Machine Learning Methods for Indicating Cultural Biases in Spoken Russian Language: Dominants and Trends of Modern Society

Anna Chizhik[a], Yulia Zherebtsova[b], Alexander Sadokhin[c]

[a] *Saint Petersburg State University, 10th line of Vasilievsky island, 49, St.-Petersburg, 199178, Russia*
[b] *Inforser Engineering, Ryazansky Prospekt, 24, building 2, 109428, Moscow, Russia*
[c] *Russian State Social University, 107076, Stromynka str., 18/28, Moscow, Russia*

Abstract

The formation of a socio-cultural layer is based on a person's understanding of himself and the world around him and the translation of this under-standing into abstraction. This inevitably leads to the emergence of cultural biases in society as an extreme form of separation of one social group from another. In fact bias is nonrandom errors in thinking. The growing cultural biases in society is preserved in the consciousness of individuals and further affects the possible interpretation and perception of the neighboring social group, and, therefore, the public mood, in other words, the level of aggressiveness of the society. Thus the problem of identifying cultural shifts is relevant for the scientific community. There are many methods based on surveys and their subsequent analysis. In this paper, we propose to use machine learning and analysis of the large collection of text data from social networks (public Telegram chat). This approach can complement the standard methodology, including helping to reveal hidden patterns by being able to cover large amounts of data.

Keywords

machine learning, natural language processing, text clustering, cultural biases, text analysis, cultural code, cultural process

1. Introduction

The view of the world formed by immersion of an individual in media space, which includes all possible communication channels (communication «one to one», «one to many», «many to many»). In the current moment, social networks accumulate the potential of the key influence on the consciousness of individuals due to their dynamic structure. A complex of stable associations, opinions and stereotypes around complex social phenomena is formed on their basis. Even where social networks are not the primary sources of information, there is a need to get an opinion of people to the news by this channel of mass communication and to compare own opinion with the opinion of others. The non-representativeness of information increases within horizontal communicative structures. Thus, social media texts not only build the information agenda, but radically affect to social mood and public opinion by spontaneously distorting the picture of objective reality.

Therefore, social media texts form a subjective and biased point of view on events, creating a certain image of reality in the mind of the information consumer and influencing the course of events in the end. The peculiarity of communication in a social media lies in the specific flow of interpersonal perception processes. People find uncertainty in interpersonal relationships unpleasant, thus they are motivated to reduce it. Stereotyping and identification by affiliation with particular social groups have a strong influence on individual mind in this way and contribute to judge a person. This provides the

IMS 2021 - International Conference "Internet and Modern Society", June 24-26, 2021, St. Petersburg, Russia
EMAIL: a.chizhik@spbu.ru (A. 1); julia.zherebtsova@gmail.com (A. 2); sadalpetr@yandex.ru (A. 3)
ORCID: 0000-0002-4523-5167 (A. 1); 0000-0003-4450-2566 (A. 2); 0000-0002-6420-6601 (A. 3)


situation in which social network participants are united by homogenized opinions by r receiving an average consciousness.

Participants of communication are keen to get approval of the social group, thus the feedback effect (verbal response, likes, emoticons) provides an important incentive. Social networks are not a platform for spontaneously emerged communities and they are not isolated from other channels of interactions with information flows. Therefore, sociocultural tendencies penetrate there from media sphere in its multifaceted understanding (including through traditional media, opinion leaders, propaganda, etc.). The cycle of comments (and replies to them) is repeated depending on the severity of the topic and the activity of the initiators of the conversation thread. Acquiring the features of recursion, this cycle of communication forms public mood and stereo-types.

The effectiveness of communication remains in question, because the existence of a rational conversation vector (constructive dialogue) is quite difficult to trace. Since the incentive to continue communication in this case is the approval of the majority, it is fair to assume that this kind of dialogue strengthens the existing cultural and social biases, and also creates the basis for the formation of an aggressive field around them.

Subjective and systemic biases of social actors influence to the information choice and content features of texts which are presented in this communicative space. Subjective biases operate at the level of individual information processing in the context of current events. Such subjective biases could arise from shared values, information overloads, and cultural preferences. Some of the subjective prejudices are transformed into systemic ones over time, which shape the consciousness of individuals at the mesoscopic level (mass consciousness). This creates patterns in the mindsets of societies. These regularities cannot be observed at the level of a specific statement of individual. However these patterns can be observed by analyzing big data. In this paper we present mathematical methods for detection these collective spontaneous information filters, which form cultural and social biases that exist in modern Russian society.

2. Method

The language model is the basis for mathematical methods of text analysis. It provides to calculate the probability that a word will follow a given word sequences as a continuation of the text. Thus a statistical language model helps to calculate the probability distribution function on a set of vocabulary sequences.

One of the first methods of constructing language models was n-gram model [1, 2]. The probability of a word sequence is considered as the product of word probabilities, given the known previous ones. Therefore only a few previous words (n words) are matter for this kind of statistical analyze. Further various architectures based on machine learning algorithms and artificial neural networks were widespread as the basis of language models [3, 4, 5]. Neural network language models are divided into two groups: word-aware NLM and character-aware NLM.

A good language model must capture two important properties of a natural language. The first one is correct syntax. Thus, a few previous words are sufficient for a relevant prediction of the next word, however the word order in a sentence becomes important item. The second property is coherence. Including large number of words is often required in order to understand the global meaning of a sentence or document (but the word order has less importance). Traditional N-gram models and neural probabilistic language models have difficulties in extracting global semantic information from text (because of a fixed-size context window), that is, polysemy and context-dependent nature of words are not taken into account. Consequently, contextualized language models are gradually gaining popularity, trying to take into account the context of the use of the word [6, 7, 8, 9]. This approach allows combining two necessary properties of a natural language (correct syntax and coherence).

The process of train a language model begins with creating a collection of texts by natural language (dataset). A language model predicts the next word in a text, thus it should have seen a lot of examples to learn the language. Essentially the model calculates the probability the appearance of a word next to the known word sequence. This prediction is based on examples from the dataset. For the representation of words into a language model it is necessary to map words or phrases from the vocabulary to vectors

of real numbers (word embeddings). Consequently the semantic similarity of words or the frequency of their joint occurrence can be detected by comparing the distances between these vectors (cosine similarity).

A classic example of this method shows the interdependency between pairs of words («king-man» and «queen-woman»). Figure 1 demonstrates that algebraic operations in this space correspond to operations on the meaning of words.

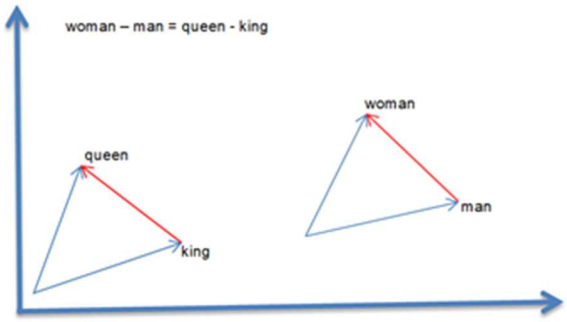

Figure 1: Semantic relationships of words in a language model

Similar calculations can be performed for sentences (Fig. 2). This allows calculating the similarity of the entered phrase (its semantic coherence or antonymic nature) to sentences from dataset that are already understandable to the language model.

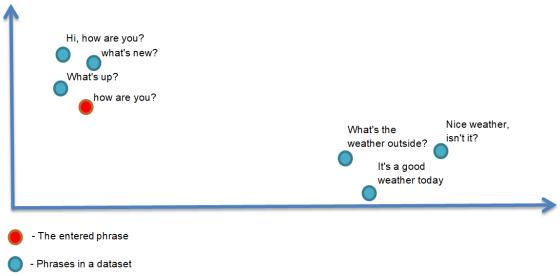

Figure 2: Matching sentences based on how the language model works

Accordingly semantic relationships between words (or sentences) are gaining mathematical meaning in a vector space. This suggests that the most productive language models reproduce text sequences that contain typical biases for social micro and macro groups. These biases initially arise as an emotional reaction to different phenomena and antagonistic groups. The detection of these biases by language model is possible because its training occurs by calculating the joint occurrence of words. In other words, a language model would know better that «freedom» is used with the word «speech», the more often such idiom («freedom of speech») would be found in the training dataset. In summing up, the biases, that was found in datasets and fallen into language models, in general, can be characterized as actual social landmarks of society.

3. Typology of cultural bias

The following types of cultural biases can be identified: national or religious ideals, social connotations, gender stereotypes and aggressive statements of a general nature.

Cultural biases in natural language can be present in a latent form or in a direct manifestation of an attitude to the object of a statement. The collected text data (on the basis of which the language model works) attracts to socio-demographic and mental stereotypes, traditions and patterns of behavior accepted in society. Therefore, it seems interesting to analyze the social connotations and contexts of cultural biases. In this paper, we provide the analysis of biases in two contexts: the ones that at the descriptive level reflect social mood [10, 11, 12] on a specific topic, and those that outline the characteristics of a social group. A group is people with the same markers (gender, nationality, etc.).

4. Experiments and Results

The public Telegram chat has been chosen as the data source (chat of the news channel Mash – «MACH», 1608 participants, not moderated). Thus, the overall text collection consists of 556 354 records, the first of which was dated 2018-07-05, and the last - 2020-12-03).

Clustering of messages provided first characteristics of the chat conversations, including the information about topics' content and their close interaction with each other. The cluster analysis of the dataset was performed with the k-means algorithm. The vector matrix was created using Word2Vec model trained on our dataset in order to obtain more unambiguous result of partitioning into cluster groups. Figure 3 shows that the clusters are very close to each other, and even sometimes intersect in space.

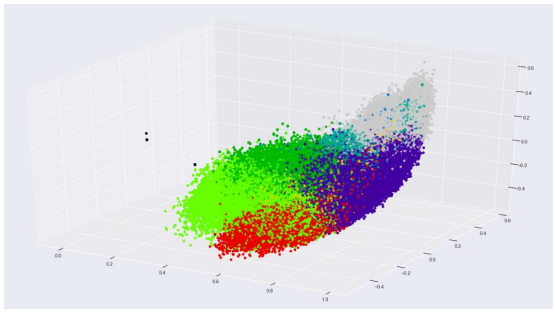

Figure 3: Cluster ratio

The next step was to identify the associative chains of terms. Pairs and triples of words were taken to take this goal. The model found the most semantically close terms from the dataset to chosen items (the theoretical algorithm of the model is shown in Fig. 1). The model finds sets of words that are close in meaning (quasi-synonyms), the meanings of which can differ in several characteristics (for example, in relation to the speaker) and change depending on the context. The closeness of the word to the term can be interpreted either as equality («she» = «her» = «girl» = «wife» = ...) or as a word very well associated with the term («she» = «girl» = «wife» => husband). Thus, only those words that are closely interrelated in the cultural code can catch into associations. An example of the resulting associations is shown in Table 1.

Table 1

Semantic associations to the words «freedom», «democracy», «Internet»

Liberty	Democracy	Internet
('recognize', 0.9017236232757568), ('corruption', 0.8983240723609924), ('punishment', 0.8904907703399658), ('citizen', 0.880867063999176), ('crime', 0.8789682388305664), ('revision_year', 0.871111273765564), ('ratio', 0.8698971271514893), ('fairness', 0.7563846111297607), ('death_punishment', 0.7523390054702759), ('monarchy', 0.7488603591918945),	('opposition', 0.9191081523895264), ('communism', 0.9145876169204712), ('equality', 0.9139655828475952), ('develop', 0.9133450984954834), ('putinsky', 0.9130712747573853), ('capitalism', 0.8889749646186829), ('monarchy', 0.8852110505104065), ('socialism', 0.8837734460830688),	('anonymity', 0.6254202723503113), ('vpn', 0.6117355227470398), ('doesn't work', 0.6087996959686279), ('free access', 0.5988985300064087), ('telegram', 0.5986903309822083), ('rkn', 0.5934107899665833), ('outage', 0.5892473459243774), ('space', 0.5862609148025513),

Liberty	Democracy	Internet
('liberalism', 0.7421742081642151), ('acting_power', 0.7411474585533142), ('civil_society', 0.7408543229103088)	('vertical', 0.8762210607528687), ('scrap', 0.8600466251373291), ('dictatorship', 0.8596632480621338), ('mentality', 0.8592602610588074)	('default', 0.5862008929252625)

People most often associate the term «democracy» with opposition. The next logical link from this word leads to «communism». Thus the evidence base for discussions around democracy is the previous political system of our country. Obviously the understanding of the term is reduced to cultural and social contexts colored by national history. The words «equality» and «develop» frequently appear in conversations on this topic. This fact can probably be classified as hopes for a brighter future. It is noteworthy that the adjective «putinskii» (the time of something in association to the period of Putin's presidential term) appears in the seven most closely related terms. It strong links the discussion of democracy with the current agenda, because this word clearly indicates a non-abstract line of reasoning).

We also investigated the reflection of the agenda through collocations. A collocation is a phrase that has signs of a syntactically and semantically integral unit (stable phrases). Highlighting of them can help delineate the social and political tendencies of the social macro groups. For example, throughout the entire data collection (more than 500 thousand messages), «Russia» most often occurs with the word «president»; such phrases as «Russians forward» and «Putin is the president» have shown themselves as stable collocations. Below is an example of identified collocations (Table 2).

Table 2
Collocation examples

Liberty	the Internet
'freedom of speech', 'deprivation of freedom', 'restriction freedom', 'release from prison.', 'right to freedom', 'internet freedom'	'bad_internet', 'sovereign_internet', 'internet_freedom', 'free_internet', 'internet_access', 'internet_passport'

The same terms as in previous example were taken for clearly interpretation of the results. Identification of stable collocations complements the ability to assess the social and cultural landscapes by the language models. If in the above example the term «liberty» was associated exclusively with criminal liability and offenses, then when searching for collocations, topics appeared that interest people most likely in connection with the term «democracy». However, there are no stable phrases with the term "democracy" in this case of text analysis, while at the previous step of the study, tendencies were identified. At the same time, the detection bounds of the term "Internet" in this way gives good results and it could be used to deep understanding the public mood.

Multidimensional space visualizations of word vectors can help in interpreting relationships between word embeddings. Figure 4 presents the visualization of the results described above for associative bounds to the words «liberty», «democracy», «Internet». The method of nonlinear dimensionality reduction (T-SNE) [13] was used for this purpose. The basic principle of t-SNE is to reduce the pairwise distance between points while maintaining their relative position. Thus the algorithm constructs a probability distribution over pairs of high-dimensional objects in such a way that similar objects are assigned a higher probability while dissimilar points are assigned a lower probability. It becomes

possible to map high-dimensional data to a low-dimensional space, while the location structure of the neighboring points is saved.

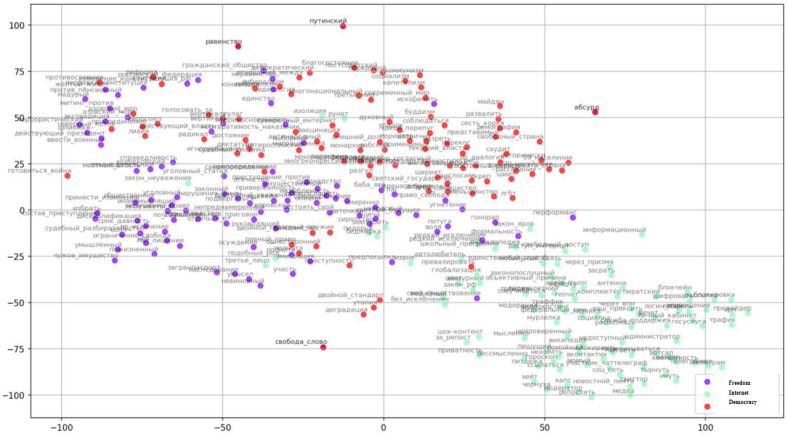

Figure 3: Semantic associations to the words «liberty», «democracy», «Internet»

Information about semantic correlations of the texts has been received by visualizing grouped data. The figure demonstrates that vector representations of words fix the semantic relations between such categories as the political system of the country, gender stereotypes, and even there are clearly seen the associations of swear words with discussions of dissatisfaction with that or another phenomenon, and etc. This kind of analysis provides direct information about cultural biases based on mathematical apparatus, that is, with a certain accuracy and impartiality. In developing this strategy of text analysis in its deeper condition, the most frequent words (the first 4 in frequency) were taken from each cluster, which were detected on the first step of the research (Fig. 5).

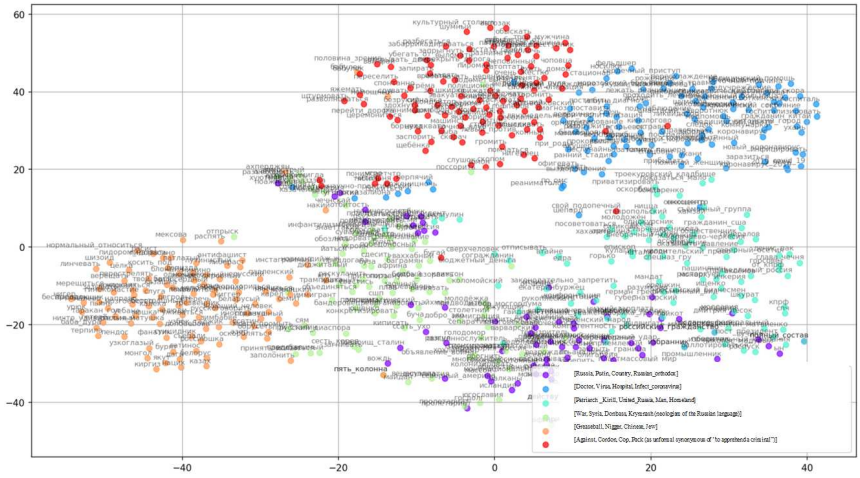

Figure 4: Sociocultural analysis of the identified clusters

The presence of sufficiently clearly grouped semantic clouds shows the tendencies of public opinion. In particular, the totality of the formed clusters and their content reveal the tendencies present in society related to the separation of male and female roles, as well as attitudes towards national minorities; at the same time, it is clearly seen the topics around COVID-19 (such as emergency medical services, general news about the virus, discussion of hospitals).

In terms of mathematical and computational linguistics the biases implies a shift from the selected item to the left or to the right along the space coordinate axes. By way of illustration, the example of the bias between the words «dictatorship» and «democracy» is shown in Figure 6. The X-axis (horizontal) is set from «democracy» to «dictatorship»: words close in meaning to democracy (within the studied texts) are on the left, and words similar to the word «dictatorship» are on the right. It was decided to use «citizen» as the anchor word.

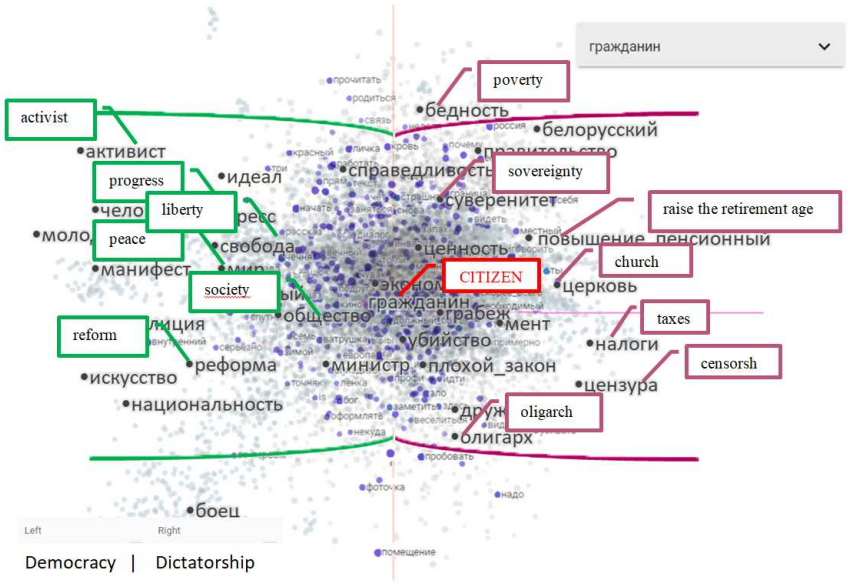

Figure 5: An example of the social and cultural biases which are detected in data collection from chat dialogs (the 0X axis is stretched between the words «democracy» and «dictatorship», point (0,0) is « civil»)

It turned out that «citizen» is a very good choice of the anchor word, because its position (in the middle of the projection, zero mark on the X-axis) indicates the neutrality of the term within the context of our research. The words that are attracted to the pole of the word «democracy» (the left of the zero mark on the X-axis) are outlined in green for convenience; words that are semantically connected with «dictatorship» (right) are purple. Accordingly, the closer a word is to the left edge of the x-axis, the more clearly it illustrates public attitudes toward «democracy». Y-axis displays the spread of words in their ideological differences and helps to detect two semantic groups that characterize this phenomenon in the public consciousness. On the one hand, an opinion clearly emerges in the mass consciousness that democracy is established with the active participation of society in this process (all words that fall into a significant sample, such as «person», «manifesto», «activist», are easily summarized in the category «civil society»). On the other hand, the words «art», «scientist», «peace» and «freedom» can also be highlighted as markers of the general idea of what democracy provides. Words that are «drawn» to the right side (to «dictatorship») are «censorship», «church», «raise the retirement age», «taxes». Obviously the clearly expressed attitude to the term is seen. Cultural biases of meanings between synonymous words become convenient to trace due to this item arrangement in space. For example, «police» - «cop» and «reform» - «law». If «police» is an attribute of a democratic society, then «cop» refers to a dictatorship in Russian language; «reform» is associated in the public perceptions with democracy, and the phrases with words «law» and «bad law» appear for the dictatorship.

5. Conclusion

Examples of aggressive cultural biases (although they have been identified) are deliberately not shown in the results. The purpose of this paper is to describe the potential of the method without deep analysis of specific stereotypes and public opinion. Thus, our experiments demonstrate that the method is applicable to interpret cultural biases, which are formed in public consciousness by active using of social media. It is important to consider a number of factors before using statistical language models, such as the definition of the subject and functional boundaries of the object under study, the nature of the object under study and possible linguocultural consequences of its use, the detection the role of the object in language (natural language as a repository of cultural code), that determines its place in the system of linguacultural universals. An important application of this method can be the identification of aggression in social groups through text data.

6. References

[1] Jelinek, F. Computation of the probability of initial substring generation by stochastic context free-grammar. Computational Linguistics. Vol. 17, № 3, 315–323 (1991).
[2] Stolcke, A. Precise n-gram probabilities from stochastic context-free grammars / A. Stolcke, J. Segal // Proceedings of the 32th Annual Meeting of ACL, 74–79 (1994).
[3] Mikolov, T. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of Workshop at ICLR (2013). https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf, last accessed 2021/03/19.
[4] Joulin, A., Grave, E., Bojanowski, P., Mikolov, T. Bag of Tricks for Efficient Text Classification. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 427–431 (2016).
[5] Pennington, J., Socher, R., Manning, C. D. GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, 1532-1543 (2014).
[6] Che, W., Liu, Y., Wang, Y., Zheng, B., Liu, T. Towards better UD parsing: Deep contextualized word embeddings, ensemble, and treebank concatenation. Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 55–64 (2018).
[7] Peters, M.E., Neumann, M., Iyyer, M. Deep contextualized word representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2227–2237 (2018).
[8] Artetxe, M., Schwenk, H. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. Transactions of the Association for Computational Linguistics. V.7, 597–610 (2019).
[9] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. Language Models are Unsupervised Multitask Learners. Technical Report OpenAi (2018) https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf, last accessed 2021/03/19.
[10] Yadov V.A. Ideology as a form of spiritual activity of society. In Russ. (1961)
[11] Porshnev B.F. Social psychology and history. In Russ. (1979)
[12] Uznadze D.N. Installation psychology. In Russ. (2001)
[13] Maaten L., Hinton G. Visualizing data using t-SNE //Journal of machine learning research. 9, 2579-2605 (2008).