KonKretiKa @ CONcreTEXT: Computing Concreteness Indexes with Sigmoid Transformation and Adjustment for Context

Yulia Badryzlova
HSE University
Moscow, Russia
`yuliya.badryzlova@gmail.com`

Abstract

The present paper is a technical report of KonKretiKa, a system for computation of concreteness indexes of words in context, submitted to the English track of the CONcreTEXT shared task. We treat concreteness as a bimodal problem and compute the concreteness indexes using paradigms of concrete and abstract seed words and distributional semantic similarity. We also conduct sigmoid transformation to achieve greater similarity to the psycholinguistically attested data, and apply dynamic adjustment of static indexes for sentential context. One of the modifications of the presented system ranked third in the task, with $r_s = .6634$ and $r = .6685$ against the gold standard.

1 Introduction

This paper is a description of the system with the working title KonKretiKa, which was submitted to the English track of CONcreTEXT, the shared task on evaluation of concreteness in context (Gregori et al., 2020) offered at EVALITA 2020, the 7th evaluation campaign of Natural Language Processing and speech tools for the Italian language (Basile et al., 2020).

KonKretiKa stems from our previous work on computation of such indexes for the purposes of metaphor identification.

Computationally obtained indexes of concreteness are extensively explored in experiments for automated metaphor identification. Application of concreteness indexes to metaphor identification relies on the assumptions made by the theories of embodied and grounded cognition (Barsalou, 2008), and primary and conceptual metaphor (Lakoff and Johnson, 1980). 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.

An established method to compute the concreteness index of a word is to collect two sets of lexemes ('seed lists', or 'paradigms') consisting of abstract and concrete words – and to measure the lexical similarity between each word in the lexicon and each of the paradigm words.

Turney et al. (2011) use concreteness indexes to identify linguistic metaphor in the TroFi dataset (Birke and Sarkar, 2006). They compute the concreteness index of a word by comparing its distributional semantic embedding to the vector representations of 20 abstract and 20 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.

Tsvetkov et al. (2013) also compute the concreteness 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 of a word is the posterior probability produced by the classifier. The Tsvetkov et al. system for metaphor identification with concreteness indexes is based on cross-lingual model transfer, when the model is trained on English data, and then the classification features are translated into other languages by means of an electronic dictionary.

Concrete	albatross, balloon, bench, bridge, catfish, cauliflower, chicken, clown, corkscrew, crab, daisy, deer, eagle, egg, frog, garlic, goat, harpsichord, lion, mattress, mussel, nightgown, nightingale, owl, ox, pants, peach, piano, pig, potato, quilt, rabbit, saxophone, sheep, shrimp, skyscraper, sofa, stoat, tulip, turtle
Abstract	affirmation, animosity, demeanour, derivation, determination, detestation, devotion, enunciation, etiquette, fallacy, forethought, gratitude, harm, hatred, ignorance, 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

Table 1. The concrete and the abstract paradigm lists.

$$\forall \, v_i, \forall s_j \, \exists \, D_i = \left\{ Sim(v_i, s_1), \; Sim(v_i, s_2), \dots , \; Sim\left(v_i, s_j\right), \; \dots , Sim(v_i, s_k) \right\}, \quad (1)$$

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

$$NN \; = \left\{ d_{i_1}' , \; d_{i_2}', \dots , d_{i_{10}}' \right\}, \quad (2)$$

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

$$I = Mean\{NN\} \quad (3)$$

**Equations 1-3. Computation of indexes with paradigm lists.**

Badryzlova (2020) explores concreteness and abstractness indexes for linguistic metaphor identification in Russian and English. The paradigm words are selected in a semi-automatic fashion: the Russian paradigm is derived from the Open Semantics of the Russian Language, the semantically annotated dataset of the KartaSlov database (Kulagin, 2019); the English paradigm is selected from the MRC database (Coltheart, 1981). The indexes of concreteness and abstractness are computed for large sets of Russian and English words (about 18,000 and 17,000 lexemes, respectively). The metaphor identification in Russian is conducted on the RusMet corpus (Badryzlova, 2019; Badryzlova and Panicheva, 2018), and the English on the TroFi dataset. The author shows that the distributions of concreteness and abstractness indexes in the two languages follow the same pattern: in the lexicon, there is a distinct group of highly concrete words, 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. Moreover, there is a general trend for abstractness indexes to increase as the corresponding concreteness indexes decrease. The author also observes statistical correlation between two Russian abstractness ratings, which may indicate that the category of abstractness is more semantically homogeneous than the category of concreteness.

The present work develops and extends the method of Badryzlova (2020) in two directions: (a) we apply sigmoid transformation to fit the curve comprised of the computed concreteness and abstractness indexes to the distribution of indexes in psycholinguistic data; and (b) we suggest a method for dynamic adjustment of the obtained indexes for sentential context, according to the requirements of the CONcreTEXT shared task (Gregori et al., 2020). The working title of the proposed system is KonKretiKa.

2 Description of the system

We demonstrate a method for evaluating concreteness on English data; however, it can be transferred to any other language provided that the following types of resources are available: (1) a lexicon with semantic (e.g. Fellbaum, 1998; Kulagin, 2019) or psycholinguistic (e.g. Brysbaert et al., 2014; Coltheart, 1981) annotation to select the paradigm words from; (2) a pre-trained distributional semantic model; and (3) a relatively large wordlist containing lexemes with different frequencies of occurrence (ipm) in order to ensure the maximum possible variation in concreteness across the lexicon.

When analyzing the distribution of psycholinguistic concreteness ratings, Brysbaert et al. (2014) observe that "concreteness and abstractness may be not the two extremes of a quantitative continuum […], but two qualitatively different
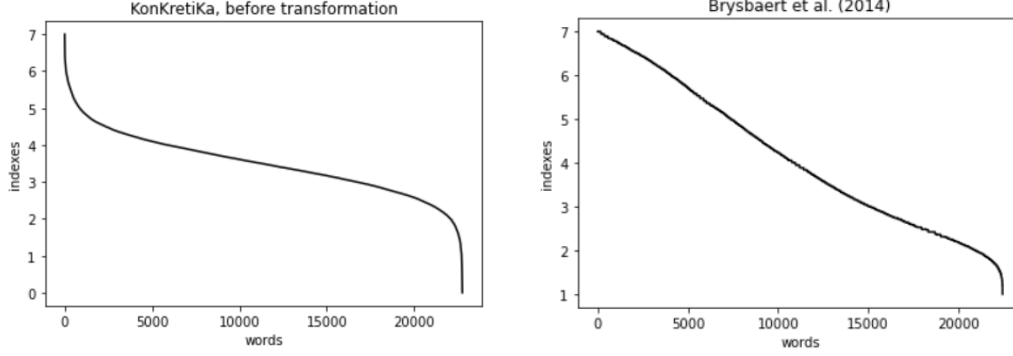
Figure 1. Distribution of computational (raw KonKretiKa) and psycholinguistic (Brysbaert et al.) indexes
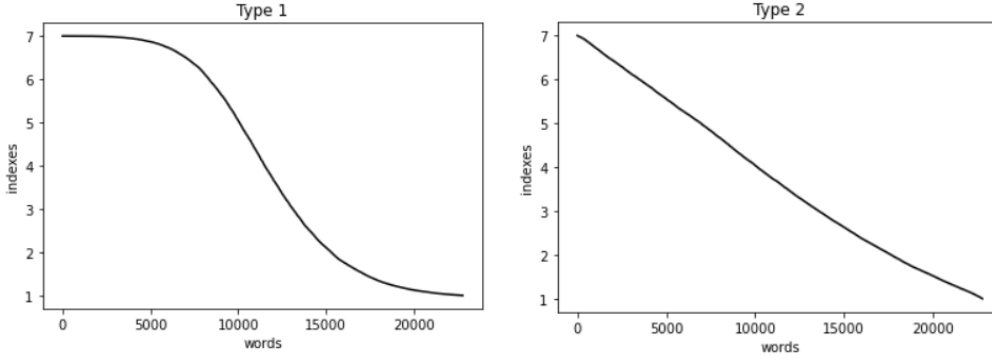
Figure 2. Sigmoid transformations

characteristics." of a word. Following this observation and the previous work in (Badryzlova, 2020), we treat concreteness as a bimodal property investing the word with two characteristics: the rate of concreteness and the rate of abstractness. Thus, we start by computing the standalone indexes of concreteness and of abstractness; then, the single aggregate index is computed as a function of these two indexes.

2.1 Computation of raw indexes with paradigm words and distributional semantic similarity

Computation of the standalone concreteness and abstractness indexes is based on paradigm lists of concrete and abstract words; we use the English concrete and abstract paradigms from Badryzlova (2020). These paradigms were compiled from the MRC Psycholinguistic Database: nouns from the top and from the end of the MRC concreteness rating were drawn to populate the concrete and the abstract paradigms, respectively. The paradigm lists are presented in Table 1.

The indexes of concreteness and abstractness were computed using a Continuous Skip-Gram model (Kutuzov et al., 2017) which had been

pre-trained on the lemmatized Gigaword 5th Edition corpus (Parker et al., 2011).

As shown in Equations 1-3, to compute a concreteness or an abstractness index (I) of a word, we measured semantic similarity (cosine distance) *Sim* between the vectors of this word and each word in the paradigm (concrete or abstract, respectively), and took the mean of the ten nearest semantic neighbors (*NN*).

In total, we computed concreteness and abstractness indexes for approximately 23,000 English words (nouns, verbs, adjectives, and adverbs); this lexicon was taken from the Brysbaert et al. (2014) ranking, which allowed us to analyze the correlation between the computational and the large-scale psycholinguistic data at the subsequent stages of the present study (see Section 3).

The obtained computational sets of concreteness and abstractness indexes were normalized to the range [1, 7][1] in order to comply with the scale set by the CONcreTEXT shared task. In order to obtain an aggregate single-value index of a word, which would be representative of both its concreteness and abstractness, we subtracted the abstractness indexes from the concreteness indexes.

---

[1] Scikit-learn's MinMax Scaler (Pedregosa et al., 2011)

System	Transformation type	Contextual adjustment (c)	Result (r_s)	Result (r)
Leader-1			0.83313	0.83406
Leader-2			0.78541	0.78682
KonKretiKa-3	2	0.5	0.6634	0.6685
KonKretiKa-1	1	0.5	0.65102	0.66652
Baseline-2			0.55449	0.56742
KonKretiKa-4	2	0.8	0.54216	0.54465
KonKretiKa-2	1	0.8	0.54089	0.54479
Baseline-1			0.3825	0.37743

Table 2. Modifications of KonKretiKa and their results in the shared task.

2.2 Sigmoid transformation of raw indexes

Figure 1 shows distributions of our raw aggregate indexes and the indexes attested in psycholinguistic research (Brysbaert et al., 2014). It is noticeable that the curve of computational indexes has a much steeper slope, resulting in lower variance; consequently, the discriminative power of such indexes will also be lower.

The raw KonKretiKa curve has the shape of a sigmoid; in generic form, the sigmoid function is described by the equation:

$$S(x) = \frac{1}{1 + \exp(-ax + a * b)}$$

where a defines the slope of the function and b defines the inflection point. Consequently, we can transform the sigmoid by changing the a and b coefficients.

In the submissions to the CONcreTEXT shared task, we experimented with two transformations of the raw KonKretiKa curve (Figure 2). In the first transformation, we applied a heuristically chosen combination of a and b which was intended to increase the slope and the curvature while preserving the *S*-shape of the sigmoid. The second transformation was intended to attain maximum resemblance of its shape to the Brysbaert et al. curve. We used grid search with different combinations of coefficients a and b to maximize the correlation between the two curves. During

this fitting, only the values of the indexes are adjusted, while their initial ranks remain intact – thus, there is no data leakage from the psycholinguistic ranking. [2]

2.3 Contextual adjustment

Since the CONcreTEXT shared task requires that the concreteness indexes of target words be dynamically adjusted to their sentential context, the following heuristic was applied in the submitted KonKretiKa models. We computed the mean concreteness of all content words in the sentence (with the target word excluded) and adjusted the concreteness value of the target word accordingly. The adjusted index A was computed as follows:

$$A_t = R_t - (M * c)$$

where t is the target word, R is the raw index from the KonKretiKa ranking, M is the mean concreteness of the sentence, and c is the adjustment coefficient. In the models submitted to the CONcreTEXT shared task, we applied two heuristically defined c coefficients: $c = 0.5$ and $c = 0.8$.

Thus, the four modifications of KonKretiKa submitted to the shared task were differentiated by the two parameters: the type of transformation and the contextual adjustment coefficient.

3 Results and discussion

The parameters of the four modifications and their results are presented in Table 2 (along with the Baselines and the Leaders). The results indicate that systems with the lower coefficient of sentential adjustment (0.5) perform better than systems with the higher adjustment coefficient (0.8) irrespective of the type of sigmoid transformation; yet, the system with Type 2 (fitted to the psycholinguistic data) transformation somewhat outperforms the system with Type 1 (*S*-shaped) transformation.

The best of our modifications, KonKretiKa-3, demonstrated Spearman correlation with the gold standard $r_s = .6634$ and Pearson correlation $r = .6685$, ranking our system third in the track, yet by a substantial margin behind the two winning system (with $r_s = .83313$ and $r = .83406$ and $r_s = .78541$ and $r = .78682$, respectively).

---

[2] The KonKretiKa ranking is available at:
https://github.com/yubadryzlova/CONcreTEXT-2020

Dataset	Gold (dynamic)	BRY (static)
KKK (static)		$r_s = .743$ $r = .751$
KKK (dynamic)	$r_s = .663$ $r = .669$	
BRY (static)	$r_s = .755$ $r = .761$	

Table 3. Pairwise correlations: KKK – KonKretiKa, BRY – Brysbaert et al., Gold – CONcreTEXT gold standard.

word	BRY	KKK	Diff
handmaiden (N)	6.45	1.54	4.91
tire (V)	7	2.18	4.82
bedrock (N)	6.18	1.55	4.63
alarm (N)	6.19	1.58	4.61
text (N)	6.89	2.31	4.58
nonreactive (ADJ)	2.25	6.82	-4.57
temptingly (ADV)	1.72	6.26	-4.55
hail (V)	5.96	1.5	4.47
stance (N)	5.53	1.11	4.42
nudge (N)	6.19	1.8	4.39
chasm (N)	5.84	1.45	4.39

Table 4. Top residuals: Brysbaert et al. (BRY) vs. KonKretiKa (KKK).

3.1 Analysis of contextual adjustment

We carried out a post hoc analysis of the contextal adjustment coefficient (c) by using grid search to maximize the correlation between KonKretiKa (Type 2 transformation) and the gold standard. Moreover, we altered the scope of the context words for which the mean sentential concreteness (M) was computed – by taking 2-3 nearest semantic neighbors (either of any part of speech, or only nouns, or only verbs); this was done in order to reduce the possible noise from the words that are not semantically related to the target in the sentence. The change of the contextual scope did not lead to a substantial difference in the result. As for the contextual adjustment coefficient, the grid search showed that $c = 0.32$ – which is lower than the most efficient coefficient from our earlier submissions ($c = 0.5$ in KonKretiKa-3) – results in a slight increase of correlations: $r_s = .678$ and $r = .688$.

A closer analysis of the test sentences suggests that contribution of contextual adjustment presumably may be increased by considering a broader context of a sentence – for instance, spanning over 1-3 adjacent sentences from the left and the right contexts; this option constitutes a possible direction for future work.

3.1 Comparison of computational and psycholinguistic data

Pairwise correlations between the computational (KonKretiKa, KKK) and the psycholinguistic rankings (Brysbaert et al., BRY and the gold standard) are shown in Table 3. It can be seen that KKK better correlates with the BRY data than with the gold standard ($r_s = .743$, $r = .751$ vs. $r_s = .663$, $r = .669$, respectively). Presumably, such difference in the two correlations is due to the much larger size of the BRY lexicon. The correlation between the two psycholinguistic datasets (BRY vs. Gold) is $r_s = .755$, $r = .761$, which is close to the correlation between KKK and BRY.

We undertook closer pairwise comparative analysis between two pairs of rankings:

1. Static KonKretiKa indexes (the indexes after Type 2 sigmoid transformation, without contextual adjustment) vs. the Brysbaert et al. ranking (which is also static): approximately 23,000 words – nouns, verbs, adjectives, and adverbs (the two wordlists are identical).
2. Indexes of the target words from the CONcreTEXT test data as presented in the dynamic version of KonKretiKa (the sigmoid-transformed Type 2 indexes with contextual adjustment coefficient $c = 0.32$) vs. the Gold standard (where the target words are also ranked dynamically in context): 436 words – verbs and nouns.

The top residuals between the KonKretiKa and the Brysbaert et al. indexes are presented in Table 4. Analysis of these discrepancies suggests that most of them stem from polysemy and the differences between its representation in distributional semantic models and in psycholinguistic reality. Thus, distributional semantic models do not discriminate between various meanings of words; if occurrences of one of the meanings substantially outnumber the other meanings in discourse and, as a consequence, in the training corpus, the resulting vector reflects the more frequent meaning.

Sentence	Target word	Gold	KKK	Diff	TEXT
399	vision (N)	6.03	1.86	4.17	Check your < vision > to see if you are seeing blurry or double.
353	vision (N)	5.97	1.82	4.15	With retinal migraine, you may experience loss of < vision > in one eye and a headache that starts behind your eyes.
155	spirit (N)	6	2.33	3.67	Gin is an alcoholic < spirit > made from distilled grain or malt.
324	pain (N)	5.2	1.59	3.61	See your doctor if you are experiencing < pain > or discomfort.
61	answer (N)	5.45	1.91	3.54	Be sure to write your final < answer > without the negative sign.
385	war (N)	5.57	2.06	3.51	They have escaped from civil < war > in Liberia or Zimbabwe.
81	answer (N)	5.32	1.92	3.4	Final < answers > for equations are considered wrong unless you have broken them down to their simplest form.
237	heart (N)	6.32	2.98	3.34	The < heart > pumps blood due to an internal electrical system.
163	pain (N)	4.97	1.63	3.34	Take your medications to ease your physical < pain >.
176	agreement (N)	5.16	1.85	3.31	After signing the indemnification < agreement >, you can sign the legally binding bond agreement.

Table 5. Top residuals: KonKretiKa (KKK) vs. Gold standard.

For example, the nearest semantic neighbors of the noun *handmaiden* in the distributional semantic model[3] are: *embodiment*, *personification*, *epitome*, and *paragon* – associating this word with its abstract, metaphoric meaning 'something that supports something else that is more important'[4], whereas for speakers of English the other, concrete meaning 'a woman who is someone's servant' apparently stands out as being more salient. Similarly, among the nearest semantic neighbors of the noun *chasm* in the distributional semantic model are: *disparity*, *schism*, *rich-poor divide*, *mistrust*, *(the) haves*, *divergence*, *antagonism*, and *inequality* – indicating that the distributional vector of *chasm* is biased towards the abstract meaning of this word ('a very big difference that separates one person or group from another') rather than the concrete one ('a very deep crack in rock or ice'), while human subjects see the latter meaning as more salient or prevalent.

As for *nonreactive* and *temptingly*, which are more concrete in the computational data, this could be explained by their perceived vagueness to human subjects, since these words do not have meanings that would be markedly juxtaposed to each other in terms of concreteness-abstractness – thus ranking them rather low in the psycholinguis-tic data. Meanwhile, the nearest semantic neighbors of *temptingly* in the distributional semantic model are: *strappy sandal*, *capelet*, *knee-length skirt*, *enticingly*, *floral-print*, *high-heeled sandal*, *lace-trimmed*, *harem pants*, and *puffed sleeve* – all rather concrete objects (or the properties of such objects).

The top residuals between KonKretiKa and the gold standard are shown in Table 5. The discrepancy between the abstract meaning of *vision* ('the ability to think about and plan for the future, using intelligence and imagination, especially in politics and business') and its concrete meaning ('the ability to see') can also be attributed to the differences between representation of meanings in distributional semantic models and in psycholinguistic reality – the reason already discussed above. Thus, the nearest distributional semantic neighbors of *vision* are: *worldview*, *ideal*, *visionary*, *thinking*, *perspective*, *idea*, *dream*, and *blueprint* – rather than terms related to eyesight.

The noun *spirit* in Table 5 (Sentence 155) is used in the sense of 'strong alcoholic drink'. However, its nearest neighbors in the distributional semantic model are *ethos*, *ideal*, *idealism*, *tradition*, *essence*, *enthusiasm*, *passion*, *faith*, *chivalric*, *zeal*, *credo*, and *compassion* – indicating that the meaning 'your attitude to life or to other people'

---

[3] Continuous Skip-Gram model (Kutuzov et al., 2017), pre-trained on Gigaword 5th Edition corpus

[4] Definitions are cited according to Macmillan Dictionary (n.d.)

is dominant in the model, and the contextual adjustment we apply is not sufficient for overcoming the abstractness of the dominant meaning.

As for the noun *war*, its nearest neighbors in the distributional semantic model are *conflict*, *warfare*, *invasion*, *1991-95 Serbo-Croatian*, *Israel-Hezbollah*, *genocide*, *Bosnia war*, *Jehad*, *civil-war*, *Croatia war*, *Cold War*, *Iran-Iraq*, *wartime*, *Vietnam-like*, etc. – that is, rather abstract concepts. The only more concrete words referring to physical combat action that occur in the distributional semantic neighborhood of *war* are *battle-field* and *bloodshed*, but this is not enough to outweigh the abstract terms. Thus, the distributional semantic model models warfare in terms of abstract rather than concrete (such as names of weapons, military equipment, military personnel, etc.) concepts. As a result, military action is not sufficiently juxtaposed to the metaphoric meaning of *war* as 'a situation in which two people or groups of people fight, argue, or are extremely unpleasant to each other'.

In the case of *answer* and *agreement*, their nearest distributional semantic neighbors in the model are fairly abstract concepts: *explanation*, *answer*, *reply*, *solution*, *unanswerable*, *query*, *TV-talkback answer*, *question*, and *yes* (for *answer*), and *accord*, *pact*, *deal*, *treaty*, *initial*, *negotiation*, *memorandum*, *compromise*, and *negotiate* (for *agreement*). Meanwhile, human subjects rank *answer* and *agreement* rather high in concreteness; presumably, this is a consequence of conflating the mental representations of the action of answering / reaching an agreement with their two modes – the spoken and the written, i.e. with the physical actions of speaking and writing. This conflation is not reflected in discourse – it largely exists in the mental representations of *answer* and *agreement* and, therefore, is not very distinguishable on the level of linguistic representation.

Of interest are the cases of *heart* and *pain*, which have much lower concreteness in KonKretiKa than in the gold standard sentences where these words are used in their physical, concrete meanings. The nearest distributional semantic neighbors of *heart* are *heart-related*, *coronary artery*, *kidney*, *liver*, *lung*, *arrhythmia*, *cardiac*, *angina*, and *aneurism*. The nearest neighbors of *pain* are *discomfort*, *ache*, *agony*, *tingling sensation*, *numbness*, *soreness*, *menstrual cramp*, *light-headedness*, *stiffness*, *nausea*, and *arthritis*. It would be quite expected for such semantic neighborhood to entitle *heart* and *pain* to higher concreteness values than what they receive in KonKretiKa. A more in-depth analysis into this contradiction revealed that it stems from the vulnerability in the semantic composition of the concrete paradigm which was used to compute the raw indexes (see Table 1). The words of this paradigm belong to the two major semantic classes – living organisms (animals and plants) and man-made artifacts. The class of words denoting human beings was intentionally excluded when the paradigm was compiled on the grounds that such nouns tend to indicate abstract social roles rather than physical humans. As a consequence, physical organic objects such as body parts and organs, or physical sensations and physiological conditions received non-uniform indexes in KonKretiKa: those that refer to humans as well as to animals (e.g. in veterinary or gastronomic discourse) ranked rather high in concreteness: e.g. *liver* (6.6), *pancreas* (6.4), *foot* (6.3), *encephalitis* (6.25), *kidney* (6.25), *entrails* (6.05), *tummy* (5.92), *womb* (5.6) – whereas those that tend to be primarily associated with humans received lower indexes, e.g. *heart* (2.63), *heartburn* (2.57), *scar* (2.53), *nausea* (2.5), *headache* (1.61), *distress* (1.5), *pain* (1.21), *queasiness* (1.12), etc. Thus, comparison of the KonKretiKa computational indexes with the psycholinguistic data of CONcreTEXT allowed us to detect a potential shortcoming in our approach to the design of the concrete paradigm. As was noted in previous study (Badryzlova, 2020), the class of concrete words seems to be more semantically heterogeneous than of abstract words; therefore, it may reasonable in future experiments to diversify the concrete paradigm and expand it in size by including words that denote human beings.

4 Conclusions

We presented KonKretiKa system for computing concreteness indexes of English words in context; the system was submitted to the English track of the CONcreTEXT shared task. The best modification of KonKretiKa ranked third in the task, with $r_s = .6634$ and $r = .6685$ against the gold standard. We treat concreteness as a bimodal problem and use paradigm lists of concrete and abstract words to compute two indexes for each word, that of concreteness and of abstractness. The single aggregate index indicative of both the word's concreteness and abstractness is computed as the function of the two respective indexes. The set of raw aggregate indexes is transformed using sigmoid transformation to increase the variance and to attain greater similarity to the psycholinguistic

data. To dynamically adjust the concreteness indexes to the context, we apply an adjustment coefficient. Post hoc analysis of the adjustment coefficient indicates that lower coefficients lead to better performance. We hypothesize that the contribution of the adjustment coefficient could be increased by expanding the scope of the context, for example, by considering one or more sentences from the left and the right contexts of the target sentence. According to our analysis, the main source of divergence between the computational and the psycholinguistic indexes lies in the different representation, or salience, of word meanings in distributional semantic models and in psycholinguistic reality. Besides, analysis of divergences between the computational and the psycholinguistic rankings prompted us a potential direction for reducing the bias in composition of the concreteness paradigm, which can be overcome by diversifying the paradigm.

References

Badryzlova, Y., 2020. Exploring Semantic Concreteness and Abstractness for Metaphor Identification and Beyond. Computational Linguistics and Intellectual Technologies 33–47.

Badryzlova, Y., 2019. Automated metaphor identification in Russian texts. National Research University Higher School of Economics, Moscow.

Badryzlova, Y., Panicheva, P., 2018. A Multi-feature Classifier for Verbal Metaphor Identification in Russian Texts, in: Conference on Artificial Intelligence and Natural Language. Springer, pp. 23–34.

Barsalou, L.W., 2008. Grounded cognition. Annu. Rev. Psychol. 59, 617–645.

Basile, V., Croce, D., Di Maro, M., Passaro, L.C., 2020. EVALITA 2020: Overview of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, in: Basile, V., Croce, D., Di Maro, M., Passaro, L.C. (Eds.), Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020). CEUR.org, Online.

Birke, J., Sarkar, A., 2006. A Clustering Approach for Nearly Unsupervised Recognition of Nonliteral Language., in: EACL.

Brysbaert, M., Warriner, A.B., Kuperman, V., 2014. Concreteness ratings for 40 thousand generally known English word lemmas. Behavior research methods 46, 904–911.

Coltheart, M., 1981. The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology Section A 33, 497–505.

Fellbaum, C., 1998. WordNet: An electronic database. MIT Press, Cambridge, MA.

Gregori, L., Montefinese, M., Radicioni, D.P., Ravelli, A.A., Varvara, R., 2020. CONcreTEXT @ Evalita2020: the Concreteness in Context Task, in: Basile, V., Croce, D., Di Maro, M., Passaro, L.C. (Eds.), Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020). CEUR.org, Online.

Kulagin, D., 2019. Opy`t sozdaniya mashinno-proveryaemoj semanticheskoj razmetki russkix sushhestvitel`ny`x [Developing computationally verifiable semantic annotation of Russian nouns]. Presented at the Annual International Conference "Dialogue," Moscow.

Kutuzov, A., Fares, M., Oepen, S., Velldal, E., 2017. Word vectors, reuse, and replicability: Towards a community repository of large-text resources, in: Proceedings of the 58th Conference on Simulation and Modelling. Linköping University Electronic Press, pp. 271–276.

Lakoff, G., Johnson, M., 1980. Metaphors we Live by, 2nd ed. The University of Chicago Press, Chicago-London.

Macmillan Dictionary, Free English Dictionary and Thesaurus [WWW Document], n.d. URL https://www.macmillandictionary.com/ (accessed 11.7.20).

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine Learning in Python Journal of Machine Learning Research.

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

Turney, P.D., Neuman, Y., Assaf, D., Cohen, Y., 2011. Literal and metaphorical sense identification through concrete and abstract context, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp. 680–690.