

Semantic Tag Categories in Corpus Linguistics: Experience and Examination

Akhmedova Dildora Bahodirovna, Bakhtiyor Mengliev

Abstract: *Two types of semantic annotation: facet and pointer are discussed in the article; where it is noted that if the facet method requires consecutive interpretation of a unit, the genealogy method requires the group, group, and field to be defined. The first method is from word to word, second method to meaning. This shows the application of the word -> meaning, meaning -> word principle in both the expert's process of searching and the interface that is reflected in the interface. Priority of such features as lexical-grammatical, lexical-thematic, methodological assessment characteristic, derivative characterization as a result of observation of existing online corpus; The constants and operator-classifier tags are also defined as permanent labels are the main points of the research.*

Keywords : *semantic coatings, semantic tags, charging system, operator, classifier, multicolor, semantic filter. Annotation: This article discusses the semantic filtering, semiconductor system, operator, classifier, multivariate, semantic filter feature and capabilities that are important categories of semantic annotation of body units.*

I. INTRODUCTION

An analysis of the semantic circuitry of the Russian, Kalmyk, Bashkir, Chuvash language is given. Just as any type of rhyme requires linguistic and software, semantic rhyme is based on lexical / lexicographic / semantic databases. The article explores and analyzes the possibilities of the lexicograph system, which is the basis for the semantic assimilation of the Russian language. It also provides information about semantic classifications, semantic dictionaries that serve as linguistic support for corpus.

II. REVIEW

Suggestions for the use of this system for semantic mapping of Uzbek language are provided. Semantic tag system: Different approaches and approaches. Corpus linguistics has developed a number of studies on the problem of semantic rhetoric, its system of labels, tag categories, problems of semantic tags, polygamy, and survival. In particular, V.K. Kukanova [5], A.A. Kretov [4], G.I.Kustova, O.L.Lyashkevskaya, E.V.Paducheva, E.V.Rakhilina, B.P.Kobritsov, T.I.Reznikova [6] has published a number of works on these issues. The

abovementioned scholars have commented on the semantic arrangement of the Russian, Kalmyk and Bashkir corpus; the search for the corpus semantic parameter on the basis of the semantic circuit. Below, we discuss scientists' approach to their semantic tag system.

V.K.Kukanova in his article "Principle of Semantic Circumstances of the Nation Corps" is based on the graphic and morphological analyzer, corpus manager, and the semantic assimilation of the body of the Russian language [15]. The Bashkir corpus [1] experience in the development of the corpus semantic system is based on the classification of taxonomic groups in the Russian semantic dictionary [3], edited by S.R. Barkhudarov.

In addressing the problem of the semantic circuitry of the Kalmykian national corpus, the semantics of the Russian language corpus served as a reference. The semantic arrangement of the Russian language is done through the linguistic database of the lexicographer system [7]. Vukhanova, V.Kukanova responds to the question "Why do we need semantic rhetoric?" The first approach shows a clear and complete description of all the existing meanings of a particular language unit and its context.

The second approach, on the contrary, aims to bring together all forms that can represent the same meaning: a set of meanings. Both approaches are equally important for linguistics and language education. In addition, as the grammar and vocabulary user / learner learn both native / foreign language, they need two different approaches. If an unknown word or construct is used, the dictionary / grammar is used; If a person has to write something or speak for himself, he / she will search for words / phrases: he knows what he means, feels, but can't find the right word / grammar form "[5].

The word can have more than one meaning, so the punctuation method does not always give the correct result, and the facet method is convenient when describing multiple meanings. For example, the word satan has the following meanings depending on the context [14]:

1. Shayton [the devil; demon, wrinkle] 1.religion. The head of the evil spirit or the evil spirits, who mislead people and lead them into sin, crime, and abomination; the devil. Turn away from this path, O great king, and do not yield to the temptations of Satan, and do not repent. The Golden Cradle. He was also misled by Satan. K. Lightning, Hamza. It is not a sin to pray, but a person who does not pray is a sin because he does not obey the devil.

Revised Manuscript Received on October 15, 2019.

* Correspondence Author

Akhmedova Dildora Bahodirovna*, Teacher of Bukhara State University, Uzbekistan

Bakhtiyor Mengliev, Professor of the chair of Uzbek linguistics at Tashkent State Technical University, Doctor of Philology.

2. portable cunning, cheerful; deceptive. There is a shopkeeper. Death to death .. R. Faizi, Spring is coming to the desert. 3. portable Shok, playful. The girl was a devil, and she was full of repentance, and she was a virgin. Tashmatov, Erk bird. "Yes, the devil's daughter," said Urunboy, "don't you even stay here?" Aydin, the bride bore a son. One devil like our Avaz shouts, "Don't let the twenty-six get shot!" S. Siyoev, Light. The despair is the devil. It is a statement that encourages a person to live with hope, good intentions and not despair. Let us see, despair is the devil!

I. Rahim, True Love. "Let it be, the devil," said Aisha, comforting her husband. H. Gulom, Mashal. Satan's cart s.t. old mirror bicycle. The boys, who had grown up in torn clothes, sat in a carriage and drew the eyes of the surrounding neighborhood children. Oybek, Selected works. The cry of Satan The cry of lies. Things will not go away until you have a little "devil cry". A. Kadyri, at Ulak. She understood it as a "devil's cry of transmitted girls." Oybek, Selected works. Depressing the Devil Avoiding any act of self-denial. [Hayitjon.] Definitely do it well, give the Devil a try, and these days will pass. Hamza, Paranj Secrets.

I made a fuss, made three reputations, and gave the devil a second. «Fist». The Devil's (or Satan's) greatness ("Satanism") is so rapid and powerful. And who knows, these girls! They say that the devil is great. I. Rahim, True Love. Launch the Devil "Devil" to demonize. He did not see Munavvar with his brother, who was hiding in the right hand. The devil's demon shouted, "Take it from the skirt!" Said the comrade. S. Anorboev, Mehr.

Clearly, the explanatory dictionary of the Uzbek language has three meanings of the word devil, and there are also three (freely and freely) united units. For example, differentiation of the meanings of multi-meaning words was done automatically by a filter, but how does the computer "understand" compound units and "answer" the query? There are several solutions to semantic tagging of such units. It is necessary to develop a separate linguistic provision for the horses of the same type and for their stable horses. The algorithm for this process will be as follows:

1) colloquial dictionaries, as defined above, combine their meanings and create a linguistic supply; 2) The combination is lemmated, which is equal to one lemma (word). For example: <nyoumid - devil>, <satan's chariot>, <devil's cry>, <devil>, <brute devil>, <lure the devil> The computer now understands it as a single word, and when it does, it finds its answer in its semantic explanation from the linguistic supply database.

The semantic annotation of the Satanic lemma is as follows:

<Satan> A wicked spirit, or leader of evil spirits, who misleads people and leads them into sin, crime, and ignorance (1); cunning, cheerful, deceitful (2); playful, playful (3).

<New Hope - Satan> A phrase that encourages people to live with hope, good intentions, and not despair.

<satanic cart> esq. mirror bicycle

<devil crying> Lying, a cry for the way.

<give up on the devil> To refrain from any action, to

quit. <satanic> The freewheeling is so rapid and powerful. <lure the devil> To provoke Satan, to demonize. If the lexical unit has some meaning depending on the context, it will be included in the desired group according to its context; those labels are also included in the same set and are separated from the other labels by a comma-separated comma. The first comment is basic, and the semicolon is the context in the footnote. Discontinuation is carried out in both word and alphabetical order. The first approach is used to describe the lexical-grammatical nature of units - to indicate which category they belong.

Tag groups in the semantics system. The Kalmyk language experience has shown that semantic rituals are executed on the basis of morphology: four more columns are added to the morphological character table, which represent the semantic nature of the word form. The following is an example of a Kalmyk corpus.

I. Lexical-Grammatical Discretion: The division of a word into a category of words belonging to a category. The lexical-semantic classification available in language grammar can be used to implement this classification. Corps authors usually form a set of tags with English abbreviations or symbols [5].

Additional semantic tags for the noun morphological symbol:

1. r: p r o p e r - a famous horse;
- 1.1. name - name;
- 1.2. s u r n - surname;
- 1.3. p a t r - otchestvo;
- 1.4. geox - toponym;
- 1.5. o r g n - organization;
2. r: a p p e l - compound horse;
- 2.1. c o n c r is the name of something;
- 2.2. a b s t r is an abstract horse.

Additional semantic tags for the morphological character of the adjective category:

1. t: q u l l - original quality;
2. t: r e l - relative quality.

Additional semantic tags for the final word morphological character:

1. o r d - number;
2. c a r d - quantity;
3. p a r is a distribution number;
4. a g e - number representing age;
5. c o l is the cumulative number.

Additional semantic tags for the final word morphological character:

1. p e r s - the leader of personality;
2. dem - display header;
3. q u a - designation clause;
4. r e f l - self-esteem;
5. i n t e r - interrogation;
6. i n d is a doubtful character.

II. Lexical-thematic object: lexical-semantic group, group defined; The units are located in taxonomic, mereological, topological relationships. According to V. Kukanova, the

semantics of the word is the language constitution. For example, one semantic group can include units of different classes. III. Characteristic of methodological evaluation: the definition of a connotation. It is defined that the word denotes a positive or negative (if any) *otten*. (*home*: *posit*; *e v*: *neg*). IV. Derivative characterization: If there is a case of decoration, the core of the word, morpho-semantic word-formation group is defined.

The lexical-semantic nature of the word is closely related to the word-building performance. For example, when a word ends with a suffix, the noun refers to the noun / adjective category, and the word ending with a suffix is a group of words that express the identity.

In Kalmyk language, animals vary according to age and sex, and so do their names. Therefore, in addition to animal name tags, *t*: animal: age, *t*: animal: gender; and when two characters are added, the *t*: animal: age & gender tags are attached.

III. DISCUSSION

Taxonomies that do not belong to all word categories the whole name of the partitioning is not included in the system of tags: only the semantic field names that apply to all units appear in the user interface. The main problem with semantic tagging is the determination of the status of a particular unit to two or more semantic groups / fields, depending on the context. The systematization of lexical units within the categories of words is based on the approach to the principle of meaning from form to meaning:

Not every word within the category is interpreted separately (such an interpretation is specific to morphological annotation), but rather a semantic field or group is defined, and words belonging to that group are collected. This approach stems from the view of the language unit as a unit of thought and communication. A.A. Kretov, based on the opinion of A.E. Kibrik, concludes: "Lexical and grammatical semantics are neither uniform nor separate. Lexical semantics - constants, grammatical semantics - variable. Lexical semantics is primary, grammatical semantics is a derivative phenomenon.

The system of semantic labels of the Russian language corpus represents a more popular view: semantic explanations are neither lexical nor grammatical. Its main taxonomy is the word series; it indicates the grammatical nature of the unit. It is complemented by semantic explanations such as subjective and non-credible nouns related to mereologic (whole, element-plural), topology for subject nouns, characterization for verbs, predetermined and non-verifiable nouns, quality and attitudinal estimates." [4] According to the authors of the corpus semantic system, each word category has its own system of semantic annotations.

Even each category differs: semantic system of things, personality, place names, and quality [8].

Let's look at an example of a basic annotation system when one word category semantic tag system can interpret another category.

Snpt: action activity, process, event (*auktion*, *night*,

election, *birthday*, *meeting*, *meeting*, *travel*) *hum hum*.

Sp t: animal animal (*cow*, *giraffe*, *crow*, *goat*, *ant*)

Snpt: color changes (*hue*, *color*, *yellow*, *shine*).

Sp t: constr buildings and structures (*house*, *attic*, *bridge*) are variable.

Snpt: disease (*angina*, *diabetes*) is variable. *Ss t*: The last name (*Pushkin*) is variable. *Sp t*: food is variable in food and drink (*pie*, *porridge*, *milk*). *Snpt*: game game (*poker*, *domino*, *volleyball*) is variable. *Snpt*: humor that changes interactions and attitudes (*mutual aid*, *war*, *birth*, *strife*). *Snpt*: param size (*height*, *weight*) is variable. *S t*: *patrn* middle name (*Sharipovich*) variable. *Ss t*: *persn* noun (*Alisher*) is variable.

Sp t: plant plant (*basil*, *roses*, *grass*) - Constant.

Sp t: space location (*space*, *city*, *taiga*, *forest*, *thicket*, *landscape*). Some of them are human intervention, some are natural, and therefore variable and constant.

Snpt: Sports (*volleyball*, *acrobatics*, *basketball*) are changing.

Sp t: stuff thing and material (*water*, *sand*, *soil*, *dough*, *silk*) Some of these are human interactions, some are created naturally and therefore variable and constants.

Snpt: Tasting Taste (*Spicy*, *Sour*, *Sour*)

Snpt: temperature changes (*cool*, *warm*, *warm*).

Sp t: text text (*story*, *book*, *ad*) is variable.

Ss t: *topon topon* (*Europe*, *Volga*, *Elbrus*, *Himalayas*, *Movarounnahr*) *t*: natural: space: *topon* (natural toponym); *Bukhara*, *Tashkent* *t*: *hum*: space: *topon* (man-made toponym) is variable and constant.

Snpt: unit of measurement (*points*, *kilograms*, *meters*, *minutes*)

T: time: age (age) and *t*: time: age: abs are also used to specify this comment along with an adequate *t*: time tag. Where the *t*: time tag denotes continuous, invariable time units, the following footnotes further clarify this comment:

A *t*: time: age: abs is absolute (three years old)

A *t*: time: age: max large (old, old, old)

A *t*: time: age: max large (old, old, old) A *t*: time: age: min small (young, small) A *t*: time: age: young (old, old) *Snpt*: time: age: age (childhood, youth, twenty years). Earlier in the work it was mentioned that the system of semantics is to include constants and operator-classifier tags. In the above-mentioned semantic system, labels are shown as variables and constants.

If the lexical unit has a single meaning irrespective of the context, a constant sign is used if the meaning is changed depending on the context. Natural toponyms are also marked with constants, and man-made toponyms are variable; in any other case, this is how the person / thing / event created by the person is defined. Analyzing the semantic mapping system of NKRYa [15], A.Kretov points to the following as the most elaborate system of labels: *Sp t*: *hum*: *ethn* - ethnonym (*Ethiopian*, *Italian*); *Sp t*: *hum*: *kin* - names denoting kinship (*aka*, *grandmother*, *amma*); *Sp t*: *hum*: *supernat* - natural creatures (*mermaid*, *alien*)



t: persn - noun (Alisher);
 t: patr - middle name (Sharipovich);
 t: famn - last name (Mahmudov);
 t: topon - toponym (Europe, Volga, Elbrus, Moscow, Movarounnahr, Tashkent).

These specific labels are important in the sense of tapping into meaningful and multi-meaning units, as these labels interpret meaningful units. Also, place descriptions are in place, loc i space. A.Kretov does not justify the use of these tags.

ADV t: place - place (middle, middle);
 A t: place - place (right, left, margin);
 V t: loc - location (lay, stand, lay);
 Snp t: loc - location (location);
 Sp t: space - space (space, gorod, taiga, meadow, hallway).
 t:color (бўёк, колорит, сарғайиш, кўқариш)

These placeholder tags initially have a sign indicating which category the unit belongs to: ADV distinguishes quality, A horse, V character, Snp and Sp. We do not agree with A.Kretov. In our opinion, these tags also serve to clarify the footnotes, and it is appropriate to include them in the syntax system. The labels listed below are of a quality category. Apparently, words in the noun category cannot be labeled with these comments: t: color - color (paint, color, yellow, bruise)A t:physq:color ранг (кизил, туссиз);

A t:physq:form (эгри, айлана);
 A t:physq:smell (хушбўй, муаттар, сассик);
 A t:physq:taste (нордон, чучук);
 A t:physq:temper (иссик, муздек);
 A t:physq:weight (оғир, енгил);
 A t:physq physical existance (юмшоқ, сиқиқ).

Snp t:be:appear (пайдо бўлиш, туғилиш, шаклланиш, ташкилот, вужудга келиш);

V t:be:appear (пайдо бўлмоқ, туғилмоқ, шаклланмоқ, вужудга келмоқ);

Snp t:be:disapp (ўлим, қатл, тугатиш/ёпиш);

In the corpus semantics a special place is given to the "Whole" field; A separate window is also placed in the interface. It is in line with the linguistic logic that all parts of the body: ears, nose, mouth, lips, hands, feet, toes, eyes, eyebrows, etc.

A.A.Kretov identifies 80 tags in the NKRY semantic tag system, 6 of which appear in two positions and 74 are non-repeating tags. For example:

Snp t: be: the beginning of the emergence (the birth, the birth, the formation, the organization, the creation);

V t: be: the beginning of the emergence (to appear, to be born, to be formed, to emerge);

Snp t: be: disapp end (death, execution, termination / closing);

V t: be: disapp end of existence (die, kill, end); Snp t: be: existence (life, existence); V t: be: exist (live, happen); Snp t: be: domestic sphere (living, emergence, death); W t: be: the domestic sphere (live, arise, kill). Apparently, these similar tags refer to different categories, except for the symbol. It's also difficult to distinguish between the comments below:

V t: be: domestic sphere (live, happen); V t: be: exist (live, happen); V t: be: disapp end of existence (die, kill, end); W t:

be: the beginning of the emergence (to appear, to be born, to be formed, to come into being). In our opinion, since these similar and repetitive labels have the same meaning units, it is preferable to leave only one (V t: be: disapp and V t: be: appear). A.A. Kretov summarizes the following on the basis of the analysis of the semantic system: [4]

1. Lexical and grammatical semantics should be an integral character / parameter. There are also places in this interpretation where the constants and the classifier are incompatible. This creates a disparity in the system of semantic tags.

2. The main function of semantic classification is to identify constants.

3. A set of semantic constants is derived from a set of key words in the language, which form the archaic picture of the world.

4. Although it is not possible to list the complete set of semantic tags, the most important of these are the signs such as nature, human, animal, flora and objects. In summary, it should be noted that the semantic roding system, along with the features of the umbilical system of rheumatism, also has several features. Similar to other linguistic systems, it is based on a linguistic supply or base. The peculiarity of a semantic jack is that there are two types of annotation: facet and a method of shredding. In practice, one or both of these methods are mixed, depending on the nature of the unit.

IV. CONCLUSION

If the facet method requires consecutive interpretation of a unit, the genealogy method requires the group, group, and field to be defined. The first method is from word to word, second method to meaning. This shows the application of the word -> meaning, meaning -> word principle in both the expert's process of searching and the interface that is reflected in the interface. Lexical-grammatical in the razmetry system as a result of monitoring the existing online corpus; lexical-thematic; characteristic of methodical evaluation; we have seen the dominance of such features as derivative characterization. Constanta and operator-classifier tags are also permanent labels on the jack system.

REFERENCES

1. Buskunbaeva L. A., Sirazetdinov Z. A. System of the national corpus of Baskir languages[electronic resource] // URL: <http://www.marlamuter.ru/forum/index.php?topic=21>.
2. Zagorulko M.Y., Kononenko I.S., Sidorova E.A. System of semantic body texture in limited predicted area // <http://www.dialog-21.ru/media/1372/94.pdf>
3. Karaulov Yu.N., Molchanov V.I., Afanasev V.A., Mikhalev N.V. Russian semantic dictionary: Опыт автоматического построения саурауса: от понятия к слову / отв. red. SG Barkhudarov. M.: Nauka, 1983. - 566 pp.
4. Kretov A.A. Analysis of semantic pometus in NKRYa // <http://ruscorpora.ru/sbornik2008/11.pdf>
5. Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. М.: Индрик, 2005. С. 155-174.



6. Кустова Г.И., Толдова С.Ю. НКРЯ: семантические фильтры для разрешения многозначности глаголов // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы / отв. ред. В.А.Плунгян. СПб.: Нестор-История, 2009. С. 258-278.
7. Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. С. 276-284.
8. Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. С. 276-284.
9. Рахилина Е.В., Кобрицов Б.П., Кустова Г.И., Ляшевская О.Н., Шеманаева О.Ю. Многозначность как прикладная проблема: семантическая разметка в Национальном корпусе русского языка // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2006». М., 2006. С. 445-450.
10. Рахилина Е.В., Кустова Г.И., Ляшевская О.Н., Резникова Т.И., Шеманова О.Ю. Задачи и принципы семантической разметки лексики в НКРЯ // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы / отв. ред. В. А. Плунгян. СПб.: Нестор-История, 2009. С.215-239.
11. Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений / Ин-т рус. яз. им. В. В. Виноградова РАН; под общ. ред. Н. Ю. Шведовой. Т. 1. М.: Азбуковник, 1998. XXV, 807с. Т. 2. М.: Азбуковник, 2000. XXXII, 762 с. Т. 3. М.: Азбуковник, 2003. 720 с. Т. 4. М.: ИРЯ РАН, 2007. 952 с.
12. Ўзбек тилининг изоҳли луғати. 5 томлик. Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2000-2006. – 4-том, 534-б.
13. <http://www.ruscorpورا.ru/>