Научная статья
УДК 811.111

Формирование терминологии в области информационных технологий и компьютерной лингвистики

Аёдижи Акинола Оладеле[1]
[1]Новосибирский государственный педагогический университет, Новосибирск, Россия

Аннотация. Потребность в общей терминологии и ее адекватном переводе привела к инкорпорированию лингвистический принципов из области терминологии в работу искусственного интеллекта и других сфер компьютерной науки.

Главная цель унификации терминологического аппарата заключается в последующем создании цифровой системы, которая будет применять социолингвистические принципы для эффективного объяснения механизмов, задействованных в формировании различных лексических единиц. Такая система также может применяться для разрешения типичных переводческих проблем (например, при переводе внутриотраслевых синонимов), повышения эффективности коммуникации и упрощения международного сотрудничества.

Лингвисты и специалисты в области теории вычислительных систем изучают способы разработки необходимых инструментов для эффективного перевода на разные языки. Основной целью является создание объемлющей лингвистической системы, основанной на большой базе данных и разных лингвистических принципов.

Ключевые слова: язык, лингвистика, социальная лингвистика, функционирование языка, искусственный интеллект, словообразование, информационные технологии.

Для цитирования: Оладеле А. А. Формирование терминологии в области информационных технологий и компьютерной лингвистики // Актуальные проблемы филологии и методики преподавания иностранных языков. 2023. Т. 17, № 2. С. 118–123.

Original Article

Formation of terminology in Information technology and Computational linguistics

Ayodeji A. Oladele[1]
[1]Novosibirsk State Pedagogical University, Novosibirsk, Russia

A need for common terminology and translations has led computer science through Artificial intelligence applied with linguistics in the study of segmental terminologies. The objective is a system that abides by a sociolinguistic approach, such a system has to provide an efficient explanation of the mechanisms applied in the formation of any lexical units. This will include the passing of negative phenomena, including examples like intra-branch synonymy the usual bottleneck of effective translations, enhanced communication, and international cooperation due to language differences. Linguistics and computer scientists have put in efforts to research the purpose of making available the needed instruments for linguistic translations. The goal will be to establish a rich knowledge-based system.

Keywords: language, linguistics, social linguistics, language function, artificial intelligence, word formation, Information technology.

For citation: Oladele A. A. Formation of terminology in information technology and computational linguistics. *Topical issues of philology and methods of foreign language teaching*, 2023, Vol. 17, no. 2, pp. 118–123. (In Russ.)

Introduction and Background of the study. Computational linguistics is an interdisciplinary field concerned with the computational modeling of natural language, as well as the study of appropriate computational approaches to linguistic questions. In general, computational linguistics draws upon linguistics, computer science, artificial intelligence, mathematics, logic, philosophy, cognitive science, cognitive psychology, psycholinguistics, anthropology, and neuroscience, among others as rightly noted by V. A. Tatarinov, knowledge of the history of science is determined by knowledge of the history of the language of science [8].

Computational linguistics is often grouped within the field of artificial intelligence but was present before the development of artificial intelligence. Computational linguistics originated with efforts in the United States in the 1950s to use computers to automatically translate texts from foreign languages, particularly Russian scientific journals, into English since computers can make arithmetic (systematic) calculations much faster and more accurately than humans, it was thought to be only a short matter of time before they could also begin to process language

ln, his Terminology Manual, Helmut Felber states the following about the relationship between artificial intelligence and the General Theory of Terminology (GTT): "Since the studies in artificial intelligence and the establishment of expert systems have to deal with concepts, systems of concepts, concept linking, conceptual mapping of reality, etc., a stronger affinity between the GTT and computer science is to be expected shortly" [1]. Human language development provides some constraints that make it harder to apply a computational method to understanding it. For instance, during language acquisition, children are largely only exposed to positive evidence. This means that during the linguistic development of an individual, the only evidence for what is a correct form is provided, and no proof of what is not correct.

This is insufficient information for a simple hypothesis testing procedure for information as complex as language, and so provides certain boundaries for a computational approach to modeling language development and acquisition in an individual. “Terminology is the most sensitive part of the vocabulary to external influences”, it clearly shows the influence of society on the language [2].

Methodology. To understand the relationship between human and artificial intelligence, it is important to highlight the difference which is the fact that the latter has to transform all of its tasks into two elementary "mental" processes:

(l) Comparison based on the identification and analysis of available data.

(2) Synthesis according to patterns and rules.

These two operations can be applied when tackling the following tasks:

Decomposition, i.e. splitting of characteristics of the same rank.

Filing, i.e. arrangement according to dissimilar characteristics.

Abstraction, i.e. selection of essential characteristics and omission of inessential characteristics.

Reduction, i.e. omission of irrelevant and equivalent characteristics.

These tasks can be performed by machines and constitute the operational framework of knowledge banks. The determination and recording of characteristics is an essential part of systematic terminology work. ln linguistic data processing it is necessary to abide by certain conditions which are set by the strictly logically oriented operation of machines. This is of particular relevance for the semantic level of the language.

Limitations. D. F. Robertson points out that these limitations constitute a particular challenge to Al-research: "Originally, AI research proceeded on the assumption that it might be possible to endow computers with symbolic reasoning capabilities similar to those of the human mind by isolating a few fundamental laws of logic, distilling those laws into sophisticated software systems and linking their operation to powerful computers" [6].

This is still a particular challenge to researchers in artificial intelligence, to coordinate and adopt a large number of linguistic, communicative, and cognitive abilities effectively. But the prize is worth the effort: if they are successful, they can achieve a much higher accessibility of knowledge than in the case of conventional databases and data banks where titles and texts of papers have to serve as titles and intermediaries.

Concerning semantic abilities there is one major difference, digital machines can only process structures, "meaning" has to be structured, therefore. This approach has already been developed by the General Theory of Terminology irrespective of any computer application. Meaning is represented in form of concepts that are parts of systems of existing concepts.

Topicality. The forerunners of strict concept systems in knowledge representation are the so-called "semantic networks" which should be correctly named "associative networks" [3]. ln, knowledge banks one encounters both fixed concept relationships and so-called associative relationships. The latter plays a very interesting role and can be described as "coming across interesting information" or as simple "browsing".

Direct access to the various categories of knowledge is only possible via concepts that are in relationship to each other or have some other kind of linkage. This means that terminological principles have to be observed.

On knowledge representation, Hayes and McCarthy [4] divided, “knowledge” into two distinctive sections:

1) The epistemological knowledge which consists of data in specific data structures.

2) Heuristic knowledge, which describes how the data has to be handled. Concepts would fall under category l and terminological principles under category 2.

As a consequence of the fact that systems become more "intelligent" and perform more demanding tasks, there is an increase in ambiguous meanings or contradictions within the special language used.

This requires explanatory dialogues between systems and users, to elucidate for instance why the systems have come to a certain conclusion and what the user could do if he is not satisfied with a certain result. Seeing this problem from the terminological point of view one can say that especially definitions and explanations contribute considerably to the transparency of a system.

Goal. H. J. Levesque mentions a further advantage of definitions and justifies terminological access systems to knowledge representations as follows: "Why should a representation scheme that is interested only in expressing knowledge about a world have to deal with statements that carry no information? The answer, I believe is that definitional mechanisms facilitate the interaction between a user (man or machine) and the Knowlege Base. The addition of terminology is not a matter of truth or falsity but a matter of convenience to the user. ln particular, the definitional mechanisms of a language should help structure and should organize what eventually will be said or asked about the world" [5].

Thus, the following preparatory work has to precede the creation of a knowledge bank:

1) Investigation of the structure of the respective subject terminologies and their applicability in a modified form as documentation languages.

2) The selection and further development of universal and hierarchically structured ordering systems.

Novelty. Now let us take a look at the development of computer programs for a user-oriented and automatic reshuffling of knowledge units. implicitly contained knowledge has to be made explicit by generating references and indexes. This is the case in information-extracting and convincing systems that have the capability to "comprehend" machine-readable texts and extract relevant information. Such capabilities are particularly important for automatic abstracting and machine translation.

Eugen Wooster has developed practice-oriented methods for conceptual ordering and economy. Especially the relationships between concepts can be used as signposts for tracing knowledge units in a bank. Furthermore, they facilitate the following functions of machines with artificial intelligence: classification (structuring), condensation, abstraction, and association. All these functions are directly dependent upon the type of characteristics that determines the criteria for subdivision and creation of schemes. lt also allows purpose-oriented abstraction which has a screening function and is a prerequisite for the operation with relevant data only [9].

The use of Expert systems as a practical application of AI Research is a factor to be also considered. Expert systems are knowledge-based systems that provide immediate access to expertise and other cognitive capabilities stored in machines. Expert systems can be distinguished from traditional information systems on the grounds of their inference functions and their storage capacity for general background knowledge.

They are designed for heterogenous and complex subject fields rather than for homogenous and mass data and serve as interactive advisory units. For this reason, it is necessary to use adequate access mechanisms such as structured terminologies and concept systems. The essential aspect in this matter is not so much the contents but the structure of the data, i.e. the relationships between them there is a deep structure of concepts that provide access to the entire system.

The role that artificial intelligence research has assumed in this matter is explained by D.F. Robertson: "When computer reasoning power was found to be a function of intensive knowledge about a limited domain of facts and heuristics, the focus of Al research naturally shifted to the development of expert systems possessing an enormous quantity of highly structured information about a well-defined set of problems" [6].

This is a very important task because only structured and categorized knowledge is of use to the expert. It goes without saying, that these complex processes have to be investigated very thoroughly and have to be made operational for the respective field of application. It has to be emphasized, however, that the further development and full implementation of the General Theory of Terminology will enrich the spectrum of possible solutions and will guarantee the practice orientation of this interdisciplinary field of research.

The use of an Intelligent user interface for terminology should be a strong consideration. The crucial point in establishing a knowledge bank is the selection of an adequate database management system. The development in this field is quite rapid and one can say that there is a general trend toward more user-friendliness.

Michael Stonebraker summarized the requirements for user-friendliness as

follows: "Hopefully, the next generation DBMSs (database management systems) should be noticeably better than the current ones. I would like to see DBMSs that deal much more intelligently with complex objects (e.g. forms, maps, icons, etc.) and relationships (e.g. generalization, inheritance, and associations)" [7].

Conclusion. In conclusion, it is constructive to say that It is typical for many advisory or consultancy situations that the expert does not only draw upon his own expertise but also on external databases which contain formatted mass data that cannot be part of his own memory. it is now up to the skill of the consultant to combine these two sources in the most efficient way and to interpret the stored facts on the grounds of his expertise.

If these functions are to be carried out by a knowledge bank, it is necessary to develop specific aids for user guidance. Users usually judge their database by the capabilities of the user interface. Therefore, it is very important to give this aspect sufficient consideration. Artificial intelligence functions are of particular importance for the creation of user interfaces, i.e. the software that mediates between a person and the programs shaping the system into a tool for specific goals, especially in linguistic tools.

The specific task that has to be kept in mind in this context is terminology work in general and as an access function to expert systems in particular. The user of such a system should be confronted as directly as possible with the network of concepts. This can be facilitated by means of a dialogue that points out the possible relationships to other concepts or subjects

Список источников

1. *Felber H.* Terminology Guide. Paris: UNESCO, 1984. 426 p.

2. *Isayeva N. S.* The role of information technology in the development of linguistics // Advances in Economics, Business and Management Research. Vol. 138. 2020. P. 698–701.

3. *Kobsa A.* Wissensrepresentation. Die Darstellung von Wis-sen im Computer (Knowledge Representation. Knowledge representation in computers). Vienna: Austrian Society for Cybernetic Research, 1982. 156 p.

4. *Hayes P. J., McCarthy J.* Some philosophical problems from the point of view of artificial intelligence // Machine Intelligence 4. Edinburgh: Edinburgh University Press, 1969, P. 463–502.

5. *Levesque H. J.* Logic of incomplete knowledge bases // On conceptual modeling. Prospects of artificial intelligence, databases, and programming languages. Springer, 1984, P. 184.

6. *Kubler S.* Introduction to Computational Linguistics [Электронный ресурс]. URL: https://www.linguisticsociety.org/sites/default/files/e-learning/intro_cl.pdf (дата обращения: 23.11.2022).

7. *Stonebreaker M.* Database Perspective // About conceptual modeling. Prospects of artificial intelligence, databases, and programming languages. Springer, 1984. 458 p.

8. *Tatarinov V. A.* The history of Russian terminology. M.: Moscow Lyceum, 1995. 336 p.

9. *Wooster E.* The general theory of terminology – the boundary area between linguistics, logic, ontology, computer science, and subject areas // Universal terminological science in Grenz-gebiet. Linguistics. 1974. № 119. P. 61–106.

Информация об авторе

А. А. Оладеле – ассистент кафедры лингвистики и теории перевода, Новосибирский государственный педагогический университет, joyfullayo@gmail.com

Information about the author

A. A. Oladele – Assistant of the Department of Linguistics and Translation Theory, Novosibirsk State Pedagogical University, joyfullayo@gmail.com

Статья поступила в редакцию 22.12.2022; одобрена после рецензирования 25.12.2022; принята к публикации 30.12.2022.

The article was submitted 22.12.2022; approved after reviewing 25.12.2022; accepted for publication 30.12.2022.