TWO-COMPONENT TEACHING SYSTEM THAT UNDERSTANDS AND CORRECTS MISTAKES

Elena BORISSOVA

Institut Russkogo yazyka im Pushkina

Moscow USSR

Abstract

This paper presents a computer-tool teaching system supplied by a language processor. Its aim is to correct mistakes in texts written by foreign students learning Russian as a second language. Since a text may include grammar mistakes, the system cannot use morphological analysis to fool extent. So one must compile a programm capable of finding and correcting mistakes without traditional means of analysis.

To solve this problem we propone a system that includes a vocabulary and rules of finding and re-writing words. So the process consists of finding word stems and than correcting word endings . Semantic and syntactic information("a model of ruling"/Mel'čuk 1974/) necessary for that is written in the vocabulary of verbs as a frame. The slots of this frame contain semantic and morphological information about words that depend on this word.

The vocabulary containing now approximately 200 lexemes is enough for beginners

1. Introduction

As a rule, computer-tool teaching programs can do without language processors. That gives us an opportunity to use Personal computers and other available and inexpensive means. But such programs can be used only for several types of exercises: mostly those that include substitution or insertion of listed words and endings in a sentence.

Such exercises train the student to use correctly words and expressions. It is impossible to modell exercises of the so called communicative type / Kostomarov et al. 1982/ that form the skills of spontaneous speech (a pupil constructs some sentences expressing his thoughts on a definite topic). While doing such exercises one cannot deal with a limited list of variants as there exists an infinite number of them(they are innumerable if a pupil has mastered even an elementary course). That is why we need a language analisis processor in a computer-assisted teaching program aimed to teach somebody to speak the language, to'communicate'.

There exists a lot of language analysis and synthesis programs of Russian applied in Automatic translation, Natural language dialogue systems etc. Nevertheless it is impossible to use them in computer assisted teaching programms.On the one hand the majority of them are oriented to the scientific or technological language. On the other - and more importantly - as a rule they cannot deal with texts containing mistakes. Those systems that'understand' a text with mistakes do not notice them or they correct them only printing them on a display /Carbonell,Hayes 1984/

So it is necessary to compile a program capable of finding and correcting mistakes. This problem is new for computational linguistics.

2. Description of language processor

To solve the problem we propose a two-component language processor that provides morphological, syntactic and (to some extent) semantic analysis of a text with mistakes, and then it synthesises correct sentences which express identical meaning to the analysed ones. This processor deals with separated sentences but some information must be used while analysing subsequent sentences of the given text (e.g. information about the sex of a speaker as in a Russian text this information is necessary for agreement of a predicate in the Past and a subject expressed by the pronoun of the 1st person singular: ya chodil 'I was going').

The 1st step. The processor executes a morphological analysis of a sentence by means of a stem vocabulary which includes variants of the stems of each verb (e.g. CHOD-, CHOJ- of the verb CHODIT' 'to go' etc.), noun, adverb, adjective, pronoun. This list includes the typical incorrect variants of these stems (e.g. ZOV-, ZAV- of the verb ZVAT' 'to call'; a typical mistake is ZAVU instead of ZOVU). The first task of the processor is to find a verb and to identify it in the vocabulary.

The 2nd step. The system uses syntactic information of the vocabulary. Every verb stem is supplied with information of the morphological, syntactic and semantic features of words which are ruled by the verb (e.g. JIT' 'to live' v chem 'place'). Since all the nouns in the vocabulary are supplied with semantic information as well (e.g; DOM 'a house' 'place'), that enables the system to find appropriate nouns for the verb.

Then in accordance with the morphological information the system synthesises a correct case form of the noun (e.g. V DOME) which is compared with the form written by the pupil. The difference is marked as a mistake that can be commented on by the list of explanations (e.g. a pupil: Jivu dom, a correct rect form, synthesised by the system is Jivu v dome, mistake: "a wrong case form")

The 3d step. Then the system accomplishes agreement of subject and predicate (e.g. Student sg,m jil sg.m 'A student lived') according to the semantic information the verb is supplied with (e.g; JIT' 'person' STUDENT 'person') and morphological information of the subject. The temporal and aspectual characteristics of verbs depend on adverbs (e.g. vchera 'yesterday' past - jil 'lived' past) and some other facts. According to this information the system synthesises verbal forms and compares them with those written by the pupil.

The 4th step. The agreement between adjectives and nouns is executed in the same way as the previously - by finding words according to the semantic features (e;g; novyi 'new' 'thing','place'... dom 'a house' 'place') and then by changing of the forms according to the morphological information (e.g. dom m - novyi m)

3. Some notes on system exploating

The result of this system's work should be a correct text with the correction of mistakes. A system based on the same principles but more complex should correct some syntactic mistakes in word order, usage if conjunctions etc.

If the result of the correction allows two possible variants of a text, the computer prints:"Do you want to say"..." or

"..."? (Possible variants - in inverted commas). If a sentence is not admissible by the given system, a computer prints: "I do not understand you, say it another way".

The system can ask other questions as well. In particular, if a pupil prints a personal name unknown to the system it asks:"Is it a male or female?" and then this name is inserted into the vocabulary with morphological characteristics fem. or masc.

Besides the grammatical information the vocabulary should include some encyclopaedical information important for a pupil. E.g. if a foreign pupul has come to Moscow already then the phrase Ja priedu v Moskvu 'I'll come to Moscow' is wrong. In order to correct such mistakes one inserts into the morphological and syntactic information an inscription : PRIYEHAT' ya 'I' past- v Moskvu 'to Moscow' which means that a prase about 1 p.sg. and Moscow is correct only in the past.

Prelimenary input of proper names which a learner may have occasion use in a text is desireable as well. Otherwise mistakes of the type Ya priehal iz kuba (instead of s Kuby) would not be corrected.

The system is intended both for testing compositions and dialogues. Since systems for advanced students would be too sophisticated and would have to include complete information about the language, nowadays we restrict ourselves to a system for beginners (150 lexemes in the vocabulary). The system will be realised on the IBM-PC

References

Mel'čuk I.A, Opyt teorii linguističeskih modeley 'Smysl⇔Text'. Moscow Nauka 1974

Kostomarov V.G.,Mitrofanova O Metodičeskoe rukovodstvo dla prepodavateley russkogo yazyka inistrantsam. Moscow Russkiy yazyk 1982 p. 7

Carbonell J.G. Hayes Ph. J. Coping with extragrammaticality. In: COLING 84 p.437-443