State, Predicatives and Idiomaticity[*]

Maria A. Todorova[1][0000−0001−5866−180]

Institute For Bulgarian Language, Bularian Academy of Sciences, Sofia, Bulgaria
maria@dcl.bas.bg

Abstract. This research represents a typological and constructional approach to the description of idiomatic predicative constructions (IPCs) in Bulgarian. The study employs a subset of 1331 idiomatic predicative constructions extracted from a large dictionary of Bulgarian multi word expressions and a corpus of Bulgarian predicative constructions extracted from the Bulgarian National Corpus. The analysis of their structure, functions and semantics is a stage of the study of the syntactic and semantic structure of the predicates in Bulgarian with a view to ontological presentation of the meanings of state.

Keywords: State · Predicatives · Constructions · Verbal Idioms

1 Introduction

The study presented in the article examines a specific group of predicates which are a subset of verbal multi word expressions (VMWEs) with specific component structure combining a predicative NP or PP and an auxiliary verb. The data observed is a subset of entries extracted from a large dictionary of compound lexical items [19] using as a criterion the typology of predicative structures in Bulgarian, described in [18]. The predicative constructions are well-known and thoroughly described phenomenon in the linguistic research of Russian [24], [6], Bulgarian ([4],[11], [9], etc.), English [20] and other languages [15], [1] mainly with a view to their syntactic functions. But with focus on idioms sharing predicative structure within the semantic domain of state and the lexical and grammatical ways of its expression its less studied. The study gives another point of view of the question of the place of predicative constructions in the continuum between the semantic integrity and syntactic structures. In the remainder of this paper, §2 presents objectives and methodology of the research, §3 - the predicative constructions in Bulgarian and the verbal MWEs as their subset, §4 describes the semantic domain of state and §5 presents the observed data, §6 is constructional and §7 is semantic analysis of the Bulgarian idiomatic predicative constructions (IPC).

2   Objectives and Methodology of the Research

The main objectives of this research is the description of idiomatic constructions with predicatives as a subtype of predicative constructions. Their typology is oriented to their place in the linguistic modeling of the ontology of state situations. We are investigating the hypothesis whether predicates with composite component structure are a semantic whole to which universal semantic features apply. Collecting data with idioms sharing common specifics is not an easy task as usually they have low frequency in corpora and are ambiguous with non-idiomatic expressions. That is why we extract idiomatic constructions from a ready made dictionary of Bulgarian MWEs (cf. § 5). In order to detect MWEs with predicative constructions we apply a list of construction types of predicative constructions in Bulgarian to the morphosyntactic description framework of dictionary entries. In this way we both extract candidates for idiomatic predicatives and apply to them the possible syntagmatic and paradigmatic forms. The possible variations in terms of component's order, optional components, and possible insertions both as modifiers of components or external phrases between the components of each of the extracted idiomatic predicative constructions are needed for the recognition of idiomatic predicative constructions in a corpus with a view to study of their argument structure.

3   Predicative Constructions in Bulgarian

Predicative constructions are a type of predicate structures, known also as copular predicate constructions, are language units whose features place them on the border between morphology and syntax. They are united by specific functions and semantics and largely enter the field of lexicology. Their characteristics in Bulgarian derive from their compositional component structure - a specific sequence of: auxiliary verb, adverb/adjective/noun/noun phrase/ prepositional phrase, obligatory or optional dative or accusative pronoun clitic. Morphosyntactically they are characterised with the grammatical categories of person, number and tense and lack synthetic conjugate forms [10]. The general semantic field to which belong the predicative structures is characterized by "the presence of the object in some unchanged state, which is not the result of direct influence of someone for a certain period of time" ([24]: 549). Additionally the predicative constructions are denoting some kind of evaluation of a state ([14], [4], [13]).The detailed typology of Bulgarian predicative constructions we are using in this study is presented in [18]

3.1   Verbal MWEs with Predicative Constructions

The interest of modern research on the lexicalization of expressions in the last 10 years has focused on the description of MWES, their components and structure ([23], [5], [19] with a view to their automatic detection in corpora. The challenges of their description has been also discussed when it comes to morphologically rich

languages as Bulgarian where VMWEs are characterized by a rich set of synthetic and analytical verb forms; with a complex and flexible word order. They combine structural features such as mandatory and optional components, the ability to insert external phrases (clitics, adverbs, nouns, etc.) and defective components ([7], [16], [17]). In this study we focus on the combination of structure and the meaning of idiomatic predicative constructions investigating their place in the structure of the lexical items expressing state semantics.

4   The Semantic Domain of State

According to the theory of Van Valin and Lapola ([22]: 82–138), state predicates represent situations defined as static, non-dynamic, and time-limited. State predicates express location, condition, position, or inner experience. Van Valin ([21]: 39) groups predicates of state into two major ontological classes - locative (to be on/in '(x, y) and non-locative with the subclasses: a) state or position (broken' (x)); b) perception (see '(x, y)) and cognition (I believe '(x, y)); c) possesion (I have possession' (x, y)); d) equality (I am '(x, y)). In the description of predicativeness as a linguistic phenomenon the authors [22] consider predicative structures as a semantic whole along with verbs.

5   The Selection of Data

5.1   Excerption of Bulgarian Idioms with Predictive Constructions

We used two POS tagged linguistic resources for Bulgarian to extract potential idiomatic predicative constructions in Bulgarian - the dictionary of Bulgarian MWEs [19] and a collection of predicative constructions' examples derived from the Bulgarian National Corps [8] comprising over 10,000 sentences [18]. The data of predicative structure groups was derived from the resources using the list of predicative constructions and their formal description in Bulgarian[18].

The dictionary of Bulgarian MWEs [19] contains over 86,373 noun and verb compound lexical units, derived from various sources: printed and electronic dictionaries of Bulgarian idioms, the Explanatory Dictionary of the Bulgarian Language [2] and the Bulgarian WordNet [8] as well as automatically extracted from the Bulgarian National Corps [8]. Each component of MWE's entries in the dictionary of the Bulgarian MWE is applied a POS tag. To extract idiomatic predicative constructions from it at first the verbal entries from the dictionary were extracted and as a result a collection of 27,902 verb idioms was obtained. Then we applied to this collection the formal structural descriptions of predicative constructions [18] and extracted all VMWE's with auxiliary verb, noun/adjective/ adverb/ and, or preposition. Then the potential predicative VMWE records from the dictionary along with the examples from BNC were manually checked. As a result a collection of 1331 idiomatic predicative constructions was obtained and the new structure groups were extracted and analysed.

6   Constructional approach to classification of IPCs

After the excerption we collected over 2500 candidates for IPCs and after manual validation 1331 IPCs were selected. The observation over the structure types led to definition of four major types of idiomatic constructions with predicatives according to their component structure: a) idioms with plain predicative constructions; b) idioms with predicative constructions with pronoun clitics; c) idioms with prepositional predicative constructions; d) idioms with conjunctional predicative constructions. The subtypes share similar structure variations. The word order of the components is semi-fixed. The position of the constituents changes depending on the sentence context. The auxiliary verb is a clitic and cannot stand at the beginning of the sentence in present tense. The distribution of the subtypes within the excerpted data and the grouping of the formal constructional types is represented in Table 1.1.

Table 1. Distribution of the Types of Verbal Idiomatic Constructions with Predicatives in the Data

Costruction types	Occurrences of the types in the collected data	Examples
Aux V + Participle/N/Adv /Adj	337	izlyazal sam ot stroya (sth can't be used any more)
Aux V + Participle/N/Adv /Adj + Acc/Dat Pron	441	pisano mi e (its my faith)
Prep + Aux V + NP	496	v plen sum (feel captured from sth)
Conj + Aux V + NP/ Participle	57	kato che sam padnal ot nebeto (s.o. is very inadequate)

6.1   Idioms with Plain Predicative Constructions

This type of constructions consists of an auxiliary verb and participle (ex. 1) or NP (ex. 3) or AdvP (ex. 2). The entries in this group express common state semantics - characteristics or attributes of someone or something.

Examples: (1) varzan sum v racete (s.o. is not very skillful); (2) van ot sebe si sam (s.o. is very furious);((3) gola voda sam (s.o. or sth. is not valuable).

6.2   Idiomatic Predicative Constructions with Pronouns

This type of constructions include 4 subtypes:

a) IPCs with an auxaliry verb, combined with nouns with constrained grammatical (fixed in singular (ex. 4)) properties and/or a dative or an accusative pronoun clitic (ex. 5, 6)

Examples: (4) grehota e (its a sin), (5) yad me e (to feel rage),(6) zhal mi e(to feel sorry for s.o.)

b) IPCs with an auxiliary verb, combined with peredicative -o words (ex. 7) and/or a dative or an accusative pronoun (ex. 8)

Examples: (7) kasno e (it's late); ((8) dobre mi e (I'm fine).

c) IPCs with an auxiliary verb, predicative noun and a dative or an accusative pronoun, most often expressing emotion (ex. 9) or physical state (ex. 10).

Examples: (9) zhal mi e (I feel sorry); (10) ne me e enya (I don't care).

d) IPCs with an auxiliary verb in 3rd person sg. and a predicative adverb, most often meaning manner (ex. 12) or state (ex. 11). The same meaning share constructions with a predicative word and an omitted auxiliary verb (ex. 13)

Examples: (11) tamno e (it's dark), (12) fasulsko e (it's easy); (13) Veche se samnalo. (It's already dawn.)

6.3   Idioms with Prepositional Predicative Constructions

This type of constructions consists of an auxiliary, preposition and an NP and/or pronoun clitic. This is the less studied group of IPCs as they are predicative constructions, characterised with difficult to reduce ambiguity, caused by the position of the predicate and the predicative function of the preposition both ([11], [3]). In these constructions the forms of the auxiliary verb function same as full paradigm verbs and represent the semantic of presence or existence [12] in combination with the semantics of the individual prepositions. The specifics of the construction type arise from the preposition. Prepositional phrase can be in the first position in the sentence followed by a pronoun clitic. The entire group can be subdivided into different semantic sub-types due to the prepositional semantic:

IPCs denoting location and direction. The most frequent constructions are IPCs with the prepositions v (in) (ex. 14), mezdy (among), and po (along) (ex.15).

Examples: (14) V kravta vi/ti e. (It is in your blood.); (15) Po pat ni e. (It is on our way.).

IPCs denoting evaluation (ex. 16, 17), characteristics or attributes of someone or something (ex. 18, 17).The most frequent constructions are IPCs with the prepositions ot (from) (ex. 15, 19), s (with) and za (for) (ex. 18).

Examples: (16) Izobshto ne im beshe do horata. (They didn't care about people at all.); (17) Tova veroyatno im e ot golyama polza. (This is probably of great benefit to them); (18) Beshe mu za prav pat. (It was his first time.); (19) Samnyavam se, che shte vi bade ot koy znae kakva polza.(I doubt it will be of much use to you.)

6.4   Idioms with Conjuctional Predicative Constructions

This type of constructions consist of a conjunction, an auxiliary verb and participle or noun phrase or adverbial phrase. The entries in this group express common state semantics - characteristics or attributes of someone or something. Within this set two subgroups with idiomatized position of participle can be observed:

a) Positive Idioms with Conjuctional Predicative Constructions - Kato che sam (As if I am) and participle/NP

b) Negative Idioms with Conjuctional Predicative Constructions - Da ne sam (As if I don't) and participle

6.5   Idioms with Predicative Constructions and an Idiomatised Subject

This type of constructions consist of an auxiliary verb, a participle and a noun phrase, idiomatised in the subject position.

Example: (20) Zhivotat mi e mil (not to dare to brake rules)

7   Idioms with Predicative Constructions and State Semantics

As discussed in §7 we group the semantic types of IPCs based on the classification of Van Valin of predicates of state ([21]) into locative and non-locative ontological classes.

7.1   Locative IPCs

Locative IPCs denote assessment of state or position, represented mainly from IPCs with prepositions.

Examples: (21) na pat sum (i'm near the goal); (22)do krivata krusa sam (to be far from the aimed location).

7.2   Non-locative IPCs

Non-locative IPCs denote several subgoups:

a) assessment of physical condition - grogi sam (feel very tired); tip top sam (feel well);

b) assessment of mental condition - kato che sam padnal ot nebeto (to be very inadequate);

c) assesment of emotional condition - izvan relsi sam (feel very confused);

d) assessment of characteristics or attributes of s.o. or sth. - gola voda sam (s.o. or sth. is not valuable).

The semantic features are distributed over the construction types of IPCs. While assessment of characteristics or attributes is typical semantics for plain IPCs (ex. 1–3), IPC's with pronouns represent: assessment of characteristics or attributes (ex. 7, 8, 11, 12), assessment of mental condition (ex. 5, 6); assesment of emotional condition (ex. 9, 10). The IPCs with prepositions represent both locative (ex. 15) and non-locative groups - assessment of characteristics or attributes (ex. 17, 18), assesment of emotional condition (16) etc. This semantic typology of idioms with predicative constructions is compatible with the traditional grouping of predicates in semantic classes and additionally they represent the semantic feature assessment. This could be an argument in support of the hypothesis that regardless of composite component structure predicates are a semantic whole to which universal semantic features apply.

8   Conclusion and Future Plans

The article offers a classification of idiomatic predicative constructions in Bulgarian with a view to study of the ontological semantics of state. A method for excerption of idiomatic predicative data based on the formal of component's structure of VMWEs and Predicative constructions both was used. The constructional analyses of the collected data led to definition of four main structural groups and their respective subtypes. Each group represents variations of state semantics. The combined constructional and semantic analyses gives thoughts on questions for the vague boundaries between predicative constructions, composed verb forms and idiomatic predicative constructions and shows arguments for the statement that regardless of composite component structure predicative constructions are a semantic whole belonging to the universal semantic features of lexemes. The collected set of examples of IPCs will be used to study their argument structure and the lexicalization patterns associated with their syntactic representations.

References

1. Anderson, G., D.S.: Auxiliary Verb Constructions (2006)
2. Andreychin, L., Georgiev, L., Ilchev, S., Kostov, N., Lekov, I., Stoykov, S., Todorov, T.: Balgarski talkoven rechnik. Dopalneno i preraboteno izdanie ot D. Popov (Bulgarian Explanatory dictionary). Nauka i izkustvo. Sofia (2005)
3. Atanasov, A.: Classification of impersonal predicates in modern bulgarian. Bulgarian Language **62**(3), 38–53 (2015)
4. Georgiev, I.: Impersonal sentences in Russian and Bulgarian. Narodna prosveta. Sofia (1990)
5. Gregoire, N.: Duelme: a dutch electronic lexicon of multiword expressions. Language Resources and Evaluation pp. 23 — 39 (2010)
6. Ivanova, E.: Dativno-predikativnyye struktury v bolgarskom yazyke: statistika eksperimenta. Russkiy yazyk za rubezhom **5**, 11–17 (2018)
7. Koeva, S.: Inflection morphology of bulgarian multiword expressions. Proceedings of Azbuki@net, International Conference and Workshop. Sofia pp. 201–216 (2006)
8. Koeva, S.: Wordnet i bulnet. Ezikovi resursi i tehnologii za balgarski ezik pp. 154 — 173 (2014)
9. Kutsarov, I.: Teoretichna gramatika na balgarskia ezik: Morfologia. Publisher of Plovdiv University Paisiy Hilendarski (1999)
10. Maslov, Y.S.: Grammar of the Bulgarian language. Moscow (1981)
11. Nitsolova, R.: Bulgarian Grammar. Morphology. University Publishing House St. Kliment Ohridski, Sofia (2008)
12. Penchev, J.: Syntax of the modern Bulgarian literary language. Plovdiv (1998)
13. Petrova, G.: Nominal predicates joining the dative experiential in the bulgarian language: semantics and syntax. Russian Abroad **5**, 24–29 (2018)
14. Rozhnovskaya, M.: Impersonal sentences in the modern bulgarian literary language. Questions of grammar of the Bulgarian literary language pp. 379–432 (1959)
15. S., M.: Complex Predicates Verbal Complexes, Resultative Constructions, and Particle Verbs in German (2002)
16. Todorova, M.: Typology and Properties of Multiword Expressions in Bulgarian. Verb idioms. PHD Thesis. Institute for Bularian Language, Sofia (2015)
17. Todorova, M.: Formalized morphosyntactic dictionary of verb idioms in bulgarian. Papers of the Institute for Bulgarian Language Prof. Lubomir Andreychin **XXIX**, 207–243 (2016)
18. Todorova, M., Dimitrova, T., Stefanova, V.: Study of the boundaries of predicates for the state in bulgarian. (izsledvane na granitsite na predikativite za sastavyanie v balgarski). Sapostavitelno jezikoznanie **XLVI**(4), 21–46 (2021)
19. Todorova, M., Stoyanova, I.: Razrabotvane na rechnitsi na sastavnite leksikalni edinitsi v balgarskiya ezik za tselite na kompyutarnata lingvistika.(compiling dictionaries of bulgarian multiword expressions for the purposes of computational linguistics). Ezikovi resursi i tehnologii za balgarski ezik (2014)
20. Van Eynde, F.: Predicative Constructions: From the Fregean to a Montagovian Treatment. CSLI Publications (2015)
21. Van Valin, R., D.: A synopsis of role and reference grammar. Advances in Role and Reference Grammar (1993)
22. Van Valin, R., LaPolla, R.: Syntax: structure, meaning and function. Cambridge: Cambridge University Press (1997)
23. Villavicencio, A., Copestake, A., Waldron, B., Lambeau, F.: The lexical encoding of mwes. Proceedings of the ACL 2004 workshop on multiword expressions: Integrating processing. Barcelona, Spain pp. 80 — 87 (2004)

24. Zimmerling, A.B.: Imennye predikativy i dativnye predlozheniya v evropejskih yazykah. Computational Linguistics and Intelligent Technologies: Based on the materials of the annual International Conference "Dialogue" **9**(16), 549 – 558 (2010)