A Corpus-driven Estimation of Association Strength in Lexical Constructions

Grigoriy Bukia, Ekaterina Protopopova, Olga Mitrofanova
St. Petersburg State University
Saint Petersburg, Russia
gregorybookia@yandex.ru, protoev@yandex.ru, o.mitrofanova@spbu.edu

Abstract—**The paper presents a method of estimating the association strength of constructions that are not observed in the corpus. The model is flexible, computationally light and easy to implement. The core idea is to aggregate 'similar' target words and propagate their selectional preferences among others. In order to describe this idea statistically, two association measures are proposed: confusion probability measured on the observed collocates only and a final association measure derived from individual counts fo all possible 'similar' words. The paper provides quantitative analysis of the data and discussion of particular cases as well as errors.**

I. Introduction

While classical trends in linguistic studies are focused on thorough description of elementary units constituting various levels of language, contemporary research pays much attention to extraction and complex description of reproducible linguistic structures occurring in texts: words, multiword expressions, collocations, idioms, etc. The core notion of such studies is the notion of construction, and lexical construction in particular.

Following the ideas of Construction Grammar [1], [2], [3], by lexical constructions we mean complex linguistic units observed in texts and constituted by a fixed lexical item (a target word) and variable slots which attract complex elements like lemmata, morphosyntactic and semantic features.. Thereby constructions can be treated as schematic templates yielding lexicalization. We admit that a target word taken in a particular sense imposes a particular structure upon a surrounding context, this structure being explained in terms of construction classes. Thus, the principal function of constructions within a text is to fix regular co-occurrence of a target word in a given sense.

Construction Grammar claims that lexical constructions reveal the unity of form and meaning. The form is maintained by fixed elements of constructions on the one hand, and on the other hand, by selectional (morphosyntactic, lexical-semantic, propositional, etc.) restrictions imposed on the slot fillers. The meaning is generally non-compositional, and allows a wide range of variations (from free co-occurrence of lexical features to highly idiomatic units). In our study constructions are treated as multilevel structures providing compressed description of collocability of a target word in a given sense and describing word combinations in terms of lemmata/tokens as well as grammatical and lexical-semantic classes. This approach to constructions reflects a crucial idea of interrelations between various linguistic levels which helps to view linguistic expressions as multilayer entities (but not as separate projections to this or that layer in the modular theories of language).

Measuring association strength in constructions defined in such a way is a crucial natural language processing task as the selectional restrictions usually are not simply described and explained. Consider, for example, Russian words *'goryachij'* and *'zharkij'*, both translated in English as *'hot'*. They seem to be synonyms while in fact they have virtually non-overlapping sets of collocations.

Formulated generally, the task is related to extraction of predicate-argument selectional preferences, idioms or WordNet-like semantic classes. As mentioned before, the degree of association in lexical constructions is an important factor in such NLP applications as paraphrase generation for machine translation, language modelling, automated and semi-automatic dictionary acquisition, semantic role labelling, word sense disambiguation etc.

The task can be formulated as follows. Let us fix a construction constisting of two slots with given morphological description (e.g., *adjective* + *noun*) and suppose that words occurring in each slot are somehow defined. So called ”collocation strength” or association measure should be derived from text data without using any complex linguistic resources.

If a construction occurs in texts, one may say, its components can be combined. The joint frequency of a pair does not exhaustively reflect the collocation strength, i.e. the degree of word relatedness, because one of them may be quite frequent itself. For instance, the pair *'good colour'* is much more frequent than *'purple colour'*, though the value of collocation strength in the latter case should be higher.

A. Stefanowitsch and S.T. Gries [5] developed a slightly more complex statistical approach which became well-known and widely accepted. They described a methodology of measuring the collocation strength assuming the occurences of words x and y are dependent events. Several statistical tests for different construction patterns were computed using a contingency table, and Fisher's exact test proved to be the most reliable.

Association measures defined on contigency tables, however, are not applicable when a pair of words is not observed in texts. In this paper we propose an extended association measure for every relevant word pair. Roughly speaking, the main assumption is that if two words (target words) relevant for a slot collocate in texts with similar words (contexts) relevant for another slot, the probability of the first target word

ISBN 978-952-68397-0-7 (paperback), ISBN 978-952-68397-1-4 (PDF)

to be combined with the contexts of the second target word and vice versa is quite high, even when some pairs are not observed in texts. The model performance is evaluated in a set of experiments, one being proposed by the authors. The paper is structured as follows: in Section 2 a brief overview of related work is presented. Section 3 describes the model in general. In Section 4 the tunable measures are explained in detail. Section 5 provides results and discussion of experiments on the pseudo-disambiguation and ranking tasks. Finally, in Section 6 we outline the conclusions and future work.

II. Related work

The ways of measuring word association and collocability are numerous and diverse because different final tasks usually require different strategies.

The sparsity of data is a well known and longly discussed problem in statistical language modelling. Even large corpora can not provide information about all possible bigrams not to mention longer word sequences. A simple statistical model estimates the probability of word co-occurence from their corpus frequency, so that the probability of unobservable combination is always zero. Several smoothing methods were proposed to overcome this issue: Additive Smoothing, Good-Turing Estimate, Jelinek-Mercer Smoothing, Discounting, Kneser-Ney Smoothing etc. (A comprehensive survey of them can be found in [6]). More recent approaches to language modelling include those based on neural networks (introduced in [7]). One of their most important aims is to overcome data sparsity problem, therefore they are constructed in such a way that the probability of unobservable word sequence is predicted. Neural language models are in some sense similar to the distributional model discussed below, because they use more context information than classical LMs to predict next word in n-gram. It is often mentioned, neural language models tend to overfitting and have high computational complexity. Moreover, as it was said above, the probability of co-occurence does not reflect the collocation strength properly.

A significant amount of work on measuring word association is devoted to verbal subcategorization frames and involves the notion of *semantic class* to describe possible slot fillers. One popular way to assess collocation strength of unobservable constructions is to make use of additional resources such as WordNet, following P. Resnik's paper [8], to induce the set of argument semantic classes that are acceptable by the given predicate. P. Resnik proposed to use selectional association measure that indicates how a given argument is related to a given predicate taking into account its possible semantic classes. Later on, several methods to induce possible semantic classes for a given predicate were employed: tree-cut model based on minimal description length (MDL) principle [9], HMM-based transformation of WordNet hierarchy [10]. Such approaches are known to have low lexical coverage and do not always outperform simple corpus-driven methods. Therefore, more up-to-date models usually derive WordNet-like linguistic knowledge from text data.

Alternative approaches are mainly based on distributional semantics model. In [11] it was proposed to cluster possible collocates (arguments) to substitute WordNet semantic classes and get rid of using external resources. More complex approaches following this paradigm use topic modelling methods so that semantic classes and clusters are replaced by the latent variables (topics) [12]. The approach is followed in [13] being applied to adjective-noun preferences. A similarity-based model was introduced in [14] to estimate the probability of such previously unobservable word combinations using available information on ”most similar” words. Such a model was then applied to the selectional preference modelling task [15], [16]. The similarity measure was computed on syntactic and semantic vector spaces and several similarity functions and feature weighting methods were compared. The recent work [17] is concerned with deriving unseen arguments from corpora using random walk on predicate-argument bipartite graph. This model is based on the same principle as the one proposed in the paper: the more the intersection of arguments between two predicates, the higher the probability of their interchange. The approach is modular: first of all, the predicate-argument bipartite graph and its monopartite projection are constructed. Then a distance function on predicates is introduced and its values transformed into transition probabilities. The random walk model arrgegates counts for close predicates and results into smoothed preferences for unseen pairs.

The approach introduced here achieves a significant accuracy and, on the other hand, is not computationally complex, easy to implement and apply.

III. Model

A. Target word and its context

Consider two non-overlapping sets of words X and Y. The frequency of some pairs $x \in X$, $y \in Y$ is defined. The task is to construct a function $F : X \times Y \to \mathbb{R}$, which characterizes collocation strength of any pair.

For some $x \in X$ consider a set of words, collocating with x in texts $c(x) \subset Y$. Denote $[x, c(x)]$ *target word* and its *context*. The target word $y \in Y$ and its context $c(y) \subset X$ is defined similarly.

Let us define a measure on a given target word's contexts which reflects the association between target word and context and also extend the number of possible contexts.

B. Basic measure

The association measure for observable constructions will be called below a *basic measure*. Following [5] the correlation between two random variables ”the first slot is filled with x” and ”the second slot is filled with y” is estimated by means of a contingency table.

	x	$X\backslash x$	
y	a	b	
$Y\backslash y$	c	d	

where

$a \;=\; \#(xy)$is a joint frequency of xy;
$b \;=\; \#(x\bar{y})$is a frequency of x paired with other contexts;
$c \;=\; \#(\bar{x}y)$is a frequency of y paired with other contexts;
$d \;=\; \#(\bar{x}\bar{y})$is the number of other pairs.

C. Measures of association

The degree of association is usually assessed by a number of coefficients $f(x, y)$.

a) Association coefficient: is computed as follows:

$$Q = \frac{|ad - bc|}{ad + bc}$$

It varies from 0 (corresponding to no assocation between variables) to 1 (for complete association).

b) Yule's coefficient of colligation: Is calculated as follows:

$$K = \frac{|\sqrt{ad} - \sqrt{bc}|}{\sqrt{ad} + \sqrt{bc}}$$

K and Q are related so that

$$Q = \frac{2K}{1 + 2K}$$

c) χ^2 contingency coefficient: is calculated as follows:

$$\chi^2 = \frac{n(|ad - bc| - \frac{n}{2})^2}{(a + b)(a + c)(b + d)(c + d)}$$

The values have a minimum of 0 (in case of no association) and increases.

d) Fisher's exact test: is computed as follows

$$p = \frac{(a + b)!(c + d)!(a + c)!(b + d)!}{(a + b + c + d)!} \times \sum_{j=0}^{a} \frac{1}{(a + b - j)!(a + c - j)!(a + d - j)!}.$$

The significance level of correlation equals to $1 - p$ so the degree of association can be defined as $f(x, y) := 1 - p$.

e) z–score: is calcualted as follows

$$z = \frac{a - b + \frac{(a+c-b-d)(a+b)}{a+b+c+d}}{\sqrt{a + b}}$$

f) G–test:

$$G = 2\Big\{(a + \frac{1}{2})\log(a + \frac{1}{2}) + (b - \frac{1}{2})\log(b - \frac{1}{2}) + (c + \frac{1}{2})\log(c + \frac{1}{2}) + (d - \frac{1}{2})\log(d - \frac{1}{2}) - (a + b)\log(a + b) - (c + d)\log(c + d) - (a + c)\log(a + c) - (b + d)\log(b + d) + (a + b + c + d)\log(a + b + c + d)\Big\}$$

g) Mutual information: $MI = \frac{a(a+b+c+d)}{(a+b)(a+c)}$

D. Target words confusion probability

In order to extend the number of possible contexts we introduce the notion of target words confusion probability (following [18]). Below the proposed approach to the confusion probability is described. Let us call two target words *confusable* and write down $x_1 \sim x_2$, if the first target word can be combined with the second one's contexts and vice versa. The posterior estimate of this value can be calculated. If $y_1 \in c(x_1)$ is a random context of target word x and $y_2 \in c(x_2)$ is a random context of target word y, then

$$\mathbb{P}\{x_1 \sim x_2\} = \mathbb{P}\{y_1 \in c(x_2) | y_1 \in c(x_1)\} \times \mathbb{P}\{y_2 \in c(y_1) | y_2 \in c(x_2)\}.$$

This value is derived from frequency data. According to Bayes' rule,

$$\mathbb{P}\{x_1 \sim x_2\} = \frac{|c(x_1) \bigcap c(x_2)|^2}{|c(x_1)||c(x_2)|}.$$

To reduce the estimate variance, it is assumed that besides the real contexts there are some fake contexts in a corpus adding the number of real contexts to 10 if needed. Thus, if the word is infrequent, the confusion probability is proportionally reduced. The necessity of such an addition was proved empirically.

Moreover, let $\mathbb{P}\{x_1 \sim x_1\} = 1$. Thus, the confusion probability is computed as follows

$$g(x_i, x_j) = \frac{|c(x_i) \bigcap c(x_j)|^2}{\max(|c(x_i)|, 10)\max(|c(x_j)|, 10)}$$

E. Measuring association

Now we can propose the association measure $F(x, y)$ for a pair of words $x \in X$, $y \in Y$. The first estimate $f(x, y)$ is the value of basic measure. Its weight is set to 1. Let us consider $f(x, y_i)$ as an estimate of $F(x, y)$ for a word $y_i \in c(x)$ from the contexts of x. This estimate is relevant when $y_i \sim y$, so its weight is set to $g(y_i, y)$ — the confusion probability of y_i and y. All words from $c(y)$ are treated similarly. Thus, three groups of estimates are obtained:

- $f(x, y)$ with weight 1;
- $f(x, y_i)$ for all $y \in c(x)$ weighted by $g(y, y_i)$;
- $f(x_i, y)$ for all $x \in c(y)$ weighted by $g(x, x_i)$;

The final measure is calculated as a weighted average:

$$F(x, y) = \frac{f(x, y) + \sum_{y_i \in c(x)} f(x, y_i)g(y_i, y) + \sum_{x_i \in c(y)} f(x_i, y)g(x_i, x)}{1 + \sum_{y_i \in c(x)} g(y_i, y) + \sum_{x_i \in c(y)} g(x_i, x)}.$$

IV. Experimental setup and results

An experiment on pseudo-disambiguation was conducted to evaluate the model performance. An error analysis is then presented. The possibility of extending the number of contexts is also tested. Finally, the confusion probability measure is also discussed.

TABLE I. PSEUDO-DISAMBIGUATION RESULTS (MEAN AVERAGE) ON NOUN CONTEXTS

$\mathbb{E}$	$S(X)$	$S(Y_1)$	$S(Y_2)$	$S(Y_3)$
Q	74,4	63,4	75,8	68,9
K	74,6	63,4	75,8	68,9
MI	**74,8**	**64,6**	**77.2**	**70.7**
χ^2	67,6	60,6	70,6	62.4
p	74,4	63,4	75,8	68,9
z	41,6	39,6	48,0	33,5
G	72,6	62,0	73,6	66,1

A. Data

We use a corpus of Russian fiction (350K sentences obtained from M.Moshkov's digital library, URL: lib.ru, later referred as corpus A). All preprocessing (tokenization, lemmatization, shallow morphological analysis) was performed by means of PyMorphy2 Python library (URL: http://pymorphy2.readthedocs.org/en/latest/). About 157K (80K unique) adjective-noun pairs was extracted from these texts. The last experiment involves an additional corpus containing 11M sentences (referred as corpus B).

B. Pseudo-disambiguation

The following lemmata registered in corpus A more than 5 times were extracted:

- 100 random nouns $N = \{n_i\}$;
- random adjectives for each noun n $A = \{a_i\}$;
- adjectives with the nearest frequency count for each a $X = \{x_i\}$
- 100 random adjectives $Y = \{y_i\}$

All pairs (a_i, n_i) are then removed from training data. The analyser processes two candicate contexts for each noun and the task is to predict the pair which was removed (by scoring it higher). Each basic measure is evaluated using the following metrics:

1) Number of times a_i was chosen against x_i paired with n_i
$$S(X) = \#(F(a_i, n_i) > F(a_i, x_i)),$$
2) Number of times a_i was chosen against y_i paired with n_i
$$S(Y_1) = \#(F(a_i, n_i) > F(a_i, y_i)),$$
3) The value $S(Y_1)$ is corrected manually, when a pair with random context (y_i, a_i) is in fact associated stronger than (a_i, n_i), resulting in $S(Y_2)$;
4) If pair (y_i, a_i) was corrected, it is then removed, and the score $S(Y_3)$ was calculated similarly.

The performance was evaluated using 5-fold cross-validation. The average results and the standard deviation are presented in tables I and II.

A similar experiment was conducted for target adjectives. We extract

- 100 random adjectives $A = \{a_i\}$;
- random noun contexts for each a $N = \{n_i\}$;

TABLE II. PSEUDO-DISAMBIGUATION RESULTS (STANDARD DEVIATION) ON NOUN CONTEXTS

σ	$S(X)$	$S(Y_1)$	$S(Y_2)$	$S(Y_3)$
Q	3,2	3,5	**1,4**	**2,0**
K	3,0	3,7	1,7	2,6
MI	2,6	3,4	1,7	2,6
χ^2	**1,9**	**1,7**	3,3	3,5
p	3,6	3,0	2,0	2,5
z	5,5	5,3	6,3	5,1
G	0,8	2,6	1,6	2,4

TABLE III. PSEUDO-DISAMBIGUATION RESULTS (MEAN AVERAGE) ON ADJECTIVE CONTEXTS

$\mathbb{E}$	$S(X)$	$S(Y_1)$	$S(Y_2)$	$S(Y_3)$
Q	76,8	63,4	66,2	64,0
K	**77,4**	64,8	67,6	65,5
MI	77,2	65,4	68,2	66,2
χ^2	70,2	63,6	66,2	64,0
p	76,2	62,8	65,6	63,4
z	25,8	38,2	42,0	38,1
G	75,6	**66,6**	**69,4**	**67,4**

- nouns with the nearest frequency count for each n $X = \{x_i\}$
- 100 random nouns $Y = \{y_i\}$

The results are presented in tables III, IV.

Two main error types are observed.

Firstly, when pair (a_i, n_i) is removed, the confusion probability of a_i and n_i, as well as of n_i and a_i, is equal to zero. This is the case when (a_i, n_i) has a figurative meaning, e.g., *hodyachij katehizis*, *dremucheje ravnovesije*, *dikoje zverstvo* (word-by-word translation: *walking cathehism*, *primeval balance*, *wild atrocity*).

Another error class contains rare co-occurences, which have low statistical significance in text data. The examples are *pylkoje begstvo*, *tonkaja vospriimchivost'*, *suhaja konvulsija* (*passionate escape*, *fine sensitivity*, *dry convulsion*).

C. Ranking new contexts

Another experiment aims to assess the ability to model new possible co-occurences tested then on corpus B. The idea is the following:

1) Two frequent nouns and adjectives are considered: *dom 'a house'*, *chelovek 'a man'*, *krasnyj 'red'*, *krasivyj 'beautiful'*.
2) All possible pairs of target words not observed in training data are ranked according to their association strength scores. Thus, new possible constructions are obtained.
3) If a pair is seen in corpus B, is is marked as possible.

TABLE IV. PSEUDO-DISAMBIGUATION RESULTS (STANDARD DEVIATION) ON ADJECTIVE CONTEXTS

σ	$S(X)$	$S(Y_1)$	$S(Y_2)$	$S(Y_3)$
Q	3,3	3,9	4,4	4,1
K	3,6	4,0	4,3	4,1
MI	3,6	3,4	4,2	3,9
χ^2	3,9	**1,7**	2,4	**2,1**
p	3,7	4,7	5,2	4,9
z	**2,9**	3,1	**2,0**	3,0
G	4,9	2,6	3,5	3,3

TABLE V. NUMBER OF POSSIBLE PAIRS AMONG 100 TOP ONES.

100	Q	K	MI	χ^2	p	z	G
dom	**83**	79	80	16	80	3	53
chelovek	89	84	91	24	**94**	10	78
krasnyj	79	78	**80**	43	77	3	77
krasivyj	71	69	70	29	**73**	1	65

TABLE VI. NUMBER OF POSSIBLE PAIRS AMONG 500 TOP ONES.

500	Q	K	MI	χ^2	p	z	G
dom	57	58	**59**	19	55	2	51
chelovek	73	73	**74**	27	73	8	70
krasnyj	57	58	59	49	**61**	3	**61**
krasivyj	51	51	**52**	33	51	2	49

Tables V and VI present information about the number of possible (seen in corpus B) pairs ranked among 100 and 500 pairs with the highest association score.

The distribution of words (seen and unseen in corpus B) in top 500 contexts is illustrated by Fig. 1, 2, 3, 4. It should be noticed, that not all really possible constructions marked as possible due to their occurence in corpus B. Some of the contexts in top 100, for examples, may not be collocable with target word, but represent a part of its meaning: consider for example, adjectives *milovidnyj* (*nice*) and *horoshenkij* (*pretty*), which are connected to target word *chelovek* (*a man*) through its hyponym *devushka* (*young woman*). This connection may be useful for word sense disambguation.

D. Word confusion

As it was mentioned in [18], the most possible confusable words represent syntagmatic relations: synonymy, hyponymy, meronymy, association. Tables VII, VIII, IX, X contain pairs with the highest confusion probability for the words from previous subsection.

V. CONCLUSION

We have presented an approach to association strength estimation in lexical constructions. The model contains two levels: first of all, a confusion probability measure is defined based on observed word pairs, then an association measure is

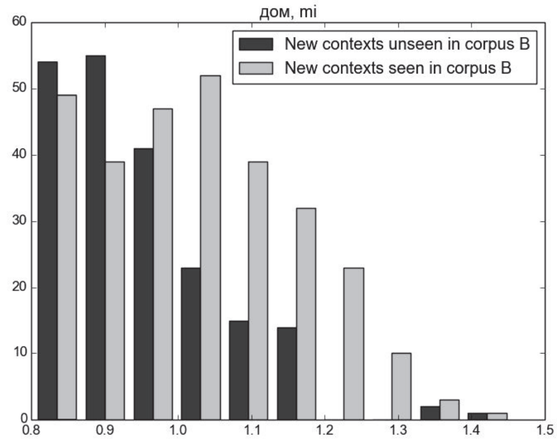

Fig. 1. The distribution of contexts for ‘dom’ (a house)

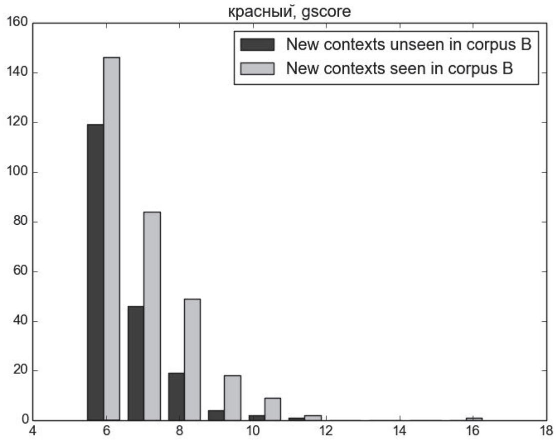

Fig. 2. The distribution of contexts for ‘krasivyj’ (beautiful)

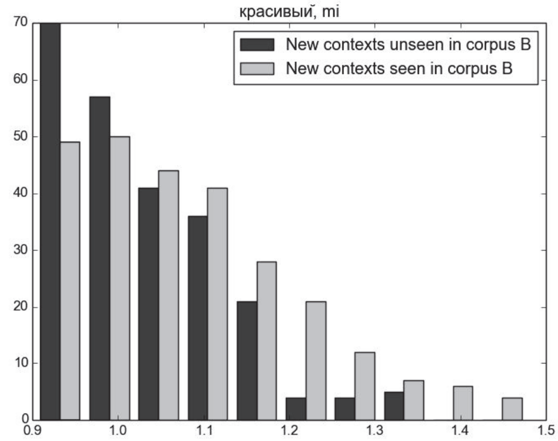

Fig. 3. The distribution of contexts for ‘krasnyj’ (red)

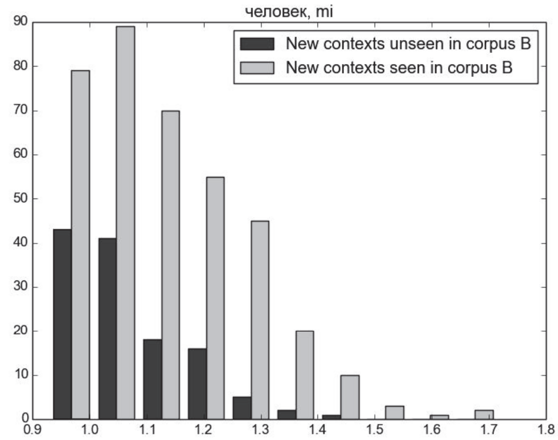

Fig. 4. The distribution of contexts for ‘chelovek’ (a human)

TABLE VII. MOST POSSIBLE SUBSTITUTES FOR WORD 'HUMAN'

chelovek 'a human'	*g*	relation
woman	0.124	hyponymy
man	0.073	hyponymy
place	0.072	association
face	0.070	meronymy
thing	0.069	association
girl	0.069	hyponymy
guy	0.066	hyponymy
bride	0.065	hyponymy
life	0.062	association
hand	0.054	meronymy
look	0.053	association

TABLE VIII. MOST POSSIBLE SUBSTITUTES FOR WORD 'HOUSE'

dom 'house'	*g*	relation
building	0.095	synonymy
flat	0.079	meronymy
room	0.056	meronymy
library	0.055	hyponymy
cottage	0.054	hyponymy
hotel	0.049	hyponymy
family	0.048	association
town	0.046	association
wall	0.045	meronymy
village	0.044	association

derived from it as a weighted average of all possible substitutes of a target word and its context.

The proposed score is based on one of known association measures for contingency tables. We have evaluated our model using several of them and noticed that the a little variance is observed in their results. However, some of them (especially mutual information) tend to perform better in different tasks.

A confusion probability measure used in the paper seems to satisfy the requirements of the given task: it extracts several syntagmatic relations simultaneously, providing the possibility of making more complex predictions about target word collocability. It may be found useful in applications like metaphor processing when is is important to capture the paradigmatic relations transferred by association.

The work may be extended in several ways. First of all, a thorough evaluation and error analysis should be conducted on different construction classes and levels. Moreover, a confusion probability measure is now very simple and may be improved.

TABLE IX. MOST POSSIBLE SUBSTITUTES FOR WORD 'RED'

red	*g*	relation
blue	0.070	cohyponymy
yellow	0.063	cohyponymy
white	0.055	cohyponymy
bright	0.047	association
pink	0.041	cohyponymy
pale	0.040	association
gold	0.036	cohyponymy
green	0.036	cohyponymy
brown	0.033	cohyponymy
dim	0.028	association

TABLE X. MOST POSSIBLE SUBSTITUTES FOR WORD 'BEAUTIFUL'

krasivyj 'beautiful'	*g*	relation
brown	0.064	association
excellent	0.063	synonymy
thin	0.047	association
swarthy	0.046	association
fragile	0.045	association
strange	0.040	–
naked	0.038	association
tired	0.038	–
foreign	0.037	–
pretty	0.035	synonymy

ACKNOWLEDGMENT

The research project is partly supported by St. Petersburg State University grant "Quantitative linguistic parameters for definition of stylistic properties and subject area of texts".

REFERENCES

[1] Ch.J. Fillmore. "The Mechanisms of Construction Grammar" In *Proceedings of the Berkeley Linguistic Society*. Vol. 14. 1988.

[2] A.E. Goldberg. "Constructions. A Construction Grammar Approach to Argument Structure". Chicago, IL/London: University of Chicago Press, 1995.

[3] M. Tomasello. "Constructing a Language: A Usage-Based Approach to Child Language Acquisition". Cambridge, MA: Harvard University Press, 2003.

[4] Ju.D.Apresjan (ed.). Prospekt aktivnogo slovarya russkogo jazyka (The prospect of active Russian dictionary). Moscow, 2010.

[5] A. Stefanowitsch and S. T. Gries, "Collostructions: Investigating the interaction of words and constructions," *International journal of corpus linguistics*, vol. 8, no. 2, pp. 209243, 2003.

[6] Stanley F. Chen and Joshua Goodman, "An Empirical Study of Smoothing Techniques for Language Modeling", Technical Report TR-10-98, Computer Science Group, Harvard University, 1998.

[7] Y. Bengio, R. Ducharme, P. Vincent and Ch. Jauvin, "A Neural Probabilistic Language Model", *Journal of Machine Learning Research*, V. 3, pp. 11371155, 2003.

[8] P. Resnik, "Selectional preference and sense disambiguation," in Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How, 1997, pp. 5257.

[9] H. Li and N. Abe "Generalizing case frames using a thesaurus and the MDL principle," *Computational Linguistics*, 24 (2), 2001. pp.217-244.

[10] S. Abney and M. Light "Hiding a semantic class hierarchy in a Markov model". In *Proceedings of the ACL workshop on Unsupervised Learning in NLP*, 1999. pp.1-8.

[11] M. Rooth, S. Riezler, D. Prescher, G. Carroll, and F. Beil, "Inducing a semantically annotated lexicon via EM-based clustering," in Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, 1999, pp. 104111.

[12] Diarmuid O. Seaghdha, "Latent variable models of selectional preference". In *ACL*, Association for Computational Linguistics, 2010.

[13] K. M. Hermann, P. Blunsom, C. Dyer, and S. Pulman, "Learning semantics and selectional preference of adjective-noun pairs" in *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, 2012, pp. 7074.

[14] I. Dagan, L. Lee, and F. C. Pereira, "Similarity-based models of word cooccurrence probabilities," Machine Learning, vol. 34, no. 13, pp. 4369, 1999.

[15] K. Erk, S. Pad, and U. Pad, "A flexible, corpus-driven model of regular and inverse selectional preferences," *Computational Linguistics*, vol. 36, no. 4, pp. 723763, 2010.

[16] V. Pekar. "Distributivnaja model sochetaemostnyh ogranichenij glagolov (A distributional model of verbal selectional restrictions)". *Computational Linguistics and Intellectual Technologies.* International Conference (Dialog2004). Moscow, 2004.

[17] Z. Tian, H. Xiang, Z. Liu, and Q. Zheng, "A Random Walk Approach to Selectional Preferences Based on Preference Ranking and Propagation.," in *ACL*, 2013, pp. 11691179.

[18] J. Weeds, D. Weir, and D. McCarthy, "Characterising measures of lexical distributional similarity," in Proceedings of the 20th international conference on Computational Linguistics, 2004, p. 1015.