



Применение изоморфизма Керри-Ховарда к семантике естественных языков

Андреев Артем Викторович

Институт лингвистических исследований РАН (ИЛИ РАН), Санкт-Петербург, Россия

Applying the Curry-Howard isomorphism to the description of the natural language semantics

Andreev Artem

Institute for Linguistic Studies, RAS (ILSRAS), St.Petersburg, Russia

Аннотация

В статье рассматривается возможность применения изоморфизма Керри-Ховарда, устанавливающего эквивалентность логических исчислений и искусственных языков, к семантике естественных языков. Предлагается подход к трактовке языковой неоднозначности через описание инвариантов классов пропозиций. Рассматривается вариант конструктивной логики высшего порядка для описания естественного языка.

Abstract

The article deals with the possibility to adapt the Curry-Howard isomorphism to describe natural language semantics. An approach to treating linguistic ambiguity is proposed based on specifying invariants for proposition classes. A variant of constructive higher-order logic

powerful enough to cover descriptions of natural languages.

Ключевые слова: изоморфизм синтаксиса и семантики, логические основания грамматики, формальная семантика.

Keywords: isomorphism between syntax and semantics, logical foundations for grammar, formal semantics

Изоморфизм синтаксиса и семантики

Принцип изоморфизма является центральным для исследований по формальной семантике, во всяком случае, после работ Р. Монтегю. Традиционно речь идет об изоморфизме между синтаксической структурой и семантической интерпретацией, т. е. о композициональности семантики



языковых форм по отношению к функции семантической оценки. Эта идея была напрямую заимствована из логики, однако следует отметить, что при всей своей элегантности, в применении к естественным языкам она сразу же начинает давать осечки. Уже сам Монтегю был вынужден вводить специальные правила трансформационного типа для адекватного описания анафорических местоимений [Montague], а позднее был обнаружен целый класс выражений (так называемых *donkeysentences*), для которых оказывается невозможным получить «естественную» интерпретацию, изоморфную их синтаксической структуре [Heim]. Следует также отметить, что и для самых обычных предложений принцип изоморфизма по-разному работает в разных языках. Так, в английском языке существует четкое различие между синтаксической категорией имен и именных групп, вполне изоморфное различию между логическими типами e и $\langle e, t \rangle$. Однако в русском языке этого синтаксического различия нет, так что любое русское имя нарицательное может само по себе с равным успехом означать как класс в целом (тип $\langle e, t \rangle$), так и отдельный индивид этого класса (тип e). С другой стороны, русские существительные изменяются по падежам, при этом разные падежные формы относятся, очевидно, к разным синтаксическим типам, но во многих случаях к одному и тому же логическому типу.

Исчисление Ламбека. Программы как доказательства

В лингвистике известен и другой тип изоморфизма, реализующийся в формализме категориальных грамматик, в первую очередь, в

исчислении Ламбека [Moot, Retoré]. Здесь речь идет о параллелизме между синтаксической структурой и очередностью применения правил вывода в логическом исчислении. Изоморфизм такого рода хорошо известен в теории формальных языков под названием изоморфизма Керри-Ховарда [Howard]. Он устанавливает соответствие между логическими **доказательствами** и программами, а именно, каждому (конструктивному) доказательству соответствует реализующая его программа, и каждой программе соответствует доказательство некоторого утверждения, причем соответствие это является **структурным**: каждому логическому правилу вывода соответствует определенная вычислительная конструкция (так, например, правилу **modusponens** соответствует применение функции к аргументу), и наоборот. По сути дела, в случае исчисления Ламбека мы имеем дело с тем же самым изоморфизмом, только как логическая система категориальная грамматика обладает слишком малой выразительной силой и поэтому на роль теории семантической структуры в целом не подходит.

Классы пропозиций

Перенесение изоморфизма Керри-Ховарда на материал естественных языков позволит по-новому взглянуть на принцип композициональности и, возможно, решить некоторые из указанных выше проблем формальной семантики, однако на пути к этому есть существенные трудности. Первая и главная из них состоит в том, что соотношение плана выражения и плана содержания в естественных и искусственных языках совершенно различно, а именно, искусственным



языкам совершенно несвойственна омонимия, а синонимия, хотя и возможна, но значительно более ограничена, чем в языках естественных. Может показаться, что это делает прямое применение изоморфизма Керри-Ховарда невозможным. В самом деле, если синтаксическая структура является, как вычислительная программа, реализацией логического доказательства, то тогда две омонимичные конструкции оказываются доказательством двух разных утверждений, что с логической точки зрения нелепо. Однако эта проблема может быть решена следующим образом: мы будем считать, что синтаксические структуры связаны не с конкретными пропозициями, а с **классами** пропозиций, удовлетворяющими некоторым инвариантам. В этом случае мы можем сопоставить каждому поверхностному выражению ровно **одно** семантическое представление, которое в случае омонимичных конструкций будет представлять собой теоретико-множественное объединение подклассов, соответствующих отдельным значениям. Такой подход позволяет устранить и другое затруднение, свойственное в целом формально-семантическому подходу, а именно, трудность или даже невозможность исчерпывающе описать значение того или иного выражения. Если же семантическими коррелятами синтаксических структур являются не отдельные пропозиции, а их классы, то исчерпывающее описание и не требуется, а речь может идти только о выявлении инвариантов, ограничивающих значение тех или иных конструкций (подробнее об этом см. [Андреев]).

Пропозициональные роли

Другая проблема, более техническая, состоит в том, чтобы выбрать подходящую логическую систему для семантического представления. Мы предлагаем использовать для этого конструктивную логику высшего порядка, известную также, как исчисление конструкций [Coquand, Huet]. Преимуществами данной логической системы являются во-первых ее наибольшая (среди других конструктивных логик) выразительность, а во-вторых — наличие развитых компьютерных инструментов для работы с ней (например, системы автоматизации доказательств Coq). Однако здесь встает вопрос о представлении модальных категорий, например, времени. Мы полагаем, что для этого достаточно оперировать не простыми пропозициями, а пропозициями, снабженными функциями выделения из них частей (ролей), причем каждая такая часть обязана быть логическим следствием исходной пропозиции. С формальной точки зрения каждая такая функция соответствует некоторому модальному оператору; подобная процедура сведения модальной логики к пропозициональной хорошо известна и использовалась, например, Д. Гэллином применительно к интенциональной логике Монтегю [Gallin, p. 58–63]. Точный список ролей еще предстоит составить, но во всяком случае он должен будет включать следующие компоненты:

- роли, соответствующие шифтерным категориям (говорящий, адресат, время совершения высказывания и т. п.)



- роли, соответствующие тематическому членению предложения
- некоторое подмножество падежных ролей, как минимум состоящее из Агенса, Пациенса и Времени
- роли, ответственные за анафору

Строго говоря, такие роли должны сопоставляться не всем пропозициями, а только своего рода атомарным. Другие пропозиции получаются из них с помощью оператора импликации (что соответствует в классическом изоморфизме Керри-Ховарда различию между объектами и функциями). В таком случае композиции синтаксических структур соответствует ровно одна логическая операция, а именно, вариант *modusponens*: $\{ X \rightarrow Y \mid P \}, \{ X \mid Q \} \vdash \{ Y \mid P \& Q \}$.

Представление отдельных языковых объектов

Теперь мы рассмотрим логические соответствия некоторых лингвистических категорий:

- основы имен существительных: $\{ X \mid \text{Class}(X) \rightarrow P \}$, где *Class* есть функция, выделяющая из пропозиции роль **класса объектов**, а *P* есть инвариант лексического значения
- основы атрибутивных имен прилагательных: $\{ X \rightarrow Y \mid \text{Class}(Y) \rightarrow P \}$ (отсюда следует, что соответствием для сочетания $N + \text{AttrAdj}$ будет $\{ X \mid \text{Class}(X) \rightarrow P \& Q \}$)
- основа прилагательного *каждый*: $\{ X \rightarrow Y \mid \forall x : \text{Class}(Y) \exists y : \text{Ext}(Y), x \in y \}$. Здесь *Ext* есть функция роли-

экстенционала, т. е. множества множеств конкретных объектов, удовлетворяющих данной пропозиции.

- анафорическое местоимение: $\{ X \mid \text{Ext}(X) = \text{Anaph}(X) \& \text{Class}(X) \}$. Отсюда следует объяснение, почему невозможно **каждый он*: у анафорического местоимения роль *Class(X)* представляет собой фактически весь универсум, так что указанный выше инвариант для *каждый* не может быть выполнен.
- основа одноместного глагола: $\{ X \mid \forall z : \text{Subj}(X) \rightarrow P(z) \}$. В отличие от стандартных моделей, в нашем формализме глаголам, как и существительным, соответствуют атомарные пропозиции, а не функции, которые являются соответствиями падежных формантов и других модификаторов. Это позволяет единообразно описывать все разновидности подчинительных конструкций.

Таким образом, предлагаемый формализм обладает по крайней мере не меньшей выразительностью, чем обычная грамматика Монтегю. Однако для того, чтобы он мог быть использован в качестве практического инструмента, требуется еще проделать значительную работу по отысканию логических соответствий для различных лингвистических категорий.

Литература

Андреев А.В. Метод достаточных оснований в формальной семантике // Материалы XLIII Международной филологической



конференции (Санкт-Петербург, 11–15 марта 2014 г.). Секция прикладной и математической лингвистики. СПб.: Филологический факультет, 2014. 62 с. — С. 23–29.

Coquand T., Huet G. The Calculus of Constructions // Information and Computation. Vol. 76 (2–3). 1988. P. 95–120.

Gallin D. Intensional and Higher-Order Modal Logic With Applications to Montague Semantics. Amsterdam: North-Holland Publishing Company. 1975. 148 p.

Heim I. E-Type Pronouns and Donkey Anaphora // Linguistics and Philosophy. Vol. 13, Issue 2. 1990. P. 137–177.

Howard W.A. The formulæ-as-types notion of construction // To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism, Boston, MA: Academic Press, 1980. 606 p. — P. 479–490.

Montague R. The Proper Treatment of Quantification in Ordinary English // Philosophy, Language, and Artificial Intelligence: Resources for Processing Natural Language. Dordrecht: Springer Netherlands, 1988. XII, 421 p. — C. 141–162.

Moot R., Retoré C. The Logic of Categorical Grammars: A Deductive Account of Natural Language Syntax and Semantics. Heidelberg: Springer-Verlag Berlin Heidelberg, 2012. X, 300 p.