Cognitive Processes in Generating and Restoring Elliptical Sentences

Xenia Naidenova[0000-003-2377-7093]

Military Medical Academy, Saint Petersburg, Russia
E-mail: ksennaidd@gmail.com
ksennaidd@gmail.com

Abstract. A new cognitive approach to resolving ellipses in geometry texts is advanced. This approach is evolving in an automated system for solving school geometry tasks expressed in natural Russian language. A classification of ellipses occurring in geometry texts is given and the rules of converting the complete sentences to their elliptical variant are formulated. The cognitive schemes are introduced as the syntactic synonyms of sentences describing planimetric configurations. The role of cognitive schemes in understanding sentences is considered. They are represented by the drawing and NL-texts generated on the principle of combining the noun, verb, and prepositional phrases corresponding with both the fragments of schemes and the expressions in real geometric texts. The cognitive schemes of geometric configurations allow to facilitate the process of syntactical and semantical parsing the tasks' text, to resolve ellipses, to visualize the task condition and to reveal hidden geometric relationships not explicitly expressed in the text of tasks.

Keywords: Cognitive Approach, Ellipsis Resolution, Planimetry, Generative Grammar, Natural Language Processing.

1 Introduction

A new cognitive approach to resolving ellipses in geometry texts is advanced. This approach is evolving in an automated system for Solving Geometric Problems (SGP) formulated in natural Russian language. Ellipsis is a natural language phenomenon where part of a sentence is missing and its information must be recovered from its surrounding context, as in "Fred took a picture of you, and Susan of me". The omission of some elements from a sentence does not imply any loss of the meaning conveyed by them. But the problem of automated recovering ellipses during natural language text processing is very difficult and it has not been completely resolved so far.

The SGP system integrates the following processes: understanding NL text, solving plane geometric problems, interactive visualizing all the stage of the system functioning, and synthesizing NL texts to explain decision making. The principles of functioning the system are described in detail in [1-4] and its general scheme is depicted in Fig. 1. The system consists of the following blocks: "Ontology", "Linguistic translator", "Solver", "Graphics+NL", and "GRF interpreter".

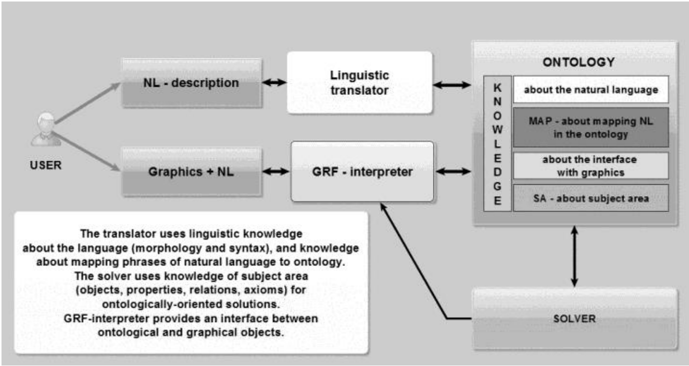

Fig.1. General scheme of the system

The ontology serves for representing knowledge necessary for all the subsystems of the system. The task of the linguistic translator is to construct the conceptual description of a given geometrical situation in terms of concepts and relations of the ontology. The solver takes the ontological description of task and searches for solution modifying the intermediate semantic representations. The language of semantic hypergraphs has been selected for ontological knowledge representation [5]. The linguistic translator converts text's fragments into corresponding ontological descriptions. The solver performs the necessary operations to solve geometry problems and uses the ontology to infer consequences from the geometric configuration formed in each current solution stage. The solver algorithm does not provide a solution guarantee, but significantly reduces the number of options. The interactive visualization is implemented based on javascript Libraries JSXGraph and MathJax [6, 7].

One of the difficulties in analyzing and understanding the planimetric texts is the ellipticity of them. The method of ellipsis resolution applied in the SGP is considered in [8, 9]. But this method does not cope with the multiple ellipsis: "the prices growth amounted to 11.9% in 2003, in 2009 – 4.4 %, in 2014 – 7.5%.

The aim of this paper is to study the elliptical structures in planimetry task texts. Our analysis involves the viewpoints of generative as well as cognitive linguistic paradigms into processing elliptical sentences. Some mental rules of generating elliptical sentences from complete ones are revealed and expressed formally. The cognitive schemes are introduced as the syntactic-semantic synonyms of sentences describing planimetric configurations. We show how the rules of generating elliptical sentences and cognitive schemes can be used in the process of analyzing the natural language texts of tasks (with one type of verb ellipses). The rules of generating elliptical sentences can be included in the set of well-known rules of equivalent text transformations. A model of understanding sentences based on cognitive schemes and equivalent transforming the task's texts as well as the descriptions of cognitive schemes is proposed.

The organization of the paper is as follows. Section 2 contains a classification of ellipses in school-level planimetry tasks. In Section 3, the cognitive rules of generating elliptical sentences are advanced. In Section 4, the role of cognitive schemes in understanding sentences is considered. Section 5 is devoted to the equivalent transformations of planimetric tasks' sentences. A short conclusion rounds off the paper.

2 Classification of Ellipses in School-level Planimetry Tasks

To study the typology of ellipses, we used a body of texts containing more than 1000 planimetric tasks from various handbooks. The following types of ellipses are revealed [9]: ellipses with dash "–": ellipses with skipped predicate (**Type 1.A**) and ellipses with skipped verb (**Type 1.B**); ellipses without "–": ellipses with skipped verb, noun, pronoun, or predicate (**Type 2)**. In general, most sentences contain several types of ellipses or/and a number of ellipses of the same type. In what follows, we shall consider ellipses only of Type 1.B. The texts of tasks are given fragmentary and their translation in English has only illustrative character (in the real translation into English texts, the dash can be absent).

The Type 1.B is known as **the verb phrase ellipsis (the VPE)**. This type of ellipses is divided into subclasses: with only one dash (**Class 1**) and with several dashes replacing the same verb (**Class 2**). An example of the ellipsis of **Class 1**:

Дана окружность и точки P и Q внутри неё. Построить вписанный в эту окружность прямоугольный треугольник, у которого один катет проходит через точку P, а другой – через точку Q. There are given a circle and two points P and Q inside it. Build a right triangle inscribed in this circle so that one leg of it passes through point P and the other – through point Q.

There exist two ellipses in this sentence. One ellipsis is the eliminating of noun (N) in noun phrase (NP) with preserving the representative of noun consistent with it [10]: "the other *leg*". The other ellipsis is the VPE with eliminating verb "*passes*". It belongs to the ellipses named "ellipses with predicative vertex" [10].

An example of ellipsis of **Class 2**:

Внутри квадрата $A_1A_2A_3A_4$ взята точка P. Из вершины A_1 проведена прямая, перпендикулярная к прямой A_2P, из вершины A_2 – к прямой A_3P, из вершины A_3 – к прямой A_4P и из вершины A_4 – к прямой A_1P; Inside a square $A_1A_2A_3A_4$ a point P is taken. From vertex A_1, it is drawn a line perpendicular to line A_2P, from vertex A_2 – to line A_3P, from vertex A_3 – to line A_4P, and from vertex A_4 – to line A_1P.

In this sentence, a whole fragment of the repeated verb phrases (VP) is missed (the verb and its direct object: "*it is drawn a line perpendicular*".

Revealing complete and incomplete NPs, VPs, PPs and other phrases of sentences is very important to restore ellipses. Revealing the NPs and PPs, for example, is realized in the system OntoIntegrator [11] in the project on creating World Digital Mathematical Library – WDML. One kind of partial parsing known as chunking [12, 13] also deals with identifying non-overlapping segments of a sentence: noun phrases (NPs), verb phrases (VPs), adjective phrases (AdjPs), adverb phrases (AdvPs), and prepositional phrases (PPs). The process of chunking for Russian Language is used in the syntactic and semantic parsing based on ABBYY COMPRENO Linguistic Technology [14]. Kobsareva T. Ju. [15] has proposed to perform a previous segmentation of sentences before the principal parsing for constructing projective fragments of NPs, PPs, and compound predicate in Russian language. It is also reasonable to suppose that the information model of elliptical sentences can be constructed in the framework of generative grammar [16].

3 Cognitive Rules of Generating Elliptical Sentences

One of the methods for resolving ellipses of Class 1 in the texts of planimetric tasks has been described in [9]. But the approach considered in this work does not cope with ellipses of Class 2.

It should be noted that the question of how to restore the complete structure of elliptical part of a sentence has not been fully solved in the conventional approach based on syntactical-semantic parsing sentences. Linguists have already realized the restriction of the approach to resolving ellipses in which syntax is separated from semantics [17]. In [18], a key problem of cognitive view on resolving ellipses is stated: understanding ellipsis does not mean that we first have to restore it, and then to turn to understanding the whole sentence. In fact, understanding the sentence also entails understanding the ellipsis in it.

Our cognitive approach to resolving ellipses rests on the following assumptions:

- Cognitive models of geometric configurations are seen as syntactic-synonymous mappings of sentences;
- Cognitive processes of designing and understanding sentences are interconnected with one another;
- Understanding sentences is based on knowing how sentences are formed;
- Sentences of Class 1 and Class 2 imply the same processes of generation.

We assume some hypotheses about cognitive operations produced mentally when generating elliptical sentences:

- **Hypothesis 1**. Repeated actions are described within the same sentence.
- **Hypothesis 2**. The complete sentence describes some cognitive (imaginable) geometric situation.
- **Hypothesis 3**. The complete sentence is mentally transformed into an incomplete (elliptical) one.
- **Hypothesis 4**. The transformation mentioned above is based on some cognitive operations performed mentally by a certain algorithm in the process of generating incomplete sentence.

We need now to introduce the concept of context for words in texts. We mean the term "context of a word" not as "accessibility" or "dedication" [19], but as a zone of action of the word, that is, a fragment of the text in which we are talking about some object or action already mentioned, or/and the situation expressed by this word. The inclusion relationship is realized between the contexts of words. We formulate some cognitive Rules 1-5 of transforming a complete sentence into an elliptical one:

- **Rule 1.** If the designation of an object is introduced in a sentence, then further in this sentence it can be used only this designation without the name of the object;

- **Rule 2.** If the designation of a figure is introduced in a sentence, then further in this sentence it can be used this designation without mention of the name of the figure (in the scope of this figure's context);
- **Rule 3.** If an action over (with) several objects is meant, then after the description of this action over (with) the first object in a sentence, further this action over (with) other objects can be described without copying the name of this action (the verb is skipped);
- **Rule 4.** An object can be expressed by its Noun Phrase, it implies the permission of missing (skipping) in a sentence the common (repeated) fragment of the Noun Phrases when describing similar objects;
- **Rule 5.** A verb can enter its Verb Phrase, it implies that if one and the same repeated action is described many times in a sentence and it has several repeated arguments, then these arguments (or their fragments) can be skipped after the first description of this action.

Let's take a look at how these rules work in the tasks.

Task 1. *A trapezium ABCD with the base AD is given. The bisectors of external angles at vertices A and B intersect in point P, and at vertices C and D – in point Q.*

In the second sentence, the part of the NP "the bisectors of external angles" and the verb "intersect" are omitted; Rule 4 and 3 were applied.

Task 2. *In a right triangle ABC, the height CK is drawn from the vertex of the right-angle C and in the triangle ACK – the bisector CE.*

In this sentence, the verb with a part of its phrase (*is drawn from the vertex of the right-angle C*) is skipped by Rule 5. The context of triangle ABC includes the contexts of "height", "the right-angle C", and the context of "triangle ACK". The context of "triangle ACK" includes the context of "bisector CE" and "the vertex of right-angle C". The context of action "is drawn" covers the entire sentence.

Task 3. *Denote the bases of perpendiculars dropped from point A to the given lines by M and N, but the bases of perpendiculars dropped from point B – by K and L.*

Action "to drop perpendicular" has two arguments: "from a point" and "to a line". When the perpendicular is dropped from the second point B, the verb "*denote*" and the argument "*to the given lines*" being the same are skipped by Rules 3 and 5.

4 The Role of Cognitive Schemes in Understanding Sentences

The using of cognitive models of geometric configurations in the framework of understanding elliptical sentences have been proposed in [8, 9]. The process of binding objects extracted from the tasks' texts is supported by creating cognitive schemes (based on cognitive models) of objects and relationships between them. The cognitive schemes combine three components: the semantic component in the form of specific relationships between objects (typical geometric situations); the corresponding natural language description; and visual component of the corresponding geometric situation. All the graphic representations of cognitive structures are supported by interactive visualization in the system of automatic solving the planimetric tasks [20].

We also assume that the cognitive schemes correspond to the profound structures of geometrical situations outlined in the tasks' texts and define the structures of the noun phrases (NPs), prepositional phrases (PPs), and verb phrases (VPs) in the description of these situations. The cognitive approach deals with modeling processes occurring in human brain during generating sentences to describe geometric structures to be analyzed.

Now two interacting processes take a part in understanding the task texts: 1) transition from the initial text of task to the cognitive scheme representing this text; 2) transition from the description of a cognitive scheme to the possible equivalent texts of task. The last transition can be implemented on the assumption that some canonical rules will be created to generate the NPs, VPs, PPs, and the other phrases based on cognitive schemes and to combine them in the description of cognitive schemes of geometrical tasks. To compare the generated cognitive scheme descriptions with the analyzed input task's texts we need to perform their equivalent transformations (among them, converting to elliptical forms). Analyzing the text of a task will be considered successful if this text coincides with a converted description of the corresponding cognitive scheme.

Consider an approximate process of creating cognitive schemes in the course of analyzing initial task texts. For example, we have the sentence: "the bisectors of angles A and B of convex quadrilateral ABCD intersect at point M, the bisectors of angles C and D – at point N". Separate (one possibility would be to use the key words) the cognitive model of "convex quadrilateral". Construct a convex quadrilateral ABCD. By the words «bisectors of angles A and B», separate the cognitive model "drawing of bisector of an angle" and construct bisectors of angles A and B in quadrilateral ABCD. The intersection point of bisectors occurs in the cognitive scheme and the following description has been added: «the bisectors of angles A and B intersect at point X». The analysis of sentence allows us to denote point X by M: «the bisectors of angles A and B intersect at point M».

By analogy with the above consideration, we add to the cognitive scheme the bisectors of angles C and D and the description «the bisectors of angles C and D intersect at point X». According to Rule 3 of the elliptic sentence generation, if the verb is repeated, then it can be missed in the second and subsequent analogical use of it; we shall get a new converted sentence with ellipsis: «the bisectors of angles C and D – at point X». Comparing this fragment with the corresponding fragment in the original sentence allows us to denote point X by N.

Consider the following sentence: "three squares are inscribed into triangle ABC: one square has two vertices lying on the side AC, the other – on BC, the third – on AB". To create the cognitive scheme (see Fig. 2) for the whole sentence, we use the following cognitive models: «triangle» and «to inscribe a square into a triangle». When creating the complete cognitive scheme, its description is generated incrementally:

"in triangle ABC, three squares are inscribed:
one\ the first square is inscribed into triangle ABC;
the second\ the other square is inscribed into triangle ABC;
the third square is inscribed into triangle ABC;
two vertices of one\ the first square lie on the side AC of triangle ABC;
two vertices of the second\ the other square lie on the side BC of triangle ABC;

two vertices of the third square lie on the side AB of triangle” ABC.

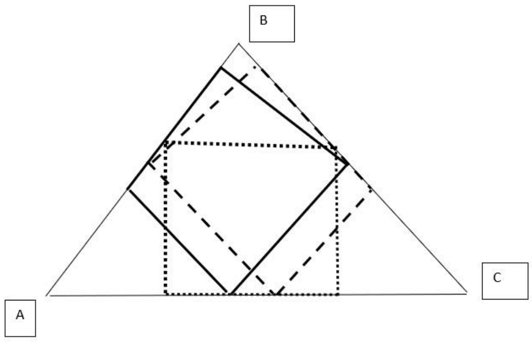

Fig. 2. Cognitive scheme for three inscribed squares

The expression “vertices of the square” can be transformed into “the square has the vertices” and “lie on” – into lying on”. And after applying this transformation to the previous sentences we obtain:

“three squares are inscribed into triangle ABC:

one\ the first square has two vertices lying on the side AC of triangle ABC;

the second\ the other square has two vertices lying on the side BC of triangle ABC;

the third square has two vertices lying on the side AB of triangle ABC”.

Since the same action is performed three times, the verb “has” is missed for second and third squares and replaced by “–“. It is possible to miss the words “of triangle ABC” because the action enters the context of it and the word “square” in the second and third previous sentences. We can also to miss the repeated part of the NP of “vertices” (two vertices lying on the side). Finally, the description of cognitive situation is converted as follows:

“One square has two vertices lying on the side AC, the other – on BC, the third – on AB”.

The omission of the word “square” is also based on referential identities [21]. The omission of the words “two vertices” is associated with constructing the squares inscribed in a triangle: any inscribed square has two vertices on one side of the triangle. Undoubtedly, the use of cognitive schemes is more difficult than we have described in our schematic examples. The problems of identical transformations of sentences and the conjunction reduction of ellipses in Russian language are thoroughly investigated in [21].

5 Inevitability of Using Equivalent Transformations of Sentences

Equivalent transformations of sentences can be divided at least in two classes: the transformations modifying the informational structure of sentences and the transformations taking place inside the phrases (NPs, VPs, PPs). Changing the order of words in sentences refers to the first kind of transformations. In some cases, the NP of a geometrical figure is transformed into prepositional phrase and precedes the NP of an element of the figure. Such information structure of sentence tells us that the expressions “in triangle ABC” or “in parallelogram” represent the theme (topic) of the sentence [20, 21], i.e., the whole sentence is associated with a certain figure – triangle or parallelogram. In our understanding, this structure gives the context of sentence. With the point of view of generative grammar, this means the possibility to move this expression in the beginning of the figure description or to copy it and to insert in the NPs of various elements of the figure without breaking (disturbing) the sentence’s content. Equivalent transformations of sentences take place on the interface between semantics and syntax [21]. More detailed study of all possible equivalent transformations in the texts of planimetric tasks requires a special consideration.

6 Conclusion

A cognitive approach to understanding task texts in planimetry is proposed. The concept of cognitive scheme as a syntactic-semantic equivalent of task’s text has been introduced. It is important in our approach to associate basic cognitive schemes with the NPs, VPs, PPs as the structural elements of sentences. The rules of cognitive transformation of full sentences into elliptical ones (design of ellipses) have been formulated. Some rules for equivalent transformations of the NPs and VPs in plane geometry sentences including the process of movement and transformation of the PPs, have been considered. Examples of cognitive-driven analysis of sentences in the texts of planimetric tasks when resolving ellipses are considered.

Acknowledgments. The research was partially supported by Russian Foundation for Basic Research, research project No. 18-07-00098A.

References

1. Khakhalin, G.: Applied Ontology in the Language of Hypergraphs. In: Proceedings of the 2th All Russian Conference with International Participation “Knowledge-Ontology-Theories” (ZONT-09), pp. 223-231, Novosibirsk, Russia (2009).
2. Khakhalin, G., Voskresensky, A.: Contextual fragmentation in linguistic analysis. In: Proceeding of 10^{th} National Conference on Artificial Intelligence with International Participation, CAI-2006, pp. 479-488, Publishing House of Physical-Mathematical Literature, Moscow, Russia (2006).

3. Kurbatov, S., Fominykh, I., Vorobyev, A.: Ontology-Controlled Geometric Solver. In: S.O.Kuznetsov et al. (eds.), RCAI 2020, LNAI 12412, pp.262-273 (2020).
4. Kurbatov, S. Naidenova, X., Khakhalin, G.: Integrating intelligent systems of analysis/synthesis of images and text: project outlines INTEGRO. Contours of the INTEGRO project. In: Proceedings of the International Scientific Conference "Open Semantic Technologies for Intelligent Systems" (OSTIS-2011), pp. 213-232 (2011).
5. Kurbutov, S., Vorobyev, A.: Ontological solver of geometric problems in natural language description. In: Proceedings of 15th National Conference on Artificial Intelligence with International Participation, CAI-2016, vol 1, pp. 56-63 (2016).
6. JSXGraph Reference: current version available http://jsxgraph.uni-bayreuth.de/docs/symbols/JXG.Board.html (2020).
7. MathJax Documentation, Release 2.7: current version available https://media.readthedocs.org/pdf/mathjax/latest/mathjax.pdf (2020).
8. Naidenova, X., Kurbatov, S., Ganapolsky, V.: Cognitive models in planimetric task text processing. International Journal of Cognitive Research in Science, Engineering, and Education, 8(1), 25-35 (2020).
9. Naidenova, X., Kurbatov, S., Ganapol'skii, V.: An analysis of plane task text ellipticity and the possibility of ellipses reconstructing based on cognitive modelling geometric objects and actions. In: A. Elizarov, N. Loukachevich (eds.), Proceedings of Computational Models in Language and Speech Workshop, CMLS 2018, co-located with the 15th TEL International Conference on Computational and Cognitive Linguistics, TEL-2018, vol. 2, pp. 70-85, the Tatarstan Academy of Sciences, Kazan, Russia (2018).
10. Kobzareva, T., Epifanov, M., Lakhuty, D.: Restoring grammatical ellipses in syntax analysis. In: Proceeding of 14th National Conference on Artificial Intelligence with International Participation, CAI-2014, vol. 1, pp. 108-116, Kazan, Tatarstan (2014).
11. Nevsorova, O., Nevsorov, V.: Intelligent tool OntoIntegrator for problem of automated text processing. In: Proceeding of the 13th National Conference on Artificial Intelligence with International Participation, CAI-2012, vol. 4, pp. 92-99, Belgorod State Technical University (BSTU), Belgorod, Russia (2012).
12. Jurafsky, D., Martin, J.: Speech and Language Processing, Constituency Parsing (Chapter 13). Prentice Hall Publisher (2020).
13. Chomsky, N: On phrases. In: R. Freidin, C. P. Otero, and M. L. Zubuzarreta (eds.), Foundational issues in linguistic theory, pp. 133-166, Cambridge, MA, MIT Press (2008).
14. Anisimovich, K., Druzhkin, K., Minlos, F., Petrova, M., Selegey, V., and Zuev, K.: Syntactic and semantic parser based on ABBYY COMPRENO linguistic technology. ABBYY, Moscow, Russia (2012).
15. Kobzareva, T.: Hierarchy of shallow syntactical analysis of Russian sentence. NTI, Series 2, 23-35 (2007).
16. Mitrenina, O., Romanova, E., Slioussar, N.: Introduction to generative grammar. 2nd ed. URSS, Moscow, Russia (2017).
17. Jurafsky, D.: A Cognitive model of sentence interpretation: the Construction Grammar approach. TR-93-077. International Computer Science Institute &University of California at Berkeley, California (1993).
18. Zhao, D.: A cognitive approach to ellipsis. Theory and Practice in Language Studies 6(2), 372-377 (2016).
19. Slioussar, N.: At the junction of theories. 2nd ed. URSS, Moscow, Russia (2011).
20. Kurbatov, S., Fominykh, I., Vorobyev, A.: Interactive visualization of cognitive structure in an integral system. In: Proceedings of the 17th National Conference on Artificial Intelli-

gence with International Participation, CAI-2019, vol. 2, pp. 222-230. Ulyanovsk State Technical University (USTU), Ulyanovsk, Russia (2019).
21. Paducheva, E.: About syntax of semantics, URSS, Moscow, Russia (2019).