

**РАСПОЗНАВАНИЕ СЕМАНТИКИ ПАДЕЖА ДЛЯ ЦЕЛЕЙ
АВТОМАТИЧЕСКОГО ПЕРЕВОДА С РУССКОГО ЯЗЫКА
НА КИТАЙСКИЙ: ТВОРИТЕЛЬНЫЙ ИНСТРУМЕНТА
VS. ТВОРИТЕЛЬНЫЙ СРАВНЕНИЯ**

**RECOGNITION OF CASE SEMANTICS FOR RUSSIAN-CHINESE
AUTOMATIC TRANSLATION: INSTRUMENTAL OF INSTRUMENT
VS. INSTRUMENTAL OF COMPARISON**

*Сунь Шуан (sunshuang@mail.ru), Кобозева И.М. (kobozeva@list.ru)
Московский государственный университет им. М.В. Ломоносова*

На базе «Онтологической семантики» строятся формальные правила распознавания семантических ролей инструмента и объекта сравнения (по форме и общего), кодируемых в русском языке творительным падежом, необходимые для интерпретации свободных синтаксисом имен артефактов в процессе их перевода на китайский язык.

Одной из серьезных проблем при автоматическом переводе (АП) с русского языка на китайский является распознавание семантики русских падежей. Особую трудность представляет перевод падежей свободных синтаксисом, то есть именных групп (ИГ), не заполняющих валентности каких-либо слов в предложении. Цель нашей работы — показать, что данная проблема может быть решена, если в системе есть семантический компонент, включающий в себя универсальную онтологию «Онтологической семантики» С. Ниренбурга и В. Раскина [5] и настроенный на нее словарь, а также правила приписывания семантических ролей. Принципиальная возможность семантической интерпретации падежа, адекватной целям перевода с русского на китайский, будет продемонстрирована на примере свободных синтаксисом творительного падежа без предлога (далее — ТВОР). При этом в связи с ограничениями объема доклада из всего множества возможных синтаксисом ТВОР рассматриваются только те, которые могут обозначать, в зависимости от контекста, либо роль *инструмента*, либо роль *объекта сравнения*. Разграничение этих значений ТВОР необходимо постольку, поскольку каждое из них переводится на китайский язык своим специальным средством. Естественно, что при формулировании правил приписывания синтаксисом этих значений мы старались учитывать и другие в принципе возможные для них значения, чтобы правильно задать условия применимости правил.

Основным средством для обозначения роли инструмента (Instr) в китайском языке является слово *yòng* (用) в синтаксисом позиции предлога¹; для выражения роли объекта сравнения (Similar) употребляется либо слово *xíng* (形) в синтаксисом позиции послелого²; либо рамочная конструкция, состоящая из предлога *xiàng* (像), с помощью которого выражается сходство или тождество одного предмета с другим и прилагательного *yīyàng* (一样) со значением ‘одинаковый, подобный’³. Выбор одного из двух способов определяется тем, какая характеристика объекта сравнения положена в основу уподобления. Если уподобление проводится по форме объекта, то ТВОР переводится при помощи *xíng* (形), если же по иным характеристикам (например, скорости движения, характеру звучания и т.п.), то используется рамочная конструкция. Следовательно, для целей АП с русского языка на китайский целесообразно разделить роль объекта сравнения на две более конкретные роли — объекта сравнения по форме (Similar-form) и объекта общего сравнения (Similar).

Таким образом, чтобы правильно перевести свободную синтаксисом ТВОР на китайский язык, в нашем случае необходимо не только установить ее общее грамматическое значение в русском языке, но и конкретизировать его. Протестировав представленные в настоящее время в Интернете системы перевода on-line с русского языка на китайский, мы убедились, что они допускают в этом деле грубые ошибки. Например, одна из систем (<http://www.russsky.net>) предлагает для предложения (1a) перевод (1б):

¹ В синтаксисом позиции предиката *yòng* (用) имеет значение ‘использовать’. Кроме этого употребляется и предлог *ná* (拿) с исходным предикатным значением ‘брать, поднимать’, но сфера его применения уже.

² В синтаксисом позиции аргумента *xíng* (形) имеет значение ‘форма’

³ Также для выражения сходства используется и конструкция *xiàng...side* (像...似的), *xiàng...yībān* (像...一般) и т.д, но не в тех типах контекстов, которые рассматриваются в данной работе.

- | | | | | | | | |
|-----|----|--------------|----|------------|----|-------------|----------|
| (1) | a. | Молотком | и | топором | | выправил | вмятину. |
| | b. | 由 锤子 | 和 | 由 轴 | 它 | 调直了 | 凹痕。 |
| | | Yóu chuǐzi | hé | yóuzhóu | tā | tiáozhíle | āohén. |
| | | PREP молоток | и | PREP топор | | он выправил | вмятина |

Мы видим, что синтаксемы ТВОР — *молотком* и *топором*, которые имеют в данном контексте значение инструмента и должны были быть переведены при помощи предлога *yòng* (用), были проинтерпретированы неверно, о чем свидетельствует выбор предлога *yóu* (由), имеющего несколько локативно-направительных значений. Причиной данной ошибки явилась неспособность системы выбрать нужное значение ТВОР в зависимости от семантики существительного: именам, обозначающим инструменты (*молоток*, *топор*) была приписана не свойственная им локативно-направительная роль. (Одну из таких ролей ТВОР может выражать, но при именах, относящихся к семантическому классу мест и некоторым другим, ср. *лесом*, *рекой*, *дорогой*, *трубопроводом* т.п.). В (1б) есть и другие ошибки, но они уже не имеют отношения к нашей теме. Очевидно, что, чтобы избежать ошибок при переводе свободных синтаксем ТВОР, система АП должна уметь распознавать их семантику.

Семантика свободных падежных синтаксем, то есть семантическая роль денотата имени в ситуации, обозначаемой предикатом клаузы, не определяется однозначно ни этим предикатом, ни каким-либо из его актантов, поскольку эти синтаксемы не входят в модели управления (МУ), которые должны быть сопоставлены именам и предикатам в словаре системы АП. Поэтому для установления того, какую семантическую роль выражает ТВОР в таких случаях, необходимы специальные правила, учитывающие семантические и формальные свойства как самого имени в форме ТВОР, так и слов, с которыми данная синтаксема синтаксически связана в предложении. Такие правила должны стать частью семантического компонента системы, обеспечивающего семантико-синтаксический анализ входного текста в необходимом для качественного перевода объеме. Предполагается, что семантико-синтаксическому анализу предложения предшествуют этапы морфологического и частичного поверхностно-синтаксического анализа, и начинается он с приписывания словам семантической информации, извлекаемой из словаря.

Необходимый для создания таких правил инвентарь семантических ролей, выполняемых свободными синтаксемами ТВОР, может быть извлечен из обобщения разных классификации значений ТВОР, существующих в русистике. Условия, при которых эта форма выражает те или иные роли, также описаны (см. [2]). При формулировании таких условий решающую роль играют семантические свойства лексемы, маркированной ТВОР, и семантические свойства предикатов и других слов, синтаксически связанных с данной лексемой. Однако в [2] эти свойства описаны неформально, без установки на автоматический семантический анализ. Для разработки формальных правил распознавания семантики падежей необходим семантический метаязык, включающий в себя достаточное количество четко определенных семантических единиц и заданных на них семантических отношений. Принципы построения такого языка и способы описания языковой семантики в его терминах изложены в [5]. В настоящее время построенный на основе этих принципов язык включает в себя около 6000 субстанциональных и реляционных концептов из самых разных предметных областей, в терминах которых могут быть выражены как парадигматические свойства слов (например, семантический класс, семантический признак), так и их и синтагматические семантические свойства (например, выражаемые падежами семантические роли). Концепты организованы в онтологию, представляющую собой иерархию, основанную на родо-видовых отношениях, благодаря чему семантическое правило, в котором упоминается концепт определенного уровня в онтологической иерархии, будет применимо ко всем его дочерним концептам вплоть до концептов терминального уровня. При необходимости онтология может пополняться новыми концептами, которые должны получить четкую неформальную дефиницию и формальное описание в терминах уже существующих концептов. Ниже при формулировании правил мы будем использовать именно этот метаязык.

В свете поставленной задачи наибольший интерес представляют те случаи, в которых имя в форме ТВОР (напомним, что речь идет только о свободных синтаксемах) в принципе способно выступать во всех трех вышеупомянутых семантических ролях. В русском языке существуют слова, которые могут выступать и в роли инструмента, ср. *Колоколом* <Instr> *жрец созывает жителей*, и в роли объекта сравнения по форме, ср. *Свод пещеры расходился колоколом*<Similar-form>⁴, и в роли объекта общего сравнения (то есть сравнения по каким-либо другим параметрам, напр., по звучанию), ср. *Его голос гремел колоколом*<Similar>. Мы сочли целесообразным начать разработку правил распознавания значений ТВОР на материале именно таких слов. При этом мы исходили из предположения, что полученные таким образом правила можно будет экстраполировать на весь класс артефактов путем ослабления семантических ограничений на имя в форме ТВОР в правилах приписывания соответствующих ролей.

⁴Здесь и далее примеры либо взяты (как правило, в сокращенном варианте) из НКРЯ, либо получены путем трансформации нефинитных конструкций из НКРЯ в финитные предложения.

Распознавание семантики надежда для автоматического перевода с русского на китайский

Для начала из Национального корпуса русского языка (НКРЯ) были выбраны все предложения со словом *колокол* в форме ТВОР. Их оказалось всего 209. Мы удалили из этого массива все контексты, в которых она либо употреблялась с управляющим предлогом, либо была связанной синтаксемой (ср. ... *хотят обеспечить знаменитого звонаря-композитора колоколами*), поскольку интерпретация предложных синтаксем — это отдельная задача, а интерпретация связанных синтаксем содержится в МУ слова-хозяина в словаре. После этого осталось 40 примеров на свободные синтаксемы ТВОР, которые и были подвергнуты семантическому анализу.

Устраняя из массива связанные синтаксемы ТВОР, мы обратили внимание на омонимию форм ТВОР у слова *колокол*, синтаксически связанного с глаголами звука (*греметь, звенеть* и т.п.)⁵. В одних случаях ТВОР выражает семантическую роль Источника звука (по [4]), как в примере (2а):

(2а) *Церковь трезвонила всеми колоколами.*

Здесь *колоколами* — косвенное дополнение, актанта предиката *трезвонить* с семантической ролью Источника звука, которая является частным случаем роли Пациенс (поскольку в общем случае объект становится источником звука не сам по себе, а в результате воздействия на него некоего Агенса-Каузатора). То, что в русском языке эта роль в предложениях с Агенси-Каузатором кодируется формой ТВОР, по мнению Е. В. Падучевой, свидетельствует о том, что Источник звука в ситуации, когда Каузатор не совпадает с Источником звука, концептуализируется как Инструмент. В китайском же языке ИГ с ролью Источника звука в аналогичной ситуации занимает позицию прямого дополнения глагола звука, как показывает перевод примера (2а) — (2б), а значит, концептуализируется как объект, подвергающийся воздействию — Пациенс:

(2б) 教堂 敲响 了 所有的 钟。
 Jiàotáng qiāoxiǎng-le suǒyǒude zhōng.
 Церковь трезвонить-PERF все колокол⁶

В других случаях форма ТВОР при глаголах, которые употреблены либо в прямом значении звука (репрезентируют концепт EMIT-SOUND), либо в переносном значении ощущения (репрезентируют концепт INVOLUNTARY-PERCEPTUAL-EVENT), является обстоятельством образа действия и кодирует роль Объекта сравнения Similar, как в (3), где глагол *гудеть* выступает в переносном значении:

(3) *Голова гудела колоколом.*

В развернутом виде смысл (3) можно представить в виде перифразы 'В голове говорящего локализовалось ощущение, подобное ощущению звука гудящего колокола'. Соответственно при переводе (3) на китайский язык используется конструкция *xiàng ... yīyàng* (像... 一样):

(3') 头 像 钟 一样 响。
 Tóu xiàng zhōng yīyàng xiǎng.
 голова подобно колокол одинаковый гудеть

Для разрешения подобной синтаксической (а часто также и сопровождающей ее лексической) омонимии необходимо, чтобы в словаре системы АП, во-первых, при ЛСВ существительных были указаны соответствующие им онтологические концепты, они же семантические классы (например, *церковь-1* — CHURCH, *церковь-2* — CHURCH-BUILDING), а во-вторых, глаголы были представлены всеми возможными для них МУ с указанием семантических ограничений по валентностям. Тогда при анализе (2) для глагола звука *трезвонить* будет выбран на МУ с диатезой типа процесс-каузация (4):

(4) *трезвонить-1* EMIT-SOUND

МУ 1

Синт. актанта	1 — ИГ им	2 — (ИГ тв)
Сем. роль	1 — Agent	2 — Patient-Source
Сем. ограничение	1 — [HUMAN, ORGANIZANION]	2 — [DEVICE, MUSICAL- INSTRUMENT]

Действительно, из двух ЛСВ вокабулы *церковь* в контексте данного глагола может быть выбран только вариант *церковь-1*, который относится к семантическому классу CHURCH (/RELIGIOUS-ORGANIZATION/ORGANIZATION)⁷, а *колокол* через сопоставленный ему концепт BELL входит в классы

⁵ Естественно, что такая омонимия свойственна не только данному слову, но всем словам, относящимся к классу артефактов, предназначенных для извлечения звуков — *труба, бубен, гудок* и т. п., если они способны выражать роль Источника звука в ситуации звучания формой ТВОР (ср. *бренчать бубенчиком*, но *бренчать на гитаре*).

⁶ Такой же перевод будет иметь и предложение *Церковь трезвонила во все колокола*, где роль Источника звука кодируется конструкцией «в + вин. пад.», выбор которой отражает концептуализацию этого участника ситуации как Конечной точки удара, при помощи которого Каузатор вызывает его звучание, ср. *бить в барабан, стучать в окно* и т.п.

⁷ Через знак «/» мы показываем путь, связывающий данный концепт с иерархически вышележащими, родовыми для него концептами.

устройств — BELL/DEVICE и музыкальных инструментов — BELL/MUSICAL-INSTRUMENT Соответственно синтаксема ТВОР будет распознана уже на этом шаге анализа как имеющая роль Patient. Та же МУ представлена в примере (5):

(5) *Я трезвонил ипорой.*

При анализе (3) для глагола *гудеть* будет выбрана МУ (6)⁸:

(6) *гудеть*-n INVOLUNTARY-PERCEPTUAL-EVENT

Синт. актант	1 — ИГ им	2 — (от ИГ род)
Сем. роль	1 — Location	2 — Cause
Сем. ограничение	1 — ANIMAL-PART (not (MOUTH, LARYNX))	2 — [EVENT]

Синт. актант 1 — ИГ им 2 — (от ИГ род) Сем. роль 1 — Location 2 — Cause Сем. ограничение 1 — ANIMAL-PART (not (MOUTH, LARYNX))⁹ 2 — [EVENT] Поскольку значение синтаксемы ТВОР на этом шаге анализа не будет распознано, она поступит на обработку в блок правил для анализа свободных синтаксем, где ей должна быть приписана роль объекта сравнения Similar.

Ниже мы представим те правила, которые были первоначально сформулированы на базе 40 примеров со словом *колокол*, а затем проверены на материале еще 680 примеров из НКРЯ с существительными ТВОР из класса артефактов. Правила имеют следующий вид. В левой части правила задается исходный шаблон, содержащей синтаксему ТВОР с незаполненной семантической ролью — <?>, а в правой части правила приводится та же синтаксема с заполненной ролью. Поясним принятые нами нотационные конвенции на примере одного из правил приписывания роли Similar-form, обозначаемого ярлыком R1-Similar-form:

- **R1-Similar-form**: V: CHANGE-POSITION + [_{NP} [N_{ins,sg}: GEOMETRICAL-OBJECT, PHYSICAL-OBJECT {FORM-TEMPLATE}]]_{NP} <?> / → NP_{ins,sg} <Similar-form>

При записи правила помимо обычных конвенций, принятых в грамматике НС, при именных и глагольных синтаксемах указывается вся необходимая информация. Во-первых, это морфологическая информация в виде ярлыков, используемых для этой цели в НКРЯ, записываемая при лексической категории как ее нижний индекс (отсутствие морфологического ярлыка означает, что соответствующий морфологический признак не играет роли). В **R1-Similar-form** это творительный падеж (ins) и единственное число (sg). Во-вторых, это семантический класс лексической категории, указываемый при ней после двоеточия и обозначаемый именем онтологического концепта (если имя может относиться к нескольким семантическим классам, то они перечисляются через запятую). В **R1-Similar-form** для N это классы геометрических объектов — GEOMETRICAL-OBJECT (*круг*, *крест* и т. п.) и физических объектов — PHYSICAL-OBJECT, а для V — класс пребывания в позиции (*сидеть*, *торчать* и т. п.) или изменения позиции (*садиться*, *вытягиваться*, *поворачивать* и т. п.) — CHANGE-POSITION¹⁰. При имени семантического класса в фигурных скобках может быть указан нетривиальный семантический признак слова [1]. В **R1-Similar-form** это признак «эталон формы» — FORM-TEMPLATE, свойственный группе слов из разных классов онтологии, напр., словам *колокол* (класс ARTIFACT), *груша* (FOODSTAFF), *змея* (ANIMAL) и многим др. [3]. Нетривиальные семантические признаки — еще один вид семантической информации, приписываемый слову в словаре системы. Слово *колокол* имеет как минимум два нетривиальных семантических признака: «эталон формы» и «эталон звука» — SOUND-TEMPLATE. Наконец, после парадигматических семантических признаков имени указывается его семантическая роль в угловых скобках, обозначаемая при помощи имен из общепринятого инвентаря семантических отношений с некоторыми необходимыми добавлениями — в **R1-Similar-form** это роль объекта сравнения по форме — Similar-form. Правило позволяет приписать свободной синтаксеме ТВОР, принадлежащей к семантическому классу геометрических объектов или физических объектов, помеченных в словаре признаком FORM-TEMPLATE, роль Similar-form, если эта синтаксема непосредственно связана с глаголом семантического класса CHANGE-POSITION.

По этому правилу устанавливается значение граммы ТВОР (ins) в 11 случаях из нашей выборки, например:

(7) а) *Плащ ниспадал колоколом почти до земли.* б) *Ветерок забрался по ее ногам под юбку и надул платище*

⁸ Та же МУ представлена в примерах *Гудели ноги; ...удовлетворенно гудят руки...* и т.п.

⁹ Данное семантическое ограничение исключит выбор данной МУ (и соответствующего ей ЛСВ гудеть) при синтактико-семантическом анализе примеров типа Все эти отверстия *рты* звенят и *гудят*, а я мысленно затыкаю их огромными кусками хлеба, где гудеть репрезентирует концепт EMIT-SOUND и имеет одновалентную модель с первым и единственным актантом в роли Пациенса-Источника звука.

¹⁰ В «Онтологической семантике» значения глаголов изменения позиции и соответствующих им глаголов пребывания в позиции (повернуть – повернуться, сажать-сидеть, вставать – стоять и т. п.) относятся к одному онтологическому концепту CHANGE-POSITION (во всех его видах).

Распознавание семантики падежа для автоматического перевода с русского на китайский

колоколом. в) *Кургузая шинель топорщилась колоколом.* г) *Лазурь колоколом стояла над его тихо работающим станком.* д) *Там стены уже напрочь расхотелись колоколом.*

Следующее правило приписывает ТВОР то же самое значение Similar-form, когда эта синтаксема распространяет ИГ, выступая компонентом имени, относящегося к семантическому классу «одежда» (CLOTHING-ARTIFACT):

- **R2-Similar-form** $[_{NP} N^1: CLOTHING-ARTIFACT \ [_{NP} N^2_{ins,sg}: GEOMETRICAL-OBJECT, PHYSICAL-OBJECT\{FORM-TEMPLATE}\}_{NP}]_{NP} <?> \rightarrow NP_{ins,sg} <Similar-form>$

По этому правилу устанавливается значение Similar-form для граммы ТВОР в 8 случаях из нашей выборки, например:

(8) а) *Вот он позирует перед кинокамерой в отражающем солнце белом мундире колоколом.* б) *Юбка колоколом выше колен.* в) *Народ ... сменил старые шубейки на пальто колоколом.* г) *Маленькая женщина в шубке колоколом ... радостно, но не громко вскричала ...*

Следующие три правила приписывают роль объекта общего сравнения — Similar (в нашей выборке представлена 12 случаями):

- **R1-Similar** $([_{NP} N^1_{nom}: NOT ANIMAL]_{NP}) + V: EMIT-SOUND +$
 $+_{NP}[N^2_{ins}: PHYSICAL-OBJECT\{SOUND-TEMPLATE}\}_{NP} <?> \rightarrow$
 $\rightarrow NP_{ins} <Similar>$
- **R2-Similar** $([_{NP} N^1_{nom}: HUMAN]_{NP}) + (V: SPEECH-ACT) + [_{NP} N^2_{ins}: PHYSICAL-OBJECT\{SOUND-TEMPLATE}\}_{NP} <?> +$ “DIRECT-SPEECH” $\rightarrow NP_{ins} <Similar>$
- **R3-Similar** $[_{NP} N^1_{nom}: ANIMAL-PART, ANIMAL-SYMPOM]_{NP} + V: INVOLUNTARY-PERCEPTUAL-EVENT +$
 $[_{NP} N^2_{ins}: PHYSICAL-OBJECT\{SOUND-TEMPLATE}\}_{NP} <?> \rightarrow NP_{ins} <Similar>$

По правилу **R1-Similar** семантика ТВОР распознается в примерах типа (9):

(9) а) ... из Кремля *разнеслось набатным колоколом*: «К вам обращаюсь я, друзья мои!». б) *Имя-то какое! - так и гудит колоколами древних русских сторожевых монастырей!* в) *И так стукнул в сердцах дверью, что весь дом загудел колоколом - бом!* г) *Она подталкивала каталку, и половник колоколом гремел в кастрюле.* д) *Но самое главное, многие иерархи не готовы к реформе церкви, хотя нужда в ней колоколами гремит над землей России.* е) ... в его голове *колоколом ударил*¹¹ *голос Каббаты...*

Правило **R2-Similar** распознает семантику ТВОР в примерах типа (10):

(10) а) - *Вспомним всех поименно! - колоколом загудел князь.* б) ... *извозчик наклонил лохматую голову и прогудел - колоколом из подземелья: - Премного благодарны.* в) И тут *колоколом* в голове: «**Шпионы**».

R3-Similar служит для интерпретации ТВОР в примерах типа (11):

(11) а) *Голова гудела колоколом...*; б) ... *у меня таки башка колоколом звенит.*

Напомним, что эти правила применяются уже после того, как с помощью МУ глаголов установлены семантические роли всех их актантов (связанных синтаксем), в том числе и тех, которые маркируются творительным падежом. Поэтому правило **R1-Similar** не приведет к ошибке в контекстах типа (2а), где NP_{ins} выступает в роли Пациенса-Источника звука. При этом приходится признать, что для некоторых глаголов невозможно в локальном контексте клаузы однозначно установить значение ТВОР. Так, в предложениях типа (12) без дополнительных знаний об описываемой ситуации невозможно снять омонимию Patient-Source / Similar:

(12) *Шторм тревожно гудел огромным колоколом.*

С одной стороны, здесь применима МУ *гудеть 1*

Синт. актант	1 — ИГ им	2 — (ИГ тв)
Сем. роль	1 — Agent	2 — Patient-Source
Сем. ограничение	1 — CURRENT, WEATHER	2 — INANIMATE

выбор которой обуславливает интерпретацию колокола как реального объекта, на который воздействует шторм, вызывая его звучание. Ср. однозначный пример с той же МУ (13):

(13) *Ветер гудит рваным листовым железом* <Patient-Source>

С другой стороны, может быть выбрана и МУ *гудеть 2*

Синт. актант	1 — ИГ им
Сем. роль	1 — Agent-Source
Сем. ограничение	1 — [CURRENT, WEATHER, INANIMATE]

и тогда NP_{ins} из примера (16) поступит в блок обработки свободных синтаксем, где будет проинтерпретирована

¹¹ В контексте N_{nom} классов MUSICAL- INSTRUMENT, TIMEPIECE глаголы, в своем прямом значении относящиеся к классам HIT или BEAT, выступают в переносном значении EMIT-SOUND., ср. *Барабаны ударили с двух сторон; Бьют часы на Спасской башне.* Можно признать такой же семантический переход и в контексте N_{nom} класса HUMAN-VOICE, как в данном примере.

как объект сравнения (Similar) по правилу **R1-Similar**. Заметим, что даже достаточно широкий контекст, в котором (16) приводится в НКРЯ не позволяет понять, какая из интерпретаций имелась в виду автором высказывания.

Ясно, что 3 рассмотренные правила для роли Similar имеют ограниченную сферу применения. Они приписывают эту роль синтаксеме ТВОР только в контексте глаголов звука, речи и неконтролируемого восприятия и только если лексема в форме ТВОР помечена в словаре признаком {SOUND-TEMPLATE}. Для приписывания той же роли в контексте других глаголов понадобятся аналогичные правила, в которых будут фигурировать глаголы других онтологических классов, а имя в ТВОР будет помечено в словаре другим нетривиальным признаком из той же «эталонной» серии (X-TEMPLATE), например, правило R4-Similar:

- **R4-Similar:** $(NP^1_{nom}) + V: MOTION-EVENT + NP[N^2_{ins}: PHYSICAL-OBJECT\{SPEED-TEMPLATE\}]_{NP} <?> / \rightarrow NP_{ins} <SIMILAR>$

для интерпретации ТВОР в примерах типа (14):

(14) а) ... она... **стрелой выскочила** из комнаты; б) Мяч **пулей влетел** в ворота; в) **Молнией кинулась** откуда-то взявшаяся бабочка.

Наконец, роль Instr приписывается свободной синтаксеме ТВОР по правилу R-Instr:

- **R-Instr:** $([NP N^1_{nom}: HUMAN, ORGANIZATION, DEVICE]_{NP}) + V + NP^2 <PATIENT, THEME> + [NP [N^3_{ins}] <?> / \rightarrow NP^3 <Instr>$

Это правило должно применяться после правил обработки свободных синтаксем ТВОР, рассмотренных выше, в соответствии с общим принципом применения правил в порядке от более частных к более общим, поскольку оно, в отличие от правил приписывания Similar и Similar-form, не ограничивает ни семантический класс имени в ТВОР, ни семантический класс глагола, требуя только присутствия при нем ИГ с ролью Пациенса или Темы (то есть, фактически, его переходности). По этому правилу роль Instr будет приписана свободной синтаксеме ТВОР в примерах типа (15), которых в нашей выборке встретилось 7:

(15) а) ... она **разбудит меня колоколом** громкого боя ... б) **Колоколом жрец созывает жителей**. в) *А я вам, сударыня, говорю, что вы **меня этими колоколами выгоните** из дому.* г) **Пробили водяную тревогу** - протяжными гудками и колоколом.

Полученные на материале слова *колокол* правила распознавания ТВОР были затем проверены на других словах класса артефактов, способных иметь все три рассматриваемые семантические роли или хотя бы две из них. Выбрав 200 первых документов со словом *труба* в ТВОР, мы убедились, что для него во всех его конкретных значениях действуют уже рассмотренные правила распознавания, но только, как и следовало ожидать, семантические ограничения на N^1 в правиле R2-Similar-form потребовалось ослабить, допустив отнесение ядерного имени к классу частей тела. В уточненном виде:

- **R2'-Similar-form:** $[NP N^1: ANIMAL-PART, CLOTHING-ARTIFACT [NP N^2_{ins}: GEOMETRICAL-OBJECT, PHYSICAL-OBJECT\{FORM-TEMPLATE\}]_{NP}]_{NP} <?> \rightarrow NP_{ins} <SIMILAR-FORM>$

это правило сможет приписать синтаксеме ТВОР роль объекта сравнения уже не только в случаях типа (8), но и в случаях типа (16):

(16) а) **Нос трубой**, тела крепенькие, как брюква, мозги жидкие, как вокзальные щипцы. б) *Без седла, китель растенулся, ..., а у лошади **хвост трубой**.*

а также во многих других случаях с эталонами формы из класса артефактов или других семантических классов, ср. *губки бантиком, грудь колесом, нос грушей* и т.п.

Анализ первых 100 примеров со словом *серп* ТВОР, подтвердив эффективность уже введенных правил, потребовал ввести еще одно правило приписывания роли Similar:

- **R5-Similar** $[NP N^1_{nom}: NOT ANIMAL]_{NP} + V: EMIT-LIGHT + NP[N^2_{ins,sg}: CUTTING-IMPLEMENT]_{NP} <?> \rightarrow NP_{ins} <SIMILAR>$

для случаев, когда *серп* или другие слова класса «режущих инструментов» выступают в роли объекта сравнения (по признаку «свечения»), как в (17):

(17) а) *Вдали лунным **серпом светилась** река;* б) *На меня **блеснул серпом** ее взгляд, полный молодых чувств, дружбы, юмора.;* в) ***Блеснула** сабельным лезвием полоска реки.*

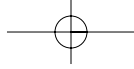
Также только одно дополнительное правило группы Similar дали первые 100 примеров со словом *мешок*:

- **R6-Similar** $NP^1_{nom} + V: CHANGE-POSITION, MOTION-EVENT + NP[N^2_{ins}: BAG]_{NP} <?> / \rightarrow NP_{ins} <SIMILAR>$

ср. примеры типа (18):

(18) а) *Митя **мешком сел** на пол;* б) *Он **мешком болтался** на турнике и не мог выполнить самого простого упражнения;* в) ... *парашюты **мешком падали** вместе с ними.* г) *Форма **сидела** на нем классическим **мешком**.* д) *Это сейчас прошлое **висит** огромным **мешком** за плечами...*

А просмотр всех 72 примеров со словом *куль* ТВОР, подтвердив правило R6-Similar, показал, что необхо-



Распознавание семантики падежа для автоматического перевода с русского на китайский

димо и правило приписывания роли Similar ИГ, распространяющей имя, аналогичное правилу R2'-Similar-form:

- **R7-Similar** [_{NP} N¹: ANIMAL-PART, CLOTHING-ARTIFACT [_{NP} N²_{ins}: BAG]_{NP}]_{NP} <?> → NP_{ins} <SIMILAR>

ср. пример (19):

(19) ... *глазки косые, спины кулем, носы блестят...*

Интересно, что именно через призму китайского языка, в котором ТВОР в примерах типа (18) и (19) не может переводиться при помощи послелога *xíng* (形) ('форма'), видно, что слова *мешок* и *куль* не являются настоящими эталонами формы, хотя по контекстам употребления и близки к таковым. Причина понятна: обозначаемые ими объекты входят в онтологический класс мягких контейнеров (BAG), который характеризуется именно отсутствием постоянной формы.

Анализ примеров с формой ТВОР слов *молоточек* (всех 107), и *лопата* (первых 100), подтвердил применимость рассмотренных выше правил для распознавания семантики свободных синтаксем, не потребовав ни их модификации, ни введения каких-либо дополнительных правил.

Ясно, что не проанализировав аналогичным образом рассмотренные 3 значения свободных синтаксем ТВОР, представленных словами других онтологических типов (животных, растений, светил и т. п.), а также другие значения свободных синтаксем ТВОР, которые встречаются у имен артефактов, мы, конечно, не можем гарантировать, что правила групп R-Similar, R-Similar-form и R-Inst в их настоящем виде обеспечат правильное распознавание соответствующих ролей для всех имен артефактов. Однако не вызывает сомнений, что предлагаемый онтологической семантикой метаязык позволяет без особого труда представить в формальном виде всю семантическую информацию, которая может потребоваться для того, чтобы задать точные ограничения в правилах семантической интерпретации падежных граммем.

После того, как при помощи этих правил значения семантических ролей ИГ в клаузе определено, они могут быть переведены на китайский язык при помощи правил, задающих переводные соответствия семантических ролей:

NP <Instr> = yòng (用) + NP

NP <Similar-form> = NP + xíng (形)

NP <Similar> = xiàng (像) + NP + yīyàng (一样)

Так, при семантическом анализе предложения (3) правило R3-Similar припишет синтаксеме ТВОР роль <Similar>, и в его переводе (3') будет использована рамочная конструкция *xiàng* (像) + NP + *yīyàng* (一样). В предложении (20a) по правилу R2'-Similar-form синтаксеме ТВОР будет приписана роль <Similar-form> и оно с учетом правил порядка слов будет переведено как (20б):

- (20) а. *Девушки носят голубые юбки колоколом.*
 б.

姑娘们	穿着	蓝色的	钟	形	裙。
Gūniángmen	chuāngzhe	lánsède	zhōng	xíng	qún.
Девушки	носитьPres	голубой	колокол	форма	юбка.

В предложении (21a) синтаксеме ТВОР будет приписана та же роль, но уже по правилу (R1-Similar-form):

- (21) а. *Шинель топорщилась колоколом.*
 б.

外套	扎煞成了	钟	形。
Wàitào	zhāshàchéng-le	zhōng	xíng
шинель	топорщиться-Past	колокол	форма.

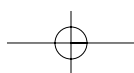
А в предложении (22a), в котором нет условий для приписывания синтаксеме ТВОР ролей объекта сравнения (равно как и ролей маршрута, количества и др., которые мы здесь не рассматривали), ей правилом R-Inst будет приписана роль Instr и в переводе использован предлог *yòng* (用):

- (22) а. *Колоколом жрец созывает жителей.*
 б.

祭司	用	钟	召集	居民。
Jīsī	yòng	zhōng	zhāojí	jūmín.
жрец	Instr	колокол	созывать	житель

Заключение

Повышение качества перевода с русского языка на китайский предполагает решение задачи распознавания семантических ролей ИГ, выражаемых в русском языке падежными показателями. Для решения этой задачи система АП должна иметь словарь, в котором русской лексеме приписаны не только ее лексико-грамматическая категория, МУ с ограничениями по валентностям и переводной эквивалент, но и семантическая информация — онтологический класс и нетривиальные семантические признаки, а также блок правил распознавания семантических ролей свободных синтаксем, формулируемых в терминах морфологических, синтаксических и семантических единиц и отношений. В качестве языка формальной репрезентации семантической информации может



Сунь Шуан, Кобозева И.М.

быть использован язык «Онтологической семантики» С. Ниренбурга и В. Раскина, при необходимости расширяемый за счет введения дополнительных субстанциональных или реляционных концептов или уточнения существующих.

Список литературы

1. Апресян Ю. Д. Типы информации для поверхностно-семантического компонента модели «Смысл <=>Текст» // Избранные труды. М., 1995. Том II.
2. Золотова Г.А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. М.: Эдиториал УРСС, 2001.
3. Кобозева И. М. Как мы описываем пространство, которое видим: форма объектов. // 2000.
4. Падучева Е.В. О парадигме регулярной многозначности (на примере глаголов звука) // НТИ. Серия 2. Выпуск 4. М., 1998.
5. Nirenburg, S., and Raskin V. *Ontological Semantics*. Cambridge, MA: MIT Press, 2004.