

АППАРАТ R-ОТСЫЛОК КАК УНИВЕРСАЛЬНОЕ СРЕДСТВО СИНТАКСИЧЕСКОГО СИНТЕЗА

(на опыте разработки системы русского синтеза RussLan)¹

THE FORMALISM OF R-ATTRIBUTES AS A UNIVERSAL MEANS OF SYNTACTIC GENERATION (based on its implementation in the RussLan system of Russian generation)

М.И. Канович,

З.М. Шаляпина (ZMShal@yandex.ru)

Институт востоковедения РАН, Москва

Аппарат R-отсылок позволяет представлять и обрабатывать структурные отношения и связанные с ними языковые правила как реляционные атрибуты сущностей, для которых они релевантны. На его основе решается целый спектр задач синтаксического синтеза, от вычисления моделей управления окказионализмов до лексико-синтаксических трансформаций.

1. Синтаксический синтез в системе RUSSLAN

Система RUSSLAN [5] разрабатывается как синтезатор русских текстов в составе экспериментального комплекса ЯРАП для исследований по японско-русскому автоматическому переводу. Однако она полностью независима от остальных модулей комплекса, будучи связана с ними только через входные данные, подаваемые на ее вход в виде текстовых файлов. Поэтому ее можно интегрировать также в другие системы обработки информации, предполагающие синтез выходных текстов на русском языке.

Система реализована на языке программирования Turbo Pascal. Она обеспечивает все основные этапы синтаксического синтеза для русского языка, а также включает модуль русского морфологического синтеза RUMORS [3], так что на выходе выдается цепочка реальных русских словоформ. По наполненности лингвистической информацией система носит пока экспериментальный характер: идет отработка формальных средств и механизмов синтеза, один из типов которых и составляет предмет данной работы.

Под синтаксическим (*Синт-*) синтезом понимается этап, промежуточный между семантическим синтезом (при переводе – межязыковым переходом) и морфологическим синтезом в узком смысле – построением синтетических словоформ по их лексико-морфологическим (*ЛМ-*) определениям. Предварительные варианты ЛМ-определений имеются уже в составе входного представления текста, сформированного семантическими или переводческими процедурами. Но эти процедуры, как правило, не ориентированы на специфику лингвистических средств выходного языка, так что входные ЛМ-определения словоформ не обязательно отвечают требованиям, которые их контекст предъявляет к ним в выходном языке. Более того, в случае перевода они могут отражать особенности другого – входного – языка, содержать лакуны и неточности, противоречить каким-то ограничениям синтаксической и морфологической систем языка синтеза и т.п. Синт-синтез должен устранить некорректности входа и свести его к цепочке полных, непротиворечивых и синтетически реализуемых ЛМ-определений словоформ.

В системе RUSSLAN эта задача решается на основе сущностного подхода к языку, который ставит в центр описания языка и текста элементарные лингвистические сущности (конкретные лексические и грамматические морфемы и обобщающие их категории), а все отношения и правила трактует как их атрибуты. Это позволяет, среди прочего, применять для представления текста на всех этапах Синт-синтеза принципиально одни и те же формальные средства. В том числе его вход и выход различаются не форматом, но лишь конкретным составом входящих в них определений словоформ и степенью эксплицированности разных компонентов этих определений. Входные определения в общем случае содержат структурно-синтаксические атрибуты словоформ, задающие их лексико-синтаксический (*ЛС-*) контекст (см. ниже), морфологические же категории могут отсутствовать либо, как и лексические единицы (*ЛЕ*), не отвечать (или неполностью отвечать) имеющемуся ЛС-контексту. На выходе ЛС-атрибуты отсутствуют, зато лексические и морфологические компоненты задают словоформу корректно и однозначно.

Таким образом, Синт-синтез предстает не как переход между глубинным и поверхностным уровнями Синт-представления текста (различающимися, по определению уровней в [1: 20], прежде всего формальными языками), но как верификация и коррекция представления текста в пределах одного и того же формального языка.

¹ Исследование выполняется при частичной финансовой поддержке РФФИ (грант № 04-01-00623).

Такая трактовка Синт-синтеза позволяет, в частности, определять и реализовать его как принципиально рекурсивный процесс.

2. R-отсылки во входном представлении синтезируемого текста

Представление текста на входе модуля Синт-синтеза системы RussLan, или *ЛС-задание*, имеет вид цепочки *ЛС-запросов*, каждый из которых определяет отдельную *ЛС-сущность* как потенциально минимальную единицу синтаксиса (= словоформу или функционально аналогичное ей словосочетание) в ее структурно-синтаксическом контексте. Формальных ограничений на объем ЛС-задания нет: входной файл может содержать сколь угодно длинную последовательность ЛС-запросов, определяющую тексты любого размера, вплоть до многоабзачных (примеры, по необходимости, небольшие, см. в разделе 4).

Порядок ЛС-запросов является значимым: при оформлении структурных связей между ЛС-сущностями приоритет имеют те средства (включая перифрастические), которые позволяют сохранить взаимное расположение участников этих связей. Понятно, что удачность результатов синтеза зависит при этом прежде всего от того, насколько полно отражена в его лингвистическом обеспечении система таких средств и насколько гибкой является сама эта система в языке синтеза. Последнее требует, очевидно, специальных исследований, так как задача рассмотрения языковых средств с точки зрения возможностей обеспечения с их помощью определенного словоупотребления ранее не ставилась.

Опорным элементом ЛС-запроса является ЛЕ, от которой предполагается образовать определяемую запросом словоформу. Ее могут сопровождать контекстные атрибуты двух типов: *признаковые (П-)* и *реляционные (R-)*. Последние вводятся буквой R, так что в общем случае ЛС-запрос имеет вид:

<ЛЕ> <П-атрибуты> R <R-атрибуты>.

В виде П-атрибутов задаются морфологические (словоизменяемые и/или словообразовательные) характеристики словоформы, а также служебные признаки. R-атрибуты во входном представлении ЛС-сущности служат для определения ее ЛС-контекста в структуре отношений зависимостей и кореферентности. Для записи таких *структурно-контекстных* атрибутов и был первоначально разработан формализм *R-отсылок*, в дальнейшем обобщенный также на ряд других аспектов синтеза. В общем случае R-отсылка имеет вид:

<R-функция>[<отсылочная ЛС-сущность C₀>]

В структурно-контекстных атрибутах R-функция задает прежде всего синтаксическую роль, выполняемую сущностью-носителем атрибута относительно вводимого им участника ее ЛС-контекста – сущности C₀. Для отношений зависимости это роль синтаксического "слуги" D<x> или "хозяина" G<x>, для отношений кореферентности – роль анафора F0. Такие R-отсылки называются поэтому *DGF-отсылками*.

Функции D и G всегда сопровождаются спецификатором <x>. Это либо одна из букв m (*modifier*), s (*specification*) или c (*co-ordination*), обозначающих, гесп., адьюнктивную (определятельную или обстоятельственную), спецификативную (апозитивную) или сочинительную связь, либо цифра – номер валентности, которую реализует определяемое отсылкой отношение зависимости. Если носителем такой валентности является зависимый участник реализующего ее отношения (адьюнкт), перед номером добавляется буква m. Если ее нет, носителем валентности является управляющий участник отношения.

Отсылочная ЛС-сущность определяется в составе R-отсылки своим ЛС-запросом (в D- и G-отсылках, возможно, предваряемым контекстной семантической интерпретацией задаваемого данной R-отсылкой отношения зависимости). Такой вложенный ЛС-запрос, в свою очередь, может включать одну или несколько собственных R-отсылок, и т.д. Тем самым в этих терминах может задаваться структурный контекст любого объема, вплоть до целого предложения. Например, словоформам *поручить* и *работу* вместе с их полным ЛС-контекстом в предложении *Сотрудница безуспешно убеждала директора поручить ей более серьезную работу* соответствуют, гесп., ЛС-запросы:

- (1) поручить R G2[работа Gm1[серьезный сравн]]
D3[убедить прош RG1[сотрудница] G2[директор]]
G3[он F0[сотрудница]] Gm1[безуспешно]]
- (2) работа R Gm1[серьезный сравн] D2[поручить R
D3[убедить прош RG1[сотрудница] G2[директор]]
G3[он F0[сотрудница]] Gm1[безуспешно]]

Как видно уже из этих примеров, отношение зависимостей O<r> между зависимой единицей X и управляющей ею единицей Y (здесь, в частности, между *работа* и *поручить*) при таком способе представления их ЛС-контекста отображается в двух разных ЛС-запросах двумя отдельными, хотя содержательно сопряженными друг с другом R-отсылками: X определяется как "слуга" единицы Y с помощью R-отсылки D<r>[Y], а Y – как "хозяин" единицы X с помощью R-отсылки G<r>[X]. Примеры показывают также, что при полном отображении ЛС-контекста единиц, входящих в одно и то же предложение, фрагменты такого контекста, совпадающие для разных единиц, дублируются при каждой из них.

Понятно, что для корректного оформления конкретной ЛС-сущности ее полный ЛС-контекст обычно не требуется. Часто достаточно единичной D-отсылки к непосредственному "хозяину" оформляемой ЛС-сущности, уже без сведений о его собственном ЛС-контексте. Так, словоформы *поручить* и *работу*, задаваемые ЛС-запросами (1) и (2), столь же точно определяются ЛС-запросами:

- (1a) поручить сов R D3[убедить прош]
 (2a) работа R D2[поручить].

Тем не менее иногда нужны данные и об элементах ЛС-контекста, отделенных в нем от рассматриваемой ЛС-сущности другими его участниками (см., напр. р.4, п. 4.5). С другой стороны, в ряде случаев достаточный ЛС-контекст может быть выявлен даже в отсутствие задающих его реляционных атрибутов – за счет лингвистических умолчаний, учитывающих линейную структуру входного представления текста. Таким образом, актуальность приобретает пока весьма мало изучившийся вопрос об оптимальном объеме ЛС-контекста для синтеза сущностей различных классов. Для разных естественных языков этот вопрос, скорее всего, должен решаться по-разному. Ниже в р. 3.3 будут отмечены некоторые из умолчаний, которые релевантны для языков типа русского и реализованы в системе RussLan.

3. Расширенное определение R-отсылок в их соотношении с R-правилами

В настоящее время понятие R-отсылки, оставаясь средством отображения ЛС-контекста, определяется более широко. Именно:

R-отсылка – это реляционный атрибут лингвистической сущности, который задает ее языковую или текстовую функцию относительно некоторой другой сущности (в частном случае относительно себя самой) и тем самым идентифицирует класс ассоциируемых с этой функцией *R-правил*, определяющих условия и способы модификации сущности-носителя данного атрибута в ходе синтеза.

В языковом описании R-правила, связанные с той или иной R-отсылкой, задаются как свойства лексических или грамматических единиц, инстанциацией которых является вводимая данной отсылкой отсылочная сущность. Соответственно, они распределяются по статьям этих единиц (*resp.*, словарным или грамматическим). Лингвистический смысл R-правила состоит в экспликации требований, предъявляемых соответствующей ему R-отсылкой к сущности, которой та приписана. Оно, во-первых, определяет ограничения, налагаемые на рассматриваемую сущность этой R-отсылкой (в случае структурно-контекстных отсылок это ограничения, связанные с выделяемыми этими отсылками фрагментами ЛС-контекста), и, во-вторых, при невыполнении требуемых ограничений задает операции над данной сущностью, служащие для устранения выявленного конфликта. Сущность-носитель R-отсылки выступает при этом как потенциальный объект преобразований – *формируемая* лингвистическая сущность $S_{\text{форм}}$.

На основе данного определения в системе RussLan, помимо DGF-отсылок, введен также ряд R-отсылок других типов, отличающихся вводящими их R-функциями и задающих другие лингвистические отношения, синтагматические или парадигматические. В отличие от DGF-отсылок, они не могут содержаться во входных ЛС-запросах, но вводятся в процессе их обработки.

При обработке R-отсылки любого типа для нее ищется такое из R-правил, задаваемых ее R-функцией, которое совместимо с другими атрибутами формируемой и отсылочной сущностей. Если такое правило обнаружено, оно применяется; в любом случае, даже если ни одного применимого R-правила не нашлось, рассмотренная R-отсылка устраняется из содержавшего ее ЛС-запроса, так что тот постепенно "очищается" от всех реляционных атрибутов и в конечном итоге сводится к сочетанию лексемы с ее словоизменительными характеристиками, определяющему некоторую синтетическую словоформу.

Перечислим все используемые в системе типы R-отсылок в том порядке, в котором они рассматриваются при обработке содержащего их ЛС-запроса.

3.1. Лексико-информационные (ЛИ-) R-отсылки

ЛИ-отсылки служат для определения языковой (в том числе семантико-синтаксической) информации к опорной лексеме ЛС-запроса, если та отсутствует в лексическом словаре системы (например, является окказионализмом), по ее (предположительному) словообразовательному составу. В настоящее время они используются только для префиксации: если отсутствующая в словаре лексема L_0 может быть представлена как сочетание префикса Пф с бесприставочной лексемой $L_{\text{бп}}$, причем Пф имеет собственную словарную статью в грамматике начальных графико-морфологических (ГМ-) элементов лексем, а $L_{\text{бп}}$ либо найдена в лексическом словаре, либо получила свою статью по грамматике концевых ГМ-элементов [2], лексеме $L_{\text{бп}}$ приписывается ЛИ-отсылка к статье префикса, вводимая R-функцией Pf:

$\langle L_{\text{бп}} \rangle R Pf[\langle \text{статья Пф} \rangle]$.

При обработке такой R-отсылки в статье префикса (напр., *вс* – см. п. 4.14 раздела 4) ищется R-правило, совместимое со свойствами лексемы $L_{\text{бп}}$, и ее статья преобразуется по данному R-правилу в окказиональную статью, сопоставляемую исходной лексеме L_0 .

3.2. Трансформационные (ТФ-) R-отсылки

ТФ-отсылки имеют вид:

$T\langle \text{тип ТФ} \rangle[\langle \text{часть речи исходной сущности} \rangle]$.

Они формируются в процессе синтеза при наличии в ЛС-запросах требующих этого П-атрибутов (в частности, словообразовательных). Такие атрибуты могут иметься уже во входном ЛС-запросе или быть введены на предыдущем шаге синтеза – например, при обнаружении несоответствия семантических или формальных свойств одной из двух сущностей, связанных D- или G-отсылкой, требованиям, предъявляемым к ней второй

сущностью. При появлении в ЛС-запросе ТФ-отсылки она обрабатывается прежде любых других (исключая только ЛИ-отсылки).

Обработка ТФ-отсылок опирается на понятие *локальных групп* обрабатываемых ЛС-сущностей. Это линейно связные именные или предложные группы, не содержащие придаточных и постпозитивных определений, группы глаголов (опять-таки линейно связные), не включающие актантов, и союзы. Границы таких групп либо задаются служебными П-атрибутами их начальных и конечных ЛС-сущностей, либо определяются по умолчанию. Если для ТФ-отсылки найдено отвечающее ей R-правило, при его выполнении может меняться не только сам текущий ЛС-запрос, но и его локальная группа: в нее могут вводиться ЛС-запросы, определяющие новые ЛС-сущности, удаляться какие-либо из ранее имевшихся ЛС-сущностей и т.п. Поэтому по выполнении любого ТФ-правила вся локальная группа данной ЛС-сущности начинает обрабатываться заново.

Примеры, в обработке которых участвуют ТФ-отсылки, приведены в разделе 4 (см. пп. 4.8-4.11).

3.3. DGF-отсылки

DGF-отсылки могут содержаться в ЛС-запросе изначально, появляться при ТФ-обработке или вводиться по умолчанию. В системе RussLan по умолчанию вводятся, например, R-отсылки от предлога к припредложному существительному и от него к предлогу, а также от прилагательного к определяемому им имени и обратно. В качестве припредложного/определяемого имени при этом берется такое существительное справа от предлога/прилагательного, что в промежутке между ними либо нет границ локальных групп, либо выделяемые такими границами группы входят в этот промежуток целиком. По умолчанию может устанавливаться и связь сказуемого с подлежащим: для личного глагола, не имеющего R-отсылки к подлежащему, таковым считается ближайшее к нему слева существительное в номинативе.

Среди DGF-отсылок, являющихся атрибутами одной и той же сущности, вначале обрабатываются F-отсылки, затем G-отсылки и последней – D-отсылка (по определению дерева, не более одной). Если DGF-отсылка содержит вложенные R-отсылки, ее обработка идет "изнутри вовне": от наиболее глубоко вложенного ЛС-запроса, не содержащего собственных R-отсылок, к ЛС-запросам последовательно меньшей степени вложенности – т.е. от элементов ЛС-контекста, максимально удаленных в структуре данного его фрагмента от самой формируемой ЛС-сущности, ко все более близким к ней единицам. R-правила ищутся при этом только для R-отсылок, не имеющих вложенных отсылок. Содержательно такие R-правила задают прежде всего формальные модели управления ЛС-сущностей (лексические или грамматические), но могут отсылать также к семантическим, лексико-функциональным, трансформационным и другим сведениям об этих ЛС-сущностях.

3.4. Лексико-морфологические (Мф-) R-отсылки

Мф-отсылка служит для верификации определения формируемой словоформы $C_{\text{форм}}$ по свойствам ее лексемы и П-атрибутов и вводится в ЛС-запрос, когда он не содержит ни других R-отсылок, ни требующих их признаков. Задается она R-функцией Mf и отсылает сущность $C_{\text{форм}}$ к самой себе, так что получаемый ЛС-запрос имеет вид:

$\langle C_{\text{форм}} \rangle \text{Mf}[\langle C_{\text{форм}} \rangle]$

Если имеющееся определение словоформы не отвечает требованиям полноты, непротиворечивости и синтетической реализуемости, и не может быть скорректировано прямо при обработке Мф-отсылки, в ЛС-запрос вводится нужный тип ТФ-отсылки. Прямую Мф-коррекцию иллюстрирует пример в п. 4.12 раздела 4, Мф-обработку с вводом ТФ-отсылок – в частности, примеры из пп. 4.10-4.11.

3.5. Морфолого-линейные (Мл-) R-отсылки

Мл-отсылки уточняют лексико-морфологическую часть ЛС-запроса с учетом его линейного контекста в именных группах с нумеративами (квантификаторами и числительными). При первичной обработке таких групп каждой участвующей в них ЛС-сущности приписываются служебные П-атрибуты, задающие тип имеющихся в группе нумеративов и их позицию относительно данной сущности. Затем группа проходит заново с построением и обработкой для каждого ее участника Мл-отсылки к нему самому:

$\langle C_{\text{форм}} \rangle \text{Ml}[\langle C_{\text{форм}} \rangle]$

Это позволяет, в частности, по-разному оформлять прилагательные, стоящие в таких группах, *resp.*, слева и справа от нумеративов (см. п. 4.13 раздела 4).

4. Лингвистические функции R-отсылки, реализованные в системе RussLan

Проиллюстрируем ряд лингвистических явлений и операций, реализуемых в системе RussLan средствами аппарата R-отсылок и R-правил, на примерах конкретных ЛС-заданий и результатов их обработки. Строго говоря, на выходе Синт-синтеза строятся только ЛМ-определения словоформ, но здесь мы для наглядности приводим уже сами соответствующие словоформы. Входные и промежуточные данные даны в примерах прямым шрифтом и отделены друг от друга и от выходных единиц стрелкой \Rightarrow . Выходные единицы представлены курсивом, элементы их ЛС-контекста, для которых во входных представлениях нет своих ЛС-запросов, даются при выходных единицах в угловых скобках (прямым шрифтом).

4.1) Уточнение номера зависимости по ее семантической интерпретации СИ в ЛС-запросе.

ребенок мн R D1[<<экспериментер>> нравиться R

G2[<<объект>> сказка мн]] \Rightarrow

ребенок мн R D2[<<экспериментер>> нравиться R
G1[<<объект>> сказка мн]] ⇒
детям <нравятся сказки>

При сохранении для ЛЕ *ребенок* роли подлежащего D1 глагола *нравиться*, а для ЛЕ *сказка* роли его дополнения G2 синтез дал бы ошибочный результат: **детям* <нравятся сказкам>. Мена D1 на D2 и G2 на G1 обеспечивается за счет сопоставления их контекстных СИ (resp., <<экспериментер>> и <<объект>>) с семантическими интерпретациями валентностей глагола *нравиться* в его словарной статье.

4.2) Влияние "хозяев" на оформление "слуг".

на
комиссия R D2[зависеть R G1[оценка]]
⇒ <оценка зависит> *от комиссии*

Предлог *на* мог появиться во входном ЛС-задании, например, в результате грубого (пословного) перевода англ. *depend on*.

4.3) Влияние "слуг" на оформление "хозяев".

убедить прош R Gm[безуспешно] G1[он ж-род]
⇒ <она безуспешно> *убеждала*

4.4) Влияние позиции "слуги" на его оформление.

приехать NN{1_4} прош RD2[знать NN{1_2} R
G1[я NN{1_1}]] G1[он NN{1_3}].

⇒ <я знаю> *что* <он> *приехал*

приехать NN{2_2} прош RD2[знать NN{2_4} RG1[я NN{2_2}]] G1[он NN{2_1}].

⇒ *то, что* <он> *приехал* <я знаю>

Атрибут вида NN{1_j} в определении ЛС-сущности задает ее исходную позицию в тексте: i – это порядковый номер содержащего ее предложения, j – ее собственный номер в этом предложении.

4.5) Влияние дистантного ЛС-контекста.

трудный RD3[считать RG1[я мн] G2[задача]]
⇒ <мы считали задачу> *трудной*

слава R D2[щадить R D2[мочь прош R Gm[не]]]

⇒ <Не мог щадить он нашей> *славы*

4.6) Явления лексической обусловленности.

начать прош doqq RG1[зал] G2[аплодисменты]
сильный doqq

аплодисменты RD2[начать прош doqq R
G1[зал] RG2[аплодисменты]]

⇒ <Зал> *разразился бурными аплодисментами*

Здесь, как и в примерах пп. 4.7-4.10, а также в примере с лексемой *молоко* из п. 4-15, при синтезе привлекаются, помимо прочих, сведения о релевантных для обрабатываемых ЛЕ лексических функциях [4].

4.7) "Двусторонняя" лексическая обусловленность.

фотография R D3[узнать R G1[дед] G2[[внук]]

⇒ <дед узнал внука> *по фотографии*

фотография R D3[смотреть R G1[лицо] G2[я]]

⇒ <это лицо смотрело на меня> *с фотографии*

Ср.: *узнал из разговора; смотрел из окна.*

Заметим, что в принципе глагол *узнать* может сочетаться и с предлогом *на*, ср.: *узнать на фотографии своего коллегу*. Но по приведенному здесь ЛС-запросу с этим глаголом не может быть синтезирована группа *на фотографии*, так как при глаголе *узнать* она является сирконстантом, а в данном ЛС-запросе формируемая единица определена как его актанта D3.

4.8) Снятие конфликта между семантическими требованиями к актанта и его семантикой.

статья RD1[изучать RG2[вопрос]]

изучать RG1[статья] G2[вопрос]

⇒ *Автор статьи изучает* <вопрос>

статья RD1[изучать RG2[вопрос]] G1[аспирант]

изучать RG1[статья RG1[аспирант]] G2[вопрос]
 ⇒ В статье <аспиранта> изучается <вопрос>

4.9) Морфолого-синтаксические перифразы.
 найти: прич.буд.3л.мн ⇒ *которые найдут*
 победить буд R G1[я] G2[соперник]
 ⇒ <я> одержу победу <над соперником>

4.10) Развертывание фразеологии.
 съесть_собаку прош R G1[мы] G2[это]
 ⇒ <мы на этом> *собаку съели*

4.11) Аналитическое формообразование.
 продумать пасс.прош.ж-род ⇒ *была продумана*

4.12) Коррекция морфологических характеристик.
 разостлать личн.изъяв.3л.ед ⇒
 разостлать нсв.наст.личн.изъяв.3л.ед ⇒
 расстилать наст.личн.изъяв.3л.ед ⇒ *расстиляет*

4.13) Влияние числительных на своих соседей.
 этот
 один
 тысяча
 пятьсот
 льготный
 билет
 ⇒ эти одна тысяча пятьсот льготных билетов

4.14) Оформление актантов "незнакомых" лексем.
 спина R D3[вскурдячить прош R
 G1[куздра] G2[мешок]]
 ⇒ <глокая куздра вскурдячила мешок> *на спину*

4.15) Синтаксическая обработка по умолчанию.
 многочисленный
 нерелевантный
 для
 окончательный
 результат
 деталь
 мы
 в
 данный
 отчет
 опустить ⇒ многочисленные нерелевантные для окончательного результата детали мы в данном отчете опускаем
 каждый
 жирный
 молоко ⇒ *каждый вид жирного молока*

В силу ограничений объема, мы не имеем возможности рассмотреть состав и форматы R-правил, обеспечивающих рассматриваемые виды обработки, как и типы, форматы и организацию других учитываемых в ней лингвистических данных (частично это описано в [4], [5]). Отметим также, что ввиду экспериментального характера системы RussLan приведенные примеры являются прежде всего иллюстративными: демонстрируются возможности описываемых механизмов, а не те или иные лингвистические решения. Что касается лингвистического обеспечения текущей версии системы, то поскольку словарная и грамматическая информация вводится в его состав в основном в рамках отладки механизмов и средств синтеза, различные аспекты русского синтаксиса отражены в нем весьма неравно. Наиболее полно описаны числительные и предлоги, а также некоторые морфолого-синтаксические и формально-синтаксические умолчания (в частности, оформление первых

двух глагольных валентностей). Формальные и семантико-синтаксические модели управления разработаны примерно для трех тысяч лексем, но большей частью еще требуют отладки. Соответственно, не приведены к окончательному виду списки используемых синтаксических и семантических признаков, семантических ролей, лексических функций и т.д. Важно, однако, что описываемый здесь аппарат R-отсылок и R-правил не налагает каких-либо ограничений на их количество и конкретный вид, но опирается только на состав и формат полей, которые предусмотрены для их записи в статьях лингвистических сущностей (и инвентарь которых является общим для всех отделов словаря и грамматики).

5. Заключение

Как видно из предыдущего изложения, аппарат R-отсылок и R-правил представляет собой многофункциональное формальное средство, позволяющее решать широкий спектр задач синтаксического синтеза. Он не имеет также жесткой привязки к тому или иному конкретному естественному языку, что позволяет говорить о его универсальности.

В то же время реализация данного аппарата в системе RussLan базируется на ряде лингвистических умолчаний (в основном касающихся взаимодействия отношений зависимости и кореферентности с линейной структурой текста), которые свойственны только языкам с определенными типологическими характеристиками. Это повышает эффективность синтезирующих процедур, но ограничивает их универсальность только языками данного класса. Наиболее жесткое из таких ограничений состоит в том, что служебные показатели управления полагаются только препозитивными, так что данная реализация рассматриваемого аппарата неприменима, скажем, к языкам типа японского, где аналогичные показатели постпозитивны.

Ряд других встроенных в систему RussLan механизмов линейно-структурного взаимодействия затрагивают не столько сам факт ее применимости к языкам, типологически отличным от русского, сколько эффективность ее применения к ним. Так, в ней предусмотрены переносы морфологических свойств существительного к предшествующим ему прилагательным, а также возможность лексической модификации прилагательных по предпочтениям следующих за ними существительных, нерелевантные, скажем, для французского языка. Перенос по умолчанию свойств существительного в номинативе к следующему за ним глаголу эффективен для языков типа SVO или SOV, но не VSO, и так далее. Тем не менее класс языков, к которым применима описываемая система, представляется достаточно широким, чтобы рассматривать ее как синтезатор принципиально многоязычного типа.

Список литературы

1. Мельчук И.А. Уровни представления языковых высказываний и общее строение модели "смысл ↔ текст" // *Актуальные проблемы лексикологии и словообразования. Вып. 1. Новосибирск: Новосибирский гос. ун-т, 1972. С. 5-35.*
2. Шаляпина З.М., Канович М.И. Русский морфологический и морфолого-синтаксический синтез по неполной и/или неточной входной информации // *Русский язык как иностранный: лингвистические проблемы. М.: ИКАР, 1997. С. 137-147.*
3. Kanovich M.I., Shalyapina Z.M. The RUMORS system of Russian synthesis // *Proceedings of COLING 94. Kyoto, 1994, I. P. 177-179.*
4. Shalyapina Z. Многофункциональность аппарата лексических функций и некоторые возможности его применения и развития // *East West Encounter: Second International Conference on Meaning↔Text Theory. М.: Языки славянской культуры, 2005. С. 447-457.*
5. Shalyapina Z.M., Kanovich M.I., et al. RUSSLAN: A System of Russian Language Generation // *Investigations into Formal Slavic Linguistics. Contributions of FDSL IV. Frankfurt am Main et al.: Peter Lang Verlag, 2003, Part I. P. 385-403.*