

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. М. Т. КАЛАШНИКОВА

Информационные технологии и письменное наследие

El'Manuscript-2012

Материалы IV международной научной конференции
Петрозаводск, 3–8 сентября 2012 года

Петрозаводск, Ижевск
2012

УДК 004.9
ББК 81.11+81.2-0
И741

Изданы при поддержке гранта РФФИ (проект № 12-06-06061-г),
гранта РГНФ (проект № 12-04-14154-г) и в рамках реализации комплекса
мероприятий Программы стратегического развития ПетрГУ на 2012-2016 г.

Ответственные редакторы:

В. А. Баранов, д-р филол. наук, проф.

А. Г. Варфоломеев, канд. физ.-мат. наук, доц.

Информационные технологии и письменное наследие [Текст] :
И741 материалы IV междунар. науч. конф. (Петрозаводск, 3–8 сентября
2012 г.) / отв. ред. В. А. Баранов, А. Г. Варфоломеев. — Петрозаводск ;
Ижевск, 2012. — 328 с.

ISBN 978-5-8021-1402-5

Сборник содержит материалы конференции, посвященной современ-
ным электронным средствам хранения, описания, обработки, исследования
и публикации памятников письменности и исторических источников.

УДК 004.9
ББК 81.11+81.2-0

ISBN 978-5-8021-1402-5

© Петрозаводский государственный
университет, 2012
© Ижевский государственный технический
университет им. М. Т. Калашникова, 2012

ПРЕДСТАВЛЕНИЕ МОДЕЛЕЙ СИНТАКСИЧЕСКОЙ СТРУКТУРЫ ПОЭЗИИ П. А. ВЯЗЕМСКОГО С ПОМОЩЬЮ ТЕХНОЛОГИИ XML

Н. Д. Москин, А. А. Лебедев, А. Г. Варфоломеев

Петрозаводский государственный университет, Петрозаводск

This article shows how XML technology can be used to represent the syntactic structure of P.A. Vyazemsky poetry.

Сложное предложение в современной лингвистике является объектом пристального научного интереса. Одна из немаловажных проблем, актуальных для данной сферы исследования — типологизация сложных предложений, описание их структуры и составляющих. Исследователи-грамматисты отмечают, что «трудности, с которыми сталкивается синтаксическая наука при решении проблем типологии сложного предложения, начинаются уже на первых этапах классификации» [Мишланов, 1996]. В связи с этим ощущается потребность в создании некоего механизма, позволяющего размечать сложные предложения с точки зрения их структуры, что облегчило бы их последующий анализ, установление взаимосвязей между частями предложения, а также определение роли того или иного сложного предложения в тексте произведения или в целом в творчестве писателя.

Механизм разметки сложных предложений разрабатывался для анализа творчества Петра Андреевича Вяземского, талантливого поэта XIX века. Выбор творчества данного поэта объясняется тем, что именно интеллектуальность поэзии, её глубина стала отличительной чертой творчества Вяземского; стремление к выстраиванию причинно-следственных, глубинных связей, нехарактерное для лирики в целом, нашло широкое отражение в творчестве этого выдающегося автора и мыслителя. Немаловажную роль в оформлении сверхсодержательности и интеллектуальности поэзии Вяземского сыграли именно сложные предложения.

Предлагаемый нами подход основан на способе связи между компонентами сложного предложения; основной критерий для разделения предложений — наличие или отсутствие союза. В том случае, если союз в предложении присутствует, предложения делятся на сложносочиненные и сложноподчиненные, в зависимости от типа союза и характера отношений между частями сложного предложения. Для русистики данный подход является традиционным, поскольку в русском языке существует особый класс служебных слов — союзов, а тип союза позволяет с большой точностью отнести предложение к сложносочиненным или сложноподчиненным.

В соответствии с этим подходом все сложные предложения делятся на:

- 1) бессоюзные (при отсутствии союза),
- 2) союзные (при наличии союза).
 - a. Сложносочиненные (сочинительная связь между частями предложения)
 - b. Сложноподчиненные (подчинительная связь между частями предложения).

Для поэтической лирики характерно употребление сложных предложений с разными видами связи, с большим количеством грамматических основ, взаимоотношения между которыми не всегда можно истолковать однозначно. Именно поэтому механизм, позволяющий выстраивать структуру сложного предложения, является прекрасным подспорьем для лингвиста, исследующего синтаксис того или иного автора, а также смежные вопросы, связанные с синтаксисом (языковая личность автора, идиостиль писателя, исследования языковой игры и т.п.). Подробнее см. работы Ю.Н. Караулова, В.В. Григорьева, В.З. Санникова и др.

В качестве примера рассмотрим фрагмент стихотворения П. А. Вяземского «Вечер на Волге» (1815 год), в котором индексами выделены части предложения:

[Их гений мужествен, как гений вод твоих]₁,
 [Когда гроза во тьме клубится над тобою]₂
 [И пеною кипят громады волн седых;]₃
 [Противник наглых бурь, он злобе их упорной
 Смеется, опершись на брег, ему покорный;]₄
 [Обширен их полет]₅, [как бег обширен твой;]₆
 [Как ты, сверша свой путь, назначенный судьбой,
 В пучину Каспия мчишь воды обновленны,]₇
 [Так славные их дни, согражданам священные,
 Сольются, круг сверша, с бессмертием в веках!]₈

Теоретико-графовая модель синтаксической структуры этого предложения представлена на рис. 1. Важным отличием подобных моделей является иерархическая организация графа. Так, вершина под номером 1 соединена с графом, образованным вершинами с номерами 2 и 3 (такие графы выделены пунктиром). Кроме этого, связи делятся на два типа: направленные (например, от 5 к 6) и ненаправленные (например, от 2 к 3). Подобные схемы можно построить для всех предложений стихотворения и объединить их общей фиктивной вершиной.

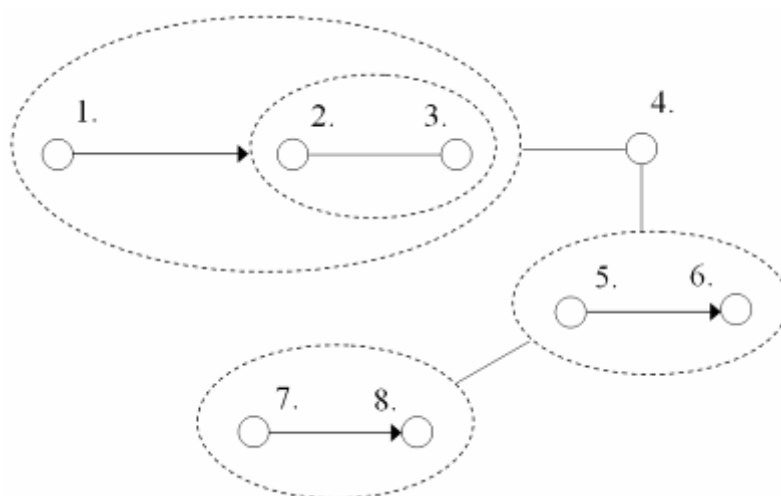


Рис. 3. Модель синтаксической структуры предложения
 «Их гений мужествен, как гений вод твоих»

Заметим, что в компьютерной лингвистике уже достаточно давно известны методы представления синтаксической структуры текстов в виде графов (деревьев). Например, в работе И.П.Севбо [Севбо, 1983] приводится описание так называемых деревьев зависимостей. Здесь при анализе предложения устанавливаются подчинительные связи между словами согласно постулатам традиционной грамматики и грамматики зависимостей. Результаты анализа представляются в виде графов, вершины которых соответствуют словам, а дуги соединяют их в соответствии с синтаксическими свя-

зьями. Такие модели обычно используются в описаниях текстов на языках со свободным порядком слов (например, русском).

Для хранения и дальнейшего анализа подобных графов мы предлагаем использовать технологию XML. В настоящее время разработано несколько XML-стандартов описания графов и графовых моделей. Одним из предшественников таких форматов является язык GML (Graph Modelling Language) [Himsolt, 1997]. Впоследствии на базе GML были созданы другие форматы, например, XGMML (eXtensible Graph Markup and Modeling Language). В 2000 году на 8-ом симпозиуме «Graph Drawing» был предложен язык описания графов GraphXML [Herman, 2001]. На этом языке могут быть описаны как абстрактные графы, так и более сложные структуры: иерархии графов, динамические графы и т. д. Позднее появились и другие спецификации: GraphML (Graph Markup Language), GXL (Graph eXchange Language), GraX и др. [Москин, 2009]

Однако данные форматы предназначены для описания произвольных графов и графовых моделей, не привязанных к тексту. Поэтому возникла необходимость разработки языка для формального описания и анализа теоретико-графовых моделей текстов — TextGML (Textual Graph Modelling Language) [Москин, 2009]. Рассмотрим, как можно применить этот язык для разметки синтаксической структуры стихотворений П.А.Вяземского. Каждый граф здесь задается с помощью тега <graph> с уникальным идентификатором id, названием name и атрибутом directed, указывающим будет ли граф ориентированным (например, на рис. 1 таких графов — пять).

Для разметки частей предложения используется тег <node>, например: <node id="id1" name="n1" type="часть предложения">Их гений мужествен, как гений вод твоих</node>. Связи определяются с помощью тега <link>, например: <link id="l6" source="n7" target="n8" type="СПП с придаточной сопоставительной частью" order="6"/>. В качестве значений атрибутов source и target используются названия инцидентных вершин (атрибут name).

Для организации иерархической структуры графа применяются фиктивные вершины с атрибутом id_graph:

```
<node id="id11" name="n11" type="graph" id_graph="g4"/>
.....
<graph id="g4" name="объединение вершин-3" directed="true">
  <link id="l4" source="n5" target="n6" type="СПП с придаточным
  собственнo-сравнительным" order="4"/>
</graph>
```

Для ввода текстов и моделей их синтаксической организации в электронную коллекцию, а также дальнейшей разметки и анализа необходимо специализированное программное обеспечение. Решением этой проблемы мог бы стать специализированный Интернет-ресурс, реализованный на языке PHP. Он позволил бы вводить XML-документы в базу данных, визуализировать структуру моделей, проводить поиск по коллекции и сравнительный анализ графов. Данный ресурс может быть полезен для организации распре-

деленных научных исследований в сети Интернет и публикации их результатов в рамках деятельности сетевых сообществ исследователей.

Работа выполнена при финансовой поддержке Программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

Список литературы

Мишланов, 1996 — Мишланов В.А. Семантика и структура сложного предложения в свете динамического синтаксиса. Пермь, 1996.

Москин, 2009 — Москин Н.Д. Решение задач визуализации и поиска мотивов в электронной библиотеке фольклорных текстов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XI Всероссийской конференции RCDL'2009. Петрозаводск, 2009. С. 465–471.

Севбо, 1981 — Севбо И.П. Графическое представление синтаксических структур и стилистическая диагностика. Киев, 1981.

Himsolt, 1997 — Himsolt M. GML: A portable Graph File Format // Technical report. University at Passau, 1997.

Herman, 2001 — Herman I., Marshall M. S. GraphXML — An XML-based graph description format // Proceedings of Graph Drawing 2000. LNCS, Vol. 1984. Springer: Berlin, 2001. Pp.52–62.