

Крылов В.С.

УДК 81.322.3:004.414.2

## КОМПЬЮТЕРНАЯ ПАРАЛИНГВИСТИКА КРЕАЛИЗОВАННЫХ ТЕКСТОВ

**Аннотация.** Информационные технологии (ИТ) сформировали глобальную инфраструктуру связи между людьми. В сетях интернета возникли и активно развиваются контактные языки, подобные пиджин и креольским языкам. Синтаксис и грамматика таких языков отличается от синтаксиса и грамматики базового языка. Эффективный анализатор структуры сообщений креализованного текста - обработчик событий.

**Ключевые слова:** компьютерная паралингвистика, событийно-ориентированный подход, креализованный язык

**Анотація.** Інформаційні технології (ІТ) сформували глобальну інфраструктуру зв'язку між людьми. В мережах інтернету виникли і активно розвиваються контактні мови, подібні пиджин і креольським мовам. Синтаксис і граматики таких мов відрізняється від синтаксису і граматики базової мови. Ефективний аналізатор структури повідомлень креализованного тексту - оброблювач подій.

**Ключові слова:** комп'ютерна паралингвистика, подієво-орієнтований підхід, креализованний мову

**Summary.** Information technology (IT) have formed a global infrastructure links between people. In the networks of the Internet emerged and are actively developing the contact languages like Pidgin and Creole. The syntax and grammar of languages differs from the syntax and grammar of the base language. Effective communications analyzer structure krealizovannogo text - the event handler.

**Keywords:** computer paralinguistics, event-oriented approach, krealizovanny language

Информационные технологии (ИТ) сформировали глобальную инфраструктуру, обеспечивающую передовые услуги за счет организации связи между людьми, и даже между вещами физическими или виртуальными. Общение в сети интернет, обмен текстами в сети повлек за собой появление новых языковых особенностей, лингвистические характеристики которых определяются соответствующим каналом коммуникации в информационном пространстве. Такими, как электронная почта, мгновенный обмен сообщениями в чатах, СМС сообщения, твиттер и так далее. Просто текстовых сообщений для передачи, например, эмоционального состояния коммуниканта оказалось недостаточным. В результате, спонтанно возникли смайлики. Текст сообщения стал дополняться кодами, не имеющими прямого отношения к алфавиту и грамматике сообщения. Синтаксис, грамматика таких сообщений значительно отличаются от синтаксиса, грамматики базового языка коммуникантов. Эти языки определяются как контактные креализованные языки, подобные пиджин и креольским языкам [5,6].

Совершенствование информационных технологий (ИТ) предоставило широкому кругу исследователей доступные аппаратные и программные средства сбора, обработки исходных данных и их анализа. Появилась возможность использовать ранее недоступные инструменты и технологии для экспериментальных и теоретических исследований в самых различных научных направлениях. В первую очередь, в междисциплинарных исследованиях. На основе этих технологий возникают новые интегрированные научные направления. Например, компьютерная паралингвистика предлагает модели, которые позволяют значительно усовершенствовать системы анализа и обработки естественной речи и текстов, обработки сообщений в социальных сетях, усовершенствовать поисковые системы [3].

**Цель настоящей статьи** - предложить модель организации и анализа сообщений креализованных текстов на основе моделей компьютерной паралингвистики и объектно-ориентированного и событийно-ориентированных подходов в программировании.

Язык, как система фонетических, лексических и грамматических средств, порождает естественную информационную реальность, через которую обеспечиваются коммуникации в обществе. Внедрения информационных технологий (ИТ) в самые различные сферы деятельности людей, стремительное развитие сети интернет первоначально рассматривалось как дополнение или расширение естественного информационного пространства. Совершенствование технических и программных средств из дополнения естественного информационного пространства превратилось в самостоятельную информационную реальность, а общество незаметно для себя стало информационным.

В каналах мгновенных текстовых сообщений интернет чатов, твиттере, форумов, СМС и других идет активное словообразование, например, через образования новых аббревиатур и сокращений. В тексты обменов активно включаются смайлики, визуальные и аудио элементы. Текст сообщения становится поликодовым, в котором все элементы не являются просто суммой, объединением, а интегрированы единое целое для передачи смысла. Такие тексты определяются как креализованные [5]. То есть тексты, состоящие из двух объединённых в единую структуру частей: вербальной (речевой) и невербальной, которая относится к другим знаковым системам. Обе эти части зависят друг от друга. Вербальную часть невозможно отделить от невербальной [5, 10].

Рассмотрим, например, обмены сообщениями в популярной социальной сети Инстаграм. Она предназначена для размещения собственных фотографий о каких-либо событиях, пейзажах и личных изображений на этом фоне [1]. В конечном итоге фото размещают для получения отзывов или комментариев, как от всех возможных участников сети, так и для определенного круга пользователей. Комментарии являются откликами, и вполне соответствуют определению понятия события, и составляют поток событий. Они сами по себе могут вызывать отклики, то есть являются событием, которое представляет определенную реакцию на данный комментарий.

Событие, как понятие, имеет различные определения в различных отраслях деятельности, ситуациях, обстоятельствах. В данном случае событие представляет собой сообщение, посылаемое объектом, чтобы сигнализировать о совершении какого-либо действия. Любой объект, способный вызывать события, является *отправителем событий*, также называемым *источником событий*. Отправителю события зачастую не известен объект, который будет получать, и обрабатывать сформированные отправителем события. Получатель должен обладать *обработчиком событий* - процедурой, исполняемой при получении соответствующего события [2, 7].

Поток комментариев представляет собой поток событий. Во всяком случае, участники сети воспринимают их как события, и принимают решение как на них реагировать. Ответ или отклик определяется в соответствии с представлениями об объекте и внутренним состоянием на момент ответа. Он может быть представлен текстом, фото, видео, смайликом, а так же какой-либо их комбинацией. Сообщение комментария представляет собой некоторый креализованный текст, отображающий вербальную и невербальную оценку события. Каждая из этих компонент формирует сообщение по своим собственным правилам. Синтаксис и грамматика вербальной части, как правило, отличаются от синтаксиса и грамматики базового языка. Существует определенная зависимость между получаемым, например, визуальным событием и структурой предложения отклика на событие [8].

Размещенные в текстах комментариев, сообщений различные знаки, символы и смайлики имеют определенные коды, поэтому возможно составлять шаблоны поиска и анализа сообщений, составленных разными знаковыми системами. То есть возможен последовательный поиск, анализ собственно текстов и отдельно последовательностей символов или знаков, отображающих невербальную компоненту сообщения. И, кроме того, возможен поиск и анализ организации объединенной вербально-невербальной последовательности символов.

Структуру сообщений на креализованном языке можно представить, как последовательности некоторых шаблонов, объединяющих разные знаковые системы. Такая модель дает возможность в едином ключе, на единой основе обеспечивать интегрированное целостное представление сообщения, объединения лингвистических паралингвистических факторов в речевой активности. Очевидно, что для выделения объединений из разных знаковых систем, шаблоны могут быть разного уровня сложности. Они являются источником моделей анализа структуры реальных сообщений на контактных языках. Например, для упрощенного креализованного языка с некоторым ограниченным набором слов и смайликов или некоторых изображений. Такие языки распространены в рекламной деятельности [5].

Для поиска и анализа синтаксической структуры самых разнообразных текстов размещенных в сети интернет активно используются программы парсеры. Эти программы разрабатываются для автоматизированного поиска, сбора и анализа различной информации: текстов кодов программ, фильмов, музыки, тематических архивов документов, отдельных документов, рецензий на события и так далее. Процесс поиска и анализа такой программой определяется как парсинг. Поэтому в основном под парсингом обычно понимается визуальный, автоматический синтаксический и лексический анализ текстов, разбор какого-либо документа для извлечения из него определенных данных. Часто востребованный результат парсинга заключается в поиске контента, содержащего не только требуемые ключевые слова, но и определенную синтаксическую организацию текста.

Особенность исполнения парсинга определяется заданием схемы отбора информации. Параметры поиска и анализа задаются заранее. Собранную информацию парсер предоставляет в определенном виде. Например, результатом работы парсера может быть база данных [9].

Наиболее часто парсинг опирается на анализ по образцам, составленным из регулярных выражений. Регулярные выражения - система обработки текста, основанная на специальной системе записи образцов для поиска [7, 9]. Образец или паттерн, задающий правило поиска, называют шаблоном или маской. Регулярные выражения используются многими текстовыми редакторами и утилитами для поиска и изменения текста на основе выбранных правил. В настоящее время сложно найти язык программирования, который не поддерживает регулярные выражения для работы со строками, составленные из различных символов. Язык, который не имеет встроенный синтаксис и механизм обработки регулярных выражений. Регулярные выражения используются для сжатого описания некоторого множества строк с помощью шаблонов, без необходимости перечисления всех элементов этого множества, и при составлении, которых используется специальный синтаксис.

Шаблоны из регулярных выражений образуют некоторый мета язык. Собственно этот мета язык и образует модель организации анализируемых сообщений представленных некоторым креализованным языком. Кроме того, такой шаблон означает некоторую структуру, которая воспринимается как событие. Анализ последовательностей встретившихся единиц сообщения, соответствующих некоторому шаблону, требует обработчика событий. Поэтому анализ структуры сообщений креализованного текста будет являться результатом работы этого обработчика событий.

Такой подход в определении интегрированных моделей компьютерной паралингвистики на основании особенностей выбранной платформы программирования, имеет не только научное или прикладное значение для развития и совершенствования существующих информационных систем. Он позволяет формулировать различной сложности задания в учебном процессе подготовки специалистов в области информационных технологий. Выполнение заданий стимулирует развитие навыков и умений приобретать дополнительные знания, осваивать как можно больше востребованных информационных технологий, умение разрабатывать программные продукты для разных отраслей, сфер и направлений практического их применения.

Включение в учебный процесс достижений прикладных исследований новых направлений, таких как компьютерная лингвистика, компьютерная паралингвистика позволяет будущему специалисту приобрести

навыки эффективного решения поставленных заказчиком задач. Именно на таком материале, на основе приобретенного опыта, молодой специалист получает необходимый набор профессиональных навыков успешной конкуренции на ИТ-рынке труда [4].

#### Источники и литература:

1. Все о программе Инстаграм (Instagram) по-русски [Электронный ресурс] / Режим доступа : <http://instagram.ru/>
2. Концептуальная модель для систем обработки событий [Электронный ресурс] / Режим доступа : <http://www.ibm.com/developerworks/ru/library/ws-eventprocessing/index.html>
3. Крылов В. С. Компьютерная паралингвистика как основа приобретения специальных профессиональных навыков в области информационных технологий / Крылов В. С. – Science and Education a New Dimension. Philology, II(4), Issue : 24, 2014 С. 41 – 44
4. Крылов В. С. Молодые специалисты способны конкурировать на рынке труда информационных технологий. / Крылов В. С. – Science and Education a New Dimension: Pedagogy and Psychology. Vol. 5, 2013 – P. 98 – 101
5. Beisenova Z. S. The Study of Creolized Texts in Written Communication / Beisenova Z. S., Bayekeyeva A. T., Sapina S. M., Dinayeva B. B., Utanova A. K. – INDIAN JOURNAL OF APPLIED RESEARCH, Volume : 4, Issue : 5, May 2014 217 p.
6. Bickerton D. Creole Languages / Bickerton D. – Scientific American, July 1983, Vol. 249, No.1, P. 116–122
7. Event-Based Programming. Taking Events to the Limit [Электронный ресурс] / Режим доступа : <http://sharebookfree.com/event-based-programming-taking-events-to-the-limit/>
8. Myachukov A. Visual Cues to Structural Choice in Visually Situated Sentence Production [Электронный ресурс] / Myachukov A., Thompson D., Garrod S., Scheepers C. – Frontiers in Psychology, Published online Jan 18, 2012 – Режим доступа : <http://journal.frontiersin.org/Journal/10.3389/fpsyg.2011.00396/full>
9. Kallmeyer L. Parsing Beyond Context-Free Grammars / Kallmeyer L. – Springer-Verlag Berlin Heidelberg 2010, – 247 p.
9. Krämer N. C. Nonverbal Communication [Электронный ресурс] / Krämer N. C. – Human Behavior in Military Contexts, P. 150 – 188, Режим доступа : <http://books.nap.edu/catalog/12023.html>

Масликова О.С.

УДК 81'373.4=161.2(043.5)

### ДО ПИТАННЯ ПРО ФУНКЦІОНУВАННЯ В УКРАЇНСЬКІЙ МОВІ ЛІТЕРАТУРОЗНАВЧИХ ТЕРМІНІВ ДЛЯ НАЗИВАННЯ СТРУКТУРНИХ ЧАСТИН ТА ПОЕТИЧНИХ ЕЛЕМЕНТІВ ХУДОЖНЬОГО ТВОРУ

*Анотація.* У науковій статті висвітлено питання про функціонування в українській мові термінологічних одиниць для позначення компонентів, структурних одиниць та поезики художнього твору на основі праць

М.С. Грушевського, який утверджував літературний варіант української мови на межі XIX – XX століть, вирішував багато проблемних питань у своїх наукових розвідках. Зазначається, що вчений часто коментує відтінки значення та аспекти вживання окремих одиниць, вказує на походження філологічних термінів та номенклатур із тематичної групи «терміни на позначення компонентів, структурних одиниць та поезики художнього твору».

**Ключові слова:** структурні одиниці (твору), вступне слово, передмова, стопа, строфа, композиція, експозиція, частина (твору), фінал, інтегральна частина, художність, контекст.

*Аннотация.* В материале научной статьи анализируется вопрос о функционировании в украинском литературном языке терминологических единиц, обозначающих компоненты, структурные единицы и поэтические особенности художественного произведения по трудам М.С. Грушевского. Ученый активно формировал литературный вариант украинского языка на рубеже XIX – XX веков. Решая много проблемных вопросов в своих научных исследованиях, М. Грушевский часто комментирует оттенки в значениях и аспекты использования терминологических единиц тематической группы «термины на обозначение компонентов, структурных единиц и поэтики художественного произведения».

**Ключевые слова:** структурные единицы (произведения), вступительное слово, предисловие, стопа, строфа, композиция, экспозиция, часть (произведения), финал, интегральная часть, художественность, контекст.

**Summary.** The scientific article examines the usage terminological units in the Ukrainian literary language, which denote the components, structural units and poetic features of literary works by M.S. Grushevskiy. The scientist formed actively the literary variant of the Ukrainian language at the turn of the XIX - XX centuries. Solving many problematic issues in his scientific researches, M. Grushevskiy often comments on the shades of the meanings and aspects about the usage of terms as well as he indicates to the origin of the individual philological terms and nomenclatures. Significant and important thematic group that stands out among the rest ones is the terminological units that serve for the "designation of the components, the structural units and the poetry of the literary works". M. Grushevskiy explores fundamentally the history of the Ukrainian literature and analyzes the literary values. Moreover, in his works he also highlights the terms that have already been recorded both in the Ukrainian periodicals: the introductory word, preface and in the scientific literature: epilogue, introduction, preface, hexameter, dactyl, foot, verse, iambic and others.