Описание языка LSPL (1.0.1)

1. Назначение и особенности языка

Язык LSPL (LexicoSyntactic Pattern Language) предназначен для формального описания конструкций (выражений) русского языка с целью их представления в системах автоматической обработки русскоязычных текстов, основанных на морфологическом и частичном синтаксическом анализе.

Ключевым в языке LSPL является понятие ***лексико-синтаксического шаблона***, рассматриваемого как структурный образец языковой конструкции. Шаблон задает ее лексический состав и поверхностно-синтаксические свойства и тем самым может быть использован для ***распознавания*** ее в тексте и последующего ***извлечения***.

Шаблон строится как последовательность *элементов*, описывающих соответствующие фрагменты языковой конструкции – в том порядке, в каком они встречаются в этой конструкции. Средства языка позволяют задавать вариативность конструкции, включая набор входящих в нее слов (лексем) и их морфологических характеристик (признаков).

Как выразительное средство, лексико-синтаксический шаблон удобен в первую очередь для формального описания именных словосочетаний русского языка, которые, как правило, являются грамматически согласованными. Для этого в языке LSPL предусмотрены средства задания синтаксической связи согласования слов.

Синтаксические связи могут задаваться в шаблоне следующим образом: согласование – путем указания равенства соответствующих морфологических признаков связанных элементов текста, управление – путем указания значений морфологических признаков подчинённого элемента.

Важной особенностью языка LSPL является возможность использовать при задании шаблона другие (уже определенные) шаблоны, что позволяет при формализации сложной языковой конструкции выделить ее составные части и описывать их по очереди в виде шаблонов, давая этим шаблонам имена и используя эти имена в других шаблонах.

В целом, язык шаблонов является достаточно гибким и мощным средством задания лексических и поверхностно-синтаксических свойств конструкций: LSPL-шаблоны могут описывать не только словосочетания, но и целые предложения и более крупные фрагменты текста на русском языке.

Синтаксис языка LSPL описан с помощью расширенных формул Бэкуса-Наура, а также таблиц с условными обозначениями – см. Приложения.

2. Общая структура LSPL-шаблона

Задание лексико-синтаксического шаблона обычно начинается с указания его ***имени***, за которым через знак равенства следует, в общем случае, описание нескольких альтернативных случаев формализуемой языковой конструкции. Эти ***альтернативы*** при записи разделяются метасимволом **|** – как, например, в шаблоне с именем *AD*, который описывает понятие адъектива, охватывающего прилагательное (*A*) и причастие (*Pa*):

```
AD = A|Pa
```

Заметим, что такие (внешние) альтернативы можно записывать на отдельных строках:

```
AD = A
AD = Pa
```

Разные шаблоны должны иметь разные имена, а в качестве имени может браться произвольная последовательность букв, причем первая буква должна быть заглавной. Если же определяемый шаблон не предполагается использовать в других шаблонах, то имя при его определении можно не указывать: *A|Pa*.

Достаточно часто описание шаблона включает только одну альтернативу, к примеру:

```
ANpattern = A N <A=N>
```

Этот шаблон с именем *ANpattern* включает ***элементы*** A (прилагательное), N (существительное), а также ***условие согласования*** *A=N* (записывается в угловых скобках). Он описывает грамматически согласованную именную группу из прилагательного и существительного: *программное обеспечение, пиратскому кораблю* и т.п. (но не *красному дома* и *шариковой ручка*, поскольку они не согласованы в падеже).

Кроме условий согласования в шаблоне может записываться ***словарное условие***, задающее проверку вхождения элементов-слов в некоторый словарь. Например, в шаблоне

```
VP = V<изучить> N <Dict_Nouns(N)>
```

с элементами *V* (глагол) и *N* (существительное) такое условие требует вхождения существительного в словарь с именем *Dict_Nouns*.

Отметим, что последовательность элементов в шаблоне строго соответствует их расположению в описываемой конструкции. Это означает, что при изменении взаимного расположения элементов в шаблоне результирующий шаблон будет соответствовать конструкции с другим порядком слов. Таким образом, шаблону *ANpattern* не соответствуют словосочетания *взгляд унылый* и *пример очередной* (но соответствуют сочетания *унылый взгляд* и *очередной пример*).

Элементами шаблона могут выступать:

- ***элемент-слово***,
- ***элемент-строка***,
- ***экземпляр шаблона,***
- ***опциональный элемент,***
- ***повторение элементов***,
- ***набор альтернатив***.

Первые три вида элементов (элемент-слово, элемент-строка, экземпляр шаблона) относятся к *простым элементам*, а остальные – к *сложным*, поскольку в их состав входят другие элементы.

При задании шаблона после записи всех его элементов могут быть указаны ***параметры шаблона*** (они записываются в круглых скобках). К примеру, в следующем шаблоне параметрами указываются все морфологические характеристики существительного *N*:

```
A N <A=N> (N)
```

Введение параметров шаблона важно, если в дальнейшем планируется использовать его при определении других шаблонов, т.е. для описания составных частей более сложных языковых конструкций и их грамматического согласования.

3. Элемент-строка

Элемент-строка позволяет записать в шаблоне нужную строку символов, в частности, конкретную словоформу, знак пунктуации или условное обозначение, встречающиеся в формализуемой языковой конструкции. Для этого они записываются в двойных кавычках: *"рамой"*, *"т.е."*, *"–"*, *"T25"*. К примеру, шаблон

```
Ex = "приведем" "пример" ANpattern
```

включает два элемента-строки, а также ранее определенный шаблон *ANpattern* согласованного сочетания прилагательного и существительного. Шаблон описывает фразы вида *приведем пример транзитивного замыкания*. Отметим, что отдельные словоформы конструкции (в данном случае *приведем* и *пример*) должны быть записаны как отдельные элементы-слова.

Приведем еще два примера шаблона с элементами-строками: шаблон *REF*

```
REF = "рассмотренный" "в" N
```

описывает фразы вида *рассмотренный в статье* (*газете, очерке* и т.п.); а шаблон `Del` задает небуквенные символы, используемые как разделители слов текста:

```
Del = ","| ";" |"-"
```

В элементах-строках можно использовать ***регулярные выражения***, записанные в синтаксисе PCRE (*Perl Compatible Regular Expressions*). Регулярные выражения строятся из произвольных символов (букв, цифр, спецсимволов) с помощью скобок и пяти *метасимволов*, обозначающих соответственно:

. – произвольный символ,
***** – повторение символов,
+ – непустое повторение,
? – опциональный элемент.
| – альтернатива.

К примеру, элемент-строка `"диплом(.)*"` описывает строки-слова, начинающиеся с шести букв *диплом*: *диплом, дипломную, дипломной, диплома, дипломник* и др.; а элемент-строка `"авиа(.)+"` описывает слова, начинающиеся с префиксоида *авиа*: *авиаполк*, *авианосцы*, *авиабилетов, авиационный* и т.п. В случае, если в элементе-строке необходимо записать символ, совпадающий с указанными метасимволами, его необходимо экранировать, т.е. поставить перед ним косую черту, например, для указания точки в виде элемента-строки следует записать `"\."`

Отметим, что язык LSPL не позволяет уточнять морфологические признаки слов, задаваемых как элементы-строки (например, невозможно указать в виде элемента-строки прилагательные мужского рода с префиксоидом *авиа*). Возможность указывать часть речи слов и их морфологические характеристики предоставляется при задании элементов-слов.

4. Элемент-слово

Элемент-слово соответствует отдельному слову, для которого в общем случае указываются:

- ***часть речи***; при этом используются символьные обозначения: `N` – существительное, `V` – глагол, `A` – прилагательное, `Pr` – предлог, `Pn` – местоимение и т.д. (см. **Таблицу 1**); буква `W` используется для обозначения слов, часть речи которых неважна или неизвестна;
- ***имя лексемы***, или *нормальная форма слова*: для глаголов это инфинитив, для существительных – форма именительного падежа единственного числа (а если единственное число отсутствует, то берется множественное: *ножницы*) и т.д.; лексема задает множество всех словоформ этого слова;
- ***морфологические характеристики*** (***признаки***) слова: падеж, число, род и др., ограничивающие множество допустимых словоформ описываемого слова.

Каждой части речи соответствует свой набор морфологических характеристик, в частности, характеристики существительных – род, одушевленность, падеж, число. У каждого слова (лексемы) определенной части речи некоторые характеристики связаны (т.е. фиксированы), например, для существительных – род, для глаголов – вид. Другие же характеристики свободны (т.е. изменяемы): для существительных – падеж, для глаголов – время. Конкретные значения свободных морфологических характеристик могут быть записаны в угловых скобках после имени лексемы в виде списка равенств вида

имя_характеристики = значение ,

при этом используются определенные имена (названия) признаков и их значений: `c` – падеж, `g` – род, `p` – лицо, `nom` – именительный падеж, `fem` – женский род и т.д. – см. **Таблицу 2**.

К примеру, элемент-слово `A<красный,c=nom,g=fem>` задаёт формы прилагательного *красный* в именительном падеже (`c=nom`) женского рода (`g=fem`), а

грамматическое число не конкретизируется – таким образом задаётся любая из словоформ *красная*, *красные*, *красна*, *красны*, *краснее*, *покраснее*, *краснейшая*, *краснейшие*. Элемент-слово `V<пониматься,t=pres,p=3>` описывает формы глагола *пониматься* в настоящем времени (`t=pres`) и третьем лице (`p=3`), т.е. его словоформы *понимается* или *понимаются*.

Обозначения частей речи, названия соответствующих морфологических признаков, а также и их возможные значения приведены в Таблицах 1 и 2.

Отметим, что при задании элемента-слова конкретная лексема и/или значения морфологических признаков могут быть не указаны, что позволяет описать:

- ✓ слово в его произвольной грамматической форме: например, элемент-слово `V<шить>` соответствует любой личной форме глагола *шить* ;
- ✓ произвольное слово определенной части речи в любой допустимой грамматической форме, к примеру, `A` – прилагательное, `Pn` – местоимение;
- ✓ произвольное слово заданной части речи с нужными морфологическими признаками: например, элемент-слово `N<c=nom,n=plur,g=masc>` задает любое существительное мужского рода (`g=masc`) множественного числа (`n=plur`) в именительном падеже (`c=nom`).

Язык шаблонов позволяет также записывать элементы-слова и без указания части речи, используя букву ***W***. Так, шаблон `V W N<c=ins,n=sing>` соответствует некоторому глаголу и существительному в творительном падеже (`c=ins`) единственного числа (`n=sing`), между которыми стоит произвольное слово (в частности, подходит сочетание *машет этим флагом, предусмотреть обмен информацией*)

В общем случае в описываемую шаблоном конструкцию могут входить как несколько слов разных частей речи, так и несколько разных слов одной части речи. При записи шаблона для их различения обычно используются цифровые ***индексы***. Например, шаблон `N1 N2<c=gen>` соответствует конструкции из двух произвольных существительных `N1` и `N2`, второе из которых должно стоять в родительном падеже. Часть речи элемента-слова и индекс, взятые вместе, образуют ***имя*** этого элемента-слова. Заметим, что элементы разных частей речи могут иметь одинаковые индексы, как в шаблоне `A1 N1 <A1=N1>`.

5. Сложные элементы шаблона

В шаблоне может быть задано ***повторение элементов***, которое записывается с помощью фигурных скобок: в них указываются элементы, которые могут встречаться в тексте несколько раз подряд. К примеру, повторение `{N<c=gen>}` задает непустую цепочку из идущих подряд существительных в родительном падеже: *анализа областей применения*. Повторение `{V<t=past,g=masc,n=sing> ",”}` задает последовательность глагольных форм единственного числа (`n=sing`) мужского рода (`g=masc`) прошедшего времени (`t=past`), разделяемых запятыми: *пришел, увидел, победил*.

Если известны ограничения на количество одинаковых элементов, т.е. ***максимальный и минимальный множитель***, то можно указать их в шаблоне, записывая в угловых скобках сразу после закрывающей фигурной скобки. Так, запись `{A}<1,3> N` с множителями `<1,3>` задает последовательность из одного, двух или трех прилагательных и одного существительного, например: *новый компактный высокопроизводительный компьютер* или *легкий синий шарф.*

Второй из множителей может быть опущен – в этом случае фиксируется только минимальное число элементов повторения. Тем самым, шаблону `{A}<3>N` соответствует только первый из приведенных выше примеров (т.к. число прилагательных должно быть больше или равным 3). Шаблон же `{A}<1>N` в отличие от `{A}N` требует присутствия по

крайней мере одного прилагательного, и в результате одиночные существительные (*компьютер*, *файл* и др.) ему не соответствуют (хотя подходят под шаблон `{A}N`).

Язык LSPL позволяет включать в шаблон ***опциональные элементы***, записываемые в квадратных скобках: к примеру, элемент `["не"]` указывает необязательность вхождения строки-частицы *не* в описываемую языковую конструкцию. Заметим, что опциональный элемент, по сути, является упрощенной записью повторения со множителями `<0,1>`. В частности, запись `["не"]` эквивалентна `{"не"}<0,1>`.

Сам опциональный элемент может состоять из ***набора альтернатив*** (альтернативных вариантов конструкции), разделенных метасимволом `|`. К примеру, шаблон

```
["в"|"на"] N <шкаф,c=prep>
```

описывает сочетание существительного *шкаф* в предложном падеже с двумя различными предлогами или без них (*в шкафу*, *на шкафу*, *в шкафе*, *на шкафе, шкафу, шкафе*).

Набор альтернатив допустимо использовать и в повторениях. Так, элемент шаблона `{Av |Ap }` задает последовательность (возможно, пустую) наречий (`Av`) и деепричастий (`Ap`). В частности, ему соответствует цепочка слов *осмотревшись неспешно тихо*.

В общем случае каждая входящая в набор альтернатива может быть последовательностью элементов шаблона, а также накладываемых на эти элементы условий (согласования и словарных). Например, набор альтернатив `A N <A=N>|Pa N <Pa=N>` задает либо грамматически согласованную пару из прилагательного и существительного, либо согласованную пару из причастия и существительного.

6. Условия согласования

Условия согласования указывают на грамматическое согласование отдельных элементов формализуемой языковой конструкции. Условия записываются в угловых скобках, и должны стоять после записи в шаблоне всех согласуемых элементов.

Условие согласования может быть выражено в виде равенства значений согласуемых морфологических признаков элементов. К примеру, шаблон

```
PV = Pn V <Pn.n=V.n, Pn.g=V.g>
```

описывает согласованную в числе (`n`) и роде (`g`) пару слов – местоимение (`Pn`) и глагол (`V`): *мы введем*, *они разработали*, *я ищу* и т.д. В такой записи используются ***составные имена***, образованные из имени элемента шаблона и имени согласуемого признака (эти имена разделяются точкой). В приведенном шаблоне `PV` используются составные имена `Pn.n`, `V.n`, `Pn.g` и `V.g`.

В случае, когда должны быть согласованы все общие морфологические характеристики двух элементов шаблона, условия согласования можно записать короче, используя только имена согласуемых элементов. Например, шаблон `ANpattern = A N <A=N>` описывает согласованную именную группу из существительного и прилагательного: *сложное доказательство*, но не *актуальные исследование* (нет согласования в числе). Указанная более короткая запись согласования не годится для случаев, когда согласуется только часть общих морфологических характеристик элементов шаблона, и тогда необходимо записывать согласование по отдельным морфологическим признакам.

Язык шаблонов допускает одновременное согласование нескольких элементов шаблона – как при согласовании отдельных их морфологических признаков, так и при согласовании всех их общих характеристик. К примеру, в шаблоне `A1 A2 N <A1=A2=N>`

согласованы оба прилагательных и существительное (при этом согласование выполняется в числе, роде и падеже: *твердым решительным шагом*). В следующем шаблоне `Ns` согласованы конкретно падеж и число трех существительных, образующих сочинительную конструкцию (например: *ложки, вилки и ножи*):

```
Ns = N1 "," N2 "и" N3 <N1.c=N2.c=N3.c, N1.n=N2.n=N3.n>
```

Отметим, что язык шаблонов позволяет указывать условия согласования и для элементов повторения. К примеру, шаблон `{A} N <A=N>` задает именную группу из нескольких прилагательных и существительного, причем все прилагательные согласованы с существительным (*краткие полезные сведения, светлой просторной комнате, адаптивная дифференциальная импульсная модуляция*):

В случаях, когда в шаблоне несколько элементов и несколько условий их согласования, их можно записывать по очереди друг за другом – требуется только, чтобы каждое условие согласования в шаблоне не опережало запись согласуемых элементов. Так, шаблон

```
ANV = {A} N <A=N> V<t=past> <V.n=N.n>
```

описывает согласованную именную группу из прилагательных и существительного, за которой идет глагол в форме прошедшего времени (`t=past`), причем глагол и существительное согласованы в грамматическом числе (`n`), например: *яркие красивые птицы пели, последнее замечание подтвердилось*.

Дополнительно, в форме согласования могут быть записаны условия равенства основ (или псевдооснов) элементов-слов – при этом используется их характеристика `st` (`stem` – основа). Укажем пример соответствующего шаблона (`Ap` означает деепричастие):

```
Ap "," V<Ap.st=V.st>
```

– этому шаблону соответствует, в частности, фраза *Уходя, уходи*.

7. Параметры и шаблона

При описании лексико-синтаксического шаблона можно указать его ***параметры***, которые записываются после всех его элементов и условий. Параметры шаблона не накладывают дополнительных ограничений на описываемую шаблоном языковую конструкцию, они служат лишь для указания морфологических признаков, которые можно конкретизировать или согласовывать при дальнейшем использовании этого шаблона.

Параметрами шаблона могут выступать только морфологические признаки входящих в него элементов-слов (а также экземпляров шаблонов – см. след. раздел), если последние не входят в его опциональные элементы и повторения. Параметры могут быть взяты только из числа тех морфологических характеристик, что не конкретизированы в рассматриваемом шаблоне (т.е. для них не были заданы конкретные значения).

Параметры записываются в конце шаблона в круглых скобках, через запятую, в виде составных имен (включающих имя элемента-слова и имя его морфологического признака). К примеру, параметрами шаблона

```
ANp = {A} N1 N2<c=gen> <A=N> (N.c, N.n)
```

заданы падеж (`c`) и число (`n`) входящего в него существительного `N`. Установление этих признаков как параметров означает, что при дальнейшем использовании шаблона `ANp` эти морфологические характеристики существительного `N` могут быть конкретизированы и/или использоваться в условиях согласования. Заметим, что в данном примере в качестве параметров шаблона не могут быть взяты признаки прилагательного `A`, поскольку он входит в элемент-повторение, а также признак второго существительного `N2.c`, т.к. он конкретизирован.

В случаях, когда параметрами шаблона выступают все морфологические признаки некоторого элемента-слова, допускается сокращенная запись: в качестве параметра записывается только имя этого элемента. К примеру, в шаблоне

```
AANp = A1 A2 N <A1=A2=N> (N)
```

который описывает согласованную именную группу из двух прилагательных и существительного, параметрами становятся все морфологические характеристики элемента-слова `N` (в том числе род, падеж, число). В дальнейшем эти параметры могут быть использованы для конкретизации экземпляров этого шаблона (см. след. раздел).

Для удобства использования параметров, заданных в виде составных имен, для них можно определить новые имена, используя ключевое слово **`as`** . Эти имена записываются строчными латинскими буквами и не должны совпадать с именами шаблонов и с именами других параметров шаблонов.

Введение новых имен необходимо в случаях, когда параметрами выступают одинаковые признаки разных элементов. Например, для шаблона двух существительных (второе – в родительном падеже):

```
ANNp =A N1 N2<c=gen>(A, N1.g as maing, N2.g as auxg)
```

заданы следующие параметры: все морфологические признаки прилагательного, род первого существительного и род второго существительного, причем последним даны новые имена: *maing* и *auxg* соответственно.

Отметим, что в шаблонах языковых конструкций, описываемых несколькими альтернативами, параметры могут быть заданы в каждой альтернативе, как в следующем шаблоне адъектива (прилагательного или причастия):

```
AP = A (A)|Pa (Pa)
```

8. Экземпляры шаблона

В качестве элемента LSPL-шаблона может быть использован другой, ранее определенный шаблон, характеризующий некоторую часть описываемой языковой конструкции. Более точно, в шаблоне может быть использован ***экземпляр шаблона***.

В общем случае экземпляр шаблона задается именем используемого шаблона и конкретизациями тех морфологических характеристик, которые входят в число его параметров. Как и для элемента-слова, конкретные характеристики экземпляра шаблона перечисляются в угловых скобках после указания его имени, которое может включать индекс. Например, после определения шаблона

```
AANp = A1 A2 N <A1=A2=N> (N)
```

можно использовать экземпляр шаблона *AANp<g=neut>* описывающий именную группу из двух согласованных прилагательных и существительного среднего рода, в частности: *яркое весеннее небо*. В данном экземпляре шаблона конкретизируется род существительного (*g=neut*), входящего в число параметров исходного шаблона *AANp* (в нем же задано согласование входящих в конструкцию слов слов).

В экземпляре *NNp<n=sing>* шаблона *NNp*, описывающего два существительных:

```
NNp = N1 N2<c=gen> (N1) ,
```

уточняется грамматическое число (единственное: *n=sing*) первого существительного (его морфологические характеристики были указаны параметрами для *NNp*). Падеж второго существительного *N2* (родительный: *c=gen*) конкретизирован в самом шаблоне *NNp* .

Следующий пример демонстрирует применение экземпляра шаблона. Шаблон

```
STP = "далее" "-" NNp<c=nom>
```

задан с помощью экземпляра *NPs<c=nom>* ранее описанного шаблона *NNp* (причем в этом экземпляре конкретизирован падеж: *c=nom*). Тем самым, *STP* описывает языковые фразы, в которых после слова *далее* через тире идет именная группа из двух существительных, в частности, описывает фразу *далее – алгоритм приведения*.

Как и элементы-слова одной части речи, экземпляры одного шаблона могут встречаться в некотором LSPL-шаблоне несколько раз – в этом случае они должны иметь разные индексы. Так, в шаблоне

```
PH = ANp1<c=acc> V<обнаружить> ANp2<c=nom> <V.n=ANp2.n>
```

заданы два экземпляра шаблона *ANp* согласованного сочетания прилагательного и существительного:

```
ANp = A N <A=N>(T)
```

Эти два экземпляра обозначают два разных словосочетания, первое – в винительном падеже (*c=acc*), а второе – в именительном (остальные грамматические признаки не конкретизированы). Тем самым шаблон *PH* описывает, к примеру, фразу *Интересную закономерность обнаружили британские учёные*. Этот шаблон показывает, что в условиях согласования могут использоваться не только имена слов-элементов, но и имена шаблонов – в данном случае эти условия определяют, что словосочетание *ANp2* должно быть согласовано с глаголом *V* в грамматическом числе (*n*). Как и ранее, при записи условий согласования шаблона используются составные имена: названия согласуемых грамматических признаков предваряются именами слов-элементов или именами элементов шаблона, к которым эти признаки относятся.

9. Словарные условия

Поскольку при распознавании конструкций в тексте довольно часто необходимо проверять наличие в них слов из некоторого словаря (например, словаря терминологических словосочетаний), в шаблонах допускается запись ***словарных условий***. Эти условия имеют вид обращения к некоторой булевской функции, проверяющей вхождение в словарь, и ее имя можно трактовать как имя словаря.

Подобно условиям согласования, словарные условия записываются в угловых скобках. Например, запись *<Dict(A1)>* означает, что элемент-слово *A1* должно входить в словарь *Dict* (точнее, *Dict* – имя функции, проверяющей вхождение в этот словарь).

Возможна также запись условий, предполагающих обращение к ***функции-словарю*** с двумя аргументами. К примеру, шаблон

```
SynPair =N1 "и" N2 <Syn(N1,N2)>
```

где *Syn* – имя словаря синонимов, описывает пары синонимов, соединенных союзом *и* (например: *жестокий и безжалостный*).

В качестве аргументов функций-обращений в общем случае допустимы шаблоны языковых конструкций, вхождение которых в словарь необходимо проверить. Например, в словарном условии *<Dict(A1 A2 N)>* проверяется вхождение в словарь *Dict* сочетания из двух прилагательных и существительного (они должны быть указаны в шаблоне до рассматриваемого словарного условия).

Заметим, что словарные условия и условия согласования могут быть записаны в шаблоне друг за другом в одних угловых скобках, например:

```
A1 A2 N <A1=A2=N, Terms(N)> (N)
```

В этом шаблоне кроме условия согласования прилагательных с существительным задается условие вхождения последнего в словарь *Terms*.

Использование в LSPL-шаблонах словарных условий предполагает предварительное подключение соответствующих словарей к программной библиотеке языка LSPL.

10. Дополнительные примеры шаблонов

Рассмотрим примеры, иллюстрирующие различные возможности языка LSPL.

1) Приведем пример шаблона, описывающего однородные сочинительные конструкции вида *горы, солнце и море* или *процессор, монитор, клавиатура, а также мышь*:

```
N1 {"," N2|"и" N3|"," "а" "также" N4}
```

Указанные в шаблоне альтернативы могут идти в произвольном порядке, например: *дамы и господа, леди и джентельмены, а также мадам и месье*. Заметим, что этому шаблону соответствует и одиночное существительное, поскольку в общем случае элементы-повторения *{}* допускают отсутствие повторяемых конструкций. Чтобы исключить это, необходимо приписать повторению множитель *<1>*:

```
N1 {"," N2|"и" N3|"," "а" "также" N4}<1>.
```

Теперь повторение альтернативной конструкции должно быть не менее 1 раза, но шаблон соответствует и частям конструкций перечисления, например, в случае фразы *горы, солнце и море* шаблону соответствуют: *горы, солнце*; *солнце и море*; *горы, солнце и море.*

Уточненный шаблон не подходит для распознавания перечисления во фразе *Дама сдавала в багаж диван, чемодан, саквояж, картину, корзину, картонку и маленькую собачонку* (из-за последнего словосочетания, включающего прилагательное). Шаблон можно подправить следующим образом:

```
N1 {"," N2|"и" A N3 <A=N3>|"," "а" "также" N4}<1>.
```

Другой способ исправления – допустить в качестве элементов перечисления согласованные пары прилагательных и существительных, используя вспомогательный шаблон `AN` :

```
AN = {A} N <A=N>(N)
PCoord = AN1 {"," AN2|"и" AN3|"," "а" "также" AN4}<1>
                                <AN1.c=AN2.c=AN3.c=AN4.c > (AN1)
```

Теперь шаблону `PCoord` соответствует также фраза *горы, яркое солнце и синее спокойное море*. В этом шаблоне дополнительно добавлено условие согласования членов перечисления в падеже (это возможно, т.к. шаблон `AN` имеет параметр `N`). Тем самым, с помощью `PCoord` можно распознать в тексте грамматически правильные сочинительные конструкции рассмотренного вида, а с помощью нижеследующего шаблона, включающего экземпляр `PCoord` (с конкретизированным винительным падежом):

```
N V<t=past> <V=N> "в багаж" PCoord4<c=acc> ,
```

можно распознать приведенную выше фразу про даму.

2) Следующий шаблон иллюстрирует возможность описания языковых конструкций с фиксированными лексемами – в данном случае описываются два конкретных терминологических словосочетания *битовый массив* и *битовый образ*:

```
A1<битовый> {N1<массив> | N1<образ>}<1,1> <A1=N1>
```

Множители `<1,1>` конструкции повторения обеспечивают обязательное присутствие тоьлько одного элемента из двух указанных альтернатив.

3) В качестве еще одного примера, поясняющего использование экземпляров шаблона, определим шаблон именного словосочетания `NP`, который затем будем использовать при описании более сложной конструкции:

```
NP = {A} N1 {N2<c=gen>} <A=N1> (N1)
```

Этот шаблон описывает именную группу, в котором сначала идет произвольное количество прилагательных `{A}`, далее – существительное `N1`, с которым согласованы все предыдущие прилагательные, затем идет произвольное количество существительных в родительном падеже `{N2<c=gen>}`, например: *<u>программа</u>*, *дипломная <u>работа</u> студента*, *<u>наименование</u> товара*, *прекрасная солнечная <u>погода</u>*, *восходящий <u>процесс</u> порождения элементов решетки* (в приведенных примерах подчеркнуто главное слово `N1` именного сочетания).

Установление параметром шаблона существительного `N1` (главного слова группы) означает, что в именной группе `NP` могут быть при необходимости конкретизированы характеристики этого существительного. Конкретизация падежа происходит, к примеру, в нижеследующем шаблоне для описания конструкции, состоящей из группы `NP` и согласованного с нею глагола:

```
NP <c=nom> V <NP=V>
```

Этому шаблону соответствует, к примеру, фраза *прекрасная солнечная погода закончилась*.

4) Язык LSPL допускает запись рекурсивных шаблонов: примером может служить шаблон именной группы более общей структуры, чем рассмотренная в предыдущем примере:

```
NG = {A} N1 <A=N1> {NG2 <c=gen>} (N1)
```

Этот шаблон допускает зависимые прилагательные не только перед главным существительным группы (*N1*), но и при всех других существительных (употребляемых в родительном падеже): *тоненькая струйка дыма далекого пожара*, *динамичность изменения доступного информационного пространства* (опять же в примерах подчеркнуто главное слово *N1* именного сочетания).

5) В качестве следующего примера приведем шаблон одной из характерных конструкций определения новых терминов, встречающихся в научно-технических текстах:

```
"под" NP1<c=ins> ["в" "общем" "случае"]
                          "будем" "понимать" NP2<c=acc>
```

В шаблоне используются экземпляры ранее определенного шаблона *NP* (можно использовать и *NG*), этому шаблону соответствует, к примеру, фраза *Под семантической связью в общем случае будем понимать отношение понятий* (подчеркнуты фиксированные слова описанной шаблоном конструкции).

11. Поиск конструкций по шаблонам

LSPL-шаблоны предназначены для автоматического поиска (распознавания) описанных ими языковых конструкций в заданном тексте на русском языке. Для этого последовательно применяется процедура ***наложения шаблона на текст*** (выполняемая LSPL-анализатором), результатом которой в общем случае являются различные *варианты наложения*. ***Вариант наложения*** – это найденный непрерывный отрезок (фрагмент) текста, соответствующий шаблону (т.е. удовлетворяющий всем его условиям) вместе с его *интерпретацией*, т.е. набором конкретных значений морфологических характеристик слов, входящих в этот отрезок (интерпретация также удовлетворяет условиям шаблона).

В результате наложения шаблона на текст всем элементам шаблона, включая элементы-слова, экземпляры шаблонов, повторения элементов и наборы альтернатив, сопоставлены определенные отрезки текста – из них и образован фрагмент текста, представляющий конкретное вхождение в текст распознанной языковой конструкции. Указанные отрезки можно считать значениями элементов шаблона, которые они получили в результате наложения шаблона.

Различные варианты наложения могут представлять различные случаи вхождения в текст этой конструкции, но не только. При поиске по заданному шаблону вследствие различных видов омонимии (совпадения форм одного или разных слов) может возникнуть несколько вариантов наложения рассматриваемого шаблона на один и тот же отрезок текста – эти варианты отличаются интерпретациями входящих в шаблон элементов. Например, при наложении шаблона `AN = {A} N <A=N>(N)` на текст *яркое солнце* получается два варианта, первый из которых соответствует именительному падежу слова *солнце*, а второй – винительному

В некоторых случаях возможно также пересечение отрезков текста, соответствующих различным вариантам наложения одного и того же шаблона. Так, при наложении шаблона

```
N1 {"," N2|"и" A N3 <A=N3>|"," "а" "также" N4}<1>
```

на текст *процессор, монитор, а также клавиатура* в число вариантов наложения входят отрезки текста: *процессор, монитор*; *монитор, а также клавиатура*; *процессор, монитор, а также клавиатура*.

Описание синтаксиса языка LSPL (в виде БНФ)

Нетерминальные символы языка LSPL выделены курсивом, а терминальные символы – синим цветом.

Шаблон языка LSPL включает имя, а также::

- левую часть – *шаблон распознавания конструкции* (в том числе – параметры шаблона);
- правую часть – *шаблоны_извлечения_текста* и *синтезируемые_шаблоны* .

описание_шаблона ::= *имя_ шаблона* = *шаблон_распознавания*
[=text> *шаблоны_извлечения_текста*]
[=pattern> *синтезируемые_шаблоны*]

имя_шаблона ::= *Загл. буква* {*буква*}

шаблон_распознавания ::= *::=* *шаблон_конструкции* [(*параметры_ шаблона*)]
{ | *шаблон_конструкции* [(*параметры_ шаблона*)] }

шаблон_конструкции ::= *последовательность _элементов* [<*условия*>]
{ *последовательность _элементов* [<*условия*>] }

последовательность _элементов ::= *элемент_шаблона* {*элемент_шаблона*}

элемент_шаблона ::= *простой_элемент* | *опциональный_элемент*| *повторение_элементов*

простой_элемент ::= *элемент-строка* | *элемент-слово* | *экземпляр_шаблона*

опциональный_элемент ::= [*набор_альтернатив*]

повторение_элементов ::= { *набор_альтернатив* }
[< *минимальный_множитель* [, *максимальный_множитель*]>]

набор_альтернатив ::= *шаблон_конструкции* { | *шаблон_конструкции* }

минимальный_множитель ::= *неотрицательное_ целое_число*

максимальный_множитель ::= *неотрицательное_ целое_число*

элемент-строка ::= "*символ* { *символ* }"| " *регулярное_выражение* "

элемент-слово ::= *имя_элемента-слова* |
имя_элемента-слова < *имя_лексемы* { , *характеристика_слова*} > |
имя_элемента-слова < *характеристика_слова* { , *характеристика_слова*} >

имя_лексемы ::= *буква* { *буква* | *дефис*}

имя_элемента-слова ::= *часть_речи* [*индекс*]

часть_речи ::= **N**| **A**| **V**| **Pa**| **Ap**| **Pn**| **Av**| **Cn**| **Pr**| **Pt**| **In**| **Nm**| **W**

индекс ::= *цифра* {*цифра*}

характеристика_слова ::= *название_ признака* = *значение_ признака*

название_признака ::= **c** | **n** | **g** | **doc** | **t** | **a** | **f** | **m** | **p** | **r** | **st**

значение_признака :: = **nom** | **gen** | **dat** | **acc** | **ins** | **prep** | **un** |
sing | **plur** |
masc | **fem** | **neut** |
comp | **sup** | **no** |

pres | **past** | **fut** | **inf**
anim | **inan** |
full | **short** |
ind | **imp** | **conj** | **cond** |
1 | **2** | **3** | **yes** | **no**

Полные названия *морфологических признаков* и их допустимые значения – см. Таблицы 1-2.

экземпляр_шаблона ::= *имя_экземпляра_шаблона*

[< *характеристика _экземпляра* {, *характеристика _экземпляра* } >]

имя_экземпляра_шаблона ::= *имя_шаблона* [*индекс*]

характеристика _экземпляра ::= *имя_признака* = *значение_ признака* |
имя_параметра = *значение_ признака*

условия ::= *условие_ согласования* | *словарное_условие*| *условия* , *условие_согласования* |
|*условия* , *словарное_условие*

условие_ согласования ::= *составное_имя* = *составное_имя* { = *составное_имя* }
| *имя_элемента* = *имя _элемента* { = *имя _элемента* }

составное_имя ::= *имя_элемента* . *имя_признака*| *имя_параметра*
имя_элемента ::= *имя_элемента-слова* | *имя_экземпляра_шаблона*

словарное_условие ::= *имя_словаря*
(*последовательность_имен* {, *последовательность_имен* })
последовательность _имен::= *имя_элемента* { *имя_элемента*}

параметры_ шаблона ::= *параметр* {, *параметр* }

параметр ::= *имя_элемента* | *имя_элемента* . *имя_признака* [**as** *имя_параметра*]
имя_параметра ::= *лат. буква* {*лат. буква*}

Условные обозначения языка LSPL

Таблица 1. Части речи и их морфологические признаки

Часть речи (part)	Обозначение	Возможные признаки
Произвольная часть речи (word)	**W**	
Существительное (noun)	**N**	Род Одушевленность (пока в реализации нет) Падеж Число
Прилагательное (adjective)	**A**	Полное, краткое, степенное или неизменяемое (пока в реализации нет) Падеж (у полного) Число (у полного и краткого) Род (у полного и краткого в единственном числе)
Глагол (verb)	**V**	Наклонение Время (в изъявительном наклонении) Лицо Число --- Род (в изъявительном наклонении, прошедшем времени, единственном числе) Возвратность
Причастие (participle)	**Pa**	Полное или краткое Падеж (у полного) Число --- Род (если единственное число) Возвратность (пока в реализации нет)
Деепричастие (adverbial participle)	**Ap**	Время Возвратность
Местоимение (pronoun)	**Pn**	Падеж Число Род (если возможно) Лицо (у личного местоимения)
Наречие (adverb)	**Av**	
Союз (conjunction)	**Cn**	
Предлог (preposition)	**Pr**	
Частица (particle)	**Pt**	
Междометие (interjection)	**Int**	
Числительное (numeral)	**Num**	

Таблица 2. Морфологические признаки и их возможные значения

Признак	Имя признака	Сокращенное имя	Возможные значения признака	Обозначение в языке
Основа	stem	**st**	Основа слова	—
Падеж	case	**c**	Именительный (nominative)	**nom**
			Родительный (genitive)	**gen**
			Дательный (dative)	**dat**
			Винительный (accusative)	**acc**
			Творительный (instrumental)	**ins**
			Предложный (prepositional)	**prep**
			Неизменяемое (uninflected)	**un**
Число	number	**n**	Единственное (singular)	**sing**
			Множественное (plural)	**plur**
Род	gender	**g**	Мужской (masculine)	**masc**
			Женский (feminine)	**fem**
			Средний (neuter)	**neut**
Степень сравнения	degree of comparison	**doc**	Сравнительная (comparative)	**com**
			Превосходная (superlative)	**sup**
			Отсутствует (no)	**no**
Время	tense	**t**	Настоящее (present)	**pres**
			Прошедшее (past)	**tpast**
			Будущее (future)	**fut**
			Неопределенная форма (infinitive)	**inf**
Одушевленность	animate	**a**	Одушевленный (animate)	**anim**
			Неодушевленный (inanimate)	**inan**
Форма	form	**f**	Полное	**full**
			Сокращенное	**short**
Наклонение	mode	**m**	Изъявительное (indicative)	**ind**
			Повелительное (imperative)	**imp**
			Сослагательное (conjunctive)	**conj**
			Условное (conditional)	**cond**
Лицо	person	**p**	1	**1**
			2	**2**
			3	**3**
Возвратность	reflexive	**r**	Невозвратный (no)	**no**
			Возвратный (yes)	**yes**