

**ИССЛЕДОВАНИЕ ЧАСТОТНОГО РАСПРЕДЕЛЕНИЯ АНАФОРЫ В РАЗНОУРОВНЕВЫХ  
АДАптиРОВАННЫХ РУССКОЯЗЫЧНЫХ ТЕКСТАХ КАК ПРИЛОЖЕНИЕ К ЗАДАЧЕ  
АВТОМАТИЧЕСКОГО РАЗРЕШЕНИЯ РЕФЕРЕНЦИИ**

А. В. Койнов, А. П. Савинов, Т. С. Петровская, Е. В. Михалёва

Национальный исследовательский Томский политехнический университет,

Россия, г. Томск, пр. Ленина, 30, 634050

E-mail: koynovav@gmail.com

**THE STUDY OF ANAPHORA FREQUENCY DISTRIBUTION IN MULTILEVEL ADAPTED RUSSIAN  
TEXTS AS A SUPPLEMENT TO THE AUTOMATIC REFERENCE RESOLUTION TASK**

A. V. Koynov, A. P. Savinov, T. S. Petrovskaya, E. V. Mikhalyova

Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

E-mail: koynovav@gmail.com

***Annotation.** The article considers the problems of automatic anaphora resolution in Russian texts and provides a brief description of reference classifications. The analysis of frequency distribution of model anaphora is performed on the training materials suggested to the international students within the learning process by the department of Russian as Foreign, TPU. The course material given in the study covers the following knowledge levels: elementary level (A1), basic level (A2) and first certification level (B2). The purpose of the given article is to provide statistical assessment of these or those types of anaphora depending on the level of the material considered, as well as to determine what types are more common and to indicate the future direction in the study of automatic reference resolution problems of Russian texts.*

Проблема референции русскоязычных текстов на данный момент остается нерешенной проблемой лингвопроцессорных систем. Одной из таких систем является интеллектуальная обучающая система КЛИОС [1], разрабатываемая в ТПУ для обучения русскому языку как иностранному. Данная система отличается от существующих электронных обучающих систем наличием интеллектуального подхода к обработке естественно-языковых (ЕЯ) текстов, в том числе и ответов студента. Одной из важнейших задач настоящего времени является извлечение знаний из ЕЯ-текстов, построение семантического графа, описание связей между объектами, их действиями, а значит, и определение когерентных связей на протяжении всего дискурса. Анафора (анафорическое выражение) является одним из средств связи предложений, которая определяется наличием её референта в тексте. Пример: *Если днем и не было дождя [референт], то казалось, что он (дождь) [анафора] вот-вот начнется.* В данном случае именная группа словоформы *дождь* является антецедентом анафорического местоимения *он*, и считается, что анафора реферирует с антецедентом. Рассмотрим другой пример: *Гражданка Иванова, попала в автомобильную аварию, которая<sub>1,2</sub> произошла сегодня утром на улице Ленина. После проведения экспертизы сотрудники ГИБДД признали её<sub>1,2</sub> виновником ДТП.* Пример наглядно показывает, что анафоры имеют неоднозначность в выборе реферирующего с ним антецедента. Алгоритмы снятия реферирующей неоднозначности были продемонстрированы в системе АОТ [2] А. В. Сокирко, в работах П. В. Толпегина по автоматическому разрешению референции местоимений третьего лица [3] и других. Предложенные авторами алгоритмы решают описываемую проблему лишь отчасти и

затрагивают только отдельные группы местоименной референции: точность разрешения анафоры - 70,6% - 85,2% [3].

В настоящем исследовании будут рассмотрены следующие виды референции [3]: референция местоимений (личные (*он*), указательные (*этот*), вопросительные (*который*) и другие), референция обстоятельственных наречий места (*здесь, там*), а также синонимические замены отдельных слов или сложных конструкций (*президент – глава государства*). Целью настоящего исследования является определение частоты возникновения того или иного типа референции в текстах адаптированных учебных материалов для разработки алгоритма решения прикладной задачи автоматического разрешения анафоры. Материалом для исследования послужили учебные тексты, разработанные на специализированной кафедре русского языка как иностранного. Для получения максимально валидных результатов и определения специфики проявления различных типов референций дифференцировано рассматривались тексты, ориентированные на разный уровень владения языком: от элементарного уровня (A1) до первого сертификационного уровня (B1) [4].

Результаты показали, что в материалах уровня A1 преобладала референция личных местоимений: 95% из общего числа выявленных 715 случаев референции. Из них около 39%, составляли местоимения 1 лица, 32% - второго и 24% - третьего лица, определен лишь единственный случай синонимии. Такой результат обоснован тем, что материал элементарного уровня строится преимущественно на простейших диалогах. При рассмотрении базового уровня A2 выявлено 1896 случаев референции, из них личные местоимения составили 87%, причем большую долю занимают личные местоимения 3 лица (57%). Референция вопросительных и указательных местоимений возросла до 8% и 4% соответственно. С увеличением сложности материала случаи синонимических замен отдельных слов существенно возросли, но они по-прежнему составляют менее 1% от общего числа референции. В материалах уровня B1 было выявлено 1757 случаев референции, из них 90% - личные местоимения 1 лица (39%), 2 лица (14%), 3 лица (37%), 3% - вопросительные местоимения, 6% - указательные и 1% - синонимия слов.

Результат исследования наглядно показывает, что во всех группах текстов большую долю референций составляет референция личных местоимений, а именно местоимений 3 лица (44% от общего числа представленных референций). Данный результат исследования определяет приоритетное направление в области автоматического разрешения референций в тексте, а также определяет направления работы по увеличению эффективности алгоритмов автоматического разрешения референции.

## **СПИСОК ЛИТЕРАТУРЫ**

1. Горисев С. А. и др. Интеллектуальный лингвопроцессорный комплекс «КЛИОС» для обучения РКИ // Современные проблемы науки и образования. — 2013. — № 6.
2. Сокирко А. В. Первичный семантический анализ [Электронный ресурс]. — Режим доступа: <http://aot.ru/docs/seman.html>. — 28.02.14.
3. Толпегин П. В. Новые методы и алгоритмы автоматического разрешения референции местоимений третьего лица русскоязычных текстов. — М.: КомКнига, 2006. — 86 с.
4. Красман В. А. и др. Давайте знакомиться! — Томск: ТПУ, 2011 — 109 с.; Михалёва Е. В. и др. Русский язык как иностранный: элементарный уровень. — Томск: ТПУ, 2011 — 378 с.; Ильиных О. В. и др. Полный вперед! — Томск: ТПУ, 2005. — 150 с.