Olga DEMYDENKO
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

GEOINFORMATICS AND ITS LINGUISTIC CHALLENGE

Abstract: *Contemporary linguistic studies are becoming more and more interdisciplinary thus reflecting the needs of the modern society. This greatly concerns the information technologies which have deeply penetrated into the human life being also currently widely used in performing different tasks of the linguistic research. At the same time, from the interdisciplinary perspective non-linguistic studies, especially those dealing with the various information processing may also open new, linguistic horizons. Geoinformatics in this context seems to be one of the most interesting and perspective field of contemporary science. The author of the article focuses her attention mostly on the possibility of combining geoinformatical and lexicographic studies, making an attempt to analyze strengths and weaknesses of such combination. Nevertheless, it is supposed to be also possible to consider the other ways of implementing geoinformatic technologies for doing other linguistic research.*

Key words: *data, information, geoinformatics, spatial analyses, lexicography, terminology.*

Nowadays, transfer to information society is the characteristic feature of the modern humanity development. Information technologies are getting into all the spheres into human lives. For geographers and other sciences` representatives, dealing with using of spatial data, geoinformatical technologies are of a particular interest, as they allow to involve the most magnificent potential of the computing appliances and recently developed space technologies into the investigation, applied activities and education.

Due to their character, geoinformatical technologies represent modern informational, geographical, geological, ecological technologies as well as the technologies of other natural sciences, which allow them to increase their potential significantly. Moreover, geoinformatical technologies are widely used nowadays, particularly in agricultural, economic, social sciences, in architecture, military and library domains, in regional management, in business and commerce, etc.

Before trying to come up with a comprehensive definition of geoinformatical technologies, an example revealing the capabilities of such systems will be presented. It is connected with working out a solution that might require availability of certain data and functions provided by geoinformatical systems. The functional capabilities of geoinformatical systems lie in the core of their essence.

Svitlychnyi provides us with a good example of how it works[1]. Scientist offers us to imagine that a family from not so distant future decided to buy a house. The real estate agency where the family addressed to has all the data on the houses on sale, and it has to come up with one or several variants for buying that would satisfy the customers` needs: reasonable price, not more than 2 floors, a large enough backyard, not far away from the city center, a mall nearby, not very close to highways, not in a particular district (as it has a bad criminal reputation), desirably in the western part of the city and with good ecological conditions.

The inquiry to the *data base* of the agency will easily allow them to choose all the houses by the first three criteria, but with all the other conditions will be unmanageable for the standard *data bases management system (DBMS)*. A special mechanism has to ensure the *spatial enquiries* processing, availability of spatial *attributes* of the house as a *spatial object.* This means that the houses as well as other objects have to be described according to their location in space (in terms of geographical or other coordinates, or at least in the form of postal address) and mutual spatial correspondences, and the informational system of the seller has to be able to perform, among the other things, the *spatial analyses.*

It is enough to find a house somewhere in the center, as the function of the spatial enquiry of any geoinformatical system will ensure the sample of the objects in search displayed in the circle-formed «window» within the 3-kilometers reach from the official center of the city (the post office, the central square, the particular monument) or from the center of gravity of the territory occupied by it (*the centroid*), by means of calculating the coordinates of the figure`s geometrical center and the distance.

The estimate of the house proximity to the malls will require data on those objects location, as well as about the nearby territories and the city planning in general, and also the algorithm of the optimal (shortest) way of calculation, that has to be present among the *network analyses* functions group. Assuming that all the routes are known, we have to distinguish those that are considered to be the noisiest and dirtiest, define the buffer zones around them (100 meters) and reject the houses within those buffer zones. Finally, the unwillingness to live in a certain district means exclusion of the houses within this district, which requires the knowledge of the city administrative division, and the preference towards living in the western part of the city implies the contouring of the area considered to be «the western part». In order to do that, the customer has to be given a chance to contour the «western part» by himself/herself on the electronic map of the city. However, the real estate agents can themselves demonstrate the most ecologically safe and convenient zones by means of referring to the environmental safety evaluation scheme, which can be borrowed from the local environmental service, and then by means of *data converting,* which is the export from the service`s system and the import to their own.

[1] О. О. Світличний, С. В. Плотницький, *Основи геоінформатики: навчальний посібник*, Суми 2006, p. 13–15.

So what does a real estate agency need in the end? For instance, the database on the objects on sale with standard, non-spatial attributes, and geoinformatical systems` layers with digital representations of the spatial objects and the reality phenomena (the land plots, roads, houses, supermarkets, administrative establishments, ecologically comfortable zones), which are collectively called the geoinformatical systems *informational support*. Also, the geoinformatical functions supported by the relevant software and lots of other systems will come in handy. As a result, the real estate agents will come up with the address of the necessary house or the list of such addresses (put it differently: the number, possibly empty, of the spatial objects).

This example illustrates the possibilities of geoinformatical systems as of a tool for spatial problems tackling. The number of issues that can nowadays be solved with the help of such systems (scientific, applied educational, personal) is endless and comprises infinite amount of the reality objects that are worth paying attention to.

From everything mentioned above it can be clearly concluded that GIS play a crucial role in tackling problems of various nature and belonging to deferent levels and spheres of out lives. So the logical question arises: what is geoinformatics?

According to Svitlychniy, geoinformatics is a 'branch of informatics dealing with spatial, distributive and coordinative information'[2]. The most commonly accepted definition is the following: Geoinformatics is a science, technology and the applied activity connected with collecting, storing, processing, analyzing and displaying of the spatial data, as well as with the designing, creation and using of geographical informational systems (GIS).

In a more specific sense, as a branch of the scientific knowledge geoinformatics can be considered to be an interdisciplinary science dealing with the means and methods of collecting, storing, exchange, processing, analyses and displaying of the spatial (or coordinative) information. The set of means and methods of the automatic collection, storing, manipulating, analyses and displaying (presenting) of the spatial data are unified under the collective term «geoinformatical technologies»[3]. In connection with the fact that today those means and methods are fully realized in geoinformatical systems (GIS), the term «geoinformatical technologies» is often substituted by the term «geographical informational systems technologies», or simply «GIS technology».

In general, geoinformatics is tightly connected with the geographical informational systems (GIS), as the main theoretical ideas of geoinformatics as a science are realized in the modern GIS on technical and technological levels. It lets us consider geoinformatics to be the «science, technology and manufacturing activity of scientific justification, projecting, creation, exploiting and using of GIS»[4].

The GIS functions can be conditionally divided into five groups, where the first three belong to the traditional functions of the geoinformatical technologies, and the last two are new ones that have been developed during the last decade.

[2] O. O. Svitlychnyi, S. V. Plotnytskyi. *Osnovy geoinformatyky*, Sumy 2006, p. 13.
[3] Е. Г. Капралов, А. В. Кошкарев, В. С. Тикунов идр. *Основы геоинформатики*, Москва 2004, p. 13.
[4] Ю. Б. Баранов, А. М. Берлянт, А. В. Кошкарев, Б. Б. Серапинас, Ю. А. Филиппов *Толковый словарь по геоинформатике* Retrieved from: http://biology.krc.karelia.ru.

1. First function can be called *informational,* and refers to the banks of spatial and coordinative information creation, among others:

– Creation of the digital maps;

– Creation and exploiting of the inventory systems.

2. *Automated mapping function*, which refers to the creation of the high-quality general geographic and thematic maps that would satisfy the modern requirements to the mapping products. The example of this function realization can be the Advanced Technologies Institute activity (Kyiv, Ukraine) where educational geographical and historical atlases of the Ukrainian territory are being designed and printed out.

3. *The function of spatial analysis and modeling* of the natural, commercial, social, and economic territory systems based on unique capabilities provided by mapping algebra, geostatistics and network analyses which comprise the basis of the analytical sides of the modern instrumental GIS with developed analytical capabilities. It is widely realized in the scientific investigations, as well as while making decisions connected with wide range of applied tasks of territory planning, projecting and managing.

4. *Processes modelling function* realized in natural, commercial, social, and economic territory systems. It is realized while estimating and forecasting of the natural and commercial systems` behavior as well as of their components, when tackling different scientific and applied tasks, including those connected with secure and rational use of natural resources.

5. *Decision-making support function* – in planning, projecting and management. In Ukraine, this direction is being developed the most actively in the domains of city-building projecting and planning. There are also some successes in the domain of geoinformatical support of the emergency situations. The range of examples here may be vast, if the defining of the term «decision-making support system» contents is being approached in a flexible way[5].

However, it is worth mentioning that the list above includes only the main GIS functions without taking into account the research where the GIS are being increasingly used. Besides, the list is not exhaustive, as the sphere of GIS usage is getting broader and more complicated. In particular, medical geography, epidemiology and tourism could also be added to the list, as GIS have been recently used in those spheres pretty much widely.

In this way, it becomes clear, that the range of GIS functions is vast and nowadays they play an important role in many aspects of our lives. However, it is worth mentioning that the effective operation of GIS systems is possible only on the basis of a powerful technical support. The GIS hardware comprises separate computers or computing systems with a network equipment and the peripherals (printers, plotters, scanners). The hardware permits:

- To input data from different sources;
- To store, manipulate and process vast amounts of data;
- Performance of difficult operations dealing with the spatial analyses of the vast amounts of data;
- Data presenting on the monitor or on the paper in a high quality;

[5] Е. Г. Капралов, А. В. Кошкарев, В. С. Тикунов идр., op. cit., p.25–27.

- Components` effective interaction[6].

The basic instrumental software support of GIS can be presented by a set of software products realizing particular functions.

As we can see, the creation and development of both technical and content parts of GIS require loads of intellectual efforts, uniting the scientists from different corners of the world. But here is where the main problem arises: working on tackling important issues connected with the development of GIS is impossible without the effective communication of the investigations from all over the world, which clearly represents a linguistic barrier.

GIS operation is impossible without the developers, service personnel and the users. The GIS effectiveness depends on all of them. The mandatory condition for GIS realization is the presence of one very important component which is the thoughtful users prepared for the earth space investigation and being able to work with geographical data, create on work with GIS-applications, and have certain applied knowledge in the sphere. To create GIS developers are needed, but in order to put them into function the users are needed, whereas the last group is far more numerous.

The point is that the majority of GIS and relevant applications (as well as the data on the subject matter) are in the English language, which creates a need for Ukrainian developers and users to become familiar with the relevant terminology. By and large, there is no comprehensive dictionary connected with the geoinformatical terminology, so we believe that this topic is of a special relevance nowadays.

From the abovementioned facts it becomes pretty obvious that a dictionary dealing with the terminology of the geoinformatical domain of science pleads to be created. But first, let us look into what kind of dictionary that should be. We all know that there exist a wide variety of types of dictionaries. Conditionally, they can be divided into monolingual (so-called explanatory) and bilingual (of different types depending on its purpose, use and many other factors)[7]. A terminological dictionary can be considered to be a separate type of dictionaries nowadays, as with the advent of new technologies and with the rapid development of science, new scientific directions are being created all the time which creates the need for their terminology (so-called metalanguage) systematization, which is the task of a linguistic science called *lexicography*.

Lexicography plays an important role in science and technology development and meantime serves a special «mirror» of the technical progress and the mean of individuals' professional skills perfection. Recently, computing technologies are becoming more and more significant in the dictionaries' creation. Here, terminography interacts with the informatics, which in its turn provides the methods and means of lexicographical material collecting, processing and transmission with the help of computers and other technical appliances, creation of the computer lexicographical data-bases which would be the basis for the creation of the dictionaries of various types and serving for various purposes.

This is where lexicography and informatics intersect and the notion of «linguistic corpus» appears. So what`s corpus? The most commonly accepted defini-

[6] В.В.Шипулін, *Основні принципи геоінформаційних систем*, Харків 2010, p.30–31.
[7] Ibidem, p. 52.

tion is the understanding of the corpus as the branch of linguistics dealing with designing, creation and use of text (linguistic) corpuses[8].

Before corpus the problem of the correct factual material was one of the most complicated ones. Later on, the individual experience of text collecting was formed in a unified object – card catalogue, and with certain conditions (namely the advent of the computer technologies) the next step towards the creation of the new form of the text investigation resource existence within the limits of linguistics was made, and that was when the corpora appeared.

Any language differs from the other one and it has its own past and future. The investigation of these fundamental aspects is nearly impossible without corpora. But this is an effective and useful tool only if it's not big in volume and full enough to cover all the material. Corpus is a collection of the texts belonging to a certain language in an electronic form accompanied by a scientific apparatus[9]. The apparatus which is built within a corpus is normally called an annotation. The better is the annotation, the better the corpus is. The science about corpora is first and foremost the science about the good annotation. It allows quickly and effectively access words and constructions needed by the user within some particular corpus. In order to do it, the search engine has to comprehend at least the homonymous characteristics of the words, which refers to at least partial understanding of the grammatical structure of a given language[10]. This comprehension is extremely important when we are looking not for the words, but for the forms. It is extremely hard to find a word processor capable of performing this kind of task. In order the grammatical forms to be able to be found in the text automatically, the text needs to be laid out, otherwise the task will have to be performed manually, which is a time-taking process. It is clear that the layout is necessary to solve many other kinds of issues. A text with a good layout is an invaluable thing for a specialist, as in their investigating activity the linguists depend first and foremost on the quantity and quality of the collected material[11]. Many linguists remember the times when the examples were being noted manually on the cards. Now, those days are gone, but the very process of the examples selection is being performed by human, as this is the process which is still pretty much difficult to be automatized. The annotated corpora are the first serious tool permitting to speed up and facilitate this procedure significantly. In other words, the process that used to take months from the investigators of the previous generations can now be performed in several minutes' time[12].

It is also well-known that during the creation of the electronic dictionaries of general of terminological lexicon both traditional lexicography achievements and some new ideas are being used, some of them not being able to be realized in the printed, or so-called hard dictionaries. It is first and foremost connected with the

[8] D. Biber, S. Conrad, and R. Reppen. *Corpus Linguistics: Investigating Language Structure and Use,* Cambridge 1998, p. 9.

[9] Ю. Н. Марчук. *Основы компьютерной лингвистики*, Москва 2000, p. 34.

[10] Дж. Кейтер, *Компьютеры – синтезаторы речи,* Москва 1985*,* pp. 194–197.

[11] Н. В. Бардина, „Проблеми сучасної прикладної лінгвістики", *Мова*, vol. 9, 2004, pp. 5–14.

[12] W. Grabe, R. B. Kaplan, *Introduction to Applied Linguistics*, Reading, MA: Addison Wesley 1992, pp. 69–71.

informational system organizing with clarifying of the lexical unit meaning, which refers to its semantization.

Lexicography, namely terminography, has acquired new quality with the advent of electronic dictionaries and new technologies of their creation. The electronic dictionaries and encyclopedia are being developed today as autonomous or network programming products. It is undisputable that the electronic dictionaries provide the user with many additional possibilities in comparison with their printed analogues:

1) storing of big amount of data due to hyperlinks: the dictionary implies the inclusion in its structure of several other dictionaries of different types (combinability dictionary, terminological dictionary, the dictionary of grammatical norms and the others);

2) availability of the dictionary due to the active search system (full-text search, simultaneous search in several dictionaries, fast search);

3) use of multimedia means for the lexicon semantization: voicing of the head words, illustrative material input together with the pictures, animations, videos);

4) use of dictionaries in the local and global networks, whereas the work with the dictionaries can be performed by several users simultaneously;

5) time and money saving while creating of electronic dictionaries[13].

Currently the question arises, is an electronic dictionary just a convenient modern version of the «paper» dictionary, or it is a new stage in the lexicography development? It is often considered that the dictionary on the computer is just an electronic copy of the printed one which is equipped with a convenient interface and automatic search engines. However, this point of view considerably underestimates all the possibilities of the computer technologies. The new form of a dictionary permits to eliminate the drawbacks of the «paper» lexicography and meantime take it to a better level of a higher quality[14].

The main peculiarity of an electronic dictionary is its capability of tackling the main problem of the printed one: the more information it offers, the more difficult it is to work with it. With regard to this problem, the printed dictionaries are divided into two categories. The first one comprises popular pocket dictionaries, which are relatively convenient but pretty primitive, and it often happens that the word the translation of which you need to know is simply absent. The second category refers to the professional editions, which are not suitable for quick search of information due to their large weight and huge format.

The electronic dictionaries combine vast volumes of information in them and their convenience. This, first and foremost, can be achieved due to the automatic search engine. To find the word you need, it is enough just to enter it in the command line, click «search» and you'll have the explanation. Some dictionaries permit words` defining right in the word processors or any other program.

The usefulness of an electronic dictionary is the main purpose of its creation, so it has to satisfy the needs of different groups of people with various linguistic experiences. This is achieved by structuration of the word article, which highlights

[13] Л. Л. Нелюбин, *Компьютерная лингвистика и машинный перевод*, Москва 1991, p. 101.
[14] Дж. Кейтер, *Компьютеры - синтезаторы речи*, Москва 1985, pp. 183–185.

the main information types concerning the word. Also, any element can serve as a hyperlink to another article. This system ensures possibility of a quick access to the needed information without exhausting searches, just by one mouse click.

Another aspect of the electronic dictionary being better is its adequate reflection of the modern lexicon, which is a subject to a constant change as the new words appear all the time. The 'paper' dictionaries lack this function just because it takes so much time to do one and it's close to impossible to update it in any way. In fact, the majority of such dictionaries are simply outdated nowadays[15].

Also, electronic dictionaries are much more convenient for those people working with a text who have to understand clearly which of the words in it can be combined, which prepositions are to be used for this purpose and if there are relevant set expressions. Performing this task with a «paper» dictionary is extremely difficult, while on a computer the user can open several word articles simultaneously, compare them, and investigate the smallest and subtlest nuances of every word that they happen to encounter within a text.

To conclude, taking into account all the facts proving the relevance of geoinformatics terminology systematization, it is necessary to highlight that creation of the corresponding dictionary (preferably an electronic version) would be of great importance to boost the effectiveness and effectivity of researches in this sphere, as lack of order within the terminological system and the absence of comprehensive and reliable source for explanation of certain terms hinders researchers in their attempts to take geoinformatical investigations to a new level.

[15] Г. Г. Почепцов. *Информационные войны*, Москва 2000, p. 15.

References

1. Biber D.,Conrad S., Reppen R., *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge 1998.
2. Grabe W., Kaplan R. B., *Introduction to Applied Linguistics*, Reading, MA 1992.
3. Баранов Ю. Б., Берлянт А. М., Кошкарев А. В., Серапинас Б. Б., Филиппов Ю. А.,*Толковый словарь по геоинформатике*. Retrieved from: http://biology.krc.karelia.ru/.
4. Бардина Н. В., „Проблеми сучасної прикладної лінгвістики", *Мова, vol. 9*, 2004, pp. 5–14.
5. Капралов Е. Г., Кошкарев А. В., Тикунов В. С. И др., *Основы геоинформатики*, Москва 2004.
6. Кейтер Дж., *Компьютеры – синтезаторы речи*, Москва 1985.
7. Марчук Ю. Н., *Основы компьютерной лингвистики*, Москва 2000.
8. Нелюбин Л. Л., *Компьютерная лингвистика и машинный перевод*, Москва 1991.
9. Почепцов Г. Г., *Информационные войны*, Москва 2000.
10. Світличний О. О., Плотницький С. В., *Основи геоінформатики: навчальний посібник*, Суми 2006.
11. Шипулін В. В., *Основні принципи геоінформаційних систем*, Харків 2010.