"Accuracy" vs "Unambiguity" in Linguistics

Volodymyr Shyrokov

Ukrainian Lingua-Information Fund NAS of Ukraine, Holosiivskyi av. 3, Kyiv, 03039, Ukraine

Abstract
The phenomenology and the concept of unambiguity in linguistics in comparison of the same aspects of accuracy are under investigation. The effect of the difference between accuracy, on the one hand, and ambiguity, on the other, is demonstrated by the example of the analysis of the texts of the legislative framework of Ukraine, namely, by analyzing of Article 4 of the Constitution of Ukraine. The use of the theory of semantic states and the apparatus of hyperchains in lexicographic structures made it possible to formally accurately fix the semantics of linguistic constructions in the studied contexts. Therefore, the concepts of accuracy and unambiguity should be distinguished when conducting semantic analysis. Accuracy here means the most accurate definition of all lexical meanings, semantic states and their superpositions in which the analyzed token can function. The necessity of creating the State Linguistic Corpus of Legislative Acts and the State Thesaurus of Ukraine was emphasized.

Keywords 1
Natural language processing, computer lexicography, unambiguity, accuracy, semantic states, hyperchains, superpositions of semantic states

1. Introduction

Once upon a time, our outstanding linguist Vitaliy Makarovych Rusanivsky said in a conversation with me: "Linguistics is a science, though humanitarian, but accurate". That was a long time ago – about 30 years ago. At that moment, I liked these words as just an aphoristic expression. However, at that time I did not think at all that I would ever have to study the accuracy of linguistic constructions, and even in connection with aspects of their unambiguity.

And really: what is accuracy in linguistics? And what are relationships between accuracy and unambiguity? And in general, in what context these questions arise? In this regard, I would like to mention the work of D. Likhachev "More on accuracy in literary criticism", which expresses interesting views on accuracy: "In literary criticism there is some kind of inferiority complex, caused by the fact that it does not belong to the cycle of exact sciences. High degree of accuracy is supposed to be a sign of 'scientificity' in any case. Hence we can see various attempts to subordinate literary criticism to the exact method of research that result in inevitable reduction of the literary criticism range to more or less determined limits. As is commonly known, any scientific theory is considered accurate when the generalizations, conclusions, and data rely on some homogeneous elements with which it would be possible to perform various operations (including combinatorial, mathematical). For this purpose the research material needs to be formalized. Since accuracy requires that the scope of the study and the study itself is to be formalized, all attempts to create an accurate method of research in literary criticism are somehow related to the desire to formalize the literary material. And in this endeavor, I want to emphasize this from the beginning, there is nothing odious. Any knowledge is subjected to formalization, and any knowledge itself formalizes the material. Formalization becomes inadmissible only when it forcibly attributes to the material the degree of

accuracy which it does not possess at all and, essentially, cannot possess. Therefore, the main objections of various kinds to excessive attempts to formalize the literary material come from the indications that the material cannot be formalized in general or, in particular, the proposed type of formalization. Among the most common mistakes is the attempt to extend formalization on the whole material, while it is suitable only some part of the material".

The present paper concerns literary studies, but its main provisions are valid for any humanitarian subject. The task of this work is to show that phenomenology and the concept of unambiguity in linguistics do not coincide with the same aspects of accuracy. We will demonstrate this effect with an example from the legal framework of Ukraine. It has accumulated a lot various text deviations, the number of which is growing rapidly and which negatively affect law enforcement practice and the legal regime of the state. Let us consider a real example in this area, where you can see and understand the effects of the interaction between accuracy and unambiguity of language structures.

The legal sphere of Ukraine was chosen as an object of research not by chance, but for certain reasons. The fact is that a feature of the legal regime of Ukraine, as well as other post-Soviet states is the extremely high intensity of legislative and other rule-making processes. For example, during the first months of the Verkhovna Rada of Ukraine of the 9th convocation, 1,490 draft laws and resolutions were registered. Of these, only 94 (6.31%) bills and 267 resolutions (17.92%) were adopted. Deputies' bills are most often taken, and the peak of the "turbo regime" came in October 2019. In 2019, the Verkhovna Rada of the new convocation adopted the most bills, the subjects of which were people's deputies, in second place – the initiatives of the President, the last - the bills of the Cabinet of Ministers. It should be added that this turbo activity has not decreased much in 2020, despite the pandemic.

As a result of such activity, the legislative (and regulatory framework in general) has already accumulated a large number of different textual deviations, the number of which is growing rapidly and which negatively affect law enforcement practice and the legal regime of the state. It is becoming increasingly clear that actions need to be taken to combat the linguistic deviations of legal documents.

2. Theoretical framework and tools

In this connection, the question inevitably arises: what is the source or sources of textual deviations and how are they related to the language system? The answer to this question, in our opinion, can be obtained approximately in such a mode as following. The system-forming relations of language, we believe, are as follows: <Subject – Object> and <Form – Content>. In this case, access to the content can be provided only through language forms. However, there are asymmetric relationships between the <Subject – Object> and <Form – Content> relations, close to those of Kartsevsky's principle of asymmetric dualism of the language sign. In fact, the linguistic sign and its meaning are not isomorphic and do not have a one-to-one correspondence. Their boundaries do not coincide at all points: one form can have several functions; one meaning can be expressed by several forms.

As a result, the same language "form" "seeks" to express as many "meanings" (in particular, polysemy, homonymy), and the same "meaning" "seeks" to be realized by the maximum number of different "forms" (in particular). This is typical for all levels of the language system.These factors serve as objective grounds for the generation of various kinds of linguistic ambiguities. These objective grounds are accompanied by numerous subjective ones. Their source, as a rule, is the fact that the language competence of the authors of texts (as well as editors, etc.) is usually not accompanied by the appropriate linguistic competence.

Unfortunately, in classical linguistics the formal apparatus for representation, interpretation and qualification (including quantitative) of linguistic ambiguities is poorly developed. This circumstance objectively determines the need for (logical) – linguistic examination of legal texts. In the Ukrainian Lingua-Information Foundation of NAS Ukraine considerable experience has been accumulated in conducting various kinds of logical and linguistic examinations, on the basis of which a methodology for performing this kind of work has been developed and effective digital linguistic technologies have been developed for their support and maintenance.

To illustrate this, we give only one example where you can see some of the relationships between accuracy and unambiguity that occur in legal texts. This example concerns the linguistic examination on the interpretation of the term "single" in the context: "Article 4. There shall be a single form of citizenship in Ukraine. The grounds for the acquisition and termination of Ukrainian citizenship shall be determined by law." (The Constitution of Ukraine).

In Ukrainian, this context is as follows: «Стаття 4. В Україні існує єдине громадянство» (Конституція України). The basis for the linguistic examination was a letter from the First Deputy Chairman of the Verkhovna Rada of Ukraine R. O. Stefanchuk (22.02.2021 р. № 02/09-2021/55635). This example relates to the task set to the author: to perform a linguistic analysis of the token "single" (єдиний) in the context of "Article 4. There is a single citizenship in Ukraine" (The Constitution of Ukraine) [«Стаття 4. В Україні існує єдине громадянство»] and to make recommendations for correct interpretation of the token "єдиний" in the context of the above Constitution Article.

To resolve solve the task above the following we applied the theory of semantic states [1] and the method of constructing the so-called hyperchains on lexicographic structures [2]. For processing linguistic material we used the software tools such as Virtual lexicographic laboratory "Dictionary of the Ukrainian language", Dictionary of the Ukrainian language in 20 volumes online (https://services.ulif.org.ua/expl/Entry/ index?wordid= 1&page=0); Integrated Lexicographic System "Dictionaries of Ukraine" (https://lcorp. ulif.org.ua/dictua/), Ukrainian National Linguistic Corps (http://unlc.icybcluster.org.ua/ virt_unlc_4.5), Language and information system "The Constitution of Ukraine".

3. Researching Unambiguity

The starting points of our study are the conclusions, justifications made by scientists during the study of the phenomenon of ambiguity in language (or speech):

1. Linguistic ambiguity is the ability of a word, expression or construction to have different meanings, i.e. it is a property of linguistic units, which is manifested in a phrase. In this case, we distinguish linguistic ambiguity and speech unambiguity. The latter can be unintentional, it can be eliminated in the process of further communication, or intentional, and in this case it is used as a literary technique [1];

2. linguistic ambiguity means the presence of several different meanings in a speech (sentence) at the same time. Depending on the nature of the meanings, the ambiguity can be lexical (when it covers a single language unit) or syntactic (if it concerns an entire utterance) [2]. At the same time, other researchers, for example, [3] emphasize that the simultaneous presence of two different understandings of a word, phrase or the whole text leads to appearing of a new meaning (sense);

3. linguistic ambiguity arises due to a random coincidence of signs (letters during the processing of a written text) and does not imply parallelism of language units [4];

4. Depending on how linguistic and speech ambiguity is exploited, the arsenal of linguistic means of linguistic experimentation is traditionally studied at different levels of language: syntax, morphology, semantics, pragmatics, etc., because resources of all language levels are used to varying degrees for language play.

So, the semantics of the token "single" (єдиний) in the phrase "single citizenship" (єдине громадянство) is determined using the theory of semantic states [5] and the method of constructing the so-called hyperchains [6] and has the structure shown in the Figure 1. Here in column 2 are the hypersems to the word "single" (єдиний). These hypersems outline the semantic properties of the word "single" (єдиний) in contexts where it acquires the lexical meanings given in column 3. In this work, digital linguistic resources [7-10] of the Ukrainian Lingua-Information Foundation of NAS Ukraine were used.

1	2	3
єдиний		
	лише один →	ніякий інший, крім названого; тільки названий (уживається при виділенні з ряду інших окремої особи, предмета і т. ін.)
	загальний →	який стосується усіх, поширюється на всіх, на все загалом
	спільний для всіх →	належить всім або багатьом
	єдність →	цілісність, неподільність
	цілісний →	має внутрішню єдність, сприймається як єдине ціле
	неподільний →	який не ділиться, не розпадається на частини

Figure 1: The semantics of the token "single".

Let us introduce the notation for these hypersems with the corresponding lexical meanings:
- C1 = (only one) → (no other than the named; only named)
- C2 = (common; common) → (relating to …)
- C3 = (indivisible → is not divisible, does not break into parts; integral → internal unity - a single whole).

According to the theory of semantic states in the context of "There is a single citizenship in Ukraine" (В Україні існує єдине громадянство), the token "single" (єдиний) functions in the superposition of partial semantic states:

$$C1 + C2 + C3, \tag{1}$$

But keep in mind that states C1 and C3 have hypersemic characteristics "Only one" and "Indivisible; Integral", respectively. Therefore, we can introduce the notation of this common semantic state by the symbol C (1 + 3). The latter are members of one synonymous series with the meaning: "which is an internal unity, is perceived as something inextricably linked". Then the general semantic state of the analyzed token in this context has the form:

$$C(1 + 3) + C2, \tag{2}$$

The first term of this sum (C(1 + 3)) has the logical-semantic qualification "constitutes" (становить), and the second (C2) – "applies" (стосується). That is, in this complex meaning, they have the specified semantic properties (features) simultaneously. We believe that this fact has been established by us quite accurately. Although, it is not necessary to speak about unambiguity here. In this case, the polysemy was precisely established, which is intrinsically relevant to the investigated context and the given semantic situation.

4. Conclusions

1. The research suggests that certain terms even in the documents of the highest level public law (including the Constitution of Ukraine) may have semantic ambiguity.
2. Therefore, the concepts of accuracy and unambiguity should be distinguished when conducting semantic analysis. Accuracy here means the most accurate definition of all lexical meanings, semantic states and their superpositions in which the analyzed token can function.
3. Thus, in the context of "Article 4. There is a single citizenship in Ukraine" (the Constitution of Ukraine), the word "single" is a combination of several meanings with semantic qualifications: {"only one"; "Common, common to all"; "Integral, indivisible"}.
4. According to the above, the term "single citizenship" implies that in Ukraine only one citizenship, common to all its citizens, is integral and indivisible in nature. In this case, the relative

weight or significance of the above semantic qualifications from Article 4 is not determined, so they are all considered equal here.

5.   Further details or semantic concretizations can be obtained from the analysis of the Law of Ukraine "On Citizenship", which determines the legal content of citizenship of Ukraine, the grounds and procedure for its acquisition and termination, the powers of public authorities involved in citizenship of Ukraine, the procedure for appealing decisions issues of citizenship, actions or inaction of public authorities, their officials and officials, as well as the definition of terms involved in the concept of citizenship.

6.   We draw attention to the relevance and importance of creating a system of linguistic expertise in Ukraine, not only legislation but also the regulatory framework in general.

7.   This problem can be solved only on the basis of modern intelligent information and linguistic technologies, the prototypes of which have already been largely developed in the Ukrainian Language and Information Fund of the National Academy of Sciences of Ukraine.

8.   Now we consider the development of the State Linguistic Corpora of normative legal acts and the State Thesaurus of Ukraine to be absolutely necessary. The creation and implementation of these tools, in our opinion, is able to raise the work of the entire legal system of Ukraine to a qualitatively higher level.

9.   There are many examples of this type, which allows us to make some generalizations:
- Namely: phenomena and concepts of accuracy and unambiguity in linguistics do not coincide. Quite the opposite: they are a certain opposition.
- Its essence is that the effects of linguistic ambiguities must be described, codified and formalized.
- The accuracy of linguistic interpretations lies in the unambiguous attribution to specific speech (text) contexts of the relevant formulas of ambiguities developed in the theory.

5. References

[1] L. V. Vlasova, Semantic diffusion, semantic ambiguity: definition of the concepts, 2014 (*in Russsian*).   URL: http://cyberleninka.ru/article/n/semanticheskaya-diffuziya-semanticheskaya-neopredelennost-opredelenie-ponyatiy.

[2] Yu. D. Apresyan, Lexical semantics: synonymous means of language, Nauka, Moscow, 1974 (*in Russian*).

[3] R. O. Jacobson, Linguistics and poetics, Structuralism "pro" and "contra" (1975): 193–230 (*in Russian*).

[4] N. V. Kopotev, Ambiguity and the ways of its resolution in the Helsinki annotated corpus HANCO, in: Proceedings of the International Conference "Corpus linguistics-2014", St. Petersburg, 2004, pp. 165–175 (*in Russian*).

[5] V. A. Shyrokov (Ed.), Computer linguistics studies: Proceedings of Ukraininan Lingua, Vol. 1: Research paradigm and main language-information structrures, Ukrainian Lingua-Information Fund, 2018.

[6] V. A. Shyrokov (Ed.), Computer linguistic studies: Proceedings of the Ukrainian Lingua-Information Fund NAS of Ukraine. Vol. 3: Explanatory lexicography. Book 2 System semantics of explanatory dictionaries, Ukrainian Lingua-Information Fund, 2018.

[7] Virtual lexicographic laboratory "Dictionary of Ukrainian Language" (SUM 20). URL: https://services.ulif.org.ua/ expl/Entry/index?wordid= 1&page=0.

[8] Integrated lexicographic system (ILS) "Dictionaries of Ukraine". URL: https://lcorp.ulif.org.ua/dictua/

[9] Ukrainian national linguistic corpus. URL: http://unlc.icybcluster. org.ua/virt unlc_4.5

[10] Language Information System "Constitution of Ukraine", Software, 2008.