NL-Processor and Linguistic Knowledge Base in a Speech Recognition System

Michael G. Malkovsky, Alexey V. Subbotin

Moscow State University, Computational Mathematics and Cybernetics Department,
119899, Vorobjovy Gory, Moscow, Russia
malk@cs.msu.su, lesha@isd.srcc.msu.su

Abstract. The linguistic methods providing grammatical and lexical control of the acoustic recognition results are described. The Parser (NL-Processor) uses original local grammar, ATN-type grammar, and comprehensive computer dictionaries (Linguistic Knowledge Base - LKB). To choose the best of the plausible strings a special meta-level is introduced that deals with the multi-criterion choice problem. An important feature of the techniques employed is that the best string does not necessarily have to be a grammatical one. The original approach to lexical n-grams dictionary correction is described too. Later this approach was generalized and now it is considered as a base of the new computer-aided technology of the LKB construction. The main component of this approach is the METAMODEL, based on UML and fuzzy mathematics. Both human and computer should participate in LKB construction process: human contributes to this process his intelligence and language intuition, computer - his speed, memory and computational capabilities.

The paper describes linguistic methods providing grammatical and lexical control of the acoustic and phonetic recognition results. These methods were elaborated for American English by the Moscow State University team (Agreement between ACCENT, Inc., a California corporation, and Computational Mathematics and Cybernetics department of the Moscow State University). The project scope was the development of the speaker-independent recognition system for translating an input speech signal to text. The commercial software product was not developed, but the experts' positive assessment of the approach and its potentiality was the motivation of the United States Patent registration (E.Komissarchik, et al. "Knowledge-based speech recognition system and methods having frame length computed based upon estimated pitch period of vocalic intervals", No. 5,799,276).

The goal of our team was the improvement of the acoustic and phonetic recognition accuracy by taking into account grammatical and semantic information. The model of the American English and the Linguistic Knowledge Base (LKB) and NL-processor (Parser) were developed. This work was done in the cooperation with the linguists. Model includes 28 main grammatical classes and about 300 grammatical grades of words.

The LKB used in the analysis process includes grammatical dictionaries, corpora of lexical bigrams (the pairs of words provided by statistical information), corpora of admissible and inadmissible sequences of syntactic classes, a grammar table, and a dependency co-occurrence dictionary. The grammatical dictionaries employed are: the grammatical dictionary of separate words, the grammatical dictionary of phrasal verbs (e.g. "get away", "get around to"), the grammatical dictionary of collocations (e.g. "on end", "as ever", "Microsoft Word"), the grammatical dictionary of the so-called "segments", i.e. sequences of words merged into one phonetic word (e.g. "isn't", "it's" for "it is" or "it has").

Entries in any of the grammatical dictionaries contain the symbols of syntactic classes assigned to the entry units (Noun, Verb, Conjunction, etc.), more particular grammatical features of these units (Number, Person, Transitivity, etc.), their semantic features (Human, PhysObject, InfoObject, etc.), and a frequency tag.

The Parser takes as an input the results of acoustic and phonetic recognition - so called "word garland" with one or more (for homophones "it's" - "its", "boys - boy's - boys'" or for quasi-homophones "boys - poise") word choices specified for each word of the utterance pronounced. Any word choice has a so-called "penalty" - a numerical estimation of the acoustic and phonetic recognition accuracy. The Parser processed grammatical and lexical properties of the words in a garland ("I have read/red this book" - "I have this red/read book") in order to transform a garland into a set of word strings as tentative representations of the utterance and to form a complex plausibility estimate for each of these strings. The grammatical and lexical properties considered are: the grammatical features of words and collocations; the grammatical structure of each string as a whole and/or of the phrases that the string might contain; the grammatical "events" found to be present in the strings (such as ellipsis, anaphora, deviant forms, lack of agreement, etc.); the frequency of words, collocations, and lexical pairs.

The general scheme of the parsing envisages two levels of processing: local grammatical analysis providing preliminary filtering of the word garlands along with their transformation into word strings; global grammatical analysis grading these strings by their preferability and choosing the ones that are more plausible as representations of the utterance.

Local grammatical analysis of the word garland associates grammatical features with the candidate word choices in the garland using triplet table to reject those of the candidate strings, which are considered to be absolutely inadmissible, and to weigh the remaining strings. The output of the local grammatical analysis is a set of word strings; each string individually ranked and weighted based on locally applied grammatical rules.

In the local grammatical analysis step the admissible and inadmissible sequences of syntactic classes are used which describe local syntactic relations. The corpora of such sequences include a table of admissible triplets, each triplet provided in them with certain grammatical features to be used in the analysis, and a table of inadmissible quadruplets, the so-called "anti-quadruplets". Each line in the triplet table includes a triplet of syntactic classes (e.g. Verb - SubordinateConjunction - Noun), an operator (so-called "demon") that checks the admissibility of this triplet, the triplet frequency tag, and its Grammatical-events marker.

Global grammatical analysis which follows the local one takes as its input the candidate word strings produced by the local analysis, and establishes the non-local syntactical relations structure of each string using the grammar table and the dependency co-occurrence dictionary . The grammar table constitutes a table-like representation of a formal ATN-type (augmented recursive transition network) grammar which determines grammatical compatibility of distantly located parts of the sentence and finds all admissible variants of the sentence's syntactic structure. It deals with the descriptions of structural phrases possible in a sentence and a set of operators providing access to the dependency co-occurrence dictionary. The dependency co-occurrence dictionary contains entries both for lexical and for various types of grammatical items, specifying for each of such items its predictions for the structural elements (dependents and/or governors) that may be co-occurrent with the corresponding text item in a sentence.

The syntactical relations structure associated with a word string in the course of its global analysis is used as a basis for the grammatical descriptor of the string, this descriptor comprising information on the properties of its structure as a whole, as well as of the structure of the phrases it might contain. The grammatical descriptor is used to verify certain non-local contextual requirements, as well as to weigh the plausibility of the word string as a representation of the input utterance.

The meta-level uses the results of all the previous steps of processing the input utterance to choose the best of the candidate word strings for its representation. It has to deal with a multi-criterion choice problem, the criteria involved being of three types: criteria resulting from the phonetic/acoustic weighting of various word choices; criteria based on the results of the local grammatical analysis of the word garlands; and criteria based on the grammatical descriptors of the candidate word strings constructed by the global analysis. An important asset of the meta-level procedures is that the best string chosen does not necessarily have to be a full and grammatical sentence. The Parser can accept both non-sentential strings, like newspaper headings, and ungrammatical expressions (to be often found in spontaneous speech).

The meta-level procedures are described by means of production rules representing statements that can be illustrated by the following one: "If the given choice for word N violates sentence grammaticality requirement of type 7 (i.e. a noun phrase after a copula does not have a determiner in it), but this choice is characterized by the acoustic and phonetic steps as absolutely reliable, then this choice is considered to be correct notwithstanding the violation mentioned". The set of such production rules may be changed at will, so that the system can be tuned in to recognize different speech styles: on the one hand, it can be made to accept only strictly grammatical sentences, on the other, the meta-level rules may be added to make acceptable various types of ungrammatical structures.

At the conclusion of the grammatical analysis step, it is expected that in most cases all words of the sentence will be defined with a confidence level exceeding a predetermined threshold value. This sentence will therefore be presented to the user as the correct spelling of the input utterance.

The main features of the methods described are:

- the utterance recognition results are controlled primarily with the help of the grammatical, rather than lexical restrictions (this makes the Parser independent of the domain of the input utterances);
- the grammatical analysis is applied to word garlands, i.e., to more complex objects than word strings;
- the grammatical analysis techniques envisage two levels of processing: local and global (this permits more accurate control of recognition results);
- local grammatical analysis provides very fast preliminary filtering of the word garlands and their transformation into word strings, using tables of triplets and quadruplets of syntactic classes;
- global grammatical analysis grades the resulting word strings by their preferability and specifies the ones that are more plausible as representations of the input utterance, using an ATN-type grammar and a dependency co-occurrence dictionary arranged within an object-oriented paradigm (this makes possible to introduce not only grammatical, but also semantic control of the recognition results);
- to choose the best of the plausible strings, a special meta-level is introduced that deals with the multi-criterion choice problem (this provides an improvement over the existing techniques in that the best string does not necessarily have to be a grammatical one).

Linguistic module (LKB and Parser) helps to improve the accuracy of acoustic and phonetic recognition from 94-95% to 95-97%.

The original approach to lexical n-grams corpora correction is described too in the paper. In particular, we have proceeded further research of methods for improvement of lexical bigrams processing issues. New methods of the bigrams corpora correction were developed for speech recognition system, when the project mentioned was closed. These methods resulted in additional improvement (1-2%) of recognition accuracy.

The bigrams are very powerful recognition improvement facility, but we can't ignore the problem of the corpora representativity. As bigrams are gathered on some text corpus, they are very suitable to be used in recognition of the text from this corpus. But for the texts from another domain they don't work so well.

There are two ways to improve the situation. The first one is to use more representative text corpus. The second one is to try to fill the bigrams corpora (dictionary). Choosing the second approach we have developed several methods of bigram dictionary improvement such as "paradigm filling", "transformations", "clusterization". The main idea of all these methods - it is necessary to analyze the current dictionary and to find what bigrams can/should be added. This analysis is based on the statistical principles, because bigram dictionary is used statistically. The main rule of thumb is the following: "It's not a trouble to add some incorrect bigrams if you add with them much more correct ones". For example, if we insert a new set of bigrams into the dictionary, some are correct and some are incorrect, the heuristic of adding is good if it adds more correct bigrams, than incorrect ones.

The main rule of the "paradigm filling" method is the following. Let one of the morphological forms WF1 of some word Wrd is in the bigram (WF1 W), contained in the dictionary. WF is a form of some other word. In this case we can add into the dictionary the bigrams (WF2 W), (WF3 W), ..., where WF2, WF3, ... - the other forms of Wrd. For example, if in the dictionary we have a bigram "brown dog", than we can add a bigram "brown dogs". For this method the heuristics in a form of productions were developed, which describe conditions and process for this method.

The principle of the "transformations" is the transformation of entire bigram, which represents some consistent word combination. For example, bigram "dog runs" can be transformed into "running dog". Rules for this kind of transformation are described by productions too and contain conditions when such transformations are possible. These rules are based on transformational grammar of American English.

The main idea of "clusterization" is to find a cluster of the "similar" words ("similar" in the sense, that they may participate in the common lexical contexts). For example, if we decide, that the days of the week are in one cluster, and in the dictionary there are bigrams "on Monday", "Sunday morning", "Saturday evening", we may add the following bigrams "on Sunday", ..., "on Saturday", "Monday morning", ..., "Saturday morning", "Sunday evening", ..., "Friday evening". If clusters are fixed, this method may be described by productions too.

The safest method is the "paradigm filling". It adds very little incorrect bigrams - only ones, which contains words having not full paradigm. The "transformations" method is more dangerous, because syntactic rules it's based on have much more exceptions than morphological ones. The "clusterization" should be used carefully involving human control over the clusters, which are the base for dictionary correction, because this method uses semantic level information. This method is applicable to small, consistent and evidential clusters.

Later the main idea of this approach - a combination of statistical and knowledge-based methods - was generalized. Now we consider it as a base of the new computer-aided technology of the LKB construction. The basic principles of this technology are the following:

- computer contributes in the process its capabilities to handle routine work, find patterns and regularities in data;
- human expert controls the process and contribute his or her intelligence and language intuition;
- for every LKB construction project unique method is required, but this method can be build as a combination of the existing proved methods;
- a lot of experiments should be conducted to find right combination of methods (statistical methods vs. linguistics-based ones, automatic methods vs. semi-automatic or manual, corpus-based methods of knowledge acquisition).

But there are some problems in implementation of this approach:

1. There exist a lot of different linguistic models, which result in different LKB construction methods. These methods may contain useful heuristics, but they cannot be used and even more combined due to the differences in the basic models.
2. Some methods have programming implementations but they cannot be used directly, as they are based on different technologies, programming languages, data formats, etc.

So, the following approach to the implementation of the proposed technology was chosen:

1. To build METAMODEL which is modeling facility permitting to map different LKB models, algorithms and heuristics into the common formalism. This METAMODEL should be based on international standards and its building, analysis and mapping should be supported by industrial technology.
2. To build LKB construction system which implements this METAMODEL and facilitates the experiments with different combination of the methods. This system should be based on standard technologies.

The METAMODEL was build as Unified Modeling Language extension. This language is standardized in Object Management Group and is de-facto a standard for object modeling languages. We build our METAMODEL as "UML profile" - standard way to extent UML metamodel. We extend the UML modeling facilities with fuzzy mathematics which permits to represent non-precise linguistic information and Petri nets formalism - for rigorious description of the distributed algorithms.

For LKB construction system the architecture was developed which is based on CORBA (as the most powerful for the present distributed object-oriented standard technology). The architecture includes such components:

- Data Access Servers, which incapsulates data sources and represent common interface based on fuzzy semantic net abstraction;
- Processor Components, which can be deployed on Data Access Servers or Application Servers and implement different processing algorithms;
- Workflow Server, which permits to describe distributed processes of data analysis in declarative form and to experiment with combination of the data sources and algorithms;
- other components, such as repository servers, user interfaces, system level services.

The mapping from the METAMODEL to this architecture was developed which permits to map models to the implementation in the LKB system. The mapping process can be automated with UML CASE tool.

This article was processed using the LaTeX macro package with LLNCS style