Abstract. The research aims to explore constructed languages as semantic and semiotic systems by analyzing various types of languages based on their lexical, syntactic, morphological and other features. In order to achieve this goal, the author examines the existing classifications of constructed languages and attempts to establish a connection between purposes of their creation and their linguistic features on various levels. The research topic relevance is determined by a substantial rise in popularity of constructed languages, the emergence of their new roles and functions as well as a multitude of new types of media available for communication in these languages and their distribution. The author argues that the recent development in technology provides constructed language creators and enthusiasts with new non-verbal ways of expression that were previously unavailable and thus facilitates communication. This hypothesis is confirmed by several case studies, including the one of “SolReSol: The Project”, an open-source computer program developed by the author, which automates and improves the implementation of semiotic systems designed back in the 19th century. Furthermore, attention is also drawn to the problem of eurocentrism in constructed languages. The research findings lead to the conclusion that on closer inspection both a priori and a posteriori constructed languages created by native speakers of European languages inevitably reveal a certain percentage of Standard Average European features in their semantic and semiotic systems.

Keywords: constructed language, conlang, auxiliary language, planned language, semantics, semiotics, eurocentrism, linguistic relativity

Article history:
Received: 01.06.2021
Accepted: 1.03.2022

For citation:
Introduction

As noted by Professor L.A. Novikov, “Due to the interconnection of various aspects of semiotic and linguistic theories, the meaning of language elements may be described not only from the semantic perspective per se, but also in terms of pragmatics, structure and paradigms” [1. P. 403]. Since all of the aforementioned aspects are also present in various types of constructed languages, albeit to different extents, general analysis of their semantic and semiotic features is deemed feasible. It should be specified that for the purpose of terminology standardization, this research uses the English term constructed language (or conlang) as an umbrella term that encompasses all types of non-natural languages (the problem of their...
A multitude of reasons serve as a foundation of this research relevance. First, it is argued that “language remains to be one of the forms of reflection, expression and comprehension as well as a thinking tool” [2. P. 30]. This statement reveals the interest in the exploration of constructed languages, which, by definition, greatly differ from the natural ones: their semantic and semiotic systems, either purposely or not, might lead to creation of new and unique thinking tools.

Secondly, the timeliness of a deeper analysis of semantic and semiotic systems of constructed languages is due to the recent development of telecommunications which leads to a twofold paradigm shift: it enables the emergence of new semiotic systems which offer previously unavailable ways of expressing ideas as well as creates an informational space for niche international linguistic projects that otherwise would not be able to reach the critical mass of followers needed for their further development.

The usual counterargument to the statement regarding the relevance of any research focused on constructed languages may start with the indication of their impracticality and the failure of even the most notable constructed languages to achieve their goals of becoming highly popular methods of communication. However, it should be argued that such a point of view is inherently narrow since it only takes into consideration the idealistic and unreachable goals that are no longer shared by the overwhelming majority of modern conlang enthusiasts that see creation and development of a new language as a linguistic and social experiment or a form of artistic expression rather than an attempt to establish a new international language that would rival the most widespread natural ones.

The aforementioned misconception formed a stigma that researchers are well aware of: “Linguists do not generally consider constructed languages to be a worthy object of study” [3. P. 10]. Moreover, learning a constructed language rather than analyzing its features may be seen as detrimental and academically discrediting as opposed to being merely counterproductive.

This awareness is shared by the scholars that focus on analyzing semantic and semiotic systems of constructed languages and contribute to various interdisciplinary projects: “one may wonder why someone would be concerned with investigating such an elusive and whimsical area of research as the translation and analysis of constructed languages” [4. P. 91].

However, the aforementioned paradigm shift in the last decade has been associated with a more positive attitude to constructed languages, as proven by the release of the book by Oxford University Press, a major mainstream publishing house. The publication is dedicated to providing the rationale for exploration of constructed languages and their beneficial use as tools of introspection and language learning facilitation. The authors maintain that “conlangs have held importance in the sociopolitical arena and in the world of literature and science fiction media” [5. P. 1].
Correlation between types of constructed languages, purposes of their creation and linguistic features

Documented attempts to construct a new language date back to the 12th century when Hildegard of Bingen described Lingua Ignota (Latin for “unknown language”), a secret ritual language, i.e., a language that is largely unintelligible to the lay people. While no evidence of its grammar has been recovered, the existing document proves that Lingua Ignota possessed two highly important and almost universal features of a constructed language: its semiotic system is based on Latin with some elements of German and Greek, which constitutes a manifestation of linguistic eurocentrism. Another point inferred by modern linguists is that the purpose of Lingua Ignota was to completely reorganize communication, either by “purifying” it through creation of an artificial state of diglossia, or to isolate the group of Hildegard’s entirely female congregation.

The latter implication is a well-established one among researchers who believe that people create constructed languages “because they are somehow dissatisfied with the set of existing languages: those are considered inadequate instruments for thought or for communication or too difficult to learn.” Both of these points will be addressed throughout the research.

Similar goals were pursued by the mystics who created Balaibalan, another early example of a constructed language, in the 14th century. This language was written with the Ottoman variant of the Arabic alphabet and comprised various elements of Persian, Turkish and Arabic languages, yet a large percentage of its vocabulary does not contain any traces of the existing languages, which also serves the purpose of obfuscation.

The two aforementioned languages were described as “secret languages” — by using a semantic system unknown to the general public, they served the purpose of security through obscurity. However, this term is not common in the modern taxonomy of constructed languages. One of the most important classifications of constructed languages includes the definitions of a priori and a posteriori languages. This dichotomy allows the linguists to separate constructed languages into two categories — languages with supposedly entirely new semantic and semiotic systems and those heavily reliant on the preexisting languages. However, it may be argued that this division should be seen as a scale rather than a binary system since all a priori languages are bound to be influenced by their creator’s linguistic worldview. Thus, the aforementioned language Balaibalan, traditionally classified as an a priori language, demonstrates a higher degree of reliance on natural languages than SolReSol, which represents a group of conlangs referred to as philosophical languages, yet in its turn will have a position different from the one of aUI with its unique semantic and semiotic systems.

It should be noted that the use of the conlang taxonomy is inconsistent and is further complicated by a lack of global terminology, as evidenced by the examples of such terms as “planned language”, “experimental language”, “artificial
language”, “fictional language”, “imaginary language”, “engineered language”, etc., some of which might be considered pejorative by the authors of the said languages.

Along with the aforementioned structural a priori / a posteriori dichotomy, there is one rather well-defined pragmatic distinction based on the initial purpose of creating a new language. Reliance on these criteria is widely accepted: “Unlike natural languages, conlangs have traceable sources, known authors, and well-defined purposes” [7].

M. Halley defines these two categories as interlangs and artlangs. While the other subcategories, including the ones mentioned in the previous paragraph, might occupy a specific place in that system, it offers a reasonable distinction — interlangs, also referred to as auxiliary languages, or International Auxiliary Languages (IAL), set the aim of connecting people that do not share a similar language. Examples of such languages include Volapük, Esperanto and Ido, Interlingua, etc. Such languages typically belong to the a posteriori category, their semiotic systems are rarely original and demonstrate a high degree of eurocentrism.

This category will also include zonal languages, one of the earliest examples of which is the Common Slavonic language created in the 17th century by J. Križanić who sought the Slavic unity both in cultural and political spheres. The highest degree of a posteriority is demonstrated by a special subcategory described as controlled natural languages: Simple English, Basic English, Special English, Globish, etc. Their semiotic systems do not usually differ from the ones of the respective natural languages and semantic systems range from the so-called lexical minimums similar to those used in foreign language teaching to such thinking devices as E-Prime that excludes all forms of the verb to be in order to promote eloquence and clarify thoughts.

These languages contrast with artlangs, i.e., artistic languages that demonstrate a much higher inconsistency in the use of semantic and semiotic systems and rarely pursue the goal of becoming a lingua franca. These languages range from those that do not possess any developed semantic systems and are only featured to create an exotic effect by using their unusual semiotic systems (e.g., the Star Wars universe features 68 languages, yet none of them have any formal description or consistency) to those used in foreign language teaching to such thinking devices as E-Prime to the well-developed ones with in-depth descriptions of inventories (such as the Star Trek’s Klingon which was used as one of the official languages of Wikipedia before its abolishment as a measure of preventing the copyright disputes).

Artlangs are not necessarily incorporated into works of fiction since their creation represents an act of art and science per se. For example, Ygyde is a language that pursues mathematical precision as its top priority and relies on the semiotic systems existing outside of natural languages: the color pink is defined as \#FFABAB, which is a unique hexadecimal expression of one of the 16,000,000 of shades while countries are only identified by using their respective geographic coordinates.

Some other examples of artlangs include Futurese, a language with a high degree of a posteriority that aims to predict the development of American English
by exaggerating its current semiotic trends and Drsk — an art language containing no vowels and using a dozanal (i.e., base-12) system as opposed to the decimal one. Expansion of semiotic systems beyond the scope of what is generally offered by modern languages is a common trait — there are also some proposals of using the binary code, base-6 and base-16 systems.

Such projects may be seen as attempts to reorganize the world — for example, Láadan was created in the late 20th century for an “international community of women seeking a way to communicate outside the constraints of languages controlled by men” [8].

A constructed language may pursue a multitude of goals — aUI, a philosophical language created by J. Weigart, who emigrated from Germany in 1939, was described by him as “the Language of Space”, a language that would be understood by the extraterrestrials attempting to establish contact with the earthlings. While this idea appealed to the young people in the Space Age of the sixties and seventies, in order to prevent it from being immediately dismissed as frivolous, the author and his successors describe the lack of semantic ambiguity as well as its simple and symbolic systems as the main features. It is added that there is a more serious purpose to the creation of aUI, which is combatting stereotypical thinking exploited by propagandists through creation of a strong and ubiquitous a priori connection between semantic and semiotic systems.

Some authors propose completely new, often specialized conlang classifications: “[…] a priori and a posteriori are unable to comprehensively analyze the relationship between fictional conlangs with game elements…” They may advocate such solutions as “constructing a new taxonomy on fictional conlang design approach that adheres specifically to video games” [9].

Despite the existence of multiple classifications of constructed languages and a large variety of their systems, it is still agreed by the academia that “natural languages are more complex than planned ones on the morphological level” [10].

**Semiotic specificity of constructed languages in the digital age**

While the World Wide Web was introduced to the general public 40 years ago, it is only recently that such technical factors as the lack of portability, limited storage, bandwidth and computational power as well as other restrictions have been minimized. This facilitation of communication led to the creation of unprecedentedly focused international niche communities, included the ones aimed at implementing various conlang-related projects.

Development of Web 2.0 in the first decade of the 21st century lowered the entry barrier to content creation and empowered users to participate in such projects regardless of their technical skills. While the English language indisputably became the language of the Internet (it accounts for more than 60 % of the World Wide
Web content according to W3Techs 2021 estimates), it did not stop the users from creating content in constructed languages and translating the existing pages, thus forming the necessary cultural foundation. This led to two major implications for constructed languages: first, conlang enthusiasts were able to use message boards and various types of social media to revive the old, long forgotten languages, popularize the more wide-spread ones and create their own; secondly, the new technology permitted them to implement some previously unavailable semiotic systems, facilitating and automating communication.

Furthermore, new forms of media are not limited to the communication-focused ones and can include various types of software such as video games. While artlangs have been an inherent part of video games (such as Gargish in the Ultima series that dates back to 1988 and Dovahzul in Skyrim), the online mode allowed developers and players to incorporate semiotic systems of auxiliary languages (such as Vōtgil with its three-letter writing system optimized for the voxel-based Minecraft game) for peer-to-peer communication as well.

In addition to the conlang-focused projects, there has been some interest in incorporating constructed languages into neural networks and using them to explore the potential of artificial intelligence and social dynamics, creating self-organized semantic and semiotic systems. The authors of one such project came to the conclusion that “development of conlangs can happen in artificial societies of simple agents” [11].

Developed in the early 19th century, Solresol is one of the first attempts at creating an a priori international auxiliary language that predates the more popular Volapuk and Esperanto. Its uniqueness is manifested in the potentially infinite number of semiotic systems: described as a language of music with its seven-tone inventory, it also incorporates such signs as solfège, the seven spectral colors, numbers, gestures, etc.

---

**Fig. 1.** Some semiotic systems used in Solresol

Рис. 1. Некоторые знаковые системы, используемые в языке сольресоль
Unreal Engine 4, a user-friendly real-time 3D creation system used in filmmaking, architectural visualization and video games, permitted the author of this research to create a self-maintained open-source computer program named SolReSol: The Project, which became the first implementation of all the semiotic systems initially designed by Francois Sudre: it augments the synesthetic effect of using the colors of the rainbow together with the high-fidelity sounds of musical instruments, allowing its users to decompose the lexis into the minimal elements of meaning and observe semantic transformations through color blending.

![Image](image.png)

**Fig. 2.** Implementation of the spectral input mode in “SolReSol: The Project”

Furthermore, it also supports direct input of sounds through the MIDI interface, providing a real-time translation of musical notes into Solresol and English. Several versions of the project have been released and the roadmap includes the plans to implement such input modes as the absolute pitch recognition through the microphone, enabling the use of non-MIDI instruments as well as optical color recognition, permitting the system to read printed or drawn color codes captured by the camera.

This program contributed to the rise in popularity of Solresol as a language with 27,000 views of its demonstration on various social media platforms, more than 1,000 installations and its inclusion into such sources as Atlas Obscura and Wikipedia. It also sparked the creation of new international SolReSol-based scientific and artistic projects such as the one by J. Lloyd from Newcastle University who used its framework as a basis for constructing a device that attempts to decipher bird vocalization.

Despite the opportunities offered by the new technology, some of the online practices have been deemed questionable by the more scrupulous members of the conlang community: Google Translate offers Esperanto as one of its non-experimental languages, Wikipedia is available in eight different constructed languages (Esperanto, Volapuk, Ido, Interlingua, Kotava, Occidental, Lingua Franca.
Nova, Novial, Lojban) with Volapuk accounting for the largest number of articles (over 117,000, which places it 17th in the global rating, above the natural languages with millions of speakers), yet the overwhelming majority of them are the examples of low-quality machine translation.

Another issue of constructed languages related to their modern state of accessibility is the lack of centralization, which leads to their forking. Creating new constructed languages based on the existing ones is not a new practice: Ido is a well-established reformed version of Esperanto that sought to be grammatically, orthographically and lexicographically regular, changing the hard-to-pronounce words (such as scii to savar) and eliminating the denotation of even the most basic female-related concepts through suffixation of their male counterparts. Lojban was derived from Loglan and SolReSol exists in at least two major versions — the original one, created by Francois Sudre and the one created many decades later by Boleslas Gajewski, who changed such basic terms as fasol from why to here.

However, purists argue that revisionism plays a detrimental role, further dividing the community that could focus on communication and content creation instead. The example of Solresol demonstrates dozens of proposals for its reform, calling for various types of changes, from the major revisions (such as introducing new semiotic systems with sharp / flat notes in order to facilitate transliteration and simplifying its grammar to the point of transforming the language into an isolating one) to the non-intrusive ones such as the expansion of vocabulary that would help to accommodate modern terms. Some of the proposed reforms seek to minimize or eliminate the eurocentrism which proved to be a widespread feature of both a priori and a posteriori constructed languages.

Eurocentrism as a semantic and semiotic feature of constructed languages

Since the inception of international auxiliary languages, their authors have been using different solutions in order to minimize advantages given to speakers of any particular language: while the vocabulary of Volapuk is based on Romance and Germanic languages, its creator purposely obfuscated the original words, often making them unrecognizable. Nevertheless, its grammar includes a variety of tenses and moods, increasing the number of paradigm elements to 234 forms, making using suffixation to make distinctions between requests, commands and demands. Nowadays this system is seen as a proof of the Standard Average European concept introduced by B. Whorf [12] and deemed unnecessarily complex for a language described as an international auxiliary one.

Similarly, Esperanto has been criticized for its 26-letter alphabet based on Polish, L. Zamenhof’s native language, and uses six diacritics, yet excludes the letters u and h, which nowadays complicates its use in different titles such
as filenames and website URLs. Its toponyms are also highly Eurocentric: the exonyms Japanio and Činio are used for Japan and China, respectively.

An in-depth examination of various semantic systems reveals the fact that eurocentrism is not limited to a posteriori languages: although Lojban is seen as a language that strives for neutrality and regularity, its lexis contains a large number of European words, e.g., mandarina — orange, blanu — blue, cyan = cicna, narju = naranja, pink = penka, etc.

While Solresol presents a unique a priori semiotic system that seemingly excludes any form of reliance on natural languages, it still bears a lot of traces of the nineteenth century French language and culture. This is demonstrated both in its grammar and vocabulary, as evidenced by the absence of words for 70 and 90, which forces the speakers to use 60+10 (soixante-dix) and 4*20+10 (quatre-vingt-dix) respectively. This Eurocentric trait dates back to the early vigesimal (base-20) systems used in French, Danish, Albanian, Welsh and other languages.

Another example of cultural relativism in Solresol is its abundance of terms describing some political structures and titles: it contains specific words for “Minister of the Marine and Colonies”, “Grand Officer” and a variety of manners of address, yet only one word for all types of celestial bodies, its inflexion system mimics the one observed in French grammar.

Some conlang creators view the eurocentrism avoidance as the main feature of their languages: Lidepla (Lingwa de Planeta) incorporates vocabulary based on the ten most spoken (at the time of its creation) languages — Arabic, Chinese, English, French, German, Hindi, Persian, Portuguese, Russian, and Spanish. Nevertheless, it can be argued that there still remains a certain preference towards Indo-European languages since the
lexis of the two non-Indo-European languages has been transformed based on its romanization.

Another attempt at avoiding eurocentrism is demonstrated in Toki Pona, an oligosynthetic polysemic language. It is written in the Latin alphabet of 14 letters, but can be transliterated into many other scripts, such as Cyrillic, Cherokee, Hangul, Hiragana, etc. [13]. Furthermore, some artlangs prove to be semantically Eurocentric despite their exotic semiotic systems: Dovahzul with its runic script and digraphs reveals a completely Anglocentric system, even borrowing such idioms as *keep (something) at bay* from the English language.

Some conlangers embrace eurocentrism instead of denying it, which leads to the creation of zonal languages, including the abovementioned attempt by J. Križanić to create Pan-Slavonic, named “Руски језик” by him. Pan-Slavonic languages are still being created and developed many centuries later, as evidenced by Interslavic language, Neoslavonic, Nowoslownica, etc. Zonal conlangs have also been designed for communication amongst speakers of Germanic languages (Folkspraak), Niger-Congo and Bantu languages (Afrihili).

Based on the extreme interpretation of linguistic relativism, it can be concluded that all constructed languages will inevitably include a certain degree of zonality in their semantic and semiotic systems and favor speakers of certain languages since the bias caused by the creator’s linguistic worldview cannot be completely avoided.

**Conclusion**

Despite the traditional skepticism expressed by the academia towards any type of research related to constructed languages, there has been a substantial rise of interest in conlang projects caused by a modern paradigm shift. Constructing new languages is not seen exclusively as an attempt to eliminate the dominance of natural languages and establish a new international auxiliary language, it might be interpreted as an act of art, a way of exploring the reality and creating new thinking tools, provoking introspection and increasing the degree of linguistic self-awareness.

Furthermore, the research findings point to the modern interdisciplinary relevance of constructed languages that along with such areas of knowledge as linguistics, poetics and culturology, contribute to artificial intelligence networks, social dynamics simulation and other types of big data projects.

The research results also reveal the lack of a uniform constructed language taxonomy and challenge the integrity of the seemingly well-established dichotomies of a priori and a posteriori, auxiliary and artistic languages. Nevertheless, there is a possibility of describing the general semantic and
semitic features of a language through the analysis of its place in the paradigm of constructed languages.

Additionally, the further exploration of constructed languages leads to the conclusion that the traces of Standard Average European features can be found in their semantic and semiotic systems, thus proving the hypothesis of linguistic relativity.

References / Библиографический список

Information about the author:
Philipp N. Novikov is PhD. in Philology, Associate Professor, Foreign Language Department, Institute of Law, RUDN University; Research interests: cognitive linguistics, corpus linguistics, constructed languages, comparative linguistics, linguistic software development; e-mail: philippnovikov@gmail.com
ORCID ID: 0000-0003-4884-3659; eLIBRARY SPIN-code 3523-2539. Web of Science ResearcherID AAH-8064-2019

Сведения об авторе:
Новиков Филипп Николаевич, кандидат филологических наук, доцент кафедры иностранных языков юридического института РУДН; сфере научных интересов: когнитивная лингвистика, корпусная лингвистика, искусственные языки, сопоставительное языкознание, разработка языковых приложений; e-mail: philippnovikov@gmail.com
ORCID ID: 0000-0003-4884-3659; eLIBRARY SPIN-код: 3523-2539. Web of Science ResearcherID AAH-8064-2019