}

Let us consider the main structure of SC-model of problem solver for conversion natural language texts into knowledge base fragments and generation natural language texts from knowledge base fragments in natural language interfaces of ostis-systems in SCn-code, respectively:

SC-model of problem solver for conversion natural language texts into knowledge base fragments
:= [SC-model of problem solver for natural language texts analysis]
⇐ *decomposition of an abstract sc-agent*:*
- {• *Abstract sc-agent of lexical analysis*
  ⇐ *decomposition of an abstract sc-agent*:*
  - {• *Abstract sc-agent of decomposing texts into segmentation units*
    • *Abstract sc-agent of marking up segmentation units*
    }
- *Abstract sc-agent of syntactic analysis*
- *Abstract sc-agent of semantic analysis*
- *Abstract sc-agent of extracting factual knowledge structures into the knowledge base*
- *Abstract sc-agent of logical inference*
}

The SC-model of problem solver for natural language text analysis is constructed on the basis of the proposed following process for factual knowledge acquisition:

- natural language text is loaded into the interface;
- lexical analysis and syntactic analysis of the input natural language text is performed;
- named entities and relations between them is extracted based on the analyzed syntactic structure and extraction rules.

In principle, this SC-model of problem solver can potentially extract structured knowledge (generally, named entities and relations between them) from texts in different language into the knowledge base of the ostis-systems for a specific subject domain, but the construction of knowledge base on the specific natural language processing, which includes rules for specific natural language processing and extraction rules, will become more complex. In turn overhead costs of construction will increase.

For generation natural language texts from knowledge base fragments the classical pipeline of natural language generation is used as the basis to develop the SC-model of problem solver for generation natural language texts from knowledge base fragments. The developed SC-model of problem solver has higher flexibility. For specific natural language, the developed problem solver can be easily modified accordingly.

SC-model of problem solver for generation natural language texts from knowledge base fragments
:= [SC-model of problem solver for natural language texts generation]
⇐ *decomposition of an abstract sc-agent*:*
- {• *Abstract sc-agent determining sc-structure*
- *Abstract sc-agent dividing determined sc-structure into basic sc-constructions*
- *Abstract sc-agent determining the candidate sc-constructions*
- *Abstract sc-agent transferring candidate sc-constructions into message triples*
- *Abstract sc-agent text planning*
- *Abstract sc-agent for micro-planning*
- *Abstract sc-agent for surface realization*
}

The SC-model of problem solver for natural language texts generation is constructed on the basis of the proposed following process for texts generation:

- a specific sc-structure (fragment of knowledge base) is selected in the knowledge base;
- the candidate basic sc-constructions from the sc-structure is determined, then is translated into a message triple (in the form of subject-relation-object);
- the resulted natural language texts is generated from the message triple as output.

It is worth noting that the composition of sc-constructs has sc-arcs that have specific meanings. Therefore sc- constructions with sc-arcs need to be converted into the corresponding message triples in form of text, which is easier to represent in the form of natural language texts.

The developed unified semantic model of natural language interface ensures the flexibility of developing a specific natural language interface and integration of various components (knowledge base on natural language processing, component for conversion natural language texts into knowledge base fragment and component for text generation) in the interface.The development of natural language interface consists in the development of individual components independently of each other. It is flexible to adjust and make extensions of linguistic knowledge and sc-agents for tasks solution in specific natural language interface. The more detailed description about function of each abstract sc-agent can be seen in [2].

IV. Imlementation of Chinise Langu INTERFACE

On the basis of unified semantic model of natural language interfaces of ostis-systems, we can implement a specific natural language interface of intelligent help systems for various subject domains. In this section we will describe the implementation of the prototype of Chinese language interface of a intelligent help system

about discrete mathematics. For developing Chinese language interface it's necessary to construct knowledge base on Chinese language processing and corresponding problem solvers for conversion Chinese language texts into sc-structures and Chinese language texts from sc-structures, which to integrate logical models on rules and neural network models for Chinese language processing. The detailed processing stage of conversion Chinese language texts into sc-structures and generation Chinese language texts from sc-structures will be shown in followings.

A.factual knowledge extraction from Chinese language texts

Currently there are some restrictions for extracting factual knowledge from Chinese language texts:

- the processed Chinese language texts are Chinese declarative sentences;
- there are specific factual knowledge (named entities and relations between them) in the Chinese declara- tive sentences;
- due to features of Chinese language, the result of decomposition of Chinese declarative sentences into segmentation units greatly influences the factual knowledge extraction.

In this section the general processing stage of conver- sion Chinese declarative sentence into sc-structure will be shown in the followings.

Step 1: From the point of view of OSTIS technology, any natural language text is a file (sc-node with content or so-called sc-file). The Chinese declarative sentence shown in our example is represented in such a node in Fig 2 and describes: "有限集合(the finite set), (comma) 严格地 (strictly) 包含 ((includes) 二元组 ((pairs) 。(full stop)".

Figure 2: The representation of the Chinese sentence

As shown in Fig 2, according to the written tradition of Chinese language texts, Chinese characters are written one after the other and there are no natural gaps between them. As we know, the lexeme is a term commonly used for lexical analysis in European processing. However in Chinese language processing the unit is considered as the smallest unit . In the "Modern Chinese word segmentation standard used for information processing", a word in Chinese language is represented as a segmentation unit. The precise definition of segmentation units is "a basic unit for Chinese language processing with certain semantic or grammatical functions".

Step 2: The Chinese declarative sentence is decomposed into separate segmentation units, lexical analysis is carried out. Afterwards syntactic structure or semantic structure of sentence is analysed, the relations between input sentence and divided segment units, as well as between these segment units in sentence are revealed. The analyzed results of input Chinese sentence is shown in the Figure 3.

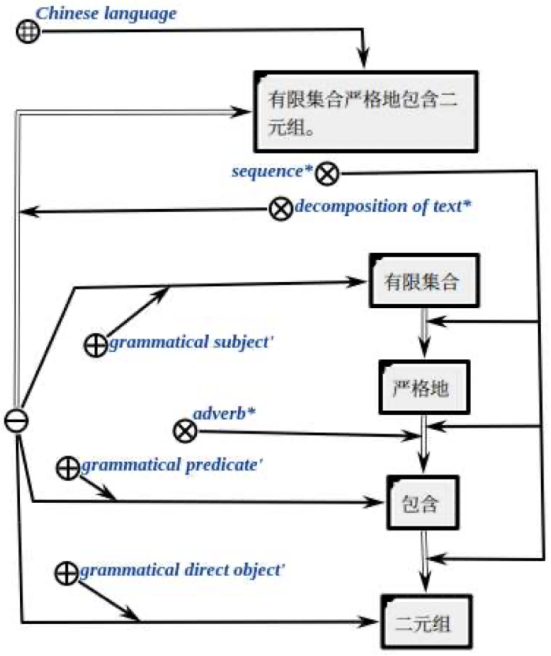

Figure 3: The syntactic structure of input Chinese sentence

Step 3: The factual knowledge that mainly consists of named entities and relations between them is extracted based on previous text analysis and extraction rules without contradiction detection. The resulted constructed knowledge base fragment (sc-structure) from input Chi- nese declarative sentence is shown in the Figure 4.

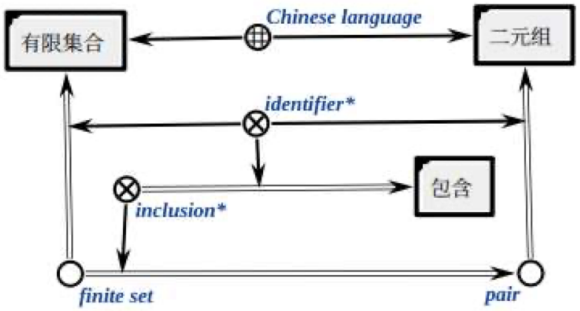

Figure 4: The constructed sc-structure from input Chinese sentence

It is important to note that in this case, a knowledge base fragment can be directly converted into knowledge base without linking extracted named entities and relations between them from the input Chinese sentence with the corresponding entities and relations defined in the knowledge base of intelligent help system.

B. text generation from knowledge base

In this section the processing stage of generation Chinese declarative sentence from sc-structure is de- scribed. The processing stage is roughly divided into two steps: firstly converting sc-structure from knowledge base into message triples; then generating Chinese declarative sentence from translated message triples. The description about concept message triple can be found in [14].

In our works there are some restrictions for Chinese lan- guage texts generation from knowledge base fragments:

- the knowledge base fragment is completed and has sc-elements with identifiers in Chinese language;
- the generated Chinese language texts are Chinesedeclarative sentences.

Step 1: The selected sc-structure is divided into standard basic sc-constructions, afterwards from which the candidate sc-construction is selected and will be converted into "message triple", then into resulted Chinese sentences. A candidate sc-construction (belong to standard basic sc- construction) is shown in SCg (Figure 5). The candidate sc-construction contains sc-elements with identifiers in Chinese language. Identifiers in Chinese language of each sc-element of sc-construction have corresponding specific segmentation units of Chinese language.

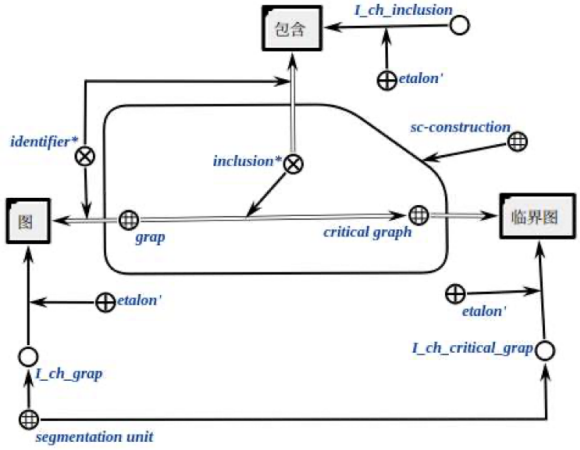

Figure 5: The determination of candidate sc-construction

Step 2: The candidate sc-construction is transferred to message triple. The converted message triple consists of sc-files (sc-node with content) containing segmentation units written by trained native Chinese speakers and verified by others. The message triple that corresponds to candidate sc-construction is generated in the Figure 6, in which each sc-element is a file corresponding to a certain segment unit in Chinese language. The contents of some sc-files (e.g. "临界图(critical graph)") correspond to the identifier of sc-element in the sc-construction, meanwhile the contents of some sc-files are added when building message triple.

It is important to note that relation of each message triple is the core. Sometimes the relation represents the specific meaning of sc-arc or sc-edge in the sc-structure in form of texts. The main task of text generation is to find suitable text fragment to explain the relation of each message triple in order to generate fluent texts. In general, the subject and object of each message triple are kept constant.

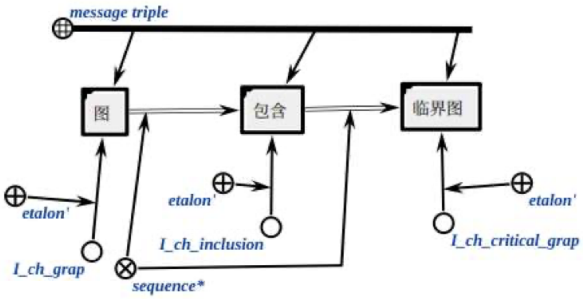

Figure 6: The message triple for candidate sc-construction

Step 3: Finally the sc-files are concatenated with certain form of that segment units to generate the resulting Chinese narrative sentence according to the permissible sequence on the constructed template for message triple with the relation "inclusion" (Figure 7). When generating result texts for some natural languages, word forms are changed according to syntactic rules (e.g. capitalizing the first word in a sentence, subject–verb agreement and others), and then added to the result texts. The relation *reference expression** is a quasi-binary relation, connecting a word to its combinatory variants.

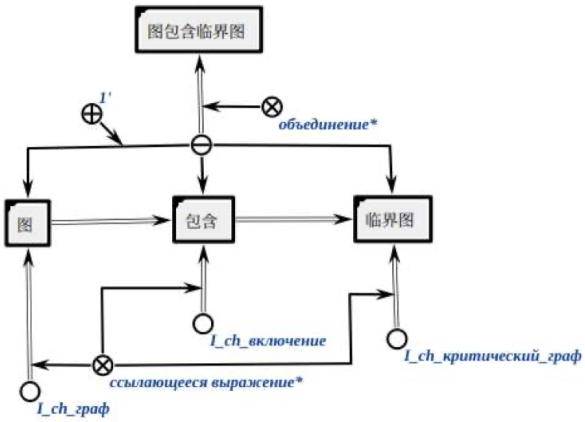

Figure 7: The generated Chinese declarative sentence

For some European languages, The inflected form of the lexical units in sc-files (e.g. singular or plural and other inflected forms) is expressed in the resulted generated texts according to the syntactic rules of a particular natural language. However, due to features of Chinese language, the processing of this step is relatively easier. In this