

А. Ф. Кургаев, С. Н. Григорьев

Нормальные формы знаний

(Представлено академиком НАН Украины А. В. Палагиньым)

В новом способе представления и использования знаний знания предельно отделены от данных. В описании и в представлении знания и данные разделены физически: знания сконцентрированы в информационной структуре, представление же и обработка данных локализованы в терминальных алгоритмах. Эти две компоненты представлены двумя разными подпространствами единого пространства решения задач. Разделение пространств и механизмов обработки знаний и данных позволяет применить для каждого из них оптимальные аппаратные средства интерпретации при сохранении единства двуединного процесса решения задач.

Ключевые слова: синтаксис, семантика, метаязык, графические средства метаязыка, формальное описание метаязыка, формальное описание простого английского предложения.

Все произвольные решения люди принимают, размышляя в пространстве смыслов, в семантическом пространстве. Более того, есть основания считать, что это же присуще и всему живому, различие между разновидностями которого состоит лишь в объеме и совершенстве механизмов мышления.

В то же время первоисточником семантики является синтаксис. Мы воспринимаем лишь то, что имеет форму, объекты без формы для нас просто не существуют. А всякую форму мы воспринимаем, лишь различая ее структуру на основании знаний о синтаксисе языка описания множества форм определенного класса.

Описание синтаксиса языков. Для строгого и точного описания синтаксиса языков используют специальные *метаязыки* (языки для описания языков) [1]. Наиболее распространены металингвистичные формы Бекуса–Наура (БНФ) и расширенные формы Бекуса–Наура (РБНФ). Метаязыком БНФ, его расширением или произвольным другим метаязыком представляют спецификацию синтаксических структур произвольного объектного языка в виде системы правил — формул, похожих на математические, которые определяют отношение между терминалами и нетерминалами. Терминалы — элементы структуры, не имеющие собственной структуры, это определенные вне БНФ-описания идентификаторы или цепочки — последовательности символов в кавычках или апострофах. Нетерминалы — элементы структуры, имеющие и собственные имена, и структуру.

Семантика правила в РБНФ: нетерминал, заданный идентификатором слева от знака “=”, определяется некоторым отношением терминалов и нетерминалов, указанным справа от знака “=”. Набор основных отношений РБНФ: конкатенация, альтернативный выбор и итерация, а дополнительных, стилистических — отношение необязательности (необязательный элемент выделяют квадратными скобками) и структурные круглые скобки (группируют элементы сложных выражений). Полным описанием структуры является набор правил, определяющих все нетерминалы так, что каждый из них может быть сведен к комбинации терминалов путем последовательного (рекурсивного) применения правил.

Метаязык РБНФ пригоден для описания произвольных языков, имеющих практический интерес. Средствами РБНФ определим и сам метаязык РБНФ, например:

- 1) описание = определение {определение};
- 2) определение = имя_понятия тело_определения тчк;
- 3) имя_понятия = идентификатор;
- 4) идентификатор = буква {буква | цифра};
- 5) тело_определения = есть_структура выражение;
- 6) выражение = элемент {отношение_ИЛИ элемент | отношение_И элемент};
- 7) элемент = имя_понятия | строка | “(“выражение”)” | “[“выражение”]” | “{“выражение”}”;
- 8) строка = кавычка {знак} кавычка.

Метаязык используют для конструирования спецификации объектного языка, согласно которой в языке реализации разрабатывают программу распознавателя (parser) — анализа и построения дерева вывода (разбора) утверждений объектного языка. Успех построения дерева разбора подтверждает принадлежность входной цепочки символов данному языку.

Однако в качестве языка представления знаний метаязык РБНФ (и все другие известные метаязыки) имеет определенные недостатки. В частности, созданный первично для узкоспециальных целей и такой, что хорошо их обеспечивает, он не является функционально полным и потому не пригоден для представления произвольных знаний.

Новый метаязык описания знаний. В [2] предложено решение задачи синтеза нового способа представления и использования знаний, адекватного процессу решения структурно сложных задач. Согласно ему, конечный программный продукт составляется из двух частей.

Первая является информационной структурой множества определений понятий (нетерминалов), связанных базовыми отношениями нормальных форм знаний (НФЗ), являющихся развитием выразительных возможностей метаязыка РБНФ. Вторая часть составляется из двух подмножеств — множества элементарных алгоритмов и множества элементарных структур данных, которые реализуют на одном из традиционных языков программирования в форме единой библиотечной системы. Каждый из терминалов имеет собственное имя и в описании структуры может использоваться наравне с нетерминальными понятиями.

Для достижения функциональной полноты в метаязык РБНФ

введено отношение инверсии, применимое к любому понятию в описании структуры (обозначается символом “^”, предшествующим имени отрицаемого понятия);

введены три элементарные операции: распознавание, распознавание со следом и порождение над информационными структурами, моделями прикладных областей. Обозначаются соответственно знаками “?”, “#” и “!”, присоединяемыми к имени понятия, вводящему некоторую информационную структуру.

Определение языка НФЗ в этом же языке выглядит так:

- 1) описание = определение (определение);
- 2) определение = инверсия имя_понятия тело_определения тзпт;
- 3) инверсия = отрицание / истина;
- 4) имя_понятия = идентификатор / целое / цепь_знаков;
- 5) идентификатор = буква (буква / цифра);
- 6) целое = цифра (цифра);
- 7) цепь_знаков = знак (^метазнак знак);

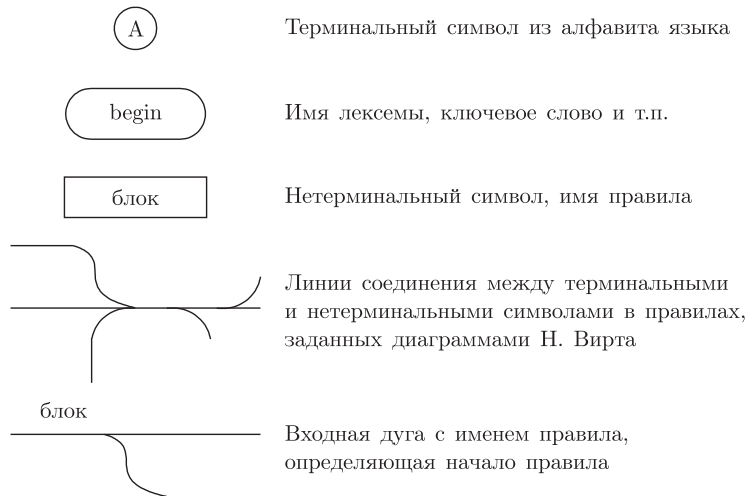


Рис. 1. Графические обозначения метасимволов диаграмм Н. Вирта

- 8) тело_определения = структура / терминал;
- 9) структура = есть_структура выражение;
- 10) терминал = есть_структура;
- 11) выражение = элемент (отношение_ИЛИ элемент/отношение_И элемент);
- 12) элемент = инверсия имя режим;
- 13) имя = итерация / строка / имя_понятия;
- 14) режим = зн_режим / истина;
- 15) зн_режим = ? / # /!;
- 16) итерация = скобка_откр выражение скобка_закр;
- 17) строка = кавычка (знак) кавычка,

где есть_структура-разделитель, изображается символом “ = ”; отношение_ИЛИ — отношение альтернативного выбора изображается символом “/”; отношение_И — отношение конкатенации, изображается символом пробел “ ”; скобка_откр — итерационная скобка, изображается символом “(”); скобка_закр — итерационная скобка, изображается символом “)”; отрицание — отношение отрицания, изображается символом “^”;

тзпт — конец определения, изображается символом “;”; кавычка — текстовая кавычка, изображается символом “ ’ ”; истина — отношение “тождественная истина”, пустой объем; пробел — разделитель, изображается символом “ ”. В этом описании понятия “буква”, “цифра”, “знак”, “метазнак” — терминальные:

буква = A/B/V/Г/.../a/б/в/г/.../я/А/В/С//.../a/b/c/.../z;

цифра = 0/1/2/3/4/5/6/7/8/9;

знак = -/)/(*/&/^/%/\$/#/@/!/~/”/’/; /: /< /> /, /.\|/| = /_/? ///;

метазнак = (/) / пробел / // = / ? / # /! /; / ’;

Наряду с текстовыми используются и графические метаязыки, среди которых наиболее популярен язык диаграмм Н. Вирта (рис. 1).

Графические средства НФЗ развивают графические средства диаграмм Н. Вирта так же, как НФЗ развивают РБНФ, и имеют вид, согласно рис. 2.

Используя эти выразительные средства получим графическое описание метаязыка НФЗ (рис. 3), эквивалентное приведенному выше текстовому.

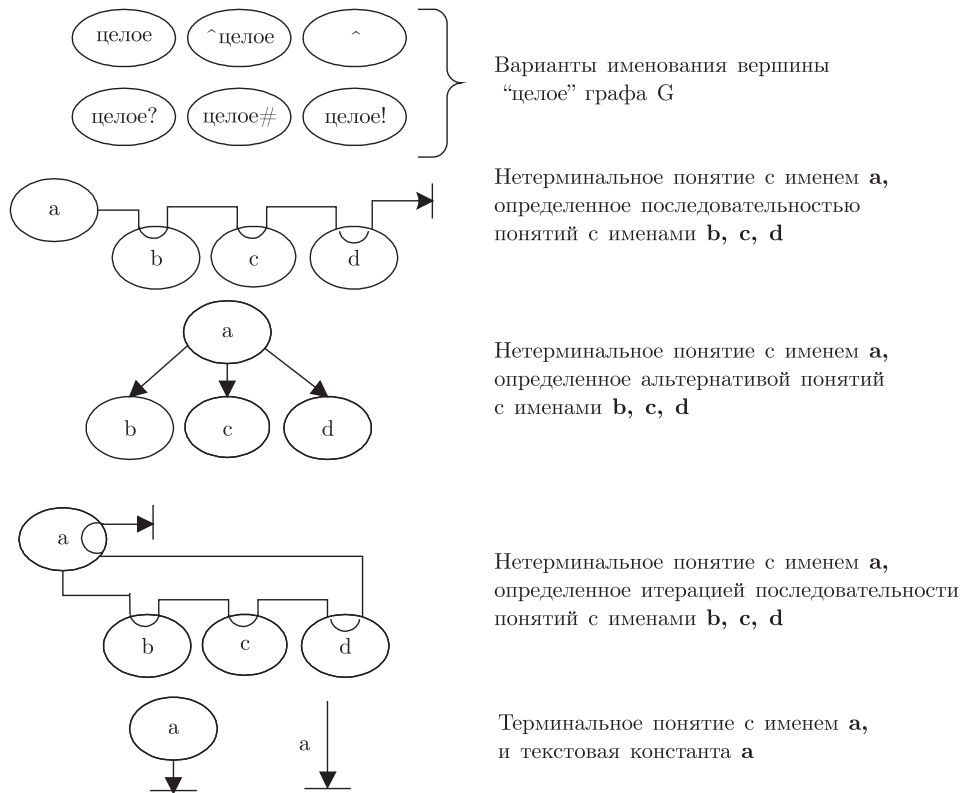


Рис. 2. Графические средства метаязыка НФЗ

Пример формального описания английского естественного предложения. Используя результаты [3], синтаксис подмножества простых предложений английского языка в нотации метаязыка НФЗ можно представить следующим образом.

1. Sentence = NP Sep1 Aux1 Sep1 VP;
2. NP = Art1 Sep1 AP1 Sep1 N Sep1 PP1;
3. Sep1 = (Sep);
4. Aux1 = Aux/1;
5. Aux = ‘will’/ ‘can’/ ‘might’;
6. VP = Verb Sep1 NP1 Sep1 PP1 Sep1 Adv1;
7. Art1 = Art/1;
8. Art = ‘the’/ ‘an’/ ‘a’;
9. AP1 = AP/1;
10. AP = Adj Sep1 PP1;
11. Adj = ‘old’/ ‘red’/ ‘slimy’/ ‘white’/ ‘new’/ ‘hungry’;
12. PP1 = PP/1;
13. PP = Prep Sep1 NP;
14. Prep = ‘at’/ ‘in’/ ‘to’/ ‘with’/ ‘out_of’;
15. N = ‘tree’/ ‘wind’/ ‘children’/ ‘toys’/ ‘toy’/ ‘box’/ ‘boy’/ ‘ball’/ ‘house’/ ‘shorts’/ ‘letter’/ ‘jon’/ ‘it’/ ‘man’/ ‘hat’/ ‘fish’/ ‘toby’/ ‘he’/ ‘book’/ ‘we’/ ‘dogs’/ ‘swimmer’;
16. Verb = ‘swayed’/ ‘put’/ ‘found’/ ‘kicked’/ ‘was_reading’/ ‘gave’/ ‘saw’/ ‘read’/ ‘fed’/ ‘was’/ ‘pulled’/ ‘have_been_reading’;

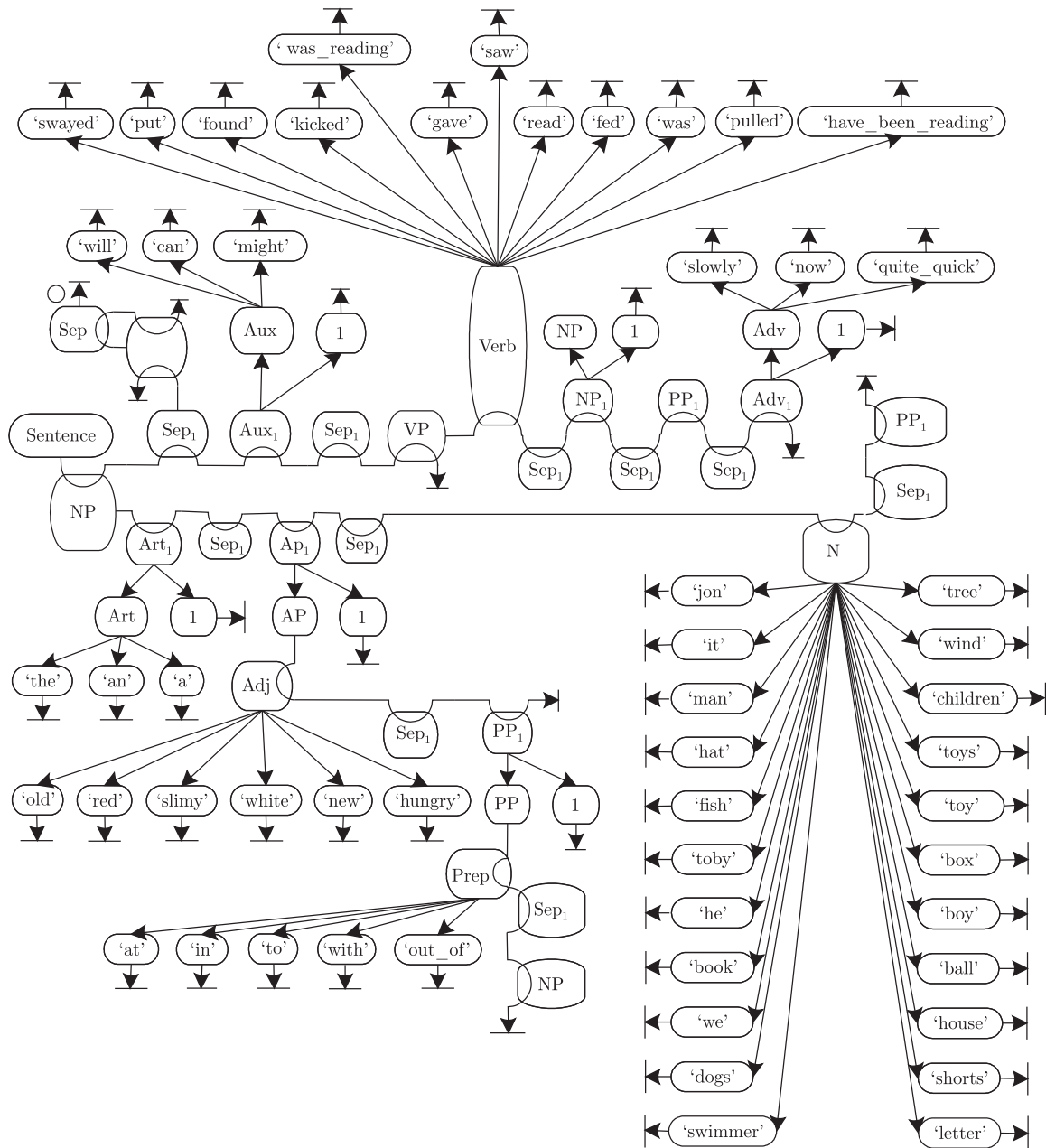


Рис. 4. Графическое описание примера синтаксиса английского предложения

- 17. NP1 = NP/1;
- 18. Adv1 = Adv/1;
- 19. Adv = 'slowly' / 'now' / 'quite_quick'.

В этом описании использованы следующие понятия: Sentence – предложение английского языка; NP – группа существительного; VP – группа глагола; PP – группа предлога; AP – группа прилагательного; N – существительное; Verb – глагол; Adj – прилагательное; Prep – предлог; Adv – наречие; Art – артикль; Aux – вспомогательный глагол; Sep – пробел и др.

Графическое описание приведенного примера синтаксиса предложения английского языка, выполненное графическими средствами метаязыка НФЗ, представлено на рис. 4. Очевидно, что оно эквивалентно приведенному выше текстовому описанию.

Таким образом, в предложенном способе знания предельно отделены от данных. В описании, в представлении знания и данные отделены физически: знание сконцентрировано в информационной структуре, представление и обработка данных локализованы в терминалах. Эти две компоненты представлены двумя разными подпространствами единого пространства решения задач. Разделение пространств и механизмов обработки знаний и данных позволяет применить для каждого из них оптимальные аппаратные средства интерпретации при сохранении единства всего двуединого процесса решения задач.

Знание — описание объективно существующей информационной структуры предметной области. На одних и тех же самых знаниях можно решать все задачи, существующие для этой области (если достаточна точность описания предметной области). То есть знания — это компонента информации, инвариантная относительно задач, решаемых в этой области. Итак, знания, когда-то осознанные, формализованные и введенные в компьютер, можно использовать многократно для решения разных задач.

Этот способ, по сравнению с известными, характеризуется большей технологической адекватностью представления и использование знаний возможностям человека.

Цитированная литература

1. Кургаев А. Ф., Григорьев С. Н. Анализ доминирующих моделей представления и использования знаний // Управляющие системы и машины. – 2014. – № 3. – С. 64–73.
2. Кургаев О. П., Григорьев С. М. Спосіб представлення і використання знань: Патент на корисну модель UA 92484 U, 2014 р. Бюл. № 16.
3. Fabb N. Sentence structure. – New York: Nigel Fabb, 2005. – 122 p.

References

1. Kurgaev A., Grygoryev S. Control Systems and Computers, 2014, No 3: 64–73 (in Russian).
2. Kurgaev A., Grygoryev S. Utility model patent UA 92484 U, 2014, Bulletin No 16 (in Ukrainian).
3. Fabb N. Sentence structure. Second ed., New York: Nigel Fabb Copyright, 2005.

*Институт кибернетики им. В. М. Глушкова
НАН Украины, Киев*

Поступило в редакцию 15.06.2015

О. П. Кургаев, С. М. Григорьев

Нормальні форми знань

Інститут кібернетики ім. В. М. Глушкова НАН України, Київ

У новому способі подання й використання знань знання гранично відділені від даних. В описі та у поданні знання й дані розділені фізично: знання сконцентровані в інформаційній структурі, подання ж і обробка даних локалізовані в термінальних алгоритмах. Ці дві компоненти представлено двома різними підпросторами єдиного простору розв'язку задач. Поділ просторів і механізмів обробки знань і даних дозволяє застосувати для кожного з них оптимальні апаратні засоби інтерпретації при збереженні єдності двоединого процесу розв'язання задач.

Ключові слова: синтаксис, семантика, метамова, графічні засоби метамови, формальний опис метамови, формальний опис простого англійського речення.

A. F. Kurgaev, S. N. Grygoryev

The normal forms of knowledge

V. M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kiev

In a new method of representation and use of knowledge, the separation of knowledge from data is maximized. The knowledge and data are physically separated both in the description and in the representation: the knowledge is concentrated in the information structure, while the representation and the data processing are located in terminal algorithms. These two components are represented by two different subspaces of the common space for problem solving. The separation of spaces and mechanisms of data and knowledge processing allows the user to apply the optimal hardware interpretation tools, while preserving the unity of the two-pronged problem solving process.

Keywords: syntax, semantics, meta-language, the graphical tools of the meta-language, the formal description of the meta-language, a formal description of a simple English sentence.