EXPLORING SEMANTIC CONCRETENESS AND ABSTRACTNESS FOR METAPHOR IDENTIFICATION AND BEYOND

Badryzlova Yu. G. (yuliya.badryzlova@gmail.com)
National Research University Higher School of Economics,
Moscow, Russia

The paper presents a method for computing indexes of semantic concreteness and abstractness in two languages (Russian and English). These indexes are used in metaphor identification experiments in both languages; the results are either comparable to or surpass pervious work and the baselines. We analyze the obtained indexes of concreteness and abstractness to see how they align with the linguistic intuitions about the corresponding semantic categories. The results of the analysis may have broader implications for computational studies of the semantics of concreteness and abstractness.

1. Introduction

1.1. The task of computational metaphor identification

Metaphor occupies a prominent place in contemporary linguistic theory: it is recognized to be one of the most powerful cognitive tools with which humans conceptualize [Lakoff & Johnson, 1980]. Metaphor is truly ubiquitous in everyday discourse, forming a fundamental part of the language system; not surprisingly, metaphor identification and interpretation pose a serious challenge to a wide range of real-world NLP applications.

Different features have been proposed to train machine learning algorithms to identify metaphor, such as lexical, morphological and syntactic, distributional semantic, and topic modelling features; features extracted from lexical thesauri and ontologies

(WordNet, FrameNet, VerbNet, ConceptNet, etc.), and psycholinguistic features including concreteness and abstractness, imageability, affect, and force; for a comprehensive overview of systems for metaphor identification, see [Veale et al., 2016].

Much of the computational metaphor identification work for Russian follows the top-down design, i.e. is aimed at identifying conceptual metaphors [Dodge et al., 2015]; [Dunn et al., 2014]; [Mohler et al., 2014]; [Strzalkowski et al., 2013]. The experiments for identification of linguistic metaphor in Russian include [Badryzlova, 2019]; [Badryzlova & Panicheva, 2018]; [Panicheva, 2019]; [Panicheva & Badryzlova, 2017]; [Tsvetkov et al., 2014, 2013]. The first four of these projects explore different sets of features for metaphor classification, including lexical and morphosyntactic co-occurrence, distributional semantic embeddings, and concreteness-abstractness indexes (the present paper is an extension of the preliminary concreteness-abstractness studies from [Badryzlova, 2019]). The two latter of the aforementioned studies are based on cross-lingual model transfer, when the model is trained on English data using English lexical resources, and then the classification features are translated into Russian and other languages by means of an electronic dictionary (for more detail, see **Section 1.2**).

1.2. Concreteness-Abstractness feature in metaphor identification: previous research

Implementation of concreteness and abstractness in metaphor identification experiments builds on the groundwork of the theory of embodied and grounded cognition, and primary and conceptual metaphor [Barsalou, 2008, 2010]; [Lakoff & Johnson, 1999]. These theories claim that human thinking is intrinsically metaphoric, since the conceptual representations underlying knowledge are grounded in sensory and motor systems, and conceptual metaphor is the primary mechanism for transferring conventional mental imagery from sensorimotor domains to the domains of subjective experience.

Since concreteness and abstractness represent two dialectically related facets of meaning, they will be discussed under the dual term 'Concreteness-Abstractness'.

The established method to compute indexes of concreteness and abstractness of a word is to collect two sets of lexemes ('seed lists') consisting of abstract and concrete words (which together constitute the 'paradigm' of concreteness-abstractness)—and to measure the lexical similarity between each word in the lexicon and each of the paradigm words [Tsvetkov et al., 2014, 2013]; [Turney et al., 2011]. In cases when researchers are unwilling to resort to the translational method, computation of concreteness and abstractness indexes becomes a nontrivial language-specific task requiring seed lists that are compiled purposefully for a given language.

[Turney et al. 2011] compute the Concreteness-Abstractness indexes of a word by comparing its distributional semantic embedding to the vector representations of twenty abstract and twenty concrete words. The paradigm words are automatically selected from the MRC Psycholinguistic Database Machine Usable Dictionary [Coltheart, 1981], a collection of 4,295 English words rated with degrees of abstractness by human subjects in psycholinguistic experiments. We are replicating the Turney et al. experiments in this work, therefore more detail on their experimental dataset and the obtained results will be reported in **Section 3**.

[Tsvetkov et al. 2013] also compute the Concreteness-Abstractness indexes of English words by using a distributional semantic model and the MRC database: they train a logistic regression classifier on 1,225 most abstract and 1,225 most concrete words from MRC; the degree of Concreteness-Abstractness of a word is the posterior probability produced by the classifier.

This paper presents two Russian paradigms which are used to compute indexes of Concreteness and Abstractness for a large Russian vocabulary. We also introduce a similar English paradigm and compute Concreteness and Abstractness indexes for a large English vocabulary. Both the Russian and the English resources are made available to the community. We use the obtained indexes in machine learning experiments for linguistic metaphor identification on representative datasets in the two languages and compare the performance with previous research and the baselines. Besides, we look into the Concreteness-Abstractness rankings to see how they align with the intuitions about the semantic domains of abstractness and concreteness. To the best of our knowledge, this is the first research of this kind on Russian data.

2.   Concreteness-Abstractness indexes

2.1. Russian Concreteness-Abstractness paradigms

We use two paradigms to compute the indexes of Russian words; each paradigm consists of a Concrete (Con) and an Abstract (Abs) seed list.

The first paradigm is a subset of the psycholinguistic database collected at the Kazan Federal University [Solovyev et al., 2019], therefore it will be referred to as the Kazan Paradigm in this paper. The psycholinguistic database consists of 500 most frequent Russian nouns which have been ranked by human judges to indicate the perceived degree of concreteness of each noun. We took nouns with the highest and the lowest rankings to populate the Con and the Abs seed lists of the Kazan paradigm, respectively; both animate and inanimate nouns were included in the Con list.

The second paradigm was selected from the Open Semantics of the Russian Language, the semantically annotated dataset of the KartaSlov database [Kulagin, 2019]. KartaSlov contains about 71,000 words with human annotations assigning them to one of the semantic categories of the ontology. The Abs seed list was selected from the category ABSTRACT; the nouns for the Con seed list were drawn from the class PHYSICAL ENTITY, namely, the subclasses INORGANIC, THING, and ORGANIC (with the exception of HUMAN and ANIMAL). The decision to discard the categories HUMAN and ANIMAL was prompted by the consideration that HUMAN lexemes mostly denote abstract social roles rather than concrete human beings (e.g. *адвокат* 'lawyer', *доброволец* 'volunteer', *мачеха* 'stepmother', etc.), and much of the ANIMAL lexicon is intrinsically metaphoric (e.g. *ворона* 'crow', *орел* 'eagle', *лиса* 'fox', *осел* 'donkey', *корова* 'cow', etc.). This second paradigm will be addressed to as the Moscow Paradigm throughout this paper.

As for the size of the paradigms, in the metaphor identification experiments described in **Section 3**, we experimented with Con-Abs indexes computed with seed lists ranging from 360 to 40 words. No substantial difference in the quality of classification was observed in relation to the size of the paradigm. Thus, we decided to follow [Turney et al. 2011] who recommend 40 as the optimal size of a seed list, since they indicated a problematic amount of overfitting when increasing the number of words. Examples from the Kazan and the Moscow Paradigms are presented in **Table 1**.

Table 1: Excerpts from the Moscow and the Kazan Paradigms

	Abs seed list (excerpt)		Con seed list (excerpt)	
Moscow Paradigm	домашность	'domesticity'	арфа	'harp'
	мечтательность	'reverie'	теплоход	'motorship'
	автоматизм	'automatism'	зеркальце	'compact mirror'
	распутство	'promiscuity'	пельмени	'dumplings'
	чрезвычайность	'extraordinariness'	электрогитара	'electric guitar'
	партийность	'party affiliation'	гранатомет	'grenade launcher'
	активация	'activation'	линейка	'ruler'
	буддизм	'Buddhism'	фуганок	'jointer plane'
	правильность	'correctness'	узелок	'knot'
	минимализм	'minimalism'	полупальто	'short coat'
Kazan Paradigm	жизнь	'life'	стол	table'
	процесс	'process'	самолет	'aircraft'
	представление	'representation / show'	мальчик	'boy'
	способность	'ability'	квартира	'apartment'
	радость	'joy'	врач	'doctor'
	соответствие	'correspondence'	девочка	'girl'
	суть	'essence'	телефон	'phone'
	положение	'situation / location'	палец	'finger'
	желание	'desire'	рубль	'rouble'
	ответственность	'responsibility'	стекло	'glass'

2.2. English Concreteness-Abstractness paradigm

To compile the English Paradigm, we used the MRC Psycholinguistic Database (similarly to the previous research presented in **Section 1.2**). Nouns from the top and from the end of the MRC concreteness rating were drawn to populate the paradigm. We use three English Concreteness seed lists: one consisting of animate (Anim), and another of inanimate (Inan) nouns, while the third is the composite concreteness (Con) list composed of animate and inanimate nouns in equal proportions. The English seed lists are presented in **Table 2**. It can be seen that the Anim list is predominantly comprised of words indicating animals rather than humans: this reflects the distribution of animate nouns in the top of the concreteness ranking in the MRC database, where animal nouns rank higher than those denoting humans. This may be due to the consideration discussed above in **Section 2.1**, namely, that lexemes denoting humans mainly indicate abstract social roles rather than physical human beings.

Table 2: The English Paradigm

Anim	ape, adder, albatross, beetle, carp, cat, catfish, chicken, clown, cow, crab, deer, eagle, frog, goat, gorilla, grasshopper, hare, hen, horse, lion, mackerel, mussel, nightingale, otter, owl, ox, pig, puppy, rabbit, rat, sheep, shrimp, skunk, skylark, sparrow, stoat, stork, turtle, walrus
Inan	balloon, banana, barn, bed, bench, bluebell, bra, bridge, camera, car, carnation, casket, cauliflower, clarinet, collar, corkscrew, cucumber, daisy, egg, garlic, harpsichord, jacket, lamp, lantern, mattress, nightgown, olive, pants, pea, peach, pencil, piano, plum, potato, quilt, saxophone, ship, skyscraper, sofa, tulip
Abs	affirmation, animosity, demeanour, derivation, determination, detestation, devotion, enunciation, etiquette, fallacy, forethought, gratitude, harm, hatred, illiteracy, impatience, independence, indolence, inefficiency, insufficiency, integrity, intellect, interposition, justification, malice, mediocrity, obedience, oblivion, optimism, prestige, pretence, reputation, resentment, tendency, unanimity, uneasiness, unhappiness, unreality, value

2.3. Computing the Concreteness-Abstractness indexes

We computed Con-Abs indexes for a Russian vocabulary with a total of about 18,000 words, and for an English vocabulary of about 17,000 words.

The Russian Con-Abs indexes were computed using a pre-trained Continuous Skip-Gram distributional semantic model [Kutuzov & Kuzmenko, 2016] based on the Araneum corpus [Benko & Zakharov, 2016]. The English indexes were computed with a Continuous Skip-Gram model [Kutuzov et al., 2017] which had been pre-trained on the Gigaword 5th Edition corpus [Parker et al., 2011].

As shown in Equation 1, we measured semantic similarity (cosine distance, Sim) between the vectors of each word in the vocabulary and each word in a seed list, and took the mean of the ten nearest semantic neighbors (NN, Equations 2–3) in order to obtain the indexes.

$$\forall\, v_i, \forall\, s_j \ni D_i = \{Sim(v_i, s_1), Sim(v_i, s_2), ..., Sim(v_i, s_j), ..., Sim(v_i, s_k)\}, \tag{1}$$

where V is the set of words in the vocabulary, S is the set of words in the seed list, k is the number of elements in S

$$NN = \{d'_{i_1}, d'_{i_2}, ..., d'_{i_{10}}\}, \tag{2}$$

where $D_i{}'$ is a linearly ordered set of D_i (in ascending order)

$$index = Mean\{NN\} \tag{3}$$

Since we had two Con and two Abs seed lists in the Russian part of the experiment, and three Con and one Abs seed list in the English part, the following rankings were generated: Moscow Con, Moscow Abs, Kazan Con, and Kazan Abs for Russian; and English Anim, English Inan, English Con, and English Abs, for English[1].

---

[1] The full English and Russian rankings computed with the described method, as well as the RusMet corpus are available at https://github.com/yubadryzlova/Concreteness-Abstractness-in-Metaphor-Identification.

3. Metaphor identification experiments

We conduct two experiments with English data and two with Russian data.

We begin by replicating the experiments with the TroFi (Trope Finder) Example Base [Birke & Sarkar, 2006] presented in [Turney et al. 2011]. TroFi is a collection of verbal metaphor: it is built around 50 polysemous target verbs (e.g. *absorb*, *assault*, *die*, *drag*, *drown*, *smooth*, *step*, *stick*, *strike*, *touch*, etc.). Each target verb has from 1 to 115 sentences which are annotated as either literal or non-literal. For example, see the sentences with the target verb *besiege*:

- (Literal) *In 1347, Mongols < besieging > the Black Sea port of Caffa began to sicken and die from the plague.*
- (Non-literal) … *Powelson began to < besiege > me with letters asking for an invitation.*

In the first of the English experiments following Turney et al. we use a subset of 25 target verbs. Each group of sentences for a given target verb is treated as a separate learning problem, by learning and testing a separate model with ten-fold cross-validation. The performance is measured as macro-averaged Accuracy and F1-score (this experiment will be referred to as *TroFi-1* below).

In our second English experiment after Turney et al., one model is trained on the entire subset of the 25 target verbs from *TroFi-1* (1,965 sentences), which is then tested on the new unseen sentences (numbering 1,772) with the other 25 verbs (this experiment will be called *TroFi-2* in the rest of this paper).

The two Russian experiments are performed on RusMet, the Russian corpus of metaphor-annotated sentences presented in [Badryzlova & Panicheva, 2018]. The corpus consists of 7,020 sentences; each of them contains one of the 20 polysemous target verbs (e.g. *бомбардировать* 'to bombard', *нападать* 'to attack', *утюжить* 'to iron (about clothes)', *взрывать* 'to explode (smth)', *взвешивать* 'to weigh', etc.) which are used either metaphorically or non-metaphorically. The number of sentences per target verb ranges from 225 to 693; each of these subsets is balanced by class. The following sentences demonstrate examples of the metaphoric and non-metaphoric classes with the target verb *укалывать* 'to prick':

- (Metaphorical) *Это поражение серьезно < укололо > меня.* This defeat seriously piqued (lit. < 'pricked' >) me.
- (Metaphorical) *Самолюбие—это наполненный ветром воздушный шар, из которого вырывается буря, лишь < уколешь > его.* 'Vanity is a balloon filled with the wind; once you < prick > it, you release a storm'.
- (Non-metaphorical) *А маникюрша, помнится, как-то до крови < уколола > мне палец.* 'I recall that a manicurist has once < pricked > my finger so that it bled'.

Here, we start by training and testing a separate model with ten-fold cross-validation for each of the 20 subsets with individual target verbs; the overall performance is measured in terms of macro-averaged Accuracy and F1-score. Then, we split the dataset into two approximately equal parts: one containing the first ten target verbs (3,617 sentences in total), and the other containing the remaining ten target verbs (3,504 sentences). We train the classifier on the first subset and test it on the

second. These two experiments will be respectively referred to as *RusMet-1* and *RusMet-2* in the discussion below.

In all the four experiments, the metaphor identification task was formulated as sentence-level binary classification. The Support Vector Machine (SVM) classifier with linear kernel was used to learn and test the models.

For each sentence, we computed the mean of the scores of its constituent content words. The decision to use entire sentences rather than syntactic dependencies of the target verbs was motivated, firstly, by the design of the Turney et al. experiments replicated in *TroFi-1* and *TroFi-2* and, secondly, following the observations by [Mu et al. 2019] and [Badryzlova & Panicheva 2018] that accuracy of metaphor identification tends to grow with the increase of the window size, suggesting that discourse-level features consistently enhance performance of conventional, non-neural classifiers making them competitive with sophisticated neural models.

In each of the experiments we compare the performance of the Con-Abs model with the baseline lexical model in which the frequencies of lemmas are represented by the ∆P indexes of co-occurrence [Ellis, 2006]. The choice of the baseline was prompted by the findings of [Klebanov, Leong, Heilman, and Flor 2014] and [Badryzlova 2019] who reported that simple lexical unigram baselines achieve surprisingly good results for some of the datasets

Since in both of the pre-trained distributional semantic models which were used to compute the Con-Abs indexes (see **Section 2.3**) the tokens are lemmatized, the TroFi and the RusMet datasets were also POS-tagged. TroFi was preprocessed with Stanford Log-linear Part-Of-Speech Tagger [Toutanova & Manning, 2000], and RusMet with the Ru-syntax parser [Droganova & Medyankin, 2016].

4.   Results

The results of the classification experiments are presented in **Tables 3** and **4** (along with the lemma baselines).

In the *TroFi* experiments, our Con-Abs models either outperform the Turney et al. models, both in Accuracy and F1-score (sometimes by a sizeable margin, as in the case of *TroFi*-1), or they fall behind, yet by a slim margin.

When comparing the Con-Abs results with the lemma baseline, we see that in all the cases the Con-Abs models either outstrip the baseline both in Accuracy and F1-score (sometimes by a substantial margin, e.g. the Accuracy in both of the *TroFi* experiments), or they fall behind, but by a very narrow margin, as in *RusMet-1*.

The overall lower results of the *TroFi* experiments, as compared to *RusMet*, are presumably due to the unbalanced setup of the former dataset.

When comparing the performance of the models learned and tested separately for each target verb (*TroFi-1* and *RusMet-1*) with the models that were trained on one set of target verbs and tested on another set of new verbs (*TroFi*-2 and *RusMet-2*), the former models, quite expectedly, produce higher results. This is not surprising considering that the second task is more challenging and ambitious, since the model has to capture semantic regularities that are common for all the verbs irrespective of the idiosyncratic patterns of their combinability.

Table 3: Results of TroFi-1 and TroFi-2 (English)

	lemma		Con-Abs	
	Acc	F1	Acc	F1
TroFi-1, Turney et al.	NA	NA	0.73	0.64
TroFi-1, this work	0.7	0.76	0.75	0.77
TroFi-2, Turney et al.	NA	NA	0.68	0.68
TroFi-2, this work	0.63	0.67	0.7	0.67

Table 4: Results of RusMet-1 and RusMet-2 (Russian)

	lemma		Con-Abs	
	Acc	F1	Acc	F1
RusMet-1	0.83	0.83	0.82	0.82
RusMet-2	0.73	0.76	0.75	0.77

Table 5 compares models based on different types of Con-Abs indexes and their combinations in the *TroFi-2* and the *RusMet-2* experiments. The uni-feature Con models outperform the Abs models twice (Mos-Con and Eng-Con vs. Mos-Abs and Eng-Abs, respectively); besides, the combination of two Con features (Mos-Con and Kaz-Con) outstrips the combination of the two Abs features (Mos-Abs and Kaz-Abs) by a safe margin—thus suggesting that indexes of Concreteness may serve as more reliable predictors of metaphoricity.

Table 5: Comparison of Con-Abs rankings and their combinations

Russian rankings	Acc	F1	English rankings	Acc	F1
Mos-Con	0.72	0.75	Anim	0.66	0.66
Mos-Abs	0.67	0.68	Inan	0.67	0.67
Kaz-Con	0.62	0.63	Con	0.68	0.68
Kaz-Abs	0.68	0.68	Abs	0.59	0.57
Mos-Con, Mos-Abs	0.74	0.77	An, Inan	0.67	0.67
Kaz-Con, Kaz-Abs	0.74	0.74	Con, Abs	0.7	0.67
Mos-Con, Kaz-Con	0.72	0.75	An, Inan, Abs	0.69	0.67
Mos-Abs, Kaz-Abs	0.69	0.69	An, Inan, Con, Abs	0.69	0.66
Mos-Con, Mos-Abs, Kaz-Con, Kaz-Abs	0.75	0.77			

The Mos-Con model substantially outperforms Kaz-Con, while both of the Russian Abs models perform on the par with each other, which may indicate that the two Abstractness rankings resemble each other. There is not much difference between Eng-Anim and Eng-Inan; besides, the combination of Eng-Anim and Eng-Inan fares

relatively similarly to the composite Eng-Con feature; this may suggest that both animate and inanimate nouns may be equally representative of the semantic category of concreteness. In the English experiments, the best result is achieved by combining only two features, Eng-Con and En-Abs, while with the Russian models the top performance is produced by combination of all the four types of indexes.

5.   Concreteness-Abstractness indexes: a reliability test

In order to put our Con-Abs indexes to test, we took 451 known abstract nouns and 571 known concrete nouns from the KartaSlov thesaurus; we computed their Moscow Con-Abs indexes as described in **Section 2** and clustered them using k-means algorithm with *N* clusters = 2.

The resulting clustering Accuracy was 0.99: **Figure 1** shows that the Con-Abs indexes of known concrete and known abstract words form two distinct clusters with little overlap. The misclassified concrete nouns are *афиша* 'poster', *бандероль* 'parcel', *бланк* 'form (document)', and *записка* 'note'; the misclassified abstract nouns are *номинал* 'nominal value' and *присловье* 'proverb'.
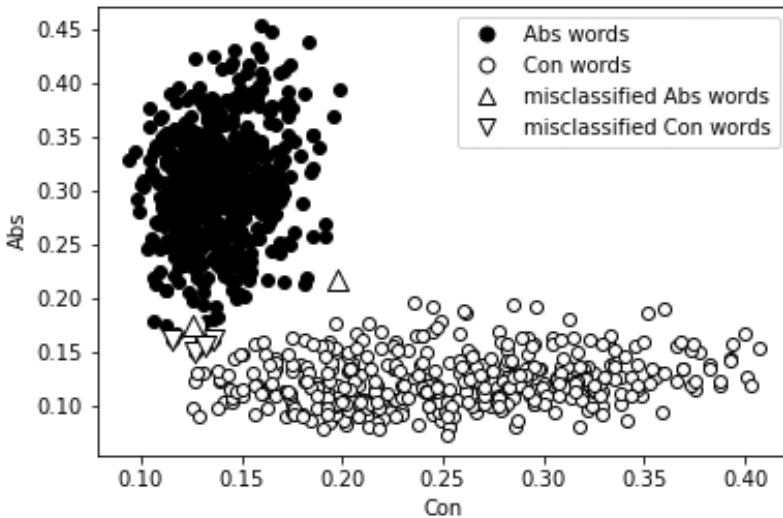

Figure 1: Reliability test: clustering known Con and Abs words (Russian)

6.   A closer look at Concreteness-Abstractness indexes

The overall encouraging results of the metaphor identification experiments and the reliability test motivate us to take a more in-depth look at the computationally generated indexes of Concreteness and Abstractness to see whether their behavior aligns with the intuitive expectations about the corresponding semantic categories.

Idealistically, one could presume that Con and Abs indexes should stand in negative correlation to each other: the higher the Concreteness of a word, the lower its Abstractness should be, and vice versa. Yet, no statistical correlation between our Con and Abs rankings has been found. Obviously, the linguistic reality is not so straightforward, especially when taking into account the issue of polysemy, when different meanings of one word may vary in Concreteness and Abstractness. A glance at visualizations of the pairs of index rankings (Con vs. Abs) reveals interesting insights.

Figure 2 shows the correspondences between the Con and the Abs indexes of the Russian and the English vocabulary. The lexemes are ordered by the Con indexes (in descending order) which thus form a smooth sigmoid curve; the small dots are the Abs indexes which correspond to each of the Con indexes in the sigmoid.

In plots A and C (depicting the Moscow and the English indexes), in the beginning of the sigmoid curve, where the Con indexes are the highest, all the Abs indexes are located below the curve, with the dots forming areas of high density. This distribution indicates that there is a well-defined group of concrete words characterized by very high concreteness and very low abstractness. At the opposite end of the curve, where the Con indexes are the lowest, all the Abs indexes are above the curve, although there is no such pronounced gap between them as in the beginning: we see that, although abstract vocabulary does form a semantically uniform group, it is not as numerous as the concrete vocabulary, and the Concreteness and Abstractness are not so distinctly contrasted in it. The long steep part of the curve shows that the Abs indexes are scattered both below and above it—yet, as the curve goes down, the number of the Abs dots above the curve increases. Thus, there is a broadly defined trend for Abstractness to grow with the decrease of the Concreteness, which conforms to the linguistic intuitions; this trend is profiled in Figure 3 which shows the number of the Abs indexes located above and below the Con line in the English vocabulary.

Table 5 demonstrates English and Russian examples of Con and Abs indexes from the beginning, the middle, and the end of the ranking. These examples are also intuitively feasible: the nouns from the top are concrete, the words in the middle are a semantic mix of the concrete and the abstract, and the words from the end of the ranking are highly abstract.

Plot B in Figure 2, which depicts the Con-Abs indexes computed using the Kazan paradigm, shows that the Abs indexes have much lower variance than in the Moscow and the English rankings. This may be a result of the fact that the lexemes in the Kazan paradigm are high-frequency words: thus, the mean frequency (ipm, according to [Lyashevskaya & Sharoff, 2009]) of the Moscow paradigm is 17.8, while the mean frequency of the Kazan paradigm is 244.23. Words with high frequency tend to co-occur with a wide range of vocabulary, so their distributional semantic vectors may be less indicative of the concreteness and abstractness expected of a paradigm word. Still, comparison of the Mos-Abs and the Kaz-Abs rankings provides an interesting observation: there is a strong positive Pearson correlation (0.73) between them—which is not the case between Mos-Con and Kaz-Con (corr = 0.5). The presence of correlation between the rankings computed with two different Abs paradigms seems to confirm the intuition that the category of Abstractness is semantically more homogeneous than the category of Concreteness. The Anim and Inan indexes of the English ranking

also show strong positive correlation (0.8), which may indicate that animate and in-
animate nouns are equally representative of the semantic category of concreteness.
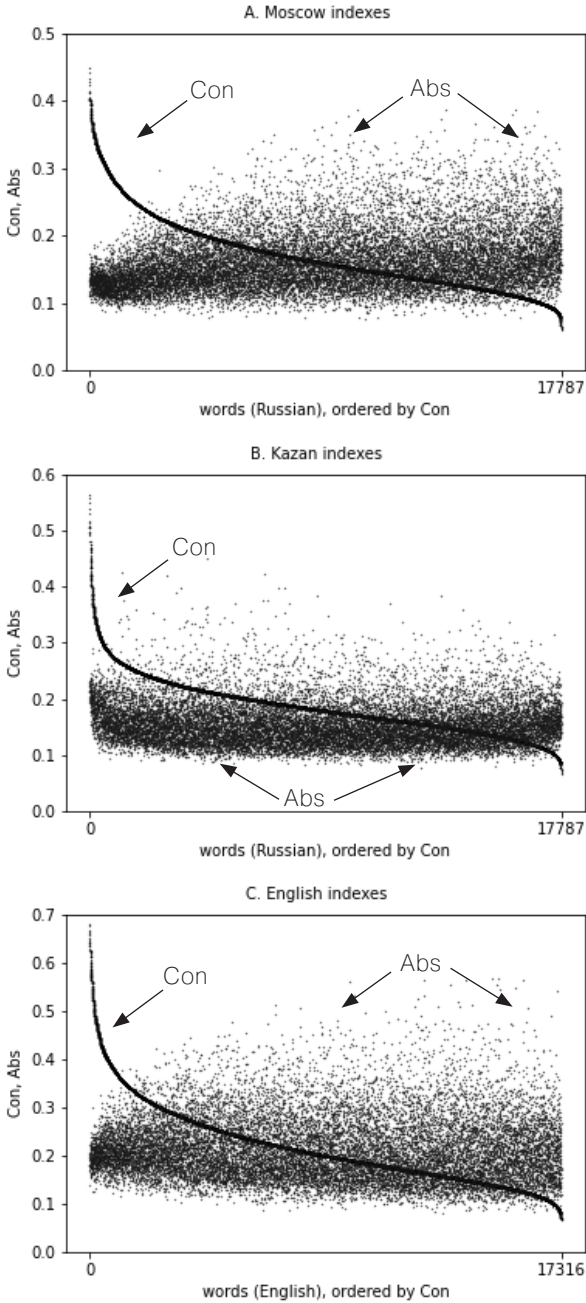

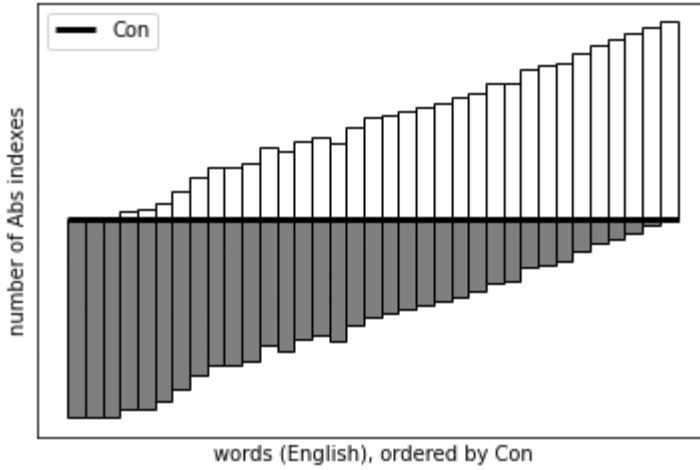**Figure 2:** Correspondences between Con and Abs indexes

Figure 3: Number of Abs indexes above
and below the Con line (English)

Table 6: Examples of indexes (Moscow and English Con-Abs)
from the top, middle, and end of the ranking

	Russian word	Translation	POS	Con	Abs	English word	POS	Con	Abs
Top	наволочка	'pillow-case'	noun	0.45	0.13	shrimp	noun	0.71	0.17
	салфетка	'napkin'	noun	0.44	0.11	crab	noun	0.68	0.18
	плед	'bed throw'	noun	0.43	0.13	tomato	noun	0.68	0.2
	омлет	'omelette'	noun	0.43	0.14	scallop	noun	0.67	0.18
	одеяло	'blanket'	noun	0.42	0.13	chicken	noun	0.66	0.14
	сумочка	'purse'	noun	0.41	0.13	mussel	noun	0.66	0.18
	скатерть	'tablecloth'	noun	0.41	0.13	lobster	noun	0.66	0.17
	винтовка	'rifle'	noun	0.41	0.13	oyster	noun	0.65	0.18
	шарф	'scarf'	noun	0.4	0.13	onion	noun	0.65	0.19
	соус	'sauce'	noun	0.4	0.15	potato	noun	0.64	0.2
Middle	торрент	'torrent'	noun	0.16	0.14	healey	noun	0.2	0.17
	нагло	'brazenly'	adv	0.16	0.19	swing	verb	0.2	0.16
	бодро	'cheerfully'	adv	0.16	0.15	carrier	noun	0.2	0.15
	раж	' zeal'	noun	0.16	0.21	smash	verb	0.2	0.11
	вальдорфский	'waldorf'	adj	0.16	0.14	unload	verb	0.2	0.14
	впереди	'in front of'	adv	0.16	0.14	stanford	noun	0.2	0.16
	продавать	'sell'	verb	0.16	0.09	north	adv	0.2	0.15
	потрениро-ваться	'practice'	verb	0.16	0.12	perilous	adj	0.2	0.21
	вековой	'centennial'	adj	0.16	0.18	enumeration	noun	0.2	0.29

	Russian word	Translation	POS	Con	Abs	English word	POS	Con	Abs
End	*согласоваться*	'correspond'	verb	0.07	0.19	vitro	propn	0.2	0.12
	дифференциация	'differentiation'	noun	0.07	0.27	open-ended	adj	0.11	0.24
	вовлечение	'involvement'	noun	0.07	0.26	inter-depart-mental	adj	0.11	0.18
	распространение	'proliferation'	noun	0.07	0.2	rejection	noun	0.11	0.36
	разделение	'separation'	noun	0.07	0.27	discredit	verb	0.11	0.31
	тотальный	'total'	adj	0.07	0.25	hold	verb	0.11	0.13
	пониматься	'be regarded as'	verb	0.07	0.22	legitimate	adj	0.11	0.28
	всесторонне	'comprehen-sively'	adv	0.07	0.16	unsuccess-fully	adv	0.11	0.15
	выявлять	'detect'	verb	0.07	0.14	long-awaited	adj	0.11	0.19
	переподчинение	're-subordina-tion'	noun	0.07	0.23	abolition	noun	0.11	0.25

7. Conclusions

We have presented a method for computing indexes of semantic Abstractness and Concreteness of words in two languages—Russian and English, and applied them to the task of linguistic metaphor identification in these two languages.

The results of the classification are either comparable with previous research or surpass it, the same holds for the baseline model. The efficiency of metaphor identification with Concreteness-Abstractness indexes suggests that metaphoric and non-metaphoric contexts are semantically different in terms of Concreteness and Abstractness, which conforms to the Conceptual Metaphor Theory. Models based on Concreteness indexes are more efficient than the models based on Abstractness, suggesting that Concreteness may be a more reliable predictor of metaphoricity.

Analysis of the Concreteness-Abstractness indexes reveals that there is a general trend for Abstractness indexes to increase as the corresponding Concreteness indexes decrease, which is in accordance with linguistic intuitions. There is a distinct group of highly concrete words in the lexicon which have very high Concreteness and very low Abstractness indexes; similarly, there is a group of distinctly abstract vocabulary, with low Concreteness and high Abstractness scores.

Presence of statistical correlation between two Russian Abstractness ratings (and absence of correlation between the respective Concreteness ratings) may indicate that the category of Abstractness is more semantically homogeneous than the category of Concreteness. Statistical correlation between the English concrete Animate and In-animate rankings suggests that both animate and non-animate nouns may be equally representative of the semantic category of Concreteness.

References

1. *Badryzlova, Y.* (2019). Automated metaphor identification in Russian texts. National Research University Higher School of Economics.
2. *Badryzlova, Y., & Panicheva, P.* (2018). A Multi-feature Classifier for Verbal Metaphor Identification in Russian Texts. Conference on Artificial Intelligence and Natural Language, 23–34.
3. *Barsalou, L. W.* (2008). Grounded cognition. Annu. Rev. Psychol., 59, 617–645.
4. *Barsalou, L. W.* (2010). Grounded cognition: Past, present, and future. Topics in Cognitive Science, 2(4), 716–724.
5. *Benko, V., & Zakharov, V.* (2016). Very large Russian corpora: New opportunities and new challenges. In Computational linguistics and intellectual technologies. Russian State University for the Humanities.
6. *Birke, J., & Sarkar, A.* (2006). A Clustering Approach for Nearly Unsupervised Recognition of Nonliteral Language. EACL.
7. *Coltheart, M.* (1981). The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology Section A, 33(4), 497–505.
8. *Dodge, E., Hong, J., & Stickles, E.* (2015). MetaNet: Deep semantic automatic metaphor analysis. NAACL HLT 2015, 40.
9. *Droganova, K., & Medyankin, N.* (2016). NLP pipeline for Russian: An easy-to-use web application for morphological and syntactic annotation. Proceedings of the Annual International Conference "Dialogue". Annual International Conference "Dialogue", Moscow.
10. *Dunn, J., de Heredia, J. B., Burke, M., Gandy, L., Kanareykin, S., Kapah, O., Taylor, M., Hines, D., Frieder, O., & Grossman, D.* (2014). Language-Independent Ensemble Approaches to Metaphor Identification. Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence.
11. *Ellis, N. C.* (2006). Language acquisition as rational contingency learning. Applied Linguistics, 27(1), 1–24.
12. *Klebanov, B. B., Leong, B., Heilman, M., & Flor, M.* (2014). Different texts, same metaphors: Unigrams and beyond. Proceedings of the Second Workshop on Metaphor in NLP, 11–17.
13. *Kulagin, D.* (2019). Opy`t sozdaniya mashinno-proveryaemoj semanticheskoj razmetki russkix sushhestvitel'ny'x [Developing computationally verifiable semantic annotation of Russian nouns]. Annual International Conference "Dialogue", Moscow.
14. *Kutuzov, A., Fares, M., Oepen, S., & Velldal, E.* (2017). Word vectors, reuse, and replicability: Towards a community repository of large-text resources. Proceedings of the 58th Conference on Simulation and Modelling, 271–276.
15. *Kutuzov, A., & Kuzmenko, E.* (2016). WebVectors: A toolkit for building web interfaces for vector semantic models. International Conference on Analysis of Images, Social Networks and Texts, 155–161.
16. *Lakoff, G., & Johnson, M.* (1980). Metaphors We Live By (2nd ed.). The University of Chicago Press.
17. *Lakoff, G., & Johnson, M.* (1999). Philosophy in the Flesh (Vol. 4). New york: Basic books.

18. *Lyashevskaya, O., & Sharoff, S.* (2009). Chastotny`j slovar` sovremennogo russkogo yazy`ka na materialax Nacional`nogo korpusa russkogo yazy`ka [A frequency dictionary of modern Russian language on the basis of the Russian National Corpus]. Azbukovnik.

19. *Mohler, M., Rink, B., Bracewell, D. B., & Tomlinson, M. T.* (2014). A Novel Distributional Approach to Multilingual Conceptual Metaphor Recognition. COLING, 1752–1763.

20. *Panicheva, P.* (2019). Analiz parametrov semantičeskoj svjaznosti s pomošč'ju distributivny xsemantičeskix modelej (na materiale russkogo jazyka) [Analysis of parameters of semantic coherence by means of distributional semantic models (on Russian data)]. St. Petersburg State University.

21. *Panicheva, P., & Badryzlova, Y.* (2017). Distributional semantic features in Russian verbal metaphor identification. Computational Linguistics and Intellectual Technologies, 1, 179–190.

22. *Parker, R., Graff, D., Kong, J., Chen, K., & Maeda, K.* (2011). English Gigaword Fifth Edition LDC2011T07 (Tech. Rep.). Technical Report. Linguistic Data Consortium, Philadelphia.

23. *Solovyev, V., Andreeva, M., Solnyshkina, M., Zamaletdinov, R., Danilov, A., & Gaynutdinova, D.* (2019). Computing Concreteness Ratings of Russian and English Most Frequent Words: Contrastive Approach. 2019 12th International Conference on Developments in ESystems Engineering (DeSE), 403–408.

24. *Strzalkowski, T., Broadwell, G. A., Taylor, S., Feldman, L., Yamrom, B., Shaikh, S., Liu, T., Cho, K., Boz, U., Cases, I., & others.* (2013). Robust extraction of metaphors from novel data. Proceedings of the First Workshop on Metaphor in NLP, 67–76.

25. *Toutanova, K., & Manning, C. D.* (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13, 63–70.

26. *Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., & Dyer, C.* (2014). Metaphor detection with cross-lingual model transfer. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1, 248–258.

27. *Tsvetkov, Y., Mukomel, E., & Gershman, A.* (2013). Cross-lingual metaphor detection using common semantic features. Proceedings of the First Workshop on Metaphor in NLP, 45–51.

28. *Turney, P. D., Neuman, Y., Assaf, D., & Cohen, Y.* (2011). Literal and metaphorical sense identification through concrete and abstract context. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 680–690.

29. *Veale, T., Shutova, E., & Klebanov, B. B.* (2016). Metaphor: A Computational Perspective. Synthesis Lectures on Human Language Technologies, 9(1), 1–160.