**Разработка компьютерной языковой системы обучения японскому письму с использованием техники обработки естественного языка: Изучение страдательного залога.**

Jie Chi YANG, Kanji AKAHORY

Эта работа описывает разработку компьютерной языковой системы обучения (computer assisted language learning system– CALL-система) японскому письму с использованием техники обработки естественного языка (natural language processing – NLP–техника). Эта система может быть использована при письме для изучения японского страдательного залога с помощью WWW. Для выявления типов ошибок, возникающих при написании страдательного залога японского языка у студентов-иностранцев, авторы проводили тест и делали обзор вопросника. В результате этих исследований авторы классифицировали типы ошибок в японском страдательном залоге на 12 категорий, 65 видов и 228 ошибок. Эта классификация используется для анализа ошибок, проводимого системой. В этой системе авторы используют средства NLP (включая анализатор морфем и синтаксический анализатор), а затем добавляют анализ ошибок и обработчик обратной связи. Таким образом, система, давая обучаемому возможность свободно вводить предложения, может обнаружить ошибку в напечатанных (введенных обучаемым с клавиатуры) предложениях и выдавать обучаемому адекватные сообщения обратной связи. Кроме того, авторы предлагают механизм исправления обучаемым, позволяющий обучаемому корректировать напечатанные предложения самому, что поможет ему лучше понять те ошибки, которые он совершил. Этот механизм может быть рассмотрен как дополнение к блоку прогнозирования системы и как усиление эффективности данной CALL-системы.

**Введение**

Эффективность CALL-систем была описана многими учеными [4],[7]. Недостатком существующих CALL-систем является то, что обучаемый не может свободно вводить с клавиатуры заданные языковые выражения, особенно это относится к компьютерным системам обучения письму. Обучаемому доступна лишь информация, сопровождающая правила учебного курса, заранее установленного на компьютер. Поэтому нельзя назвать такой тип систем полностью интерактивными. В результате этих причин проводится все больше и больше исследований [1],[3],[7]в области применения NLP-техники в CALL-системах. Исследования [2],[5],[9] анализа ошибок в системах обучения письму были особенно усилены превосходными результатами эксперимента [6] в области выносливости (устойчивости к ошибкам) средств NLP.

В CALL-системах анализ напечатанных предложений является необходимостью, так как обучаемый должен иметь возможность свободно формулировать предложения без каких-либо предописанных правил. Поэтому мы используем средства NLP для анализа напечатанных предложений.

К сожалению, почти все теории грамматики и техника NLP, предложенные до сегодняшнего дня могут анализировать только грамматически корректные предложения. Однако обучаемые японскому языку, использующие CALL-систему, скорее всего, напечатают неправильные предложения. Таким образом, мы можем или добавить анализатор ошибок в существующие, нескорректированные NLP-средства, или скорректировать NLP-средства для решения этой проблемы. В этой статье мы применили первый метод. Однако, это сопряжено с проблемой неопознания правильной морфемы или правильной части речи слова. Трудность заключается в обнаружении правильных морфем даже в грамматически корректных предложениях, так как в японском языке предложения пишутся непрерывно, без каких-либо промежутков1. Поэтому в случае неправильно введенных предложений они будут интерпретированы как грамматически корректные. Неправильная морфема будет признана правильной существующими средствами NLP. Например, правильное предложение с правильными морфемами «Jyon san wa kaku san ni nagura re mashita.» (Джон ударен мистером Каку) будет анализировано как «Jon san wa kaku san nina kura re mashita.» (Джон перевернут мистером Каку), если правильное слово «naguraremashita» будет введено как неправильное слово «inakuraremashita». Вот, почему мы представляем разработанную систему. Настоящая система может обнаруживать правильные морфемы, даже если было введено неправильное предложение.

В нашей системе мы использовали метод, который прогнозирует структуры неправильного ввода, а затем сохраняет их с целью анализа ошибок. Другими словами, он сводится к исследованию возможных структур некорректных предложений в страдательном залоге, логически вытекающих из понимания правил определения корректности предложений.

Настоящая работа описывает CALL-систему, использующую NLP-технику, которая может помочь устранить недостатки существующих CALL-систем. Представленная система позволяет обучаемому свободно вводить с клавиатуры предложения, она может обнаруживать ошибки в напечатанных предложениях и выдавать обучаемому адекватные сообщения обратной связи и может быть использована для обучения написанию японского страдательного залога с помощью любого WWW браузера.

**Структура и реализация системы**

2.1. Cхема и алгоритм системы

На рисунке 1 изображена схема системы. Система состоит из интерфейса, то есть, WWW браузера, системы анализа предложений, системы обратной связи, словаря и грамматических правил. Система анализа предложений включает в себя анализатор морфем (JUMAN 3.0), синтаксический анализатор (KNP 1.1) и анализатор ошибок. Средства NLP – JUMAN и KNP – были разработаны лабораторией Нагао университета Киото (Япония). Эти средства могут анализировать только грамматически корректные японские предложения. Система обратной связи включает в себя генератор сообщений обратной связи, базу знаний и истории всех обучаемых, работавших с системой. Кроме того, в системе используются электронные словари EDR1.52, IPAL3 и Bunrui goi hyou4.

Перед началом использования системы необходима регистрация, включающая в себя указание имени обучаемого и опыта изучения им японского языка. После регистрации обучаемый управляет системой в соответствии с инструкциями, написанными на каждой Web-странице. Страницы обучения состоят из большого количества картинок, которые могут быть выбраны обучаемым. Когда обучаемый выбирает одну из них, то загружается новой страница с увеличенной копией картинки и текстовым окном под ней. Затем обучаемый может наглядно просмотреть картинку и ввести по-японски предложение в страдательном залоге в текстовом окне. Текст может вводиться в буквах обоих алфавитов («канджи» или «кана») японского языка, как это принято в большинстве японских текстовых редакторов. Таким образом, система выдает обучаемому адекватные сообщения обратной связи в соответствии с введенными фразами, если были обнаружены какие-либо ошибки.

Сноски:

Однако выражения на «ромайи» (романский (латинский) алфавит японского языка) пишутся с пробелами между словами. В случае «ромайи» идентифицировать правильное или неправильное предложение не составит труда.

EDR1.5: Japan Electronic Dictionary Research Institute, Ltd - Исследовательский Институт Электронных Словарей).

IPAL: IPA (The Information-technology Promotion Agency) Lexicon of Japanese Language for computers. - (Агентство Развития Информационных Технологий) Лексикон Японского Языка для компьютеров.

Bunrui goi hyou: один из тезаурусов издательства Syuuei publication, Токио.

Алгоритм системы:

Обучаемый печатает по-японски предложение.

Напечатанное предложение анализируется «Системой Анализа Предложений», включающей в себя анализатор морфем (JUMAN), синтаксический анализатор (KNP) и анализатор ошибок.

Код ошибки, определенный системой анализа предложений, посылается системе обратной связи.

Если была ошибка, система заставляет обучаемого исправить ее с повторением всех шагов.

Рисунок 1 Схема системы

**Система обратной связи и исправление обучаемым**

Информация об ошибках поступает на систему обратной связи от системы анализа предложений в виде кода ошибки. Адекватные сообщения обратной связи выдаются в соответствии с кодом ошибки. В этой системе авторы предлагают механизм исправления ошибок обучаемым, позволяющий обучаемому лучше понять те ошибки, которые он совершил. «Исправление обучаемым» означает, что когда анализатор ошибок обнаруживает ошибку, обучаемому выдается не готовый ответ, а соответствующее введенному предложению сообщение обратной связи. Обучаемый может или прочитать сообщение обратной связи сразу и исправить введенное предложение, или он может обратиться к грамматическим темам, освещающим его ошибку и исправить предложение позже. Сообщения обратной связи с каждым исправлением становятся более детальными. Например, если это ошибка в спряжении глагола, то будет выдано примерно такое сообщение обратной связи «Ошибка в глаголе», и несколько примерно таких ссылок «Таблица глаголов», «Практика спряжения глаголов», «Что такое страдательный залог» и т.п. - внизу сообщения. Обучаемый может обратиться к этим грамматическим темам и исправить ошибку самостоятельно. Конечно же, предложение можно исправить и непосредственно после выдачи сообщения обратной связи.

Рисунок 2 иллюстрирует сообщения обратной связи о неправильном спряжении глагола в страдательном залоге.

Рисунок 2. Сообщение обратной связи о неправильном спряжении глагола в страдательном залоге.

**Заключение**

В этой работе авторы описывают разработку CALL-системы, основанной на средствах NLP, целью которой является обучение страдательному залогу японского языка. Эта система позволяет обучаемому свободно вводить с клавиатуры предложения по-японски и может обнаруживать ошибки в напечатанных предложениях и обеспечивать обучаемому выдачу адекватных сообщений обратной связи. Обучаемый может использовать эту систему где угодно и когда угодно, так как эта система запускается на любом WWW-браузере. Таким образом этот подход может быть рассмотрен как новый метод в обучении языку (особенно японскому языку в нашем случае). Как уже было описано, мы не используем существующие средства NLP в том виде, в котором они есть. Вместо этого мы применяем анализатор ошибок для анализа неправильно введенных предложений. Кроме того, мы используем метод, прогнозирующий структуры неправильного ввода и сохраняющий их внутри системы для анализа ошибок. С этой целью авторы проводили обзор для выявления типов ошибок, возникающих при написании страдательного залога японского языка у студентов-иностранцев. Хотя это и утомительная работа, но она помогает в накоплении данных об ошибках, которые могут быть позднее использованы для анализа ошибок.

Мы планируем расширения системы, например: увеличение базы ошибок, различные сообщения обратной связи разным людям и более гибкие правила анализа ошибок при использовании словаря понятий EDR.

**Список литературы**

1.Holland, V.M., and Kaplan, J.D. (1995). Natural Language Processing Techniques in Computer Assisted Language Learning: Status and Instructional Issues. Instructional Science, 23, 351-380.

2.Juozulynas, V. (1994). Errors in the Compositions of Second-Year German Students: An Empirical Study for Parser-Based ICALI. The CALICO Journal, 12(1), 5-15.

3.Loritz, D. (1992). Generalized Transition Network Parsing for Language Study: The GPARS System for English, Russian, Japanese and Chinese. The CALICO Journal, 10(1), 5-22.

4.Lam, F.S., and Pennington, M.C. (1995). The Computer vs. the Pen: A Comparative Study of Word Processing in a Hong Kong Secondary Classroom. Computer Assisted Language Learning, 8(1), 75-92.

5.Lee, K.K., Konishi, T., Takagi, A., Shirai, K., and Ohara, H. (1994). Composition Error Analysis and Advice Strategy Using a New Japanese-Korean Composition Training ICAI System. Transactions of Information Processing Society of Japan, 35(7), 1223-1234. In Japanese.

6.Matsumoto, Y., and Imaichi O. (1994). Current Issues in Robust Natural Language Processing. In Proceedings of the First SIG-SLP Conference of the Information Processing Society of Japan. In Japanese.

7.McEnery, T., Baker, J.P., and Wilson, A. (1995). A Statistical Analysis of Corpus Based Computer vs. Traditional Human Teaching Methods of Part of Speech Analysis. Computer Assisted Language Learning,

8(2-3), 259-274. 8.Nagao, M. (1996). Natural Language Processing. Tokyo: Iwanami syoten. In Japanese.

9.Nagata, N. (1995). An effective Application of Natural Language Processing in Second Language Instruction. The CALICO Journal, 13(1), 47-67.

10.Yamamoto, H., Sakayauchi, M., and Yoshioka, R. (1996). Development System of Japanese Teaching Material Using CASTEL/J. In Proceedings of the 20th Annual Conference of the Japan Society for Science Education, 105-106. In Japanese.

11.Yang, J.C., and Akahori, K. (1997). Development of Computer Assisted Language Learning System for Japanese Writing Using Natural Language Processing Techniques: A Study on Passive Voice. In Proceedings of the 8th World Conference on Artificial Intelligence in Education (AI-ED97).