

# Language-Independent Prediction of Psycholinguistic Properties of Words

Yo Ehara

y-ehara@aist.go.jp <http://yoebara.com/>

Artificial Intelligence Research Center (AIRC),

National Institute of Advanced Industrial Science and Technology (AIST)

2-4-7 Aomi, Koto-ku, Tokyo, Japan

## Abstract

The psycholinguistic properties of words, namely, word familiarity, age of acquisition, concreteness, and imagery, have been reported to be effective for educational natural language-processing tasks. Previous studies on predicting the values of these properties rely on language-dependent features. This paper is the first to propose a practical language-independent method for predicting such values by using only a large raw corpus in a language. Through experiments, our method successfully predicted the values of these properties in two languages. The results for English were competitive with the reported accuracy achieved using features specific to English.

## 1 Introduction

The psycholinguistic properties of words, namely, word familiarity, age of acquisition, concreteness, and imagery, are measured real values of human responses in cognitive experiments in which participants are presented with the written or spoken form of words (Coltheart, 1981). They are not only important for psycholinguistics but for natural language processing (NLP) because they are effective features for educational applications such as lexical simplification (Jauhar and Specia, 2012). In spite of their importance, dictionaries describing them are rare and small. To enlarge these dictionaries, previous methods have been used to predict the values of these properties using supervision from a small dictionary and features from other language resources. The predicted values can be further used as features for other NLP tasks and provide excellent results (Mohler et al., 2014; Köper and Im Walde, 2016; Paetzold and

Specia, 2016).

However, all previous studies relied on language-specific features; thus, their methods cannot be applied to other languages. When predicting the psycholinguistic properties, considering the word *domains* is quite effective. For example, “bread” and “onion” have high concreteness values of 622 and 632, respectively, while those of “economy” and “finance” are low, i.e., 284 and 371, respectively<sup>1</sup>. Evidently, capturing word domains, such as “food” and “economics”, is effective for roughly predicting the range of values. To capture word domains, previous studies used combinations of semantic features, such as WordNet (Miller, 1995), and word frequencies from corpora in various domains. Since both are language-specific, previously proposed methods are language-dependent.

In this study, we propose a simple but practical language-independent method for predicting the psycholinguistic properties of words. It involves using only a large raw corpus of the target language. Our key idea is two-fold. First, instead of using the combination of semantic features and word frequencies, we first decompose the raw corpus by using latent Dirichlet allocation (LDA) and use the probability of words given each topic to capture the word domains. Second, we apply a multi-task Gaussian process regression (GPR), which enables the joint prediction of these properties. This captures the relations among the properties and can improve predictive accuracy. Our experimental results are competitive with those in which language-dependent features are used.

Our method is also useful with linear models for analyzing the obtained prediction models. The

---

<sup>1</sup>The values are taken from (Coltheart, 1981), which encodes all properties within fixed ranges. The larger values indicate the more concrete, or physical, what the word signifies is.

weight of each topic indicates how well the obtained prediction models capture domains. This characteristic is useful for error analysis and further improving the prediction models.

## 2 Task Setting

### 2.1 Dataset

First, we briefly introduce the available psycholinguistic databases. The Machine Readable Dictionary (MRC) psycholinguistic database (Coltheart, 1981) is one of the largest for English and also used in psycholinguistic social studies (Schwartz et al., 2013). The 27 psycholinguistic properties of words in the database also contain easily obtainable lexical properties<sup>2</sup>. By excluding these properties, 4 of the 27 properties are considered important for NLP applications: familiarity, age of acquisition, concreteness, and imagery. Each property is available for a different set of vocabulary. Familiarity is the frequency with which a word is seen, heard, or used daily and available for 9,392 words. Age of acquisition is the age at which a word is learned and available for 3,503 words. Concreteness is the degree of how palpable the object to which the word refers is and available for 8,228 words. Imagery is the intensity with which a word arouses images and available for 9,240 words.

These properties are measured through questionnaires given to adult native speakers of the language. For Japanese, we can use a word familiarity and imagery database for Japanese (Amano and Kondo, 1998).

### 2.2 Formalization

Let  $T$  be the number of psycholinguistic properties, e.g., the MRC database has  $T = 4$  properties. Let  $V$  be the set of all the vocabulary. We have supervision for some words in  $V$ . Let  $S \subset V$  be the set of words with supervision. Then, we have training data  $\mathcal{D} = \{(\mathbf{y}_v, \mathbf{x}_v) | v \in S\}$ , in which  $\mathbf{y}_v$  is a  $T$ -dimensional vector filled with the values of the  $T$  properties of word  $v$ , and  $\mathbf{x}$  is a  $K$ -dimensional feature vector where  $K$  is the number of features. Then, the goal of the task is, given new word  $v' \in V \setminus S$ , to predict the vector of its properties, namely  $\mathbf{y}_{v'}$ .

<sup>2</sup>The full list of the 27 properties can be found in [http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa\\_mrc.htm](http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm)

For  $\mathbf{x}$ , the choice of features to use has been extensively studied for predicting familiarity. Tanaka-Ishii and Terada (2011) investigated the relation between corpus frequency and familiarity and found that high correlation can be achieved using the logarithm of frequencies of various corpora because each corpus is focused on different domains. Unlike their study, we have only one large raw corpus for each language.

## 3 Proposed Method

As mentioned above, the key idea of our method is two-fold. First, we use LDA (Blei et al., 2003) to calculate the  $p(\text{word}|\text{topic})$  probability from a large raw corpus. The number of topics  $K$  is a hyper-parameter of LDA. This probability can be regarded as (and used as) a substitute of word frequencies from various corpora. Although we have only one raw corpus, LDA enables us to use  $K$  probabilities. These enriched features enable us to effectively capture domains of words.

Second, we use multi-task GPR (Bonilla et al., 2008) with which we can predict  $\mathbf{y}$  jointly. Previous studies built a predictor for each element of  $\mathbf{y}$ , i.e., each property, independently. Joint prediction can capture the relations among the psycholinguistic properties. This enables us to take the values of easy-to-predict properties into account when predicting the values of difficult-to-predict properties. Thus, joint prediction can boost predictive accuracy.

## 4 Experiments

We conducted experiments on English and Japanese. The proposed method requires only one large raw corpus for each language. Wikipedia (Wiki) can be used for this thanks to its availability in many languages. For comparison, we used general corpora, i.e., British National Corpus (BNC) by The BNC Consortium (2007) for English, and BCCWJ (Maekawa, 2007) for Japanese. In each language, we extracted the top 100,000 words in frequency on Wikipedia and ignored other words in the experiment using **gensim**<sup>3</sup>. For Japanese word segmentation and lemmatization, we used (Kudo, 2005).

We used the datasets described in §2 as the psycholinguistic database. For each property from the word set of these candidates, we chose words that

<sup>3</sup><https://radimrehurek.com/gensim/wiki.html>

Model	Feature set	Familiarity		Age of Acquisition		Concreteness		Imagery	
-	-	$\rho$	$r$	$\rho$	$r$	$\rho$	$r$	$\rho$	$r$
FREQ(Wiki)		0.681	0.667	0.391	0.412	0.041	0.049	0.215	0.244
SVR-RBF	LDA(Wiki)	0.814	0.804	0.750	0.754	0.766	0.760	0.737	0.726
SVR-RBF	w2v(Wiki)	0.692	0.659	0.562	0.563	0.819	0.818	0.693	0.689
SVR-RBF	All	0.838	0.821	0.774	0.776	0.823	0.820	0.762	0.752
Ridge	LDA(Wiki)	0.836	0.820	0.763	0.759	0.773	0.766	0.741	0.722
Ridge	w2v(Wiki)	0.660	0.635	0.550	0.547	0.833	0.830	0.702	0.697
Ridge	All	0.849	0.823	0.770	0.772	<b>0.843</b>	<b>0.837</b>	0.767	0.757
GPR-RBF	LDA(Wiki)	0.829	0.820	0.766	0.764	0.777	0.771	0.747	0.735
GPR-RBF	w2v(Wiki)	0.683	0.653	0.557	0.555	0.827	0.826	0.692	0.687
GPR-RBF	All	0.845	0.829	0.781	0.782	0.819	0.818	0.769	0.759
JGPR-RBF	All	<b>0.854</b>	<b>0.838</b>	<b>0.793</b>	<b>0.789</b>	0.818	0.814	<b>0.772</b>	<b>0.762</b>
FREQ(BNC)		0.777	0.749	0.339	0.365	0.062	0.062	0.045	0.050
SVR-RBF	LDA(BNC)	0.860	0.840	0.754	0.767	0.610	0.602	0.648	0.646
SVR-RBF	w2v(BNC)	0.697	0.683	0.641	0.631	0.858	0.857	0.796	0.786
SVR-RBF	All	<b>0.874</b>	<b>0.860</b>	0.807	0.809	0.862	0.859	0.817	0.807
GPR-RBF	LDA(BNC)	0.855	0.836	0.757	0.770	0.601	0.589	0.648	0.646
GPR-RBF	w2v(BNC)	0.698	0.687	0.657	0.650	0.850	0.846	0.785	0.777
GPR-RBF	All	0.869	0.856	0.824	0.825	0.871	<b>0.866</b>	0.826	0.816
JGPR-RBF	All	0.867	0.852	<b>0.833</b>	<b>0.839</b>	<b>0.871</b>	0.865	<b>0.831</b>	<b>0.820</b>
(Paetzold and Specia, 2016)		0.863	0.846	0.871	0.862	0.876	0.869	0.835	0.823

Table 1: Prediction Results of English. Larger values imply better predictive accuracy.

completely matched the spelling of those that appear in the database and used these words as the vocabulary set. As a result, we obtained  $|V| = 1,842$  words for the MRC database<sup>4</sup>.

From the 1,842 words, we took 500 for the test data. We used the other 1,342 words for training and development, over which the parameters of methods are tuned using 5-fold cross validation.

We compared the following feature sets. **FREQ(corpus name)** is the log of word frequency in the *corpus name*, and **LDA(corpus name)** is the log of word probability given each topic calculated by applying LDA to *corpus name*. We used the **gensim** implementation and fixed the number of topics to 150 for both English and Japanese. For all 150 topics, we calculated  $\log p(\text{word}|\text{topic})$  and used all the 150 log-probabilities as features. **w2v(corpus name)** are word-embedding features obtained using the Word2Vec toolkit (Mikolov et al., 2013) trained on *corpus name*. We used the Word2Vec setting for

<sup>4</sup>The number of these target words was lower than that given in Paetzold and Specia (2016) because 1) we only used the words that had all four properties, and 2) many words that share the same spelling are doubly registered for verbs and nouns in the MRC database.

each property according to p. 438 of Paetzold and Specia (2016). **All** is the concatenated features of FREQ, LDA, and w2v.

We compared the following regression models<sup>5</sup>. Two are linear models: support vector regression (SVR) (Smola and Vapnik, 1997) with a linear kernel and Ridge regression (Tikhonov, 1963), denoted as **Ridge**, a linear regression with penalties (regularization) added to keep parameters from taking extreme values. They have a weight for each feature; thus, each feature’s importance can be obtained from its weight. In contrast, methods using radial-basis function (RBF) kernels do not provide weight vectors, via which we cannot obtain each feature’s importance. However, we used **SVR-RBF**, SVR with a radial-basis function (RBF) kernel, **GPR-RBF**, GPR with an RBF kernel, and **JGPR-RBF**, GPR with an RBF kernel and joint prediction (§3) since these models can take into account combinations of features using the RBF kernels, which are useful for combining

<sup>5</sup>The results of SVR with a linear kernel and Ridge in BNC were lower than the other models and were omitted from Table 1 due to space limitations. We used scikit-learn (<http://scikit-learn.org/>) to implement all models and will release them after acceptance.

both domain and semantic features.

#### 4.1 Quantitative Results

For evaluation measures, as done by Paetzold and Specia (2016), we used Pearson’s correlation coefficient ( $r$ ) and Spearman’s rank correlation coefficient ( $\rho$ ) between the predicted and target properties of a word in the test set. Intuitively, the former shows accurateness in predicting the values of the target property, and the latter shows that of predicting the ranking of that property.

The experimental results are listed in Table 1. The bold values are the largest in each column in a section. When predicting familiarity and age of acquisition, we can see that **LDA** consistently outperformed **w2v**. This suggests that domains are more informative than semantic features for predicting these two properties. In contrast, when predicting concreteness and imagery, **w2v** performed better than **LDA** except when predicting imagery with **Wiki** features. This suggests that semantics is more important than domains for predicting these. This matches our intuition of psycholinguistic properties because familiarity and age of acquisition mainly reflect the difficulty of words, while concreteness and imagery have little to do with difficulty and more to do with the semantic aspects of words.

We can also see that **BNC** roughly performed better than **Wiki**. This shows that BNC, a general corpus, is better for predicting psycholinguistic properties than Wikipedia. One possible explanation for this phenomena could be that BNC is manually tuned to be general and to include typical usage of the language, while Wikipedia, a collection of user content, is noisy.

**All** performed consistently better than **LDA** and **w2v**. Thanks to joint prediction, **JGPR-RBF** performed better than **GPR-RBF** in almost all cases and performed the best out of the all models in most cases, especially when used for **Wiki**. This suggests that joint prediction is robust against the noise in Wikipedia. At the bottom of Table 1, we cite the results by Paetzold and Specia (2016)<sup>6</sup>,

<sup>6</sup>Their results are not directly comparable to ours because their test set used in the experiments was not known; thus, different from ours. We are also interested in the difference of verbs and nouns with the same spelling. We also re-implemented and applied their bootstrapping method to our language-independent features. This re-implementation slightly (below 0.01) outperformed the Ridge regression, on which their method is based, but performed worse than **JGPR-RBF**.

Model	Feature set	Familiarity		Imagery	
		$\rho$	$r$	$\rho$	$r$
FREQ(Wiki)		0.225	0.224	0.119	0.076
GPR-RBF	LDA(Wiki)	0.512	0.492	0.554	0.643
GPR-RBF	w2v(Wiki)	0.576	0.576	0.600	0.682
GPR-RBF	All	0.607	0.610	0.650	0.727
JGPR-RBF	All	<b>0.609</b>	<b>0.612</b>	<b>0.650</b>	<b>0.727</b>
FREQ(BCCWJ)		0.272	0.234	0.187	0.120
GPR-RBF	LDA(BCCWJ)	0.477	0.485	0.517	0.539
GPR-RBF	w2v(BCCWJ)	0.592	0.608	0.624	0.700
GPR-RBF	All	0.620	0.623	<b>0.662</b>	0.727
JGPR-RBF	All	<b>0.626</b>	<b>0.624</b>	0.659	<b>0.727</b>

Table 2: Prediction Results for Japanese

Weight	Top words	Interpretation
2.387	the, he, and, in, his, to, of, was, as, at	General topic
1.851	food, rice, restaurant, fruit, sugar, beer, milk, meat, tea	Food
0.919	john, st, william, sir, american, de, thomas, bishop, henry, charles	Peoplefs names

Table 3: Top 3 highly weighted topics

who used language-dependent features. Our results were competitive with theirs, although our method uses features obtained only from the raw corpus, i.e., language independent.

##### 4.1.1 Prediction Results for Japanese

In Japanese, only familiarity and imagery are available (Amano and Kondo, 1998). The number of words whose familiarity and imagery were annotated was 2,475. Among those, we used 2,030 words for training and development. A disjoint set of 445 words were used for test.

For simplicity, we show the results of the best two models, **GPR-RBF** and **JGPR-RBF** in Table 2. We can first see that frequencies of Japanese corpora have lower correlation values with familiarity and imagery in Japanese compared with English. This implies that, overall, in this experimental setting, Japanese psycholinguistic properties were more difficult to predict than those of English. We discuss this reason in §5. Similar to English, **All** consistently performed the best in each corpus, and the general corpus (**BCCWJ**) performed better than Wikipedia. We can also see that **JGPR-RBF** outperformed **GPR-RBF** in almost all cases, presumably thanks to joint prediction.



## 4.2 Qualitative Results

Table 3 lists the top 3 topics highly weighted using **Ridge** with **LDA(Wiki)** in Table 1 and words in the topics when predicting word familiarity. The most weighted topics are called “general” topics and contain words that appear in most of the documents in the dataset. Since words that appear in every document tend to have high frequency, this result is also consistent with that by [Tanaka-Ishii and Terada \(2011\)](#), in which word familiarity roughly correlates with word frequencies.

The next weighted topics are those related to food and people’s names. This also matches our intuition of “familiarity” because we use these words in daily life, and they usually do not have negative connotations. In Table 3, stop words are omitted from these topics’ top-word lists except for the general topic.

## 5 Discussion

Frequency is a good estimator for difficulty-related properties, namely familiarity and age of acquisition. Specifically, familiarity was previously reported to be one ([Tanaka-Ishii and Terada, 2011](#)). Since  $p(\text{word}|\text{topic})$  is the frequency of words in the topic except for the normalizing constant, it can naturally be a good estimator for the properties for which frequency is a good estimator. A corpus is a collection of documents in various domains, and the proportion of the domains of the collected documents varies corpus to corpus. By directly using  $p(\text{word}|\text{topic})$  as features, we can adjust the proportion of the domains of the given corpus to the proportion to which the target psycholinguistic property tends to correlate. Also,  $p(\text{word}|\text{topic})$  is practically easy to use: preparing 150 different corpora to use their frequencies is impractical, whereas preparing 150 different  $p(\text{word}|\text{topic})$  probabilities is easy.

Removing the stop words before applying LDA would be appropriate if we do not need to predict psycholinguistic properties for stop words. However, both English and Japanese psycholinguistic databases include the properties for words usually considered as stop words. Thus, we included these words when we ran LDA so that we could obtain  $p(\text{word}|\text{topic})$  for stop words.

English and Japanese correlation values differ greatly. This difference may be explained by the difference in the original psychological experimental settings or the difference in writing sys-

tems for English and Japanese. We focused on predicting word properties when participants respond for written language. The Japanese writing system involves Chinese characters, in which many characters are ideographic. This may have resulted in the difference.

Our experimental results shown in Table 1 and Table 2 indicate the predictive performance of each model under a fixed training-data size. We also conducted experiments on smaller training-data sizes, for example, half the experiments in §4. Overall, the **JGPR-RBF** produced the best or competitive results when compared to other models for smaller sizes as well. For example, with half the training size, **JGPR-RBF** performed the best among the models listed in Table 1 for familiarity and age of acquisition and both types of correlation coefficients.

## 6 Conclusion

We proposed a language-independent method for predicting the psycholinguistic-property values of words. It involves using only a large raw corpus for a language. To predict these properties, capturing the word domains is important. We capture them with word probability given each topic obtained by applying LDA to the raw corpus. Jointly predicting multiple properties also leads to better prediction. Experiments showed that our method improves predictive performance by joint prediction and is competitive when language-dependent features are used. When used with linear models, our method provided interesting insights between word familiarity and daily life, which can be used for further error analysis.

Predicting psycholinguistic properties of words has broad application: other than lexical simplification, which we mainly focused on, as mentioned in §1, we can use word familiarity and age of acquisition as features indicating word difficulty. Such features are valuable for the vocabulary-prediction task in which learners’ vocabulary knowledge is predicted ([Ehara et al., 2010, 2012, 2013, 2016](#)). We focused on lexical simplification as the direct application of predicting the psycholinguistic properties of words. Future work includes leveraging confidence values that GPR can produce for graph-based weakly supervised learning, as in ([Ehara et al., 2014](#)).

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 15K16059.

## References

- Shigeaki Amano and Tadahisa Kondo. 1998. Estimation of mental lexicon size with word familiarity database. In *Proceedings of the 5th International Conference on Spoken Language Processing (IC-SLP)*.
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Edwin V Bonilla, Kian M. Chai, and Christopher Williams. 2008. Multi-task gaussian process prediction. In *Advances in Neural Information Processing Systems 20 (NIPS)*, pages 153–160.
- Max Coltheart. 1981. The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, 33(4):497–505.
- Yo Ehara, Yukino Baba, Masao Utiyama, and Eiichiro Sumita. 2016. Assessing translation ability through vocabulary ability assessment. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3712–3718.
- Yo Ehara, Yusuke Miyao, Hidekazu Oiwa, Issei Sato, and Hiroshi Nakagawa. 2014. [Formalizing word sampling for vocabulary prediction as graph-based active learning](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1374–1384, Doha, Qatar. Association for Computational Linguistics.
- Yo Ehara, Issei Sato, Hidekazu Oiwa, and Hiroshi Nakagawa. 2012. [Mining words in the minds of second language learners: learner-specific word difficulty](#). In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, Mumbai, India.
- Yo Ehara, Nobuyuki Shimizu, Takashi Ninomiya, and Hiroshi Nakagawa. 2010. Personalized reading support for second-language web documents by collective intelligence. In *Proceedings of the 15th international conference on Intelligent user interfaces (IUI 2010)*, pages 51–60, Hong Kong, China. ACM.
- Yo Ehara, Nobuyuki Shimizu, Takashi Ninomiya, and Hiroshi Nakagawa. 2013. Personalized reading support for second-language web documents. *ACM Transactions on Intelligent Systems and Technology*, 4(2).
- Sujay K. Jauhar and Lucia Specia. 2012. Uow-shef: Simplex – lexical simplicity ranking based on contextual and psycholinguistic features. In *\*SEM 2012: 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 (SemEval 2012)*, pages 477–481.
- Maximilian Köper and Sabine Schulte Im Walde. 2016. Automatically generated affective norms of abstractness, arousal, imageability and valence for 350 000 german lemmas. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*.
- Taku Kudo. 2005. Mecab: Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.net/>.
- Kikuo Maekawa. 2007. Kotonoha and bccwj: development of a balanced corpus of contemporary written japanese. In *Corpora and Language Research: Proceedings of the First International Conference on Korean Language, Literature, and Culture*, pages 158–177.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems (NIPS)*, pages 3111–3119.
- George Armitage Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Michael Mohler, Marc Tomlinson, David Bracewell, and Bryan Rink. 2014. Semi-supervised methods for expanding psycholinguistics norms by integrating distributional similarity with the structure of wordnet. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, pages 3020–3026.
- Gustavo Paetzold and Lucia Specia. 2016. Inferring psycholinguistic properties of words. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 435–440.
- H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E. P. Seligman, and Lyle H. Ungar. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 8(9):1–16.
- Alex Smola and Vladimir Vapnik. 1997. Support vector regression machines. In *Advances in neural information processing systems (NIPS)*, volume 9, pages 155–161.
- Kumiko Tanaka-Ishii and Hiroshi Terada. 2011. [Word familiarity and frequency](#). *Studia Linguistica*, 65(1):96–116.
- The BNC Consortium. 2007. The british national corpus, version 3 (bnc xml edition). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>.

Andrey Tikhonov. 1963. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 5:1035–1038.