

Г. Г. Корбетт и Н. М. Фрэзер
(Университет Суррея, Гилдфорд, Суррей, Великобритания)

Компьютерная лингвистика и типология

Введение

В данной работе мы демонстрируем плодотворность применения формальных методов компьютерной лингвистики при решении типологических проблем. Подобные методы иногда представляются несколько удаленными от занятий типологов. Мы считаем это ошибочным: описания, на которых базируется типология, должны быть точными и адекватными; их нужно рассматривать наряду с возможностями естественных языков.

Для эксплицитного описания данных, в особенности проблематичных с типологической точки зрения, мы используем язык как представление лексических знаний DATR, который далее будет кратко описан (раздел 1)¹. Затем будут обсуждены два типологических вопроса: приписывание родовых характеристик (раздел 2) и синкретизм (раздел 3).

1. Язык DATR

Сначала мы рассмотрим понятие наследования по умолчанию (default inheritance), к которому подойдем, используя таксономическую иерархию, изображенную на рис. 1. Линии на схеме обозначают скорее создание экземпляра (instantiation), чем дальнейшую классификацию (sub-classification). Так, орел является птицей, так же как малиновка и пингвин; Эдвина — это орел, Руперт — малиновка, а Перси — пингвин.

Используя подобного рода иерархию создания экземпляра, наследование по умолчанию делает возможным наследование всех атрибутов данного узла в иерархии (подобно ПТИЦЕ) узлами, которые их инстанцируют (подобно ОРЛУ). Так обычно происходит, за исключением тех случаев, когда нижний узел уже имеет значение для некоторого атрибута и в данном случае значение по умолчанию (т. е., наследуемое) для этого атрибута переопределяется. В нашем примере ПТИЦА имеет оперение и может летать. Эти факты наследуются ОРЛОМ и МАЛИНОВКОЙ и, опосредованно, Эдвином и Рупертом. Атрибут наличия оперения наследуется также ПИНГВИНОМ и через

него — Перси. Однако специфическая информация о том, что ПИНГВИНЫ не могут летать, блокирует наследование более общей информации о ПТИЦАХ. Хотя ПЕРСИ — это ПТИЦА, она не может летать.

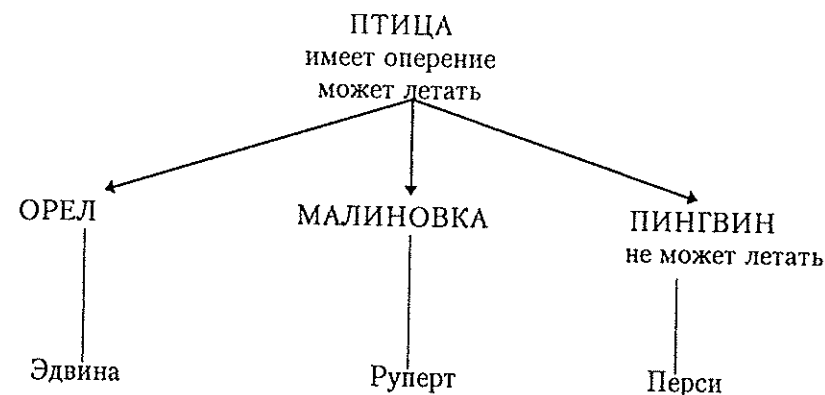


Рис. 1. Простой пример иерархии создания экземпляра

Наследование по умолчанию дает возможность, однажды выразив общие свойства на высоком уровне, затем автоматически использовать их везде, где есть наследование. Правила и исключения могут быть описаны легко и экономично. Данный подход имеет еще и дополнительные преимущества при описании исключений, подобных примеру с ПИНГВИНОМ на рис. 1. Если Перси был бы на самом деле исключительным пингвином, который мог бы летать, эта крайняя исключительность была бы видна, поскольку исключительный факт (ПЕРСИ может летать) переопределял бы исключительный факт (ПИНГВИНЫ не могут летать), переопределяющий наследование по умолчанию (ПТИЦЫ могут летать).

В нашей работе, относящейся к Сетевой Морфологии (Corbett and Fraser, 1993), понятие наследования по умолчанию является фундаментальным. Мы хотим отделить вопросы лингвистической теории (Сетевой Морфологии) от используемого нами формализма (DATR). Это разделение фокусирует внимание скорее на действительном содержании теорий, чем на их нотации, что способствует сопоставительному анализу (Shieber, 1987). Выбор языка DATR для формального выражения Сетевой Морфологии не случаен². Язык представления знаний DATR, разработанный Р. Эвансом и Дж. Газдаром, является основанным на логике формальным языком для описания сетей наследования (Evans and Gazdar, 1989a; 1989b). Так как компьютерная реализация языка DATR доступна, можно проверить, что желаемая классификация реализуется при помощи сети наследования языка DATR.

Информация на рис. 1 может быть выражена на языке DATR следующим образом:

(1) ПТИЦА:

<имеет оперение> == да
<умеет летать> == да

ОРЕЛ:

<> == ПТИЦА

МАЛИНОВКА:

<> == ПТИЦА

ПИНГВИН:

<> == ПТИЦА
<умеет летать> == нет

Эдвина:

<> == ОРЕЛ

Руперт:

<> == МАЛИНОВКА

Перси:

<> == ПИНГВИН

Метки после двоеточия являются 'узлами'; выражения в угловых скобках слева от символа '==' — 'пути доступа'; слова справа от непустых путей доступа являются 'значениями'. Таким образом, значение пути <может летать> для узла ПИНГВИН — 'нет'. Символом '<>' обозначены случаи, когда путь доступа пуст (следовательно, максимально неспецифицирован). Это, например, делает возможным наследование узлом ОРЕЛ всех доступных значений узла ПТИЦА. Так как мы имеем дело с наследованием по умолчанию, ПИНГВИН наследует от ПТИЦЫ все значения, за исключением того, которое переопределяется.

Должно быть очевидно, как подобный формализм может быть полезен при выражении лингвистических обобщений. Рис. 2 показывает сеть наследования для русских имен, которая была представлена в нашей более ранней работе (Corbett and Fraser, 1993).

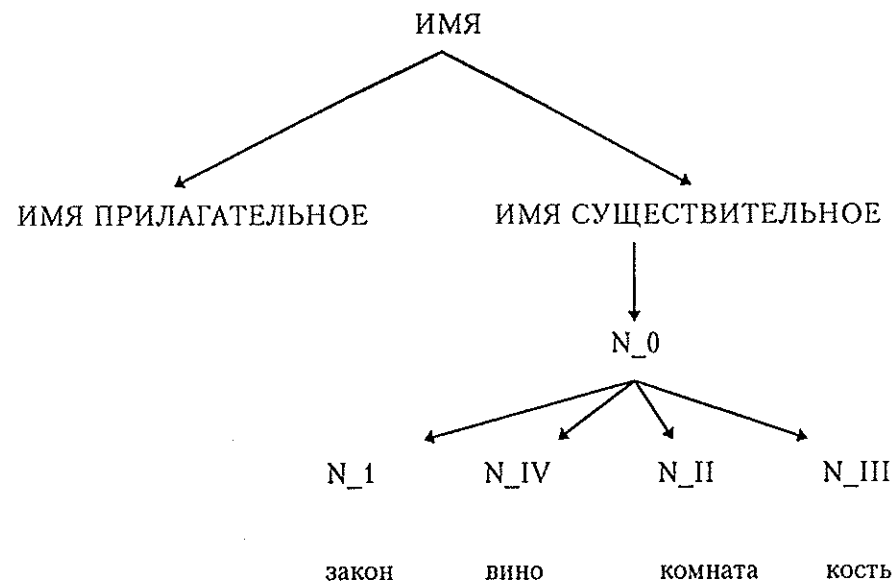


Рис. 2. Структура наследования для русских имен

Узлы N_I — N_IV — это типы склонения. (Заметим, что мы устанавливаем существование узла N_0, от которого наследуют узлы N_I и N_IV, объединяя таким образом преимущества тех описаний русского языка, которые постулируют три типа склонения, и тех, которые постулируют четыре.) Следующий неполный фрагмент взят из нашего более раннего анализа³:

(2) NOUN:

<mor loc sg> == "<stem>"_e
<mor nom pl> == "<stem>"_и.

N_III:

<> == NOUN
<mor loc sg> == "<mor dat sg>".

Кость:

<> == N_III
<stem> == кость.

Первая строка означает, что предложный падеж единственного числа включает основу и следующее за ней окончание -е. Путь доступа, заключенный в кавычки на языке DATR, используется, чтобы восстановить определенное значение для этого пути в нужном узле. Если

мы хотим найти форму именительного падежа множественного числа от слова *кость*, нам нужно наследовать выражение $\langle \text{mor nom pl} \rangle == \langle \text{stem} \rangle$ и. Перед тем, как идти дальше, нам следует узнать, что такое $\langle \text{stem} \rangle$ от слова *кость*. Так как ответом является *кость*, именительный падеж множественного числа будет *кости*. Путь доступа в кавычках обозначает, что мы берем основу слова *кость* (но не NOUN, у которого нет основы). Если, однако, мы хотим узнать предложный падеж единственного числа слова *кость*, нам не следует наследовать определение предложного падежа единственного числа имени существительного, потому что это переопределяется в узле N_III, из которого идет наследование слова *кость*. Определение предложного падежа единственного числа в узле N_III устанавливает асимметричную идентичность между формой предложного падежа единственного числа узла N_III и дательным падежом единственного числа этого имени существительного.

Часто требуется не только наследовать как можно больше информации из одного источника, но и иметь доступ к определенному роду информации, которая хранится в других местах. Язык DATR располагает подобными средствами поддержки множественного наследования. Рассмотрим следующий фрагмент⁴:

```
(3) N_II:
    <> == NOUN
    <mor gen sg> == "<stem>"_и.
```

```
N_III:
    <> == NOUN
    <mor gen sg> == N_II.
```

Это значит, что узел N_III может наследовать свою схему для формирования родительного падежа единственного числа от узла N_II, несмотря на то, что узел N_III (подобно узлу N_II) наследует непосредственно от имени существительного. Более эксплицитно это может быть выражено следующим образом:

```
(4) N_III:
    <mor gen sg> == N_II:<mor gen sg>.
```

На этом мы заканчиваем наше короткое введение в метод наследования по умолчанию и в формализм языка DATR. Более подробное введение в метод наследования по умолчанию можно найти в работах Газдара (1987), Делеманса, де Смита и Газдара (1992), а о введении в язык DATR — в работе Газдара (1990; в печати). В следующих разделах мы покажем, как формальное исследование, которое мы обозначили выше, сочетается с типологическими исследованиями.

2. Приписывание родовых характеристик

Согласование является определяющей характеристикой родовых систем. Сгруппировав однажды имена существительные по роду в соответствии с согласовательными признаками, мы можем затем задать себе вопрос, как носитель языка, который производит это согласование, определяет род различных имен существительных. Окажется, что в подавляющем большинстве случаев отнесение имен существительных к определенному роду всегда возможно из информации, которую в любом случае требует лексический вход (Corbett, 1991: 7–69; 1992). Определенный тип информации, который может быть использован в различных языках, дает нам типологию родовых систем. Мы находим семантические системы (где требуется только семантическая информация) и семантические + формальные системы (где семантическая информация дополнена морфологической и/или фонологической информацией). Чисто формальные системы (где род предсказывался бы формальными средствами, но где различные согласовательные классы имен не имели бы семантического значения) не найдены.

Примером языка с системой семантического родового распределения может служить годоберинский язык (Kibrik, 1994, Gudava, 1967). Годоберинский язык — это бесписьменный язык андийской подгруппы аварской группы дагестанских (северо-восточных кавказских) языков. На нем говорят около 2500 человек в двух селах (большее село называется Годобери) в Ботлихском районе Дагестана (данные см. Kibrik, 1994). В этом языке существует три рода, которые называются 'мужским', 'женским' и 'средним'. С другой стороны, вследствие семантической прозрачности этих классов имен кавказоведы иногда предпочитают называть их именными классами (I, II, III). Распределение имен по трем родам/именным классам показано в табл. 1.

Таблица 1. Приписывание родовых характеристик в годоберинском языке

критерий	род	пример	перевод
мужские разумные существа	мужской (I)	ima	отец
женские разумные существа	женский (II)	ila	мать
другие	средний (III)	hamaXi	осел

Данная родовая система проста и последовательна. Зная значение имени существительного, можно без обращения к форме предсказать его род. Можно быть уверенным, что имя существительное, которое обозначает женское разумное существо, будет женского рода и что имя существительное, которое женского рода, обозначает женское разумное

существо. Подобные системы иногда называют «системами с естественно-родовым согласованием».

Другой основной тип родовых систем также использует семантическую информацию, но дополняет ее формальной. Пример (в котором морфологическая информация дополняет семантическую) представлен в русском языке.

В русском языке, как и в годоберинском, важна семантическая информация:

Правила семантического распределения

1. Имена существительные, различающиеся по роду, которые обозначают мужские существа (людей и высших животных), — мужского рода: *сын, дядя, лев*;
2. Имена существительные, различающиеся по роду, которые обозначают женские существа, — женского рода: *дочь, тетя, львица*.

Имена существительные, которые различают род, — это те имена существительные, где род имеет отношение к человеку (как в случае людей и домашних животных) и где это противопоставление ярко выражено (как в случае со львами). Существует небольшое количество исключений из этих правил, но многие слова ими не покрываются. В отличие от годоберинского языка, в русском всем существительным, попадающим в подобный семантический остаток, не приписывается одна и та же родовая характеристика. Они распределены по следующим правилам:

Правила морфологического распределения

1. Имена существительные первого типа склонения (слова типа *закон*) — мужского рода;
2. Имена существительные второго (*комната*) и третьего (*кость*) типов склонения — женского рода;
3. Все остальные имена существительные — среднего рода.

Рассмотрим теперь пример, в котором семантические правила дополнены фонологическими, например, в кафарском (афарском) языке, одном из языков восточно-кушитской группы (кушитские языки включаются в афро-азиатские языки). На кафарском языке говорит примерно 250 000 человек в северо-восточной Эфиопии и в Джибути (Parker and Hayward, 1985). Имена существительные, которые обозначают людей мужского пола и тех животных мужского пола, для которых это различие значимо, являются именами существительными мужского рода, например: *baqla* 'муж'. Существа женского пола (люди и животные) — женского рода: *barra* 'женщина, жена'. Заметим, что символом ` обозначена акцентная позиция, которая маркирует потенциально высокий тон. Данные правила семантического согласования аналогичны рассмотренным выше русским. Рассмотрим теперь фонологические правила:

Правила фонологического согласования

1. Имена существительные, чья форма оканчивается на ударный гласный, — женского рода: *cato* 'помощь', *katma* 'осень'.
2. Остальные — мужского рода. Существуют две возможности:
 - (а) имена существительные, которые оканчиваются на согласный: *ceder* 'обеденное время', *gilal* 'зима';
 - (б) имена существительные, оканчивающиеся на гласный, но не ударный: *tamu* 'вкус', *baanta* 'труба'.

Из этих фонологических правил существует несколько исключений. Как мы уже видели, семантические побеждают, когда морфологические и фонологические правила вступают в конфликт. Так, *abba* 'отец' мужского рода из-за своего значения, хотя он и оканчивается на ударный гласный, что обычно указывает на женский род.

Как мы отмечали ранее, логическая возможность системы, основанной на формальной информации без семантических отношений, исключена.

Семантические и формально-фонологические системы относительно просты. Более сложными являются формально-морфологические системы (см., например, Aronoff, 1994, p. 73–4). Эти системы часто анализировались по-другому; вместо того, чтобы предсказывать род (и, следовательно, не нуждаться в том, чтобы указывать его в лексическом входе), некоторые трактуют его как спецификатор и через него пытаются предсказать морфологический класс имени существительного. Когда количество родов и флективных классов полностью или частично совпадает (как в русском), нельзя сразу сказать, какой анализ следует предпочесть.

Здесь традиционно приводятся аргументы двух типов. Первые, наиболее важные, имеют отношение к конкретному языку, в нашем случае к русскому. Можно показать, что предсказание рода на основе типа склонения является простым и включает в себя меньше исключений, чем попытка предсказать тип склонения на основе рода. Подобные доводы были обстоятельно рассмотрены в работе (Corbett, 1982), и мы не будем повторять их здесь. Во-вторых, существует типологический аргумент: так как во многих языках род непосредственно предсказуем, можно просто заявить, что он предсказуем во всех языках с типологическими изменениями, которые ограничиваются типом информации, используемой для предсказания.

Используя язык DATR в качестве средства, у нас появляется возможность третьего типа аргументации: поскольку существует компилятор (компьютерный интерпретатор) с языка DATR, можно продемонстрировать, что наш анализ (который является эксплицитным отчетом о взаимодействии семантики, рода и классов склонения в русском языке) действительно дает правильные результаты. Это значит,

что, получив наш анализ и лексические входы для русских имен существительных, компьютер может быть использован для того, чтобы верифицировать тот факт, что наш анализ правильно предсказывает род.

В работе (Fraser and Corbett, 1995) мы детально описали приложение Сетевой Морфологии к приписыванию рода в русском языке. Как уже отмечалось, Сетевая Морфология является эксплицитной морфологической теорией, которая использует формализм языка DATR. В качестве иллюстрации мы представим здесь только часть нашего анализа грамматического рода. В нашей работе каждое имя существительное наследует информацию из узла NOUN. Следовательно, каждое имя существительное наследует следующее выражение пути доступа, если это не переопределяется более конкретным выражением в лексическом входе:

(5) NOUN:

```
<syn gender> == GENDER: < "<sem sex>" >
```

Для того, чтобы найти значение пути <syn gender>, надо вычислить путь <sem sex> и путь доступа, который включает в себя только род в узле GENDER:

(6) GENDER:

```
<male> == masc  
<female> == fem  
<undifferentiated> == "<mor formal_gender>"
```

Другими словами, на основе этого восстанавливается пол референта имени существительного и определяется род. Интерпретация этого фрагмента на языке DATR очевидна в случае имен существительных, различаемых по полу. Если референт имени существительного мужского пола, тогда род — мужской; если референт женского пола, тогда род — женский. Однако если род не различается, тогда необходимо рассмотреть дополнительный формальный критерий, и подобное делается при вычислении пути <mor formal_gender>. Формальный род определяется для каждого основного типа склонения. Так, например, следующее выражение находится в узле для имен существительных второго типа склонения, N_II (имена существительные типа *комната*):

(7) N_II:

```
<formal_gender> == fem.
```

Это используется при вычислении пути <mor formal_gender> (функция начального *mor* не задержит здесь нашего внимания). Так, имя существительное второго типа склонения, обозначающее существо

мужского пола (подобно слову *мужчина*), получит мужской род; имя существительное второго типа склонения, обозначающее существо женского пола (*учительница*), получит женский род; имя существительное второго класса, обозначающее референта, не различающего род (подобно слову *комната*), получит женский род, в соответствии с родом по умолчанию для второго именного типа склонения. Существуют дополнительные сложности с неизменяемыми именами существительными, где род также должен быть включен в рассмотрение (Smirnova, 1979).

В качестве иллюстрации рассмотрим следующий лексический вход:

(8) Комната:

```
<> == NOUN  
<declensional_class> == N_II:<>  
<gloss> == room  
<infl_root all> == комнат.
```

Заметим, что *комната* здесь не более чем метка — с таким же успехом может быть и номер '397', так как форма недоступна. Имея этот минимальный вход, мы получаем следующую готовую информацию:

(9)

```
Комната: <gloss> = room.  
Комната: <mor nom sg> = комнат _a.  
Комната: <mor acc sg> = комнат _y.  
Комната: <mor gen sg> = комнат _ы.  
Комната: <mor dat sg> = комнат _е.  
Комната: <mor inst sg> = комнат _ой.  
Комната: <mor loc sg> = комнат _е.  
Комната: <mor nom pl> = комнат _ы.  
Комната: <mor acc pl> = комнат _ы.  
Комната: <mor gen pl> = комнат.  
Комната: <mor dat pl> = комнат _а _м.  
Комната: <mor inst pl> = комнат _а _ми.  
Комната: <mor loc pl> = комнат _а _х.  
Комната: <syn gender> = fem.  
Комната: <syn animacy> = inanimate.
```

Заметим, что в этой транскрипции *и* и *ы* трактуются как аллофоны. Здесь нам важно то, что морфологические формы и род правильно предсказаны. Наш полный анализ (Fraser and Corbett, 1995) охватывает более сложные случаи взаимодействия семантики, родовых харак-

теристик типа склонения и фонологии (во многих примерах морфологический класс предсказуем из семантической или фонологической информации).

Задача этого раздела состояла не в объяснении анализа. Скорее, мы хотели подчеркнуть, что наш анализ, который является теоретическим лингвистическим исследованием в рамках Сетевой Морфологии, может показать, как он работает, используя компьютерные методы. Другие описания присваивания родовых характеристик в русском языке не поддерживаются подобными точными демонстрациями. Независимо от частного взгляда на типологию родовых систем, основное наблюдение состоит в том, что средство, подобное языку DATR, позволяет нам разъяснять основополагающие для типологических целей случаи.

3. Синкретизм

В нашем первом примере типологические данные привели нас к необходимости построения формального описания. Сейчас мы движемся в обратном направлении, пытаюсь создать формальное описание, которое приведет нас к типологии.

Когда мы анализировали сложные флективные системы подобно русской, с использованием языка DATR (Corbett and Fraser, 1993; Brown and Hippisley, 1994), мы удивлялись разнообразию средств нейтрализации. Интуитивно они разного типа и различие в естественных путях их обработки на языке DATR подтверждает эту точку зрения. Некоторые из них соотносятся с понятием 'синкретизм', где отдельная флективная форма соответствует более чем одному морфо-синтаксическому описанию (Spencer, 1991, p. 45), или, неформально, где морфологии 'не хватает синтаксиса'.

Работа с языком DATR привела нас к типологии этих нейтрализаций, сначала к сфере их действия и потом к их природе.

3.1. Сфера действия синкретизма

Рассмотрим парадигму слова *кость*:

(10) ЕДИНСТВЕННОЕ ЧИСЛО

nom(inative)	кость
acc(usative)	кость
gen(itive)	кости
dat(ive)	кости
inst(rumental)	костью
loc(ative)	кости

Среди прочего мы хотим отметить, что родительный и дательный падежи единственного числа идентичны. Используя язык DATR, мы

достигаем этого при определении лексического входа для слова *кость* следующим образом:

(11) Кость:

```
<> == N_II
<gloss> == bone
<mor dat sg> == <mor gen sg>
...
```

Однако, если мы выражаем идентичность подобным способом, это применимо к лексическому входу только для одного этого слова. Можно представить такую ситуацию: синкретизм включает в себя отдельную лексическую единицу и образует первую часть нашей типологии. Однако нам не известны подобные случаи синкретизма для русского языка.

Подобный синкретизм действительно имеет место для всех членов класса узла N_III, имен существительных, подобных слову *кость*. Мы можем выразить этот факт на языке DATR на более высоком уровне в иерархии наследования — как узел для имен существительных третьего типа склонения, от которых наследует слово *кость*:

(12) N_III:

```
...
<mor dat sg> == "<mor gen sg>"
```

В обоих примерах идентичность управляется единственным узлом и нет нужды в множественном наследовании. Значение двойных кавычек, как это обсуждалось ранее, в том, что дательный падеж единственного числа есть любой «собственный» родительный падеж единственного числа: другими словами, дательный падеж единственного числа приобретает значение родительного падежа единственного числа того узла, в котором вычисляется путь <mor dat sg>.

Подобное установление идентичности должно быть помещено как можно выше в дереве наследования (язык DATR способствует установлению все более и более высоких обобщений), и это дает нам первый параметр нашей типологии, названный сферой действия синкретизма. Это может быть:

- отдельное слово
- отдельный флективный класс
- подмножество флективных классов в классе слов

(В русском языке дательный и предложный падежи единственного числа идентичны для двух флективных классов. Однако они одинаковы в пределах флективных классов и раз-

лично между флективными классами. Это приведет нас к другому параметру в типологии.)

между всеми флективными классами в классе слов

(В русском языке все имена прилагательные имеют родительный падеж множественного числа, совпадающий с предложным падежом множественного числа.)

между двумя и более классами слов

(В словенском языке именительный и винительный падежи двойственного числа идентичны для всех имен существительных и прилагательных, Priestly, 1993: 399)

между всеми потенциально релевантными классами слов

(В словенском языке дательный и творительный падежи двойственного числа идентичны для всех форм, в которых может быть этот показатель, Priestly, 1993: 399)

3.2. Природа синкретизма

Рассмотрим теперь данные, приведенные в табл. 2.

Таблица 2. Влияние категории одушевленности на морфологию имени

	студент одуш.	закон неодуш.	учительница одуш.	комната неодуш.
ЕДИНСТВЕННОЕ ЧИСЛО				
nom	студент	закон	учительница	комната
acc	студента	закон	учительницу	комнату
gen	студента	закона	учительницы	комнаты
МНОЖЕСТВЕННОЕ ЧИСЛО				
nom	студенты	законы	учительницы	комнаты
acc	студентов	законы	учительниц	комнаты
gen	студентов	законов	учительниц	комнат
	мышь одуш.	кость неодуш.	чудовище одуш.	вино неодуш.
ЕДИНСТВЕННОЕ ЧИСЛО				
nom	мышь	кость	чудовище	вино
acc	мышь	кость	чудовище	вино
gen	мыши	кости	чудовища	вина
МНОЖЕСТВЕННОЕ ЧИСЛО				
nom	мыши	кости	чудовища	вина
acc	мышей	кости	чудовищ	вина
gen	мышей	костей	чудовищ	вин

Рассмотрим сначала первую колонку. Здесь наблюдается синкретизм винительного и родительного падежей единственного числа (в зависимости от одушевленности-неодушевленности имени). Как мы увидим ниже, это можно выразить средствами языка DATR следующим образом⁵:

(13) <acc sg animate masc> == "<mor gen sg>"

Основное, что следует отметить, — это то, что винительный падеж совпадает с родительным. Если мы посмотрим более внимательно на различные парадигмы, мы заметим, что они имеют общую модель идентичности, хотя и различаются отдельными флексиями. Было бы явно неоправданно устанавливать идентичность форм внутри каждой парадигмы в отдельности; из этого можно заключить, что рассматриваемые случаи могут различаться от парадигмы к парадигме. Эта регулярность может быть описана средствами языка DATR (более высокий узел дерева наследования):

(14) ИМЯ:

```
<acc> == "<mor nom>"
<acc pl animate> == "<mor gen pl>"
<acc sg animate masc> == "<mor gen sg>"
<mor acc $number> == <acc $number "<syn animacy>" "<syn
gender>">
```

Ожидается, что именительный и винительный падежи должны различаться в синтаксической мотивации. Поэтому можно было бы ожидать, что термин 'синкретизм' будет здесь уместен. Но мы имеем дело со значительно более общими моделями, чем в предыдущих типах, и поэтому более точным может быть термин 'суперсинкретизм'.

Заметим, что присваивание значений категории одушевленности в русском языке объясняется с помощью как формальных средств вычисленного пути (Gazdar, 1990: 10–11), так и использованием переменной (\$number имеет значение sg и pl). Модели этого общего типа могут быть проще. Например, в церковнославянском языке в двойственном числе имена существительные имели три формы: одну для именительного и винительного падежей, одну для родительного и предложного и одну для дательного и творительного. Конкретные формы различаются от парадигмы к парадигме, но совпадение двух форм остается постоянным (Huntley, 1993: 139–141).

Вернемся к нашему выражению (13) на языке DATR, для удобства повторив его:

(15) <acc sg animate masc> == "<mor gen sg>"

Заметим, что это несимметричное отношение. Форма родительного падежа «правильна», а винительный падеж копирует ее. Это можно установить сравнением со второй колонкой: любое имя существительное этого типа будет иметь показатель родительного падежа единственного числа на -а, а винительный падеж приравнивается к родительному в тех случаях, когда имя существительное одушевленное, и к именительному падежу, когда неодушевленное. Актуальным является вопрос о направленности. Правила, которые описывают, как одна морфологическая форма реализуется идентично другой, часто называют 'ссылочными' правилами согласно Zwicky (1985: 372). Их можно сравнить с 'правилами предсказания' Перлмуттера и Орешника (1973).

Аронов (1994: 83) критикует использование ссылочных правил в некоторых аналитических работах как раз из-за их направленности. Можно продемонстрировать, что существуют примеры синкретизма, которые определенно несимметричны. Русские примеры представляются в этом смысле удачными. Однако существуют еще более чистые случаи.

Рассмотрим следующие данные из словенского (южно-славянского) языка.

Таблица 3. Парадигма склонения слова *kót* 'угол' в словенском языке (Priestly 1993: 400—2)

	единств. число	двойств. число	множеств. число
nominative	<i>kót</i>	<i>kóta</i>	<i>kóti</i>
accusative	<i>kót</i>	<i>kóta</i>	<i>kóte</i>
genitive	<i>kóta</i>	<i>kōtov</i>	<i>kōtov</i>
dative	<i>kótu</i>	<i>kótoma</i>	<i>kótom</i>
instrumental	<i>kótom</i>	<i>kótoma</i>	<i>kóti</i>
locative	<i>kótu</i>	<i>kōtih</i>	<i>kōtih</i>

В этой парадигме можно найти различные случаи синкретизма, некоторые из них менее значимы, чем кажется. Важными являются те, в которых родительный падеж двойственного числа совпадает с родительным множественного, и те, где предложный двойственного числа совпадает с предложным множественного. Это сохраняется для всех имен существительных и прилагательных, но неверно для местоимений (Priestly, 1993: 399).

Вопрос состоит в том, должны ли мы устанавливать симметричную идентичность или несимметричную. Оба варианта возможны в формализме языка DATR, но они различаются по форме. Если взять пример

родительного падежа, симметричный анализ будет выражен в соответствующем узле следующим образом:

(16) $\langle \text{form1} \rangle == \langle \text{stem} \rangle_ov$
 $\langle \text{genitive dual} \rangle == \langle \text{form1} \rangle$
 $\langle \text{genitive plural} \rangle == \langle \text{form1} \rangle$

Асимметричный анализ будет выражаться проще (допустим, что двойственное число 'паразитично' от множественного):

(17) $\langle \text{genitive plural} \rangle == \langle \text{stem} \rangle_ov$
 $\langle \text{genitive dual} \rangle == \langle \text{genitive plural} \rangle$

Это показывает, что для представления симметричного синкретизма требуется дополнительное, по сравнению с асимметричным, выражение. Но в примерах, где симметричные синкретические формы могут устанавливаться как обычный падеж (умолчание, которое переопределяет другие формы), это выражение может быть соответственно проще (ср. Stump, 1993: 450—451).

Возникает вопрос, в каких случаях предпочесть тот или иной формализм. Рассмотрим следующий пример супплетизма (супплетивные формы маркируют число).

Таблица 4. Парадигма слова *človek* 'человек, персона' в словенском языке (Priestly 1993: 401)

	единств. число	двойств. число	множеств. число
nominative	<i>člóvek</i>	<i>človéka</i>	<i>ljudjê</i>
accusative	<i>človéka</i>	<i>človéca</i>	<i>ljudî</i>
genitive	<i>človéka</i>	<i>ljudî</i>	<i>ljudî</i>
dative	<i>človéku</i>	<i>človékoma</i>	<i>ljudêṁ</i>
instrumental	<i>človékom</i>	<i>človékoma</i>	<i>ljudmî</i>
locative	<i>človéku</i>	<i>ljudêh</i>	<i>ljudêh</i>

Ясно, что синкретизм может быть асимметричным. Нужно ли исследовать, всегда ли синкретизм асимметричен⁶.

Так как возможные типы подобных нейтрализаций не могут быть ограничены рамками формального языка, мы должны искать причины, привнесенные извне. Это часть общего подхода, который мы назвали Сетевой Морфологией. Находить подобные причины для доказательства синкретизма оказывается неожиданно проблематичным.

Однако основная идея состоит в том, что использование формального языка DATR, который заставляет нас определять часто плохо определенные различия, приводит нас к формулировке типологических правил.

4. Заключение

В нашем первом примере, где обсуждались проблемы приписывания родовых характеристик, мы видели, как типологическая работа приводит нас к использованию формальных методов для того, чтобы сделать более ясным анализ спорных лингвистических примеров. Во втором примере, посвященном синкретизму, мы проследили, как различные типы выражения на формальном языке приводят нас к типологии. Из этого мы заключаем, что компьютерная лингвистика и типология являются не противоположными областями лингвистики, а, скорее, они сближаются друг с другом. Будем надеяться, что с началом следующего столетия мы увидим дальнейшее сближение различных подходов в лингвистике.

Библиография

- Aronoff M. 1994. Morphology By Itself: Stems and Inflectional Classes (=Linguistic Inquiry Monograph 22). Cambridge, MA.: MIT Press.
- Brown D. and Hippiisley A. 1994. Conflict in Russian genitive plural assignment: a solution represented in DATR. *Journal of Slavic Linguistics* 2. P. 30–48.
- Corbett G. G. 1982. Gender in Russian: an account of gender specification and its relationship to declension. *Russian Linguistics* 6. P. 197–232.
- Corbett G. G. 1991. Gender. Cambridge: Cambridge University Press.
- Corbett G. G. 1992. = Корбетт, Гревилл Г. 1992. Типология родовых систем. Вопросы языкознания № 3. P. 21–30.
- Corbett G. G. and Fraser, Norman M. 1993. Network Morphology: A DATR Account of Russian Nominal Inflection. *Journal of Linguistics* 29. P. 113–142.
- Daelemans W., De Smedt, K., and Gazdar G. 1992. Inheritance in natural language processing. *Computation Linguistics* 18. P. 205–218.
- Evans R. and Gazdar G. 1989a. Inference in DATR. *Proceeding of the 4th Conference of the European Chapter of the Association for Computational Linguistics*. P. 66–71. Manchester, England.
- Evans R. and Gazdar G. 1989b. The semantics of DATR/A. G. Cohn (ed.) *Proceedings of the Seventh Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*. P. 79–87. London: Pitman/Norgan Kaufmann.
- Fraser N. M. and Corbett G. G. 1995. Gender, Animacy and Declensional Class Assignment: a Unified Account for Russian. To appear in: G. Booij and J. van Marle (eds) *Yearbook of Morphology 1994*. Dordrecht: Kluwer.
- Gazdar G. 1987. Linguistic applications of default inheritance mechanisms/ P. Whitelock, M. McGee Wood, H. L. Somers, R. L. Johnson and P. Bennett (eds) *Linguistic Theory and Computer Applications*. P. 37–68. London: Academic Press.
- Gazdar G. 1990. An introduction to DATR/R. Evans and G. Gazdar (eds). *The DATR Papers*. Cognitive Science Research Paper CSRP 139. P. 1–14. School of Cognitive and Computing Science, University of Sussex.
- Gazdar G. forthcoming. Ceteris paribus. To appear in J. A. W. Kamp and C. Rohrer (eds) *Aspects of Computational Linguistics*. Berlin: Springer.

- Gudava T. E. 1967. = Гудава Т. Е. Годоберинский язык//Языки народов СССР: IV; Иберийско-кавказские языки, Е. А. Бокарев и К. В. Ломташидзе (ред.). P. 307–321. Москва: Наука.
- Huntley D. 1993. Old Church Slavonic/B. Comrie and G. G. Corbett (eds). *The Slavonic Languages*. P. 125–187. London: Routledge.
- Kibric A. E. (ed.) 1994. Godoberi's noun phrase. *EUROTYP Working Papers VII/21*. Konstanz: University of Konstanz.
- Parker E. M. and Hayward R. J. 1985. *An Afar-English-French Dictionary (with Grammatical Notes in English)*. London: School of Oriental and African Studies, University of London.
- Permuter D. M. and Oresnik J. 1973. Language-particular rules and explanation in syntax/S. R. Anderson and P. Kiparsky (eds). *A Festschrift for Morris Halle*. P. 419–459. New York: Holt Rinehart.
- Plank F. 1994. Homonymy vs. suppletion: A riddle (and how it happens to be solved in...) Agreement gender number genitive *EUROTYP Working Papers VII/23*. P. 81–86. Konstanz: University of Konstanz.
- Priestly T. M. S. 1993. Slovene/B. Comrie and G. G. Corbett (eds). *The Slavonic Languages*. P. 388–451. London: Routledge.
- Shieber S. M. 1987. Separating linguistic analyses from linguistic theories/ P. Whitelock, M. McGee Wood, H. L. Somers, R. L. Johnson and P. Bennett (eds). *Linguistic Theory and Computer Applications*. P. 1–36. London: Academic Press.
- Smirnova G. A. 1979. = Смирнова Г. А. 1979. Категория рода несклоняемых существительных. В сборнике: *Лингвистика и поэтика/Под ред. В. П. Григорьева*. М.: Наука. С. 86–105.
- Spencer A. 1991. *Morphological theory*. Oxford: Blackwell.
- Stump G. T. 1993. On rules of referral. *Language* 69. P. 449–479.
- Zwicky A. 1985. How to describe inflection/M. Niepokuj, M. van Clay, V. Niki-foridou and D. Feder (eds). *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*. P. 372–386. Berkeley, California: B. L. S., University of California.

Примечания

- ¹ Разделы 1 и 2 написаны на основе материала, который содержится в работе (Fraser and Corbett 1995). Это исследование поддерживалось Советом Экономических и Социальных Исследований (грант R000233633) и организации Leverhulme Trust (грант F.242M); обеим организациям мы выражаем большую признательность.
- ² Язык DATR является одним из основных источников развития структуры Сетевой Морфологии и мы хотим выразить нашу благодарность за это Р. Эвансу и Дж. Газдару. Привлекательность языка DATR состоит в понятию простом синтаксисе и семантике, а также в существовании ряда компьютерных интерпретаторов. Язык DATR не является акронимом.
- ³ В работе морфологические формы даны в фонетической транскрипции.
- ⁴ В узлах N_II и N_III в примере (3) приведены лишь те данные, которые непосредственно касаются предмета обсуждения. В нашем полном анализе русской именной морфологии каждый из этих узлов богаче данными, чем это показано здесь. Данные для узла N_III отличаются от подобных для узла N_II, однако анализ формы 'N_III: <> == N_II' выходит за рамки нашей работы. Более полный анализ русского языка см. в работах (Corbett and Fraser 1993), (Brown and Hippiisley 1994).

⁵ Родовые характеристики являются доступными, как это описано в разделе 2.
⁶ Важные словенские данные другого плана были независимо отмечены Планком (1994).

Resume

The fruitfulness of applying formal methods of computer linguistics in solving typological problems is demonstrated in the present paper. Similar methods are sometimes seen as somewhat remote from typologists' studies. But this is erroneous: descriptions on which typology is based must be precise and adequate; they should be considered side by side with the possibilities of natural human languages.

The language DART designed for representing lexical knowledge is used for the explicit description of the data, especially those ones which are problematic from the typological point of view. This language is described in the article. Two typological questions are also discussed: assigning gender characteristics and syncretism.

МАТЕРИАЛЫ И СООБЩЕНИЯ

Е. И. Тимошенко (Гомель, Белоруссия)

К вопросу о лексическом влиянии старобелорусского языка на великорусский в период XV—XVII вв.

После опубликования рецензии академика А. И. Соболевского на «Словарь иностранных слов, вошедших в русский язык в эпоху Петра Великого» Н. А. Смирнова (1903)¹ в науку получило распространение положение о том, что одним из важных путей проникновения — из лексики классических, западноевропейских и некоторых западнорусских языков в русский язык Петровской эпохи, а также в язык более раннего времени XV—XVII вв. — было посредство западнорусского языка. После А. И. Соболевского об этом посредничестве более или менее подробно писали В. В. Виноградов, Л. А. Булаховский, Ф. П. Филин, М. Г. Булахов, И. С. Козырев, Г. П. Пивторак и другие исследователи.

Действительно, исторические условия функционирования и контактирования русского и белорусского, а также украинского языков в XV—XVII вв., с одной стороны, и время фиксации в каждом из них лексики классических и западноевропейских языков — с другой, дает основания предложить, что в указанный период старобелорусский и староукраинский языки² выполняли посредническую функцию в заимствовании русским языком иноязычной лексики. При этом многими исследователями иноязычные слова, встречающиеся раньше в памятниках старобелорусской и староукраинской письменности, чем в памятниках великорусского языка, квалифицируются как заимствованные через западнорусское языковое посредство³. На наш взгляд, вопрос требует более тщательного и глубокого рассмотрения.

Контакты старобелорусского и великорусского языков в XV—XVII веках. (особенно во 2-й половине XVI—XVII вв.) осуществлялись как с сфере письменности, так и в устной речи. Западная (или юго-западная) языковая традиция оказывала определенное влияние на книжный язык Московской Руси, прежде всего через распространение старопечатных белорусских книг на территории Московского государства.

Известно, что белорусский язык позднего средневековья имел благоприятные условия для своего развития. Являясь языком основной части населения и официальным языком Литовского княжества, старобелорусский язык уже с XIV в. активно функционирует в области судопроизводства и дипломатии; на нем издаются судебники, создает-