х)** Вместо второго компонента в этой группе (Сущ.) могут быть Субстантивированные Прилагательные ("рабочий, безработный, прачечная, выходной, отпускные" итп.): их надо задать списком, если в словаре они только в форме Прилаг.: по большей части они просто и являются Существительными; или

// Тут правило может быть в будущем, например, такое: если слова опознаны как Неоднород\_прилаг. ("московский безработный"), но не опознаны (не включены в состав) никакой из Согласов\_групп, т.е. не найдено среди них слово-"хозяина"), то надо попробовать (последнее из них справа) уже в роли Субстантивир\_прилаг.

**у)** на втором месте в этом правиле могут стоять также Личные\_Местоимения, Относительные, Неопределенные и Отрицательные Местоимения ("обескураженный я; этот некто; полнейшее ничто")

## II. Правила перевода русских слов и конструкций в английскую фразу

**0. идеология, или общие слова** 1. Наследование Частей речи 2. Изменение порядка слов в англ. простом Предложении 3. Вставка Глагола-связки 3.1. Однородные сказуемые 4. Согласование Глагола (Сказуемого) с Подлежащим 5. Страдательный залог 6. Образование и употребление (английских) временных форм 7. Артикль 8. Вопрос 9. Отрицание 10. Группа Приблизительного количества (...)

### 0. идеология, или общие слова:

После нормализации по морфологическому словарю и пословного перевода каждого из нормализованных слов мы получаем переводы этих слов, а после перевода еще и словосочетаний по словарям имеем в результате: 1. набор отдельных (вообще не опознанных морфологически) слов, (о них можно строить гипотезы, как их опознать – и морфологически, и, далее, синтаксически), 2. набор отдельных не переведенных (т.е. хоть и опознанных морфологически, но не имеющих пока перевода) слов, 3. набор отдельных слов с их (наборами) переводов, 4. набор именных словосочетаний с их (тоже наборами) переводов, 5. набор глагольных сочетаний с их (наборами) переводов и еще (может быть) 6. набор смешанных глагольно-именных сочетаний с их переводами, а также 7. набор знаков препинания внутри предложения.

Затем нам бы надо было разбить слова из классов (3-6) на Члены предложения, прежде чем вставлять их в английские фразы. Это значит соотнести полученные исключительно словарными методами, по-"белоноговски" (<http://www.viniti.ru/russian/vintrans.html>) сочетания с теми, которые получатся в результате синтаксического анализа. Согласно моему разбору простейшие Именные сочетания (Соглас\_гр); (Сущ1\_упр\_Сущ2.Род); (Предл\_гр) итп. (более мелкие и менее существенные – Гр\_числит, Причаст\_об(орот), Дееприч\_об, Гр\_приложения) могут претендовать на следующие места в дереве

Предложения: Подлежащ(его); Прям\_доп(олнения); Косв\_доп; Обстоят(ельства) или же Сказуем(ого) и самого Предложения (последние два варианта возможны при определенных, дефектных, условиях, которые были оговорены в правилах синтаксиса) в зависимости от той роли, которую они играют в отношениях друг с другом или с Глаголом-Сказуемым данного предложения.

Простейшие глагольные сочетания (Глагол + Предлог); (Глагол + Союз); (Глагол + Падеж(Сущ)); (Глагол + Нареч); (Глагол + Глагол) могут претендовать на следующие роли (или не иметь роли вообще, как нетерминальный узел Глагол + Предлог, например): Сказуем(ого), Прост(ого)\_предл(ожения), или Придаточ\_предл. В результате синтаксического разбора формируются соответствующие гипотезы о покрытии данного Предложения – Членами\_предложения. Возьмем фразу: *Мама мыла раму.* – Пословно по словарю она будет распознана как-нибудь, например, так: [*ma; mama; mother*] [*soup; wash*] [*frame*].

Естественно, что “белоноговский”, т.е. словарно-сочетательный поиск, не даст нам тут ничего (если не найдет какого-нибудь, предположим, “*мамино* *мыла*” или (особого сорта) “*мыла для рамы*”), и вот тут-то, на втором этапе, нам и окажется нужен синтаксический анализ, чтобы понять, что, например, возможная гипотеза принять слова “*мама мыла*” за группу (Сущ1 + Сущ2.Род.п) безусловно проигрывает перед гипотезой считать их вместе – группой (Подлеж + Сказуем). Это становится ясно, конечно, не сразу, а потому, что в первом случае третье в предложении слово (“*раму*”) повисает, будучи незавязанным с другими, а во-втором случае оно законно встраивается в структуру (и соответственно в дерево) – на основании шаблона (Глагол + Сущ.Вин.п): “*мыла раму*”, или (Сказуем + Прям\_доп) и весь шаблон Предложения становится более насыщенным. Ведь когда все распознанные морфологически слова в Предложении находят себе то или иное место в дереве, это безусловно лучше, чем когда остаются ни с чем не связанные слова-островки. (А кроме такого правила можно было бы применять и что-нибудь вроде следующего: все “дефектные” шаблоны построения Предложения надо было бы выстроить по вероятности и эту вероятность как-то соотносить с вероятностью отнесения Неопознанного слова к той или иной Части речи.)

Главным при оценке гипотез о реальности тех или иных возможных (априорно-синтаксических) сочетаний в данной фразе является найти вершину дерева, в наиболее простом варианте – пару Главных членов, или Подлежащее и Сказуемое, а все остальные слова увязать в те или иные Второстепенные члены, подчинив кому-то из этой пары. Например, во фразе: *Мама мыла раму после дождичка в четверг.* – Если не найдется на первом этапе, при поиске по словарю, вариант (контактного) сочетания “*рама после дождичка*” и “*дождичек в четверг*” (а такого в словаре, скорее всего, и не должно быть), то Предложные\_группы “*после дождичка*” и “*в четверг*” будут привязаны (уже на втором этапе) к узлу Сказуемое (т.е. к слову “*мыть*”) как его Обстоятельства – на основании моих синтаксических правил (“*мыла после дождичка*”; “*мыла в Четверг*”), поскольку Предлож\_группа, если не связана с хозяином-Существительным, всегда относится к Глаголу и считается Обстоятельством при Сказуемом.

Теперь пойдем еще далее: *Грустная мама мыла раму тряпкой, а папа весело пил из чайника денатурат.* – Здесь “*грустная мама*” свяжется в Согласов\_группу; “*весело пил*” – в группу (Нареч + Глагол), а если не будет найдено по словарю сочетания “*рама тряпкой*” (наверно и не будет), то естественно привязать Сущ.Твор.п “*тряпкой*” к Глаголу “*мыть*”, – но опять вначале надо будет предварительно проверить, нет ли такого сочетания (“*мыть тряпкой*”) уже в словаре (что вполне вероятно); если даже и нет, то его просто надо будет породить с помощью синтаксических правил, ведь это вполне законная группа (Сказуем +

Сущ.Твор.п), для перевода которой с трансформациями (Прош\_время; Творительный падеж дополнения Неодуш\_сущ; Артикль) получаем: wash<ed> + <with> <the> rag. Во всяком случае словарно-закрепленные соответствия должны иметь приоритет перед синтаксическим порождением сочетания. “Пил из чайника” – опять просто: если нет в словаре в целостном виде, то собираем по частям, согласно уже имеющемуся синтаксическому шаблону (Глагол + Предл\_гр) – а до этого конечно же проверяем, нет ли такой Предложной группы (“из чайника”) в словаре: если нет, то собираем ее: from <the> teapot. (Кстати, коли она отсутствует в словаре, но не запрещена синтаксически, то тут и можно было бы применить статистику – определяя, хотя бы, то, с каким из англ. вариантов Предлога “из” более сочетается “чайник” – [from]; [of]; или [out of].) То же самое проделываем и с группой “пил... денатурат” – если нет в словаре словосочетания *пить денатурат*, то лепим по шаблону (Сказуем + Прям\_доп): drunk + <the> + methylated spirits. Итд. итп. То есть здесь я нарушаю твой принцип, что анализируются всегда только контактные группы (после “проглатывания” стоящих между ними слов – внутри одной из них), но, думаю, что и при проработке с твоим подходом результат должен быть тот же самый.

### 1. Наследование Частями речи и синтаксическими группами слов – Членов Предложения

Те Члены\_предложения, которые установлены в рус. Предложении в результате его синтаксического разбора, автоматически переносятся на те Слова (Группы\_слов), которые им соответствуют в результате пословного перевода (или перевода через конструкции-словосочетания – пытаться же устанавливать соответствия, ища переводы, выходя за установленные в ходе синтаксического разбора группы, если он однозначен для данного Предложения, не следует). // Конечно, по ходу дела будут накапливаться и исключения из этого, но тогда и будем приписывать их к этому правилу.

### 2. Изменение порядка слов в Простом предложении

Какой бы ни был Порядок слов в русском Простом предложении, в переводе на английский он меняется на следующий:

1) Подлежащее, 2) Сказуемое, 3) Дополнение 3а) Косвенное дополнение без предлога (me, us), 3б) Прямое дополнение, 3в) Косвенное дополнение с предлогом (to Peter), 4) Обстоятельство (every day). (Отделяемый\_предлог здесь не учитывается, для него отдельные правила.)

за исключением следующих случаев (но можно и вообще без нижеперечисленных исключений):

2.1. // [В одну неделю] все дело оборудовали = [in a week] it was all settled.

**условие:** если в русском Предложении на первом месте стоит Обстоятельство или Группа\_обст., то

**вывод:** тогда и в англ. переводе оно тоже ставится на 1) место (сохраняется в препозиции) – выше дан пример для Обстоятельства времени.

2.2. // "Обед готовила моя жена; Обед жена готовила; Протокол составляли полчаса";

**условие:** если по-русски подчеркнут объект воздействия – т.е. когда Прям\_доп. стоит на первом месте в Предложении, Подлежащее следует за ним (или вообще Предложение не имеет Подлежащего), а Сказуемое – Глагол в Действ\_залоге, то

**вывод:** чтобы подчеркнуть объект действия в англ., делаем Пр\_доп рус. Предложения – Подлежащим английского и переводим Глагол Сказуемого – в Страд\_заялок, а Подлежащее (если оно есть в рус.) делаем Косв\_доплением (с использованием трансформации пассивизации, см. ниже) // *The diner was cooked by my wife.*

2.3. // "Письмо написано мной; Протокол составляется следователем",

**условие:** или же Сказуемое – это Глагол (или Причастие) в Страдат\_залоге (или Возвр\_глагол на "-ся") и при нем есть Дополнение в Твор\_п, то есть в русском Порядок слов (инверсивный) таков:

Прям\_дополнение А (Сущ, Вин\_пад.) + Сказуем. Б (Глаг., Действит\_залога) + Подлеж. В (Сущ.,Им\_пад),

**то тогда** в англ. он становится таким: (для случаев, когда В – Одуш.):

Подлеж. А + Сказуем. Б (Глаг., Страд\_залога) + by + Дополнение В  
или (для случаев, когда Б – Неодуш.):

Подлеж. А + Сказуем. Б (Глаг\_Страд\_залога) + with + Дополнение В.

### III. [Предполагавшийся далее порядок работы ]

Далее предполагалось, что будет собран словарь так называемых «переводных эквивалентов» или «прецедентных переводов». Например, фраза из одного из английских переводов рассказа А.Платонова «Государственный житель» и соответствующая ему – из оригинала: *The steam locomotive snuffled with its thick, peaceful force and slowly pulled off carloads of public substances: bottles of sulphuric acid, mounds of ropes, institutional luggage, and unmarked bags with something useful.* = *Паровоз, сопя гущей своих мирных сил, медленно осаживал вагоны, полные общественных веществ: бутылей с серной кислотой, бугров веревок, учрежденской клади и необозначенных мешков с чем-то полезным.*

После проведения синтаксического анализа русская и английская фразы членятся на фрагменты и получается (неоднозначное разбиение допускается) приблизительно следующее: (The (steam locomotive)) (snuffled [with] ((its thick, peaceful) force)) {and} (slowly [pulled off] carloads] (of (public substances)): ((bottles [of] (sulphuric acid))), (mounds of ropes), (institutional luggage), (and (unmarked [bags]) (with (something] useful)). = (Паровоз), (сопя [тв.п] [гущей (своих мирных) сил], (медленно [осаживал) вагоны], {полные [род.п.]} (общественных веществ): ((бутылей [с] (серной кислотой))), (бугров веревок), (учрежденской клади) и (необозначенных мешков) (с (чем-то полезным))).

Отсюда и извлекаются параллельные фрагменты (их возможно привести к нормализованной форме):

*Паровоз* = the steam locomotive; *сопя* + тв.п, = snuffled with... and; *гущей своих мирных сил* = its thick, peaceful force; *своих... мирных сил* = its... peaceful force; *мирные силы* = peaceful force; *свои силы* = its... force; *гуща своих... сил* = its thick, peaceful force; *медленно осаживал вагоны* = slowly pulled off carloads; *осаживал вагоны* = pulled off carloads; *медленно осаживал* = slowly pulled off; *общественных веществ* = of public substances; *бутылей с серной кислотой* = bottles of sulphuric acid; bottles of = *бутылей с*; of sulphuric acid = *с серной кислотой...*; *бугров веревок* = mounds of ropes; *учрежденской клади* = institutional luggage; *и необозначенных мешков с чем-то полезным* = and unmarked bags with something useful; *необозначенных мешков* = unmarked bags; with something useful = *с чем-то полезным...* итд.

То же проделывается и с оригинальными английскими текстами, скажем, рассказом Конан Дойля (о Шерлоке Холмсе) «Скандал в Богемии». (Как и прежде,) он (был (глубоко увлечен)) (расследованием преступлений). Он (отдавал (свои (огромные способности))) и ((необычайный дар) наблюдательности) (поискам (нитей к выяснению)) тех тайн, (которые (официальной полицией) ((были признаны) непостижимыми)). = *He (was (still, as ever,) (deeply attracted) (by (the study of crime))), and (occupied ((his (immense faculties)))*

*and (extraordinary (powers of observation)) in ((following out) (those clues)), and ((clearing up) (those mysteries)) (which ((had been abandoned) as hopeless) (by (<the> official police))).*

В результате обработки подобных параллельных текстов должно было быть собрано большое количество квазифразеологических сочетаний, то есть кандидатов в переводные эквиваленты. При этом, естественно, с обеих сторон будут накапливаться большие списки синонимов («паровоз» можно перевести, кроме указанного, в разных контекстах, еще и как *steam-engine, railway engine* итд.). Можно соответственно увеличивать или уменьшать длину сочетания (и контекста), а при выборе переводного эквивалента учитывать статистические критерии. Все вообще одиночные Части речи можно просто выкинуть, если они не вносят ничего нового по сравнению со старым словарем.

### Примеры простейших сочетаний словаря:

1. Соглас\_гр: *вредное пространство; регистрирующее устройство; средняя статистическая поляризуемость; парашютнодесантные войска*, то есть (Прилагат + Сущ) или (Причаст + Сущ) или сколько угодно таковых: (Прилаг + Причаст + Прилаг + ... + Сущ); а также, возможно, присовокупив сюда и (Сущ + Запятая + Причаст), т.е. Обособленный Причастный оборот

2. Сущ\_упр\_род: *крышка люка; люк доступа; доза радиации / доза облучения / дозы излучения; регистрация параметров* или: (Сущ + Сущ.Род.п). Иногда второе Сущ только в Ед.ч., иногда только во Мн.ч., но можно наверно Число вообще не фиксировать: так, если и встретится «люк доступов» или «регистрация параметра» в тексте, то и они скорее всего отождествятся по тому нормализованному варианту, какой у нас будет в словаре (а можно давать и несколько вариантов, если словарь их предлагает)

1-2. Числит\_порядковое + Сущ: *третье поколение; third international conference* (Numeral\_ordinal + Adj + Noun). Здесь же и тип сочетаний (Согл\_гр) и менее частый тип – (Сущ\_упр\_род): *четыре экземпляра договора*, хотя можно этот тип не считать самостоятельным

3. Сущ + Наречие (именно так, когда Наречие с правой стороны): *поворот налево; шаг вперед; соединение внахлестку; бросок переворотом*; 3а. Глагол + Наречие: *двигаться вперед*

4. Предлож\_гр, или (Сущ + Предл): *под напряжением; на диван; на диване*; 4а.

Глагол + Предлог: *наметить на ; зажимает в* (группы нетерминальные); 4б.

Причаст + Предлог: *обнаруженным в; отличающийся от*; 4в. Дееприч + Предлог (то же самое)

5. Однородные\_члены, т.е. (Сущ + Сущ) или (Прил + Прил) или (Глагол + Глагол) или (Предл\_гр + Предл\_гр) итп., где на месте (+) могут быть Знаки\_препинания или Союзы (подробнее они были описаны в правилах выше).

[Варианты глагольных групп:]

Сущ + Глагол (или Подл + Сказуем): *мама мыла*

Отриц + Глагол: *не утроил*

6. Глагол + Сущ\_Вин.п.: *получать доступ; проявлять инициативу*

[Терминальные варианты (трехсловные)]

7. Сущ + Предлож\_гр: *литье под давлением; задержки при регистрации*, 7а. Глагол +

Предлож\_гр: *наметить на четверг; зажимать в тиски*;

[Комбинаторные (парные) варианты (трехсловные) с наложением]

8. Соглас\_гр + Сущ\_упр\_род: *(высокая (стоимость аренды)); сводный проект резолюции*  
// правда, уже и тут возможны расхождения в членении или «кройке тулупа»

словосочетания: поди разбери, то ли это, на самом деле, ((сводный проект) резолюции), то ли: (сводный (проект резолюции)). Вообще-то мы можем взять любой из вариантов, например, остановившись на последнем (то же и во всех последующих случаях – фиксировать наиболее вероятный вариант разбиения). Или же вообще ничего внутри подобных сочетаний не членить, а брать только целое (Соглас\_гр + Сущ.Род.п).

9. Сущ\_упр\_род (Соглас\_гр): *пучок последнего выбора*; // это “обратный” вариант комбинации этих двух групп, и здесь уже расхождений не возникает, потому что Согласованная группа – контактное сочетание, а управляющая Существительным в Род.п. – разорванное. Наверно можно обозначать их сочетание как (Сущ + Согл\_гр.Род.п)?

10. Сущ\_упр\_род (Сущ\_упр\_род): *меню оригиналов псевдонимов; система проверки кредитоспособности* // Тут снова возникает расхождение (если необходимо внутри это сложное целое разбить на простенькие сочетания), но пусть так и будет: то есть мы сознательно отбросим вариант (*меню оригиналов*) или (*система проверки*) – в пользу только (*оригиналов псевдонимов*) и (*проверки кредитоспособности*): они могут, конечно, возникнуть в словаре, но только лишь заново, как самостоятельные сочетания (через другие единицы, если таковые будут попадать в словарь в дальнейшем). Или же другой вариант: мы вообще не проставляем никаких скобочек внутри сочетаний такого типа, а лишь фиксируем, что Заглавным словом всего сочетания (его “Хозяином”) является первое по порядку Сущ (оно и может быть в любом Падеже), а все остальные (Сущ и Прил) – в Родительном. 10а. Сущ\_упр\_род (Числит + Сущ): *метод трех баз* 10б. Сущ\_упр\_твор + (Сущ\_упр\_род): *пайка волной припоя* 10в. (Сущ\_упр\_дат) – *письмо брату*

[Тройные и тд. варианты с наложениями]

11. Сущ\_упр\_род (Сущ + Предл\_гр): (*люк (доступа к аккумулятору)*) // по-видимому, нет смысла накапливать все промежуточные возможные формы, позволяющие “склеивать” это сочетание воедино – как здесь: [доступа к]; <к аккумулятору>; но ограничиться только парой, обозначенной в имени данной группы при помощи плюсики. Иначе говоря, плюс означает, что все “наезжающие” друг на друга варианты просто отброшены.

12. Сущ\_упр\_род (Согл\_гр + Согл\_гр): ((*цепочка данных*) (*неограниченной длины*)) // для них встает возможность «перескока» при согласовании через группу (Сущ\_упр\_род): цепочка... длины, – но мы ее отбрасываем, как как будто не существующую

13. Предл\_гр (Сущ + Предл + Соглас\_гр): (*подробности по (рассматриваемому делу)*)

14. Сущ\_упр\_род (Согл\_гр + Согл\_гр): ((*солнечных батареях*) (*космического назначения*));

15. (Согл\_гр + Предл + Согл\_гр): ((*бортовая опалубка*) для (*дорожного покрытия*)) 15а. Согл\_гр + Предл\_гр (Предл + Сущ\_упр\_род): (*гидравлических жидкостей*) (*на (основе воды)*)

16. Сущ\_упр\_род (Сущ\_упр\_род + Сущ\_упр\_род) + (Предл\_гр): ((*крышка люка доступа*) (*к аккумулятору*))

17. (Сущ\_упр\_род + Предл + Согл\_гр): ((*порядок подготовки*) к (*длительной остановке*))

// или, в другой записи: Сущ + Сущ.Род.п. + Предл\_гр (Согл\_гр)

18. Соглас\_гр (Сущ + Причаст\_об (Причаст + Предл\_гр)): (*усталость, (возникающая на кромках)*) // но также и в обратном порядке: 18а. Соглас\_гр (Причаст\_об (Причаст + Предл\_гр) + Сущ): *возникающая на кромках усталость*, 18б. Соглас\_гр (Сущ + Причаст\_об (Прич + Сущ.Твор.п.): *микрорунок меченных биотином*, 18в. Соглас\_гр (Сущ + Причаст\_об (Причаст + Предл\_гр)): *бульдозер, снимающий слой грунта* или же: (Сущ +

Запятая + Причаст\_об); // но тут могут быть и нагромождения: 18вв. Соглас\_гр (Сущ + Запят + Причаст\_об (Причаст + Предл\_гр (Согл\_гр + Предл + Согл\_гр))): *поведение, дающее (правильную ориентацию) (в (окружающей среде))*

[Комбинаторные варианты глагольных групп: ]

18г. (Отриц + Вспомог\_глагол.Буд\_врем или Прош\_врем) + Крат\_страд\_причаст: *не будут высушены; не был надут; не был вбит* 18д. (Отриц + Модал\_гл.Сослагат\_накл) + Инф: *не мог бы забывать* // Но вообще, наверное, не стоит собирать в словаре сочетания с отрицанием. 18е. ((Отриц + Модал\_гл) + Инф) + (Нареч + Нареч): *не мог тратить чрезмерно много* 18ж. (Модал\_гл + Инф) + Крат\_причаст: *должны быть сбалансированы; должен быть разбавлен* 18з. ((Модал\_гл + Сослагат\_накл) + Инф) + Кратк\_причаст: *должен был бы быть разбавлен* 18и. Глагол + Вин.п. + Твор.п.(Согл\_гр): *решать дело третейским судом* 18к. Кратк\_прич + Союз + Кратк\_прич: *приостановлен или отложен* 18л. (Отриц + Модал\_глагол.Прош\_врем.Мн.ч.) + Крат\_страд\_прич: *не должны были быть взорваны; не должен был быть приписан*

[Комбинаторные варианты глагольно-именных групп:]

19. Глагол + (Соглас\_гр): *совершать одиночный полет* // и наверно все типа (Глагол + (Сущ\_упр\_род)) или (Глагол + (Предл\_гр))

20. Глагол + Согл\_гр(Сущ\_упр\_род): *делать местную выправку пути*

21. Глагол + (Соглас\_гр (Сущ\_упр\_сущ (Сущ + Соглас\_гр))): *продали большую часть своих акций* // Вообще говоря, мы не обязаны идти на поводу у такого многообразия комбинаций сочетаний и все их заносить в словарь. Так, например, мы не должны заносить туда целые фразы, вроде: *больной жалуется на уменьшение продолжительности эрекции* –мы просто разложим это на: (Глагол + Предлог): *жалуется на;* (Сущ1 + Сущ2.Род.п. + Сущ3.Род.п.): *уменьшение продолжительности эрекции* и возьмем из правил синтаксиса сочетание (Сущ + Глагол): *больной жалуется.*