

Article

Transferring Sentiment Cross-Lingually within and across Same-Family Languages

Gaurish Thakkar * , Nives Mikelić Preradović and Marko Tadić

Faculty of Humanities and Social Sciences, University of Zagreb, Ivana Lučića 3, 10000 Zagreb, Croatia; nmikelic@ffzg.hr (N.M.P.); marko.tadic@ffzg.hr (M.T.)

* Correspondence: thak123@gmail.com or gthakkar@m.ffzg.hr

Abstract: Natural language processing for languages with limited resources is hampered by a lack of data. Using English as a hub language for such languages, cross-lingual sentiment analysis has been developed. The sheer quantity of English language resources raises questions about its status as the primary resource. This research aims to examine the impact on sentiment analysis of adding data from same-family versus distant-family languages. We analyze the performance using low-resource and high-resource data from the same language family (Slavic), investigate the effect of using a distant-family language (English) and report the results for both settings. Quantitative experiments using multi-task learning demonstrate that adding a large quantity of data from related and distant-family languages is advantageous for cross-lingual sentiment transfer.

Keywords: sentiment analysis; language models; transfer learning

1. Introduction

Classification of sentiment is essential to text analysis. It is concerned with the automatic extraction of subjective data from text sources. The data provide a clear picture of the entities of interest, such as people, products, aspects, or concepts. The process assigns labels with varying granularity depending on the task. For instance, labels can be positive–negative [1], positive–neutral–negative [2], positive–negative–mixed–other [3], or positive–negative–neutral–mixed–other. Earlier work concentrated primarily on extracting well-designed features [4–6]. Recent work with neural networks simplifies the feature engineering required to extract features from input text [7]. The main challenge posed by such deep neural networks is the requirement for training data (supervision). The availability of such supervised resources is a challenge for languages with limited resources. Cross-lingual sentiment analysis (CLSA) makes use of resources from high-resource languages to construct a sentiment analyzer for low-resource languages. For instance, the simplest configuration translates instances of data from the target language to the source and applies the classifier trained on the high-resource source language [8]. Using machine translation (MT) systems to translate the resources from the source language (annotated datasets or lexicons) into the target language is an alternative method [9,10]. However, such an accurate translation system is not always available for language pairs with limited resources. Similarly, a prior attempt employed parallel data [11].

Recent research utilizing word embeddings and context-sensitive representations, such as GPT [12], ELMO [13], BERT [14], and RoBERTa [15], has improved overall classification performance. Pretraining on large corpora is used to acquire these representations. In a multilingual environment, multiple languages are trained collectively on a single model. Downstream tasks like Named Entity Recognition and Classification (NERC) [16] or Question Answering (QA) [17] refine the pre-trained language models (PLMs). The primary challenge with multilingual pre-trained language models is the effect of similar languages and representation in the learned space. For instance, the Multilingual BERT (MBERT) [14],



Citation: Thakkar, G.; Preradović, N.M.; Tadić, M. Transferring Sentiment Cross-Lingually within and across Same-Family Languages. *Appl. Sci.* **2024**, *14*, 5652. <https://doi.org/10.3390/app14135652>

Academic Editor: Chilukuri K. Mohan

Received: 7 May 2024

Revised: 20 June 2024

Accepted: 26 June 2024

Published: 28 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

which has been trained in 104 languages, does not proportionally represent each language in terms of training corpus. Fine-tuning PLMs for under-represented languages results in poor performance for the target language. In addition to under-representation, many languages are absent from these PLMs. All of these conditions are a result of the lack of data for languages with limited resources.

Sentiment classifiers trained on pre-trained language modeling tasks have demonstrated cutting-edge performance [18,19]. Even though these approaches have been investigated in a cross-lingual context, their applicability to low-resource languages, particularly languages within the same language family, remains to be investigated. Our analysis investigates the transfer of knowledge between languages of the same language family. We seek the optimal means of combining source-language and target-language data sources. This article describes all the techniques and experimental analyses for combining high-resource languages with low-resource languages.

Throughout the past decade, cross-lingual sentiment classification has remained an active field of study. Das and Sarkar [20] classify cross-lingual processing approaches as either model transfer or annotation adaptation. Model transfer utilizes language-independent features. One of the ways to learn these characteristics is through adversarial training [21,22]. These cross-lingual representations are optimized for the final task, such as the recognition and classification of parts of speech or named entities. Methods for annotation projection utilize massive parallel corpora between the source and target languages. They exploit the semantic similarity between the parallel corpora. Using the source-trained classifier on a machine-translated view of the target dataset is the simplest approach. As previously observed [23], machine translation introduces noise into the translation, altering the final output's meaning. The classification of noisy input does not guarantee its conformance to the target instance class. The second class of methods [24] combines model transfer and annotation adaptation into a single unit. The configuration simultaneously trains the shared encoder and parallel corpora for alignment and classification tasks.

2. Research Questions and Hypotheses

Empirically, we have the following question for our proposed study.

Q. What is the effect of language similarity and available resources inside PLMs? We hypothesize that:

1. A cross-lingual transfer is more successful for typologically similar languages than for typologically different languages.
2. A large annotated dataset in a distant-family language can overcome typological differences, unlike a small annotated dataset in a close-family language.

To answer the research question, which is to examine the effect of typology on the performance of cross-lingual sentiment analysis, we trained models using English and Slavic language datasets. The training involved the combination of diverse language datasets. We calculated the effect of utilizing a language during training and its effect on final performance in several different combinations. The outcomes were compared to previously published research. We determined the optimal language combination for sentiment transfer.

This paper's contributions are as follows:

1. Initially, we propose a framework for unified deep learning that utilizes existing data labels from high-resource languages on low-resource datasets. We conduct rigorous experiments on languages within the same language family. We investigate how effectively sentiment classification abilities could be transferred.
2. Second, we demonstrate that, given multiple large-scale training datasets, our framework is superior to a straightforward setup for fine-tuning. Finally, we devise the optimal method for jointly training sentiment analysis systems in order to address the issue of insufficient resources for target languages.

3. Languages in the Study

A language family is a collection of languages that share a common ancestor. English, for instance, is a member of the Indo-European (IE) language family. Our target languages are South Slavic languages with very few labeled examples for sentiment analysis tasks. Languages within the same family typically share a subset of vocabulary and typological characteristics. In the context of languages from the same family tree, typological characteristics, such as word order, morphology, and phonology, refer to shared structural features that these languages exhibit despite their historical divergence. Cognates [25], which are sets of words in different languages that have been directly inherited from an etymological ancestor in a common parent language, are one such phenomenon. For instance, the Proto-Slavic word *noktь* (night) has equivalents in other languages, such as *ночь* (*noč'*) (Russian), *ніч* (*nič*) (Ukrainian), *ноч* (*noč*) (Belarusian), *noc* (Polish, Czech, Slovak), *noč* (Slovene), *ноћ/ноћ* (Serbo-Croatian), *нощ* (*nosht*) (Bulgarian), and *ноќ* (*noќ*) (Macedonian).

The language family is subdivided into branches that are categorized as subsets. For instance, one of the branches of IE, Balto-Slavic, has a Slavic branch that is subdivided into West, South, and East subgroups [26]: Russian, Belarusian, and Ukrainian (in the East group); Polish, Czech, and Slovak (in the West group); Bulgarian and Macedonian (eastern dialects of the South group); and Serbo-Croatian and Slovene (western dialects of the South group). We chose to concentrate on three West Slavic languages (Czech, Slovak, and Polish), three South Slavic languages (Croatian, Slovene, and Bulgarian), and one East Slavic language (Russian). Czech and Slovak have the highest degree of mutual intelligibility, followed by Croatian and Slovenian [27]. Except for Bulgarian and Russian (which use the Cyrillic script), all languages use the Latin alphabet. Russian has a complex case system, whereas Bulgarian has lost almost all of its case declensions [28]. The Slavic language family is illustrated in Figure 1.

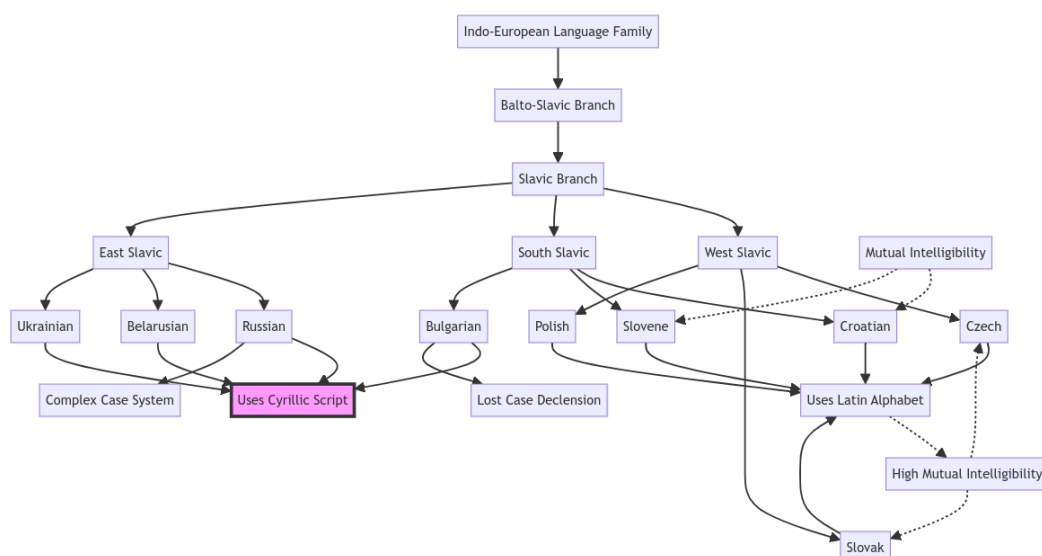


Figure 1. Family tree of Slavic languages.

4. Related Work

4.1. Sentiment Analysis

Turney [29] extracted phrases containing adverbs and adjectives by focusing on consecutive words within the context. Patterns were applied to this phrase extraction to eliminate the influence of proper names. Excellent and poor were used to calculate the semantic orientation (SO) of the phrase. The final review score was determined by averaging the phrases' semantic orientation. The author noted that text from a particular domain has a distinct writing style that can mislead the final grade.

Vanilla sentiment lexicon-based methods employ either the presence or absence of words or the scoring of individual words in the text [30], ultimately averaging the final score.

The authors chose a list of verbs, adjectives, and nouns as a starting point and expanded it using Wordnet. Using the synsets from Wordnet, a word's polarity score was calculated. The final class was derived from emotionally charged words. Using negation, intensifiers, and diminishers, the lexicon-based technique [31] was investigated. The combination of positive and negative words inverts the overall evaluation. In contrast, negative phrases and negation result in a positive final evaluation. The use of modal operators establishes a context for the possibility of necessity. Therefore, realis and irrealis events should be treated differently, as irrealis situations do not necessarily reflect the true attitude of opinion holders toward a concept like they do in the realis context. Other linguistic structures mentioned by the authors include presuppositional items (such as it is barely sufficient), connectors, and irony.

The earliest attempts were rule-based methods with a high degree of precision [32] that relied heavily on subjective lexicons and patterns. Results were obtained using two classifiers that relied on the presence and absence of subjective clues for subjective and objective classification. The initially classified sentences were then subjected to pattern extraction and iterated in a bootstrapping process to increase the classifier's lexicon size and coverage. The training dataset was used to train a naïve Bayes classifier for ranking unlabeled text corpora and passed through the initial pattern extraction procedure to enhance the self-training procedure.

Several sentiment lexicons include SentiWordnet [33], General Inquirer [34], SenticNet [35], and AFFIN [36]. Traditional machine learning models such as naïve Bayes and support vector machines (SVMs) have played essential roles in classification. These methods [6,37] utilize feature engineering. Mullen and Collier [37] used Turney's [29] features and lemma to conclude that the calculation of pointwise mutual information (PMI) could be supplemented with domain information when searching the web for the context window, assuming that domain information did not reduce the hit count.

Wilson et al. [6] compiled a list of subjectivity clues and expanded it using additional lexicons, including General Inquirer, a dictionary, and a thesaurus. The methodology was based on the prior lexicon-based polarity classifier. This was refined through a two-step process based on intensive feature engineering to distinguish contextual polarity.

McDonald et al. [38] conducted experiments with cascading sentences and document labels. Together, the document and sentences are trained for the classification task. The sentence classification feature space included unigram, bigram, trigram, and POS tags. The inference is performed using the Viterbi algorithm to calculate the document's final score based on the scores of its sentences.

Paulus et al. [39] integrated phrase-level predictions into global belief recursive neural networks to provide feedback to words. This is accomplished by incorporating a backward pass that propagates from the parse tree's root to its leaves. The GB-RNN employs both forward and backward parent nodes, whereas the Bi-RNN employs only forward parent nodes. This method necessitates a parser for the tree structure. In addition to supervised and unsupervised techniques, research also focuses on semi-supervised methods.

Read and Carroll [40] created domain-independent polarity classifiers using word similarity techniques in a semi-supervised setup. The authors described numerous matrices of word similarity. First, the lexical association is calculated using PMI to determine the similarity between two words. Second, semantic spaces represent a collection of conceptually similar words. Last but not least, distributional similarity defines the similarity between two words based on the words in their context. A large unsupervised dataset was utilized to compute the co-occurrence and occurrence frequencies required for the aforementioned matrices.

Moraes et al. [41] compared the performance of SVMs and ANNs (artificial neural networks). The authors discovered that ANNs statistically outperformed SVMs when combined with the information gain-based feature selection method. Nonetheless, the results demonstrated that SVMs are less susceptible to noisy terms in the presence of data imbalance.

Other authors [42–44] investigated recursive-style neural networks for learning vector representation for a sentence. These methods abandon single-word features in favor of a vector-based strategy. The authors' proposed recursive neural network learns the vector representations of phrases in a tree structure. It assigns a vector and a matrix to each node in a parse tree in order to capture its influence on the surrounding words. The recursive neural tensor network computes higher node representation using leaf-level word vectors. These procedures utilized parse trees.

CNN's semantic modeling of sentences was investigated [45,46]. The CNNs presented by the author are not parse tree-based. Utilizing filter pooling operations, relations between discontinuous phrases were captured. In addition to using a single neural schema such as unidirectional LSTM [47] or bidirectional LSTM [48], authors have mixed and matched networks such as CNN–LSTM [49], CNN, and RNN [50]. CNN is used to acquire regional characteristics, while the recurrent network learns the interdependencies between these regional characteristics. These methods consistently outperform feature engineering techniques. During back-propagation, which retrofits these representations for sentiment analysis, the word embedding used as input layers is also fine-tuned. The task-specific knowledge is eventually helpful during the time of inference. To prevent overfitting, they require an extensive training set.

4.2. Sentiment Analysis in Slavic Languages

The Kapukaranov and Nakov [51] dataset of film reviews with fine-grained scores was a significant contribution to Bulgarian sentiment analysis. Georgieva-Trifonova et al. [52] compiled a dataset containing customer feedback derived from online store reviews. Lazarova and Koychev [53] classified film reviews using a semi-supervised multi-view genetic algorithm. Osenova and Simov [54] described the creation of a corpus of Bulgarian political speech. A classification of Bulgarian tweets was performed by Smailović et al. [55]. Hristova [56] provides a concise overview of the text-analytic work in Bulgarian.

Steinberger et al. [57] created sentiment dictionaries for multiple languages, including Czech, that are multilingual and comparable. Veselovská [58] compiled a corpus of annotated opinion articles from the Aktualne.cz news website. This was supplemented with supplementary data derived from domestic appliance reviews on the Mall.cz retail website. A dataset of Czech film reviews was compiled by Habernal and Brychcín [59]. The authors iteratively examined the Maximum Entropy classifier and Gibb's sampling to determine the desired probabilities. Čano and Bojar [60] evaluated supervised machine learning algorithms using the Mall.cz and Facebook datasets. BERT-based models for Czech sentiment have also been attempted [61–64]).

Agić et al. [65] developed grammar-based rules for determining the overall sentiment of Croatian financial news. Agić and Merkle [66] have created rule-based techniques for detecting sentiment in horoscopes published on news portal websites. Jakopović and Mikić Preradović [67] evaluated a lexicon-based method for analyzing user comments in the transportation domain. Glavaš et al. [68] presented aspect-based, domain-specific sentiment analysis for the Croatian language. Mozetič et al. [69] and Rotim and Šnajder [70] have studied the sentiment analysis of Croatian social media text. Robnik-Šikonja et al. [71] compared the Slavic and Germanic language families for a Twitter sentiment analysis task. Lula and Wójcik [72] discussed theoretical and practical aspects of Polish consumer opinions. Haniewicz et al. [73] presented the first attempt to create a polarity lexicon that is accessible to the public. They utilized readily available resources such as dictionaries, thesauri, and existing open-source initiatives. Other attempts at solving SA in Polish primarily include lexicons [74], Wordnet features [75], and unigrams/bigrams [76]. Numerous authors, such as Kocoń et al. [77], Wawer and Sobiczewska [78], have compared and contrasted machine learning and deep learning techniques for sentiment recognition, including naïve Bayes, SVM, BiLSTM, and BERT.

Rules [79], machine learning techniques [80], and deep learning approaches have been described in previous work on the Russian language. Using various neural techniques, Gol-

ubev and Loukachevitch [81] improved the scores on multiple Russian sentiment datasets. This work proposed sentiment classification as a task of natural language inference and improved final scores. Golubev and Loukachevitch [82] continued the same work with three-step sequential training and achieved state-of-the-art results. Smetanin and Komarov [83] identified multiple datasets and baselines for a sentiment analysis task in Russian. Machová et al. [84] translated an English lexicon into Slovak and combined it with a particle swarm optimization algorithm to construct a lexicon-based sentiment categorization system. Bučar et al. [85] annotated and evaluated five distinct classifiers for Slovenian web media content. Various attempts have been made at sentiment analysis in Slovenian news texts [86–89]. The corpus of web commentary was examined by Kadunc and Robnik-Šikonja [90]. Offensive language detection in Slovene [91,92] is an active area of research.

4.3. Cross-Lingual Sentiment Analysis

In a cross-lingual multi-task learning setup, Cotterell and Heigold [93] performed morphological tagging and language identification by jointly training a BiLSTM with character embeddings. The tagger shared the same tagsets for all languages. Lin et al. [94] studied optimal transfer language selection but did not include sentiment transfer in their setup. In the earliest work in cross-lingual sentiment analysis, Mihalcea et al. [95] utilized resources such as bilingual dictionaries, subjectivity lexicons, and manually translated parallel corpora. Rather than relying on manually translated parallel corpora, Banea et al. [9] investigated this further with automatic translation and cross-lingual projections of subjectivity annotations. It was observed that translating the target dataset into the source language was the preferred approach to training a classifier with translations of source language data into the target language. Feng and Wan [96] employed adversarial training and multilingual language modeling. The English and French language representation models were shared, and language-specific decoders, sentiment classifiers, and language discriminators were trained jointly (DVD and books).

Earlier cross-lingual sentiment analysis research focused primarily on translation. In such scenarios, the objective was to translate the target language instances into the source language and perform inference using the source language classifier. The translated instances were also used to train a language tagger with limited resources. Kanayama et al. [97] introduced the machine translation methodology. Galeshchuk et al. [98] demonstrated the efficacy of using machine translation systems when there is insufficient data for the target language. These translations necessitate the existence of a reliable translation system. It has been demonstrated that such systems introduce semantic modifications and errors [23,99,100]. Subjectivity indicators used by humans can be lost in translation. Wan [8] merged two distinct perspectives by using Chinese and English translations for a co-training setup. For the task of bilingual lexicon extraction, Vulic and Moens [101] used language models trained on comparable corpora to identify and extract words with similar meanings. This was based on the theory that two words are identical if their top semantic word responses are identical. In lexicon-based approaches where supervised resources are scarce, such words are crucial resources.

According to Conneau et al. [102], multilingual pre-trained models utilizing shared transformers are superior to shared softmax, shared BPE, and anchor points for cross-lingual representations. Cross-domain sentiment analysis research focuses on the acquisition of shared representations across domains and is closely related to cross-lingual sentiment analysis. Li et al. [103] performed domain-independent feature extraction using domain classification and sentiment classification. Conditional domain adversarial networks [104] incorporated multilinear conditioning of features to enhance the discriminator's performance. Using multi-view representations and a six-layered transformer model with shared encoder and decoder and adversarial training, Fei and Li [105] aligned data from two distinct languages. The configuration also captures the cross-lingual and cross-domain aspects. The authors used Wikitext to train the model. Compared to the romance languages, the model's performance for Japanese was the worst.

Previous research [93,106,107] has demonstrated that selecting a hub language from the same language family or one that is closer to the target language in the language family tree facilitates knowledge transfer. Dong et al. [106] utilized fewer instances from Latin languages (French, Spanish, and Portuguese) to improve the performance of machine translation using large parallel English corpora. This also improved performance in the Germanic Dutch language. They did not, however, investigate the correlation with a distant family language. The selection of a transfer language based on the linguistic properties pertinent to the specific task is another important consideration. Lin et al. [94] identified many heuristics for choosing a transfer language. A couple of indicators include lexical overlap and the quantity of available training data.

5. Data

Our supervised resources include datasets in eight distinct languages, seven of which are official EU languages. We considered English to be the source language for all pairs of languages. Bulgarian, Croatian, Czech, Polish, Slovak, and Slovene are the target languages. A single dataset was selected for each language in the study. In Table 1, we present the sizes of the datasets' training, development, and test splits.

Table 1. Distribution of sentiment analysis datasets.

Language	Dataset	Train	Validate	Test
Bulgarian	Cinexio	5520	614	682
Croatian	Pauza	2277		1033
Czech	CSFD	63,966	13,707	13,707
English	MARC	200,000	5000	5000
Polish	all2	28,581	3572	3572
Russian	ROIMP 2012	4000	260	5500
Slovak	Reviews3	3834	661	1235
Slovene	KKS	3977	200	600

Sentiment Analysis Datasets

- Bulgarian: The Cinexio [51] dataset is composed of film reviews with 11-point star ratings: 0 (negative), 0.5, 1, ... 4.5, 5 (positive). Other meta-features included in the dataset are film length, director, actors, genre, country, and various scores.
- Croatian: Pauza [68] contains restaurant reviews from Pauza.hr4, the largest food ordering website in Croatia. Each review is assigned an opinion rating ranging from 0.5 (worst) to 6 (best). User-assigned ratings are the benchmark for labels. The dataset also contains opinionated aspects.
- Czech: The CSFD [108] dataset was influenced by Pang et al. [109]. It includes film reviews from the Czech Movie Database (<http://www.csfd.cz> accessed on 10 September 2023). Every review is classified as either positive, neutral, or negative.
- English: The Multilingual Amazon Reviews Corpus (MARC) is a large collection of Amazon reviews [110]. The corpus contains reviews written in Chinese, English, Japanese, German, French, and Spanish. Each review is assigned a maximum of five stars. Each record contains the review text, the title, the star rating, and product-related metadata.
- Polish: The Wroclaw Corpus of Consumer Reviews Sentiment [77] is a multi-domain dataset of Polish reviews from the domains of schools, medicine, hotels, and products. The texts have been annotated at both the sentence level and the text body level. The reviews are labeled as follows: [+m] represents a strong positive; [+s] represents a weak positive; [−m] represents a strong negative; [−s] represents a weak negative; [amb] represents ambiguity; and [0] represents neutrality.
- Russian: The ROMIP-12 dataset [80] is composed of news-based opinions, which are excerpts of the direct and indirect speech published in news articles. Politics,

economics, sports, and the arts are just some of the diverse subject areas covered. This dataset contains speech classified as positive, neutral, or negative.

- Slovak: The Review3 [111] is composed of customer evaluations of a variety of services. The dataset is categorized using the 1–3 and 1–5 scales.
- Slovene: The Opinion corpus of Slovene web commentaries KKS 1.001 [90] includes web commentaries on various topics (business, politics, sports, etc.) from four Slovene web portals (RtvSlo, 24ur, Finance, and Reporter). Each instance within the dataset is tagged with one of the three labels (negative, neutral, or positive).

Language models: XLM-RoBERTa is a language model that has been pre-trained in 100 different languages. We chose this model as the foundation for the fine-tuning process because all of the target languages were present during the model's pre-training.

6. Methodology

Phylogenetic similarity, typological properties, lexical overlap, and the size of the available data all contribute to the final performance of cross-lingual transfer. Lin et al. [94] posed the selection of optimal transfer languages as a ranking issue. Previous research [112] has demonstrated that single or multiple similar languages provide adequate performance in languages with limited resources. For the final performance metric, we carefully analyzed the various datasets and their presence in the training phase. We examined single-source versus multiple-source transfer in zero-shot and few-shot situations. The following training regimens were implemented.

For each study language, a dataset from the target language is:

1. Used directly to train the model. Here, the source language serves as the target language as well (like Bulgarian).
2. Combined with a single dataset from a distant language family (like English).
3. Combined with a single dataset from a different sub-branch of the same language family (like Russian, Polish, or Czech).
4. Merged with a number of low-resource language datasets (Croatian, Slovak, and Slovene).

We completed another training session by converting Bulgarian and Russian from Cyrillic to Latin. The datasets were merged with other language-specific datasets.

6.1. Model Details

Transformer-based neural networks are the current gold standard for classification tasks [19,113,114]. Taking cues from previous work [115,116], each of the fine-grained labels (1 (worst) to 5 (best)) and their corresponding coarse-grained labels (positive, neutral, and negative) were treated as two distinct tasks. The model was trained to perform both tasks simultaneously. Not all datasets in our study employ the same annotation scheme. This prompted us to conduct an annotation projection from fine-grained labels (such as 5-star or 11-star ratings) to coarse-grained labels (three-class, i.e., positive, negative, and neutral). Our model is based on the multi-task transfer learning setting [117] for training a sentiment classifier with multiple datasets. The model is a hierarchical network that performs end-to-end training and stacks two classifiers on top of one another. The encoder is shared by all classifier layers. We framed cross-lingual sentiment classification as a problem of multitask learning. We aimed to jointly learn a set of neural network parameters for classifiers in the source and target languages. This was accomplished by jointly optimizing a loss function that took coarse and fine-grained labels and resources from both languages into account. A transformer-based model fits a parameterized model to maximize the conditional probability of a target label y given a source sentence x , i.e., $z = \underset{x}{\operatorname{argmax}} p(y|x)$ given a training instance x, y . By combining training data from various sources and languages, learning is extended to multiple languages. The objective function we optimized is the sum of the conditional probabilities of different datasets

from different languages based on the representations obtained using a shared pre-trained language model.

$$\begin{aligned}
 L_{multi}(\theta) = & \underbrace{\sum_{D_s} \log P_{\theta}(y^5|x^{L1}, \theta, \omega)}_1 + \\
 & \underbrace{\sum_{D_t} \log P_{\theta}(y^3|x^{L1}, \theta, \phi)}_2 + \underbrace{\sum_{D_s} \log P_{\theta}(y^5|x^{L2}, \theta, \omega)}_3 + \\
 & \underbrace{\sum_{D_t} \log P_{\theta}(y^3|x^{L2}, \theta, \phi)}_4
 \end{aligned} \tag{1}$$

In objective Equation (1), four distinct loss terms share a common parameter, θ . In addition, language-independent classifiers share the parameters ω and ϕ , which are label-specific. The first and third terms optimize the source language loss for coarse and fine-grained labels, respectively. Similarly, the second and fourth terms enhance performance in the target language. At two points, parameterization is performed. First, we modified a PLM for sentiment classification in source and target languages jointly. Second, there are two distinct parameters for labels. The global loss function is capable of both cross-lingual and hierarchical classification.

Consider the training examples x and y^5 , where y^5 is a five-class label (1–5). The labeled five-class dataset is also realizable as a three-class dataset. One and two are mapped to the negative category, three to the neutral category, and four to the positive category, in that order. For a particular training instance, $\langle x, \langle y^3, y^5 \rangle \rangle$, where y^3 is a three-class label that can be positive, negative, or neutral based on the five-class label, i.e., 1, 2, 3, 4, 5, and the objective is to jointly maximize the conditional probability such that instances that belong to the negative class also receive a lower rating and vice versa. This is performed to optimize the model uniformly for various languages and labeled datasets. Two feed-forward neural networks with a softmax output layer define the architecture. Consequently, we have two classifiers trained on the same text but with distinct labels. The first label is coarse, while the second is fine. This is known as “pseudo-multitask learning” because two tasks are simultaneously trained on a shared representation from a single training instance.

6.2. Training

A typical dataset is divided into training, testing, and validation sets in the ratio of 8:1:3. This partitioning is preferred when the dataset used to train the system is extensive. Low-resource languages have a paucity of examples (thousands of examples). By separating the few training data samples into test and validation sets, the training set is reduced further. We therefore conducted cross-validation. K -fold cross-validation randomly divides all dataset instances into K groups, where K is a predetermined number. Folds are used to denote each group. One-fold from K is selected as the test set, while the remaining $K-1$ subsets are used for training. This process is repeated until each fold in the dataset has been utilized as a test set. The training process will therefore be repeated k times. In our case, K was assigned the value 5. The most prevalent pattern is the transfer of knowledge to a low-resource language task using data from high-resource tasks. We investigate the use of multiple datasets from low-resource languages to enhance the performance of target languages. We conducted experiments in the following environments:

1. Using only source-language data for fine-tuning. This is the conventional transfer learning setup performed by a source-language fine-tuning classifier. A zero-shot test is administered to the trained model using a test of the target language. We guided the training process using the target language’s validation set. We projected labels

- from a fine-grained class of 5 classes to a coarse-grained class of 3 classes due to the possibility that the target language dataset labels do not match the source language.
2. Fine-tuning with a single source and target language. We sampled training sets from multiple languages and jointly trained the classifier. We utilized datasets from distantly related languages and vice versa.
 3. Fine-tuning using multiple datasets derived from a single source and target language. This is a multilingual environment with multiple sources.
 4. Fine-tuning with the Latin versions of the Bulgarian and Russian datasets.

In Tables 2 and 3, we list an assortment of experiments. The first section of Table 2 depicts the combined instruction of Slavic and English languages. The second section substitutes Russian for English. The third section only utilizes language data from a single source. The following section illustrates the compilation of various low-resource languages. The next two sections combine Czech and Polish with Slavic languages with limited resources. Bulgarian is ultimately selected as the source language for the combinations. In Table 3, the study’s Latin transliterations of Bulgarian and Russian with other language combinations are displayed.

Table 2. Language pairs in various combinations for joint training.

Source Languages			
1st	2nd	3rd	4th
Bulgarian	English		
Croatian	English		
Czech	English		
Polish	English		
Russian	English		
Slovak	English		
Slovene	English		
Bulgarian	Russian		
Croatian	Russian		
Czech	Russian		
Polish	Russian		
Slovak	Russian		
Slovene	Russian		
Bulgarian			
Croatian			
Czech			
Polish			
Russian			
Slovak			
Slovene			
Croatian	Slovene		
Croatian	Slovene	Slovak	
Croatian	Slovene	Slovak	Bulgarian
Czech	Bulgarian		
Czech	Croatian		
Czech	Slovak		
Czech	Slovene		
Polish	Bulgarian		
Polish	Croatian		
Polish	Slovak		
Polish	Slovene		
Bulgarian	Croatian		
Bulgarian	Slovak		
Bulgarian	Slovene		

Table 3. Language pairs with Latin Bulgarian and Latin Russian.

1st	Source Languages 2nd	3rd
Bulgarian (Latin)		
Russian (Latin)		
Bulgarian (Latin)	Russian (Latin)	
Russian (Latin)	Croatian	
Bulgarian (Latin)	Croatian	
Russian (Latin)	Slovak	
Russian (Latin)	Slovene	
Bulgarian (Latin)	English	
Russian (Latin)	English	
Bulgarian (Latin)	Polish	
Russian (Latin)	Polish	
Bulgarian (Latin)	Czech	
Russian (Latin)	Czech	
Bulgarian (Latin)	Slovene	Slovak
Russian (Latin)	Slovene	Slovak

7. Experimental Setup

Training Details

Figure 2 illustrates the neural network's schematic diagram. The model was trained on a 24 GB NVIDIA RTX 3090, (Palit, Munich, Germany) using the PyTorch-lightning Python library. To ensure reproducibility, the standard techniques for fine-tuning as described by Devlin et al. [14] were utilized, along with a constant seed of 0. A 10% proportion of the training data was used as a warm-up alongside the Adam optimizer with a 1×10^{-5} learning rate. Each run of 5-fold cross-validation utilized a batch size of eight. The experiment was terminated early when the validation loss did not improve after three iterations. A [CLS] token was extracted from the encoder and passed through a fully connected network (FC1) three-class softmax layer for each training instance. The encoder's output is routed through a second fully connected network (FC2) and then a five-class softmax layer. The encoder's features were discarded with a probability of 0.2. We sampled a mini-batch from each of the datasets for each training step, namely four English samples and four Bulgarian samples. The instances were then sent through the network, and the error was calculated separately for five and three classes before being summed. To update the parameters, the calculated error was back-propagated throughout the network. In cases where the lengths of the two datasets did not match, the smaller dataset was duplicated to match the larger length. Slovene was the only language for which we performed any preprocessing, replacing user mentions with placeholders and removing all URLs from the text. To compute the effect of having two classifiers, a simple three-class classifier was added to the pre-trained language model as a baseline. No additional hyperparameters were altered. The results are displayed in Table 4. We observe that cascading classifiers lead to improvements over our baseline in all languages except Polish.

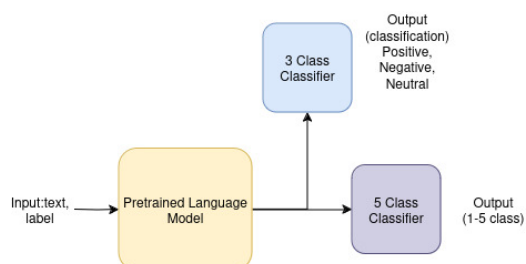


Figure 2. A neural network diagram showing the multi-task fine-tuning process on the pre-trained language model (PLM).

Table 4. Baseline three-class classification scores are averaged over five-fold runs. The standard deviation is presented in the brackets to the right. * ($p < 0.05$) denotes statistically significant results and indicates no significant improvement was achieved when combined with other languages during MTL training.

Language	Acc-3	F1-3
Bulgarian	67.80 (0.0076)	69.42 (0.0046)
Croatian	62.37 (0.004)	57.47 (0.0053)
Czech	83.82 (0.0037)	83.76 * (0.0033)
English	68.15 (0.0076)	67.85 (0.0100)
Polish	87.70 (0.0033)	87.57 * (0.0039)
Russian	71.43 (0.0013)	70.20 (0.0030)
Slovak	81.60 (0.0057)	79.75 (0.0017)
Slovene	59.13 (0.0180)	59.97 (0.0307)

8. Results and Discussion

8.1. Results

We conducted experiments on each of the datasets described in Section 5 using the methodology described in Section 6. We reported on the accuracy and macro F1 evaluation of five-class and three-class classifications. To verify the performance of one model over the other, we performed statistical testing using the almost stochastic order significance test [118,119] implemented by Del Barrio et al. [120] and the random approximation test [121]. We ran the two tests for each model, as well as the corresponding five- and three-class metrics (F1) scores.

We only have three-class scores for Czech, Russian, and Russian (Latin) because the datasets use three-label tagging schemes. Table 5 displays the results for the best-performing language pairs that are statistically significant. Each combination that outperformed the others on any of the four metrics has been listed. Bulgarian + English performed best for the five-class metric, while Czech + Bulgarian performed best for the three-class metric. Bulgarian + English and Bulgarian + Czech have no significance over each other for 3-class F1 but are statistically significant over the other combinations, like Bulgarian + Croatian. We did not find any decently performing non-Bulgarian combinations in the top 10 list. In a five-class setting, the combination of Croatian + English produced higher scores. The combination of Croatian + Czech, Croatian + Polish, and Croatian + Bulgarian data proved advantageous in a three-class setting. One observation was made that the combination of Croatian + Bulgarian performed statistically similar to Croatian + Bulgarian (Latin) data. The baseline in Czech performs statistically better as compared to all the other cases. Combining Czech with other languages does not help Czech, which is a language with a lot of training instances. Similar to Czech, adding other languages to English does not help in a 5-class setting, but we see a significant improvement in the 3-class F1 with the Czech 3-class data combination.

For the 5-star classification, we see no significant improvement for the combinations of Bulgarian (Latin) + Polish and Russian (Latin) + Polish. For the 3-class, the Polish baseline performs better, and it too belongs to a larger data instance; adding more training instances does not help. The combination of other language data with Polish during training leads to a drastic drop in performance. When Russian is combined with either Croatian or Bulgarian, the results are about the same in both metrics, and statistically, this combination does better than others. The combination of Slovak + English performs better for 5-class F1, while Croatian + Slovak + Slovene works best for 3-class F1. For Slovene, a combination with a high-resource language provided better performance compared to the baseline. The best results were obtained when Russian data were converted to the Latin script and combined with English data. Russian (Latin) + English, Russian (Latin) + Polish, and Russian (Latin) + Czech perform similarly but significantly better statistically than other combinations. However, the Latin version of the Bulgarian dataset did not outperform its Cyrillic counterpart. Bulgarian (Latin) + English was the highest scoring combination for the

5-class metric in Latinized Bulgarian. Similarly, for the 3-class metric, Bulgarian (Latin) + English, Bulgarian (Latin) + Polish, and Bulgarian (Latin) + Croatian perform significantly better than others. Three-class metrics for Czech and Polish, two high-resource languages, did not improve from their initial scores.

In the case of Polish and Czech, the addition of the English dataset had no positive effect. While all other languages, i.e., Bulgarian, Croatian, Russian, Slovak, and Slovene, had improved performance with a large English dataset, we noticed that combining Slavic languages had slightly lower performance than English combinations. It was also observed that combining multiple languages (such as Bulgarian, Croatian, Slovene, and Slovak) did not outperform the five-class metric. A further observation is that, except for Russian (Latin) and Slovak, none of the model combinations that scored over 80% on the three-class F1 value utilized English during their training phase. The performance of Bulgarian (Latin) is inferior to that of the Cyrillic version. In contrast, Russian (Latin) achieved the highest scores in each of the four metric classes. This may be due to the lack of training data in Bulgarian. When the languages are combined with English, it has resulted in superior performance in the majority of instances. Slovene was found to be the dataset/language with the lowest performance. This is because the Slovene dataset is derived from informal sources, such as news commentaries, which are noisy in nature. The Slovak dataset has fewer examples for training than the Slovene dataset, but these examples come from customer reviews. The scores that were previously reported in the respective target language are presented in Table 6.

Table 5. Classification scores are averaged over 5-fold runs. The standard deviation is presented in brackets to the right. Language pairs with Bulgarian (Latin) and Russian (Latin). The highest scores for a specific target language are indicated by numbers in bold. The statistically significant scores are indicated by the asterisk (*) ($p < 0.05$).

Target Language	Source Languages	5-Class Accuracy	5-Class F1	3-Class Accuracy	3-Class F1
Bulgarian	Bulgarian English	53.37 (0.0123)	54.60 * (0.0097)	72.73 (0.0142)	74.22 * (0.7422)
Bulgarian	Bulgarian Czech	52.18 (0.0070)	53.14 (0.0106)	72.79 (0.0098)	74.11 * (0.0081)
Croatian	Croatian English	54.12 * (0.0186)	53.80 * (0.0163)	74.07 (0.0121)	74.12 (0.0097)
Croatian	Croatian Czech	50.88 (0.0094)	50.12 (0.0251)	74.69 (0.0107)	75.82 * (0.0106)
Czech	Czech Croatian			82.29 (0.0035)	82.24 (0.0036)
English	Czech English	56.22 (0.0099)	55.36 (0.0123)	69.09 (0.0035)	69.06 * (0.0043)
English	Bulgarian (Latin) English	56.91 (0.0031)	56.78 (0.0042)	68.36 (0.0086)	68.05 (0.0103)
Polish	Bulgarian (Latin) Polish	52.34 (0.0017)	52.28 (0.0012)	87.05 (0.0028)	87.15 * (0.0016)
Polish	Russian (Latin) Polish	52.19 (0.0010)	52.15 (0.0005)	86.92 (0.0016)	87.00 * (0.0007)
Russian	Bulgarian Russian			71.84 (0.0035)	71.31 (0.0022)
Slovak	Slovak English	68.87 (0.0351)	68.03 * (0.016)	83.51 (0.0182)	82.14 (0.0076)
Slovak	Slovak Croatian Slovene	64.47 (0.0135)	58.71 (0.0441)	85.36 (0.0046)	83.44 * (0.0064)
Slovene	Slovene English			69.52 * (0.0203)	68.97 * (0.0154)
Slovene	Slovene Czech			68.24 * (0.0084)	69.56 * (0.0078)
Bulgarian (Latin)	Bulgarian (Latin) English	50.73 (0.0094)	51.76 (0.0075)	70.30 (0.0093)	72.01 (0.0071)
Russian (Latin)	Russian (Latin) English			88.14 * (0.0299)	87.95 * (0.0290)

Table 6. Previously reported results for the languages in the study. ACC—Accuracy.

Language	Metric	5-Class	3-Class	2-Class
Bulgarian [51]	MSE	0.666	0.141	
Croatian [68]	F1			91.1
Czech [122]	F1		87.08 ± 0.11	96.00 ± 0.02
English [110]	ACC	56.5		
Russian [82]	F1		72.69	87.04
Slovak [123]				
(http://ar16.library.sk/nlp4sk/webapi/analyza-sentimentu , accessed on (23 April 2023))	F1		81.5	
Slovene [90]	F1		65.7	

8.2. Error Analysis

We calculated the confusion metric for each fold for all models with the highest performance. The Bulgarian + English model for target Bulgarian misclassified more neutral and negative instances than positive ones. The same effect occurs when predicting five classes, where zero to two classes are incorrectly predicted. The number of neutral and positive classes was overestimated by the Slovak + Slovene + Croatian model. In the scenario involving five classes, labels for classes two and three were exchanged. The negative instances were assigned to the neutral and positive classes by the model trained in Czech and Bulgarian. The neutral instances were incorrectly categorized as negative and (mostly) positive, followed by the negative class drifting into neutral. The Czech with Croatian training performed the best in two cases, namely Croatian and Czech. The negative class instances in Czech were miscategorized into the neutral class. The neutral category instances were misclassified as negative and positive. The same can be said of Croatian. In Slovene, the negative was predicted to be neutral or positive. The neutral and positive comments were grouped with the negative ones.

8.3. Language Representations in XLM-RoBERTa

The training setup consists of three components: the shared encoder, training data, and classifier heads. Using the training data, the classifier heads are trained. The shared model is a black box component of the entire system for representing multiple languages. The XLM-RoBERTa model was trained using 2.5 TB data from 100 languages. The various training dataset sizes for the languages under study are listed in Table 7. The text is divided into tokens using a sentence-piece tokenizer. We conducted a simple study to examine these representations in different languages. We ran each dataset's training set through the XLM-R tokenizer. For the obtained sentence-piece tokens, we calculated the intersection of all possible language combinations. Table 8 indicates the number of common tokens in various languages. We observed that the best-performing language combinations have many shared tokens for a given target language's sentence fragments. In the case of Croatian, it shares 5075 sub-tokens with Czech, allowing it to advance under the joint-training system. We would like to note that the Slovene dataset consists of comments from a news website and is therefore highly informal and noisy. Consequently, we hypothesize that, when combined, it adversely affects the Croatian performance metric. The performance of the Czech language decreases when it is combined with other languages. When English is combined with Czech, we observe a slight improvement over the baseline and other combinations. Russian (Latin), Bulgarian (Latin), and English combinations have higher scores for Polish. In the case of Slovak, training alongside Czech led to results that were comparable to those obtained with Slovak, Slovene, and Croatian combined. Czech and Slovene shared the second-greatest number of tokens. In addition, distant high-resource languages (such as English) do not aid in the improvement of the performance of high-resource languages. Adding English data improves performance in five classes, whereas adding same-family language data improves performance in three classes. Although we observe that Bulgarian shares numerous sub-words with Russian, Czech, and English, the languages with the most shared tokens, the precise classification behavior of tokens requires further investigation.

Table 7. Data size used for training XLM-RoBERTa.

Language	Size (GB)	Tokens (Million)
Bulgarian	57.5	5487
Croatian	20.5	3297
Czech	16.3	2498
English	300.8	55,608
Polish	44.6	6490
Russian	278.0	23,408
Slovak	23.2	3525
Slovene	10.3	1669

Table 8. Languages and number of shared tokens in their training set. Croatian-Hr, Czech-Cs, Polish-Pl, Russian-Ru, Slovak-Sk, Slovene-Sv, Bulgarian-Bg, English-En. The language combinations that perform the best are indicated by bold numerals.

Languages	Hr	Cs	Pl	Ru	Sk	Bg Latin	Ru Latin	Sv	En
Bulgarian	130	235	90	2919	123	261	126	122	215
Croatian		5075	2881	432	2215	1778	3014	4420	3256
Czech			9656	1300	6035	3573	8733	10,075	15,122
Polish				690	2927	2207	5075	5417	6931
Russian					371	314	1522	733	1207
Slovak						1616	2923	3412	2774
Bulgarian (Latin)							2689	2655	2416
Russian (Latin)								5799	5702
Slovene									6352
English									

9. Conclusions

We have presented our framework for multitask cross-lingual sentiment classifier transfer. We evaluated seven official Slavic languages using a model trained with multiple language resources. We discovered that the transfer of sentiment knowledge is enhanced within the same language family, i.e., the closer the language, the easier the transfer, given a large dataset. We also discovered that a large training dataset from a distant language family can outperform smaller datasets from similar languages. Consequently, datasets from the same language family and distant language families can be utilized to combat the issue of inadequate resources. Furthermore, we transform the fine-grained sentiment problem into a 3-class problem by multi-task training, leading to improved scores.

A potential study limitation may arise from the limited amount of the dataset utilized to evaluate the effectiveness of the trained model. Although substitutes cannot replace the test set in the target language, low-resourced languages inherently have fewer supervised examples. This also leads us to the conclusion that ensuring the presence of supervised resources in each language is crucial, and efforts should be focused on providing datasets and corpora for languages that lack sufficient resources. Furthermore, these findings may be specific to classification tasks and may not be generalizable to other tasks, although this can only be confirmed by empirical verification.

This study suggests that languages with limited data can depend on datasets from other languages that have more resources. Our work employs the XLM-R language model, which can be readily trained on a consumer-grade GPU (or Google Colab). A recent study [124] has demonstrated that large language models such as Llama 2 and ChatGPT, while exhibiting comparable performance, possess substantial hardware demands and deployment constraints that are absent in our situation.

Future work can be divided into four distinct paths. Firstly, the performance enhancement might be computed by adding more datasets from high-resourced languages. Additionally, an examination of several language families could be conducted to determine if the inclusion of a specific distant family language enhances the overall performance. Furthermore, another element related to the dataset is the quality of the text and annotations. A study conducted independently could calculate the information value contained within a specific supervised instance and its impact on transfer learning. Lastly, our hypothesis is that the process of classification is indirectly influenced by the sharing of sub-word tokens and, as a result, certain dataset combinations facilitate sentiment transmission in languages that are part of the same language family. We intend to empirically verify this assertion through additional experiments.

Author Contributions: Conceptualization, G.T. and N.M.P.; methodology, G.T. and N.M.P.; software, G.T.; validation, G.T. and N.M.P.; formal analysis, G.T. and N.M.P.; investigation, G.T.; resources, G.T.; data curation, G.T. and N.M.P.; writing—original draft preparation, G.T.; writing—review and editing, G.T., N.M.P. and M.T.; visualization, G.T.; supervision, N.M.P.; project administration, M.T.; funding acquisition, N.M.P. and M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 812997 (H2020-MSCA-ITN 2018-812997), with the project name CLEOPATRA (Cross-lingual Event-centric Open Analytics Research Academy).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

ANN	artificial neural network
BERT	Bidirectional Encoder Representations from Transformers
CLSA	Cross-lingual sentiment analysis
IR	Indo-European
NERC	Named Entity Recognition and Classification
NLP	Natural language processing
PLM	Pre-trained language model
PMI	pointwise mutual information
SO	Semantic orientation
SVM	Support vector machines
QA	Question Answering

References

1. Go, A.; Bhayani, R.; Huang, L. *Twitter Sentiment Classification Using Distant Supervision*; CS224N Project Report; Stanford University: Stanford, CA, USA, 2009.
2. Nakov, P.; Rosenthal, S.; Kozareva, Z.; Stoyanov, V.; Ritter, A.; Wilson, T. SemEval-2013 Task 2: Sentiment Analysis in Twitter. In Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, GA, USA, 14–15 June 2013; pp. 14–15.
3. Saif, H.; Fernández, M.; He, Y.; Alani, H. Evaluation datasets for Twitter sentiment analysis: A survey and a new dataset, the STS-Gold. In Proceedings of the 1st Interantional Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013), Turin, Italy, 3 December 2013.
4. Wilson, T.; Wiebe, J.; Hoffmann, P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Comput. Linguist.* **2009**, *35*, 399–433. [[CrossRef](#)]
5. Agarwal, A.; Xie, B.; Vovsha, I.; Rambow, O.; Passonneau, R.J. Sentiment analysis of twitter data. In Proceedings of the Workshop on Language in Social Media (LSM 2011), Portland, OR, USA, 23 June 2011; pp. 30–38.
6. Wilson, T.; Wiebe, J.; Hoffmann, P. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, BC, Canada, 6–8 October 2005; pp. 347–354.
7. Socher, R.; Lin, C.C.Y.; Ng, A.Y.; Manning, C.D. Parsing natural scenes and natural language with recursive neural networks. In Proceedings of the ICML, Bellevue, WA, USA, 28 June–2 July 2011.
8. Wan, X. Co-training for cross-lingual sentiment classification. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Singapore, 2–7 August 2009*; Su, K., Su, J., Wiebe, J., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2009; pp. 235–243.
9. Banea, C.; Mihalcea, R.; Wiebe, J.; Hassan, S. Multilingual subjectivity analysis using machine translation. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, HI, USA, 25–27 October 2008; pp. 127–135.
10. Balahur, A.; Turchi, M. Multilingual sentiment analysis using machine translation? In Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, Jeju, Republic of Korea, 12 July 2012; pp. 52–60.

11. A.R., B.; Joshi, A.; Bhattacharyya, P. Cross-Lingual Sentiment Analysis for Indian Languages using Linked WordNets. In Proceedings of the COLING 2012: Posters, Mumbai, India, 8–15 December 2012; pp. 73–82.
12. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
13. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep contextualized word representations. In Proceedings of the NAACL-HLT, New Orleans, LA, USA, 1–6 June 2018; pp. 2227–2237.
14. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186. [[CrossRef](#)]
15. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
16. Grancharova, M.; Dalianis, H. Applying and Sharing pre-trained BERT-models for Named Entity Recognition and Classification in Swedish Electronic Patient Records. In Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa), Reykjavik, Iceland, 31 May–2 June 2021; pp. 231–239.
17. Wang, Z.; Ng, P.; Ma, X.; Nallapati, R.; Xiang, B. Multi-passage BERT: A Globally Normalized BERT Model for Open-domain Question Answering. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 5878–5882. [[CrossRef](#)]
18. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *Advances in Neural Information Processing Systems, Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019*; Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.
19. Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Zhao, T. SMART: Robust and Efficient Fine-Tuning for Pre-trained Natural Language Models through Principled Regularized Optimization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 6–8 July 2020; pp. 2177–2190. [[CrossRef](#)]
20. Das, A.; Sarkar, S. A survey of the model transfer approaches to cross-lingual dependency parsing. *ACM Trans. Asian Low-Resour. Lang. Inf. Process. (TALLIP)* **2020**, *19*, 67. [[CrossRef](#)]
21. Chen, X.; Awadallah, A.H.; Hassan, H.; Wang, W.; Cardie, C. Multi-Source Cross-Lingual Model Transfer: Learning What to Share. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 3098–3112. [[CrossRef](#)]
22. Kandula, H.; Min, B. Improving Cross-Lingual Sentiment Analysis via Conditional Language Adversarial Nets. In Proceedings of the Third Workshop on Computational Typology and Multilingual NLP, Online, 10 June 2021; pp. 32–37. [[CrossRef](#)]
23. Pintu Lohar, M.P.; Way, A. Building English-to-Serbian Machine Translation System for IMDb Movie Reviews. In Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing BSNLP 2019, Florence, Italy, 2 August 2019; pp. 105–113.
24. Chen, X.; Sun, Y.; Athiwaratkun, B.; Cardie, C.; Weinberger, K. Adversarial deep averaging networks for cross-lingual sentiment classification. *Trans. Assoc. Comput. Linguist.* **2018**, *6*, 557–570. [[CrossRef](#)]
25. Crystal, D. *A Dictionary of Linguistics and Phonetics*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
26. Sussex, R.; Cubberley, P. *The Slavic Languages*; Cambridge University Press: Cambridge, UK, 2006.
27. Golubović, J.; Gooskens, C. Mutual intelligibility between West and South Slavic Languages. *Russ. Linguist.* **2015**, *39*, 351–373. [[CrossRef](#)]
28. Townsend, C.E.; Janda, L.A. *Common and Comparative Slavic: Phonology and Inflection: With Special Attention to Russian, Polish, Czech, Serbo-Croatian, Bulgarian*; Slavica Pub: Bloomington, IN, USA, 1996.
29. Turney, P. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 417–424. [[CrossRef](#)]
30. Kim, S.M.; Hovy, E. Determining the sentiment of opinions. In Proceedings of the COLING 2004: 20th International Conference on Computational Linguistics, Geneva, Switzerland, 23–27 August 2004; pp. 1367–1373.
31. Polanyi, L.; Zaenen, A. Contextual valence shifters. In *Computing Attitude and Affect in Text: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1–10.
32. Riloff, E.; Wiebe, J. Learning extraction patterns for subjective expressions. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, Sapporo, Japan, 11–12 July 2003; pp. 105–112.
33. Esuli, A.; Sebastiani, F. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06), Genoa, Italy, 22–28 May 2006.
34. Stone, P.J.; Hunt, E.B. A computer approach to content analysis: Studies using the general inquirer system. In Proceedings of the Spring Joint Computer Conference, Detroit, MI, USA, 21–23 May 1963; pp. 241–256.
35. Cambria, E.; Speer, R.; Havasi, C.; Hussain, A. Senticnet: A publicly available semantic resource for opinion mining. In Proceedings of the 2010 AAAI Fall Symposium Series, Arlington, VA, USA, 11–13 November 2010.
36. Nielsen, F.Å. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In Proceedings of the Workshop on ‘Making Sense of Microposts’: Big Things Come in Small Packages, Heraklion, Crete, Greece, 30 May 2011; pp. 93–98.

37. Mullen, T.; Collier, N. Sentiment analysis using support vector machines with diverse information sources. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 25–26 July 2004; pp. 412–418.
38. McDonald, R.; Hannan, K.; Neylon, T.; Wells, M.; Reynar, J. Structured Models for Fine-to-Coarse Sentiment Analysis. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 25–27 June 2007; pp. 432–439.
39. Paulus, R.; Socher, R.; Manning, C.D. Global Belief Recursive Neural Networks. In *Advances in Neural Information Processing Systems, Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; Volume 27.
40. Read, J.; Carroll, J. Weakly Supervised Techniques for Domain-Independent Sentiment Classification. In Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion (TSA '09), New York, NY, USA, 6 November 2009; pp. 45–52. [[CrossRef](#)]
41. Moraes, R.; Valiati, J.A.F.; Gavião Neto, W.P. Document-Level Sentiment Classification: An Empirical Comparison between SVM and ANN. *Expert Syst. Appl.* **2013**, *40*, 621–633. [[CrossRef](#)]
42. Huang, E.H.; Socher, R.; Manning, C.D.; Ng, A.Y. Improving word representations via global context and multiple word prototypes. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Jeju Island, Republic of Korea, 8–14 July 2012; pp. 873–882.
43. Socher, R.; Huval, B.; Manning, C.D.; Ng, A.Y. Semantic compositionality through recursive matrix-vector spaces. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Republic of Korea, 8–14 July 2012; pp. 1201–1211.
44. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.Y.; Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1631–1642.
45. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A Convolutional Neural Network for Modelling Sentences. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 655–665. [[CrossRef](#)]
46. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1746–1751. [[CrossRef](#)]
47. Wang, X.; Liu, Y.; Sun, C.; Wang, B.; Wang, X. Predicting Polarities of Tweets by Composing Word Embeddings with Long Short-Term Memory. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 1343–1353. [[CrossRef](#)]
48. Dong, L.; Wei, F.; Tan, C.; Tang, D.; Zhou, M.; Xu, K. Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 49–54. [[CrossRef](#)]
49. Wang, J.; Yu, L.C.; Lai, K.R.; Zhang, X. Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, 7–12 August 2016; pp. 225–230. [[CrossRef](#)]
50. Wang, X.; Jiang, W.; Luo, Z. Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts. In Proceedings of the Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–16 December 2016; pp. 2428–2437.
51. Kapukaranov, B.; Nakov, P. Fine-grained sentiment analysis for movie reviews in Bulgarian. In Proceedings of the International Conference Recent Advances in Natural Language Processing, Hissar, Bulgaria, 7–9 September 2015; pp. 266–274.
52. Georgieva-Trifonova, T.; Stefanova, M.; Kalchev, S. Customer Feedback Text Analysis for Online Stores Reviews in Bulgarian. *IAENG Int. J. Comput. Sci.* **2018**, *45*, 560–568.
53. Lazarova, G.; Koychev, I. Semi-supervised multi-view sentiment analysis. In *Computational Collective Intelligence*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 181–190.
54. Osenova, P.; Simov, K.I. The Political Speech Corpus of Bulgarian. In Proceedings of the LREC, Online, 21–27 May 2012; pp. 1744–1747.
55. Smailović, J.; Kranjc, J.; Grčar, M.; Žnidaršič, M.; Mozetič, I. Monitoring the Twitter sentiment during the Bulgarian elections. In Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, France, 19–21 October 2015; pp. 1–10.
56. Hristova, G. Text Analytics in Bulgarian: An Overview and Future Directions. *Cybern. Inf. Technol.* **2021**, *21*, 3–23. [[CrossRef](#)]
57. Steinberger, J.; Ebrahim, M.; Ehrmann, M.; Hurriyetoglu, A.; Kabadjov, M.; Lenkova, P.; Steinberger, R.; Tanev, H.; Vázquez, S.; Zavarella, V. Creating sentiment dictionaries via triangulation. *Decis. Support Syst.* **2012**, *53*, 689–694. [[CrossRef](#)]
58. Veselovská, K. Sentence-level sentiment analysis in czech. In Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, Craiova, Romania, 13–15 June 2012; pp. 1–4.
59. Habernal, I.; Brychcín, T. Semantic spaces for sentiment analysis. In Proceedings of the International Conference on Text, Speech and Dialogue, Pilsen, Czech Republic, 1–5 September 2013; pp. 484–491.

60. Čano, E.; Bojar, O. Sentiment Analysis of Czech Texts: An Algorithmic Survey. In Proceedings of the 11th International Conference on Agents and Artificial Intelligence, Prague, Czech Republic, 19–21 February 2019; Rocha, A.P., Steels, L., van den Herik, H.J., Eds.; SciTePress: Setúbal, Portugal, 2019; pp. 973–979. [\[CrossRef\]](#)
61. Klouda, I.K.; Langr, L.; Ing, D.V. Product Review Sentiment Analysis in the Czech Language Student. Bachelor's Thesis, Czech Technical University in Prague, Prague, Czech Republic, 2019.
62. Sido, J.; Prazák, O.; Pribán, P.; Pasek, J.; Seják, M.; Konopík, M. Czert—Czech BERT-like Model for Language Representation. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), Online, 1–3 September 2021; Angelova, G., Kunilovskaya, M., Mitkov, R., Nikolova-Koleva, I., Eds.; INCOMA Ltd.: Hyderabad, India, 2021; pp. 1326–1338.
63. Straka, M.; Náplava, J.; Straková, J.; Samuel, D. RobeCzech: Czech RoBERTa, a monolingual contextualized language representation model. *arXiv* **2021**, arXiv:2105.11314.
64. Vysušilová, P.; Straka, M. Sentiment Analysis (Czech Model). LINDAT/CLARIAH-CZ Digital Library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. 2021. Available online: <http://hdl.handle.net/11234/1-4601> (accessed on 23 April 2023).
65. Agić, Ž.; Ljubešić, N.; Tadić, M. Towards sentiment analysis of financial texts in croatian. *Bull. Mark.* **2010**, *143*, 69.
66. Agic, Z.; Merkler, D. Rule-Based Sentiment Analysis in Narrow Domain: Detecting Sentiment in Daily Horoscopes Using Sentiscope. In Proceedings of the 2nd Workshop on Sentiment Analysis Where AI Meets Psychology, Mumbai, India, 15 December 2012; pp. 115–124.
67. Jakopović, H.; Mikelić Preradović, N. Identifikacija Online Imidža Organizacija Temeljem Analize Sentimenata Korisnički Generiranog Sadržaja na Hrvatskim Portalima. *Med. Istraž.* **2016**, *22*, 63–82. [\[CrossRef\]](#)
68. Glavaš, G.; Korenčić, D.; Šnajder, J. Aspect-oriented opinion mining from user reviews in Croatian. In Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing, Sofia, Bulgaria, 8–9 August 2013; pp. 18–23.
69. Mozetič, I.; Grčar, M.; Smailović, J. Multilingual Twitter sentiment classification: The role of human annotators. *PLoS ONE* **2016**, *11*, e0155036. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Rotim, L.; Šnajder, J. Comparison of short-text sentiment analysis methods for croatian. In Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing, Valencia, Spain, 4 April 2017; pp. 69–75.
71. Robnik-Šikonja, M.; Reba, K.; Mozetič, I. Cross-lingual transfer of sentiment classifiers. Slovenščina 2.0: Empirical. *Appl. Interdiscip. Res.* **2021**, *9*, 1–25. [\[CrossRef\]](#)
72. Lula, P.; Wójcik, K. Sentiment analysis of consumer opinions written in Polish. *Econ. Manag.* **2011**, *16*, 1286–1291.
73. Haniewicz, K.; Rutkowski, W.; Adamczyk, M.; Kaczmarek, M. Towards the lexicon-based sentiment analysis of polish texts: Polarity lexicon. In Proceedings of the International Conference on Computational Collective Intelligence, Craiova, Romania, 11–13 September 2013; pp. 286–295.
74. Rybiński, K. Political sentiment analysis of press freedom. *Stud. Medioznawcze* **2018**, *2018*, 31–48. <https://bibliotekanauki.pl/articles/484372> [\[CrossRef\]](#)
75. Zaško-Zielińska, M.; Piasecki, M.; Szpakowicz, S. A large wordnet-based sentiment lexicon for Polish. In Proceedings of the International Conference Recent Advances in Natural Language Processing, Hissar, Bulgaria, 7–9 September 2015; pp. 721–730.
76. Bartusiak, R.; Augustyniak, L.; Kajdanowicz, T.; Kazienko, P. Sentiment Analysis for Polish Using Transfer Learning Approach. In Proceedings of the 2015 Second European Network Intelligence Conference, Karlskrona, Sweden, 21–22 September 2015; pp. 53–59. [\[CrossRef\]](#)
77. Kocoń, J.; Zaško-Zielińska, M.; Miłkowski, P. Multi-level analysis and recognition of the text sentiment on the example of consumer opinions. In Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP 2019), Varna, Bulgaria, 2–4 September 2019.
78. Wawer, A.; Sobiczewska, J. Predicting Sentiment of Polish Language Short Texts. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), Varna, Bulgaria, 2–4 September 2019; pp. 1321–1327.
79. Kuznetsova, E.S.; Loukachevitch, N.V.; Chetviorkin, I.I. Testing rules for a sentiment analysis system. In Proceedings of the International Conference Dialog, Metz, France, 22–24 August 2013; Volume 2, pp. 71–80.
80. Chetviorkin, I.; Loukachevitch, N.V. Evaluating Sentiment Analysis Systems in Russian. In Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing, BSNLP@ACL 2013, Sofia, Bulgaria, 8–9 August 2013; Piskorski, J., Pivovarov, L., Tanev, H., Yangarber, R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 12–17.
81. Golubev, A.; Loukachevitch, N.V. Improving Results on Russian Sentiment Datasets. *arXiv* **2020**, arXiv:2007.14310.
82. Golubev, A.; Loukachevitch, N.V. Transfer Learning for Improving Results on Russian Sentiment Datasets. *arXiv* **2021**, arXiv:2107.02499.
83. Smetanin, S.; Komarov, M. Deep transfer learning baselines for sentiment analysis in Russian. *Inf. Process. Manag.* **2021**, *58*, 102484. [\[CrossRef\]](#)
84. Machová, K.; Mikula, M.; Gao, X.; Mach, M. Lexicon-based Sentiment Analysis Using the Particle Swarm Optimization. *Electronics* **2020**, *9*, 1317. [\[CrossRef\]](#)
85. Bučar, J.; Povh, J.; Žnidaršič, M. Sentiment classification of the Slovenian news texts. In Proceedings of the 9th International Conference on Computer Recognition Systems CORES, Wrocław, Poland, 25–27 May 2015; Springer: Berlin/Heidelberg, Germany, 2016; pp. 777–787.

86. Bučar, J. Manually Sentiment Annotated Slovenian News Corpus SentiNews 1.0. Slovenian Language Resource Repository CLARIN.SI. 2017. Available online: <http://hdl.handle.net/11356/1110> (accessed on 23 April 2023).
87. Žitnik, S. Slovene Corpus for Aspect-Based Sentiment Analysis—SentiCoref 1.0. Slovenian Language Resource Repository CLARIN.SI. 2019. Available online: <https://www.clarin.si/repository/xmlui/handle/11356/1285> (accessed on 21 April 2023).
88. Pelicon, A.; Pranjić, M.; Miljković, D.; Škrlj, B.; Pollak, S. Sentiment Annotated Dataset of Croatian News. Slovenian Language Resource Repository CLARIN.SI. 2020. Available online: <https://www.clarin.si/repository/xmlui/handle/11356/1342> (accessed on 20 April 2023).
89. Pelicon, A.; Pranjić, M.; Miljković, D.; Škrlj, B.; Pollak, S. Zero-Shot Learning for Cross-Lingual News Sentiment Classification. *Appl. Sci.* **2020**, *10*, 5993. [[CrossRef](#)]
90. Kadunc, K.; Robnik-Šikonja, M. Opinion corpus of Slovene Web Commentaries KKS 1.001. Slovenian Language Resource Repository CLARIN.SI. 2017. Available online: <https://www.clarin.si/repository/xmlui/handle/11356/1115> (accessed on 20 April 2023).
91. Ljubešič, N.; Fišer, D.; Erjavec, T.; Šulc, A. Offensive language dataset of Croatian, English and Slovenian comments FRENK 1.1. Slovenian Language Resource Repository CLARIN.SI. 2021. Available online: <https://www.clarin.si/repository/xmlui/handle/11356/1462> (accessed on 20 April 2023).
92. Evkoski, B.; Pelicon, A.; Mozetič, I.; Ljubešič, N.; Kralj Novak, P. Slovenian Twitter Dataset 2018–2020 1.0. Slovenian language resource repository CLARIN.SI. 2021. Available online: <https://www.clarin.si/repository/xmlui/handle/11356/1423> (accessed on 20 April 2023).
93. Cotterell, R.; Heigold, G. Cross-lingual Character-Level Neural Morphological Tagging. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, 9–11 September 2017; Palmer, M., Hwa, R., Riedel, S., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 748–759. [[CrossRef](#)]
94. Lin, Y.; Chen, C.; Lee, J.; Li, Z.; Zhang, Y.; Xia, M.; Rijhwani, S.; He, J.; Zhang, Z.; Ma, X.; et al. Choosing Transfer Languages for Cross-Lingual Learning. In Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, 28 July–2 August 2019; Korhonen, A., Traum, D.R., Màrquez, L., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; Volume 1: Long Papers, pp. 3125–3135. [[CrossRef](#)]
95. Mihalcea, R.; Banea, C.; Wiebe, J. Learning Multilingual Subjective Language via Cross-Lingual Projections. In Proceedings of the ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, 23–30 June 2007; Carroll, J.A., van den Bosch, A., Zaenen, A., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2007.
96. Feng, Y.; Wan, X. Towards a unified end-to-end approach for fully unsupervised cross-lingual sentiment analysis. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), Hong Kong, China, 3–4 November 2019; pp. 1035–1044.
97. Kanayama, H.; Nasukawa, T.; Watanabe, H. Deeper Sentiment Analysis Using Machine Translation Technology. In Proceedings of the COLING 2004: 20th International Conference on Computational Linguistics, Geneva, Switzerland, 23–27 August 2004; pp. 494–500.
98. Galeshchuk, S.; Qiu, J.; Jourdan, J. Sentiment Analysis for Multilingual Corpora. In Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing, BSNLP@ACL 2019, Florence, Italy, 2 August 2019; Erjavec, T., Marcinczuk, M., Nakov, P., Piskorski, J., Pivovarov, L., Snajder, J., Steinberger, J., Yangarber, R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 120–125. [[CrossRef](#)]
99. Lohar, P.; Afli, H.; Way, A. Maintaining Sentiment Polarity of Translated User Generated Content. *Prague Bull. Math. Linguist.* **2017**, *108*, 73–84. [[CrossRef](#)]
100. Lohar, P.; Afli, H.; Way, A. Balancing Translation Quality and Sentiment Preservation. In Proceedings of the 13th Conference of the Association for Machine Translation in the Americas, Boston, MA, USA, 17–21 March 2018; pp. 81–88.
101. Vulic, I.; Moens, M. Cross-Lingual Semantic Similarity of Words as the Similarity of Their Semantic Word Responses. In Proceedings of the 2013 Conference of the North American Chapter of the Association of Computational Linguistics: Human Language Technologies, Atlanta, GA, USA, 9–14 June 2013; Vanderwende, L., Daumé, H., III, Kirchhoff, K., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 106–116.
102. Conneau, A.; Wu, S.; Li, H.; Zettlemoyer, L.; Stoyanov, V. Emerging Cross-lingual Structure in Pretrained Language Models. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, 5–10 July 2020; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 6022–6034. [[CrossRef](#)]
103. Li, Z.; Zhang, Y.; Wei, Y.; Wu, Y.; Yang, Q. End-to-End Adversarial Memory Network for Cross-domain Sentiment Classification. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, 19–25 August 2017; Sierra, C., Ed.; AAAI Press: Washington, DC, USA, 2017; pp. 2237–2243. [[CrossRef](#)]
104. Long, M.; CAO, Z.; Wang, J.; Jordan, M.I. Conditional Adversarial Domain Adaptation. In *Advances in Neural Information Processing Systems, Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2018; Volume 31.

105. Fei, H.; Li, P. Cross-lingual unsupervised sentiment classification with multi-view transfer learning. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 5759–5771.
106. Dong, D.; Wu, H.; He, W.; Yu, D.; Wang, H. Multi-task learning for multiple language translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 1723–1732.
107. Johnson, M.; Schuster, M.; Le, Q.V.; Krikun, M.; Wu, Y.; Chen, Z.; Thorat, N.; Viégas, F.; Wattenberg, M.; Corrado, G.; et al. Google’s multilingual neural machine translation system: Enabling zero-shot translation. *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 339–351. [[CrossRef](#)]
108. Habernal, I.; Ptáček, T.; Steinberger, J. Sentiment analysis in czech social media using supervised machine learning. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, GA, USA, 14 June 2013; pp. 65–74.
109. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), Philadelphia, PA, USA, 6–7 July 2002; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; pp. 79–86. [[CrossRef](#)]
110. Keung, P.; Lu, Y.; Szarvas, G.; Smith, N.A. The Multilingual Amazon Reviews Corpus. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 4563–4568. [[CrossRef](#)]
111. Pecar, S.; Simko, M.; Bielikova, M. Improving Sentiment Classification in Slovak Language. In Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing, Florence, Italy, 29 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019.
112. McDonald, R.; Petrov, S.; Hall, K. Multi-Source Transfer of Delexicalized Dependency Parsers. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 27–31 July 2011; pp. 62–72.
113. Thongtan, T.; Phienthrakul, T. Sentiment Classification Using Document Embeddings Trained with Cosine Similarity. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Florence, Italy, 28 July–2 August 2019; pp. 407–414. [[CrossRef](#)]
114. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 16–20 November 2020; pp. 38–45. [[CrossRef](#)]
115. Wu, Z.; Saito, S. HiNet: Hierarchical Classification with Neural Network. *arXiv* **2017**, arXiv:1705.11105.
116. Thakkar, G.; Preradovic, N.M.; Tadic, M. Multi-task Learning for Cross-Lingual Sentiment Analysis. In Proceedings of the 2nd International Workshop on Cross-Lingual Event-Centric Open Analytics Co-Located with the 30th The Web Conference (WWW 2021), Ljubljana, Slovenia, 12 April 2021; Demidova, E., Hakimov, S., Winters, J., Tadic, M., Eds.; Sun SITE Central Europe: Aachen, Germany, 2021; Volume 2829, CEUR Workshop Proceedings, pp. 76–84.
117. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P.P. Natural Language Processing (Almost) from Scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
118. Dror, R.; Shlomov, S.; Reichart, R. Deep Dominance—How to Properly Compare Deep Neural Models. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; Korhonen, A., Traum, D., Màrquez, L., Eds.; pp. 2773–2785. [[CrossRef](#)]
119. Ulmer, D.; Hardmeier, C.; Frelsen, J. deep-significance-Easy and Meaningful Statistical Significance Testing in the Age of Neural Networks. *arXiv* **2022**, arXiv:2204.06815.
120. Del Barrio, E.; Cuesta-Albertos, J.A.; Matrán, C. An optimal transportation approach for assessing almost stochastic order. In *The Mathematics of the Uncertain*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 33–44.
121. Yeh, A. More accurate tests for the statistical significance of result differences. In Proceedings of the COLING 2000 Volume 2: The 18th International Conference on Computational Linguistics, Saarbrücken, Germany, 31 July–4 August 2000.
122. Příbáň, P.; Steinberger, J. Are the Multilingual Models Better? Improving Czech Sentiment with Transformers. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), Online, 1–3 September 2021; pp. 1138–1149.
123. Pikuliak, M.; Grivalský, Š.; Konôpka, M.; Blšták, M.; Tamajka, M.; Bachratý, V.; Simko, M.; Balážik, P.; Trnka, M.; Uhlárik, F. SlovakBERT: Slovak Masked Language Model. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, 7–11 December 2022; Goldberg, Y., Kozareva, Z., Zhang, Y., Eds.; pp. 7156–7168. [[CrossRef](#)]
124. Příbáň, P.; Šmíd, J.; Steinberger, J.; Mištera, A. A comparative study of cross-lingual sentiment analysis. *Expert Syst. Appl.* **2024**, *247*, 123247. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.