

MACHINE TRANSLATION OF THE KAZAKH LANGUAGE ON THE BASIS OF AUGMENTED ATTRIBUTE GRAMMAR

U.A. Tukeyev, D.R. Rakhimova

Al-Farabi Kazakh National University, Almaty, Kazakhstan

E-mail: ualsher.tukeyev@kaznu.kz, diana.rakhimova@kaznu.kz

In this paper we propose an approach to machine translation of the Kazakh language using the proposed work of augmented attribute grammars.

The essence of this offer method is to create a sentences for each source (Kazakh) language of ontology, which is used to form the text of sentences of the target (Russian and English) language.

For representation of ontology of the offer in the device of the expanded attribute grammar which feature is inclusion of special semantic rules taking into account features of the Kazakh language at level of representation of words, phrases and whole sentences.

The attribute grammar was offered by Donald Knut in which to each rule of context-free grammar of language the semantic rule of determination of value of terminal and non-terminal symbols of grammar is attributed. The augmented attribute grammar for the description sentences of the Kazakh language is represented in the following look, is defined as:

$$AAG = \langle G, A, R^W(A), R^F(A), R^S(A) \rangle,$$

where G — context-free grammar of sentences of the Kazakh language, A – a final set of semantic attributes; $R^W(A)$ – a set of semantic rules at level of words, $R^F(A)$ - a set of semantic rules at level of phrases of the sentence, $R^S(A)$ - a set of semantic rules of sentence level.

To determine the semantic rules to the level of words used to describe the attribute of the *Value* of the semantic meaning of the word. At the level of lexical analysis to account for the peculiarities of the Kazakh language, we introduce two kinds of objects: real-world objects, which we denote by O^r and linguistic objects, which we denote by O^l .

For example, the word "оқулық" (textbook, book) consists of real-world object O^r "оқу" (to read) and the linguistic object O^l "лық" (affix of kazakh language).

Then the word W - "оқулық" will be determined by the semantic formula:

$W.Value = O^r.Value \bullet O^l.Value$, where the \bullet -operator of semantic connections.

For determination of semantic values of phrases and sentences semantic rules for group of a noun, a verb, circumstance, and also a sentence structure taking into account features of grammar of the Kazakh language are entered.

In general, the proposing augmented attribute grammar of sentences Kazakh languages allows to create a private ontology proposal describing all knowledge of the proposal of the Kazakh language, algorithms that allows the machine translation of her sentences is to synthesize the target language to meet the requirements of the target language grammar.

REFERENCES

1. Knuth, D. E. (1968) Semantics of context-free languages. *Mathematical Systems Theory* 2, 2, pp. 127--145.
2. D. E. Knuth: (1990), The genesis of attribute grammars. *Proceedings of the international conference on Attribute grammars and their applications LNCS*, vol. 461, pp. 1--12. Some informal, historical information.

МАШИННЫЙ ПЕРЕВОД КАЗАХСКОГО ЯЗЫКА НА ОСНОВЕ РАСШИРЕННОЙ АТРИБУТНОЙ ГРАММАТИКИ

Тукеев У.А., Рахимова Д.Р.

Казахский Национальный университет имени Аль-Фараби, Алматы, Казахстан

E-mail: ualsher.tukeyev@kaznu.kz, diana.rakhimova@kaznu.kz

В работе предлагается подход к машинному переводу казахского языка с использованием предлагаемой в работе расширенной атрибутивной грамматики. Сущность предлагаемого подхода заключается в формировании для каждого предложения исходного(казахского) языка онтологии, которая используется для формирования текста предложения целевого(русского, английского) языка. Для представления онтологии предложения в используется аппарат расширенной атрибутивной грамматики, особенностью которой является включение специальных семантических правил с учетом особенностей казахского языка на уровне представления слов, фраз и всего предложения.

Атрибутную грамматику предложил Дональд Кнут, в которой каждому правилу контекстно-свободной грамматики языка приписывается семантическое правило определения значения терминальных и нетерминальных символов грамматики.

Расширенная атрибутивная грамматика для описания предложений казахского языка представляется в следующем виде, определена в виде:

$$AAG = \langle G, A, R^W(A), R^F(A), R^S(A) \rangle,$$

где G — контекстно-свободная грамматика предложений казахского языка, A — конечное множество семантических атрибутов; $R^W(A)$ — множество семантических правил на уровне слов, $R^F(A)$ — множество семантических правил на уровне фраз предложения, $R^S(A)$ — множество семантических правил на уровне предложения.

Для определения семантических правил на уровне слов используется атрибут *Value* для характеристики семантического значения слова. На уровне лексического анализа для учета особенностей казахского языка введем два вида объектов: объекты реального мира, обозначим их через O^r и лингвистические объекты, обозначим их через O^l . Например, слово "оқулық" состоит из объекта реального мира O^r "оқу" и лингвистического объекта O^l "лық". Тогда значение слова W - "оқулық" будет определяться по семантической формуле: $W.Value = O^r.Value \bullet O^l.Value$, где \bullet -оператор семантического соединения.

Для определения семантических значений фраз и предложений вводятся семантические правила для группы существительного, глагола, обстоятельства, а также структуры предложений с учетом особенностей грамматики казахского языка.

В целом предлагаемая расширенная атрибутивная грамматика предложений казахского языка позволяет формировать частную онтологию предложения, описывающей все знания о данном предложении казахского языка, что позволяет алгоритмам машинного перевода по ней синтезировать текст предложения на целевом языке с учетом требований грамматики целевого языка.

СПИСОК ЛИТЕРАТУРЫ

3. Knuth, D. E. (1968) Semantics of context-free languages. *Mathematical Systems Theory* 2, 2, pp. 127--145.
4. D. E. Knuth: (1990), The genesis of attribute grammars. *Proceedings of the international conference on Attribute grammars and their applications LNCS*, vol. 461, pp. 1--12. Some informal, historical information.