

Article

Keystroke Dynamics as a Language Profiling Tool: Identifying Mother Tongue of Unknown Internet Users

Ioannis Tsimperidis ^{1,*}, Denitsa Grunova ¹, Soumen Roy ² and Lefteris Moussiades ¹

¹ Department of Computer Science, International Hellenic University, 65404 Kavala, Greece; ntgkrou@teiemt.gr (D.G.); lmous@cs.ihu.gr (L.M.)

² Department of Computer Science and Engineering, University of Calcutta, Acharya Prafulla Chandra Roy Siksha Prangan, JD-2, Sector-III, Saltlake City, Kolkata 700106, India; soumen.roy_2007@yahoo.co.in

* Correspondence: tsimperidis@cs.ihu.gr; Tel.: +30-2510462321

Abstract: Understanding the distinct characteristics of unidentified Internet users is helpful in various contexts, including digital forensics, targeted advertising, and user interaction with services and systems. Keystroke dynamics (KD) enables the analysis of data derived from a user's typing behaviour on a keyboard as one approach to obtain such information. This study conducted experiments on a developed dataset that recorded samples of typing in five different mother tongues to determine Internet users' mother tongue. Based on only a few KD features and machine learning techniques, 82% accuracy was achieved in recognising an unknown user's mother tongue. This research highlights the potential for KD as a reliable method for identifying the mother tongue of Internet users, with implications for various applications such as improving digital forensic investigations, targeted advertising strategies, and optimising user experiences with online services.

Keywords: mother tongue determination; keystroke dynamics; user classification; machine learning



Citation: Tsimperidis, I.; Grunova, D.; Roy, S.; Moussiades, L. Keystroke Dynamics as a Language Profiling Tool: Identifying Mother Tongue of Unknown Internet Users. *Telecom* **2023**, *4*, 369–377. <https://doi.org/10.3390/telecom4030021>

Academic Editor: Sotirios K. Goudos

Received: 28 April 2023

Revised: 23 June 2023

Accepted: 28 June 2023

Published: 3 July 2023



Copyright: © 2023 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/>).

1. Introduction

The definition of “mother tongue” varies across different sources and continues to evolve to encompass the nuances of language use by individuals. One commonly accepted definition is that it refers to the language a person learns through their interactions with family and society during the early years of their life [1]. According to UNESCO, there are over 7000 known mother tongues in the world, with approximately 3000 of them facing the risk of extinction in the near future [2]. Hundreds of millions of people speak some mother tongues, such as Chinese, Hindi, Spanish, English, Arabic, Japanese, and Russian. Others like Turkish, Korean, French, German, Bengali, and Italian also have many speakers globally.

As the Internet continues to expand its reach across the globe, becoming accessible even to less economically developed populations, the diversity of languages used in the digital world is also increasing. English used to be the dominant language on the Internet; however, this is changing as more non-English speakers access online resources. The ability of people from different countries and cultures to communicate and share information in their mother tongue has resulted in a proliferation of diverse languages being used on the Internet.

The exponential increase in the global Internet user base has expanded the market reach for companies; however, the diverse linguistic landscape online presents a formidable challenge in effectively engaging with people from different language backgrounds. Communicating and marketing successfully with individuals speaking different languages is crucial to understanding and utilising their mother tongue. The mother tongue of Internet users serves as a defining characteristic, and knowledge of this aspect can be leveraged in various ways to enhance business strategies and user experiences.

Understanding a user's mother tongue can have practical applications in various domains. For instance, Internet service providers (ISPs) can customise their services to align with users' language preferences, thereby enhancing user experience. Similarly, online businesses can improve their targeted advertising strategies by considering customers' mother tongues, as different language preferences may entail distinct consumer needs. Additionally, in digital forensics, knowledge of a suspect's mother tongue can serve as valuable evidence in criminal investigations, allowing investigators to narrow the pool of potential suspects. Investigators often need to sift through substantial amounts of data and digital evidence to identify perpetrators when dealing with cybercrimes. Information about the suspect's mother tongue can help focus investigative efforts on a smaller subset of suspects. Another practical application is automatically modifying the interface of a website or application based on the user's mother tongue, making it more accessible and user-friendly, thus enhancing user satisfaction and engagement. Overall, leveraging the knowledge of a user's mother tongue can have diverse applications in fields such as ISP services, targeted advertising, digital forensics, and website/application design to enhance user experiences and streamline processes.

This paper aims to identify the mother tongue of an unknown user by leveraging KD, which involves analysing how individuals type on a keyboard, including factors such as timing, speed, and pressure of key usage. KD has been utilised in various applications such as authentication, biometrics, and user behaviour analysis [3], and offers a promising avenue for research in diverse fields, ranging from cybersecurity to psychology. The ability to identify individuals based on their unique typing patterns has significant implications for authentication, user behaviour analysis, and other related areas [4].

The remainder of this paper is organised as follows: The next section provides a comprehensive review of the relevant literature that contextualises the topic of this study. Subsequently, the methodology employed in this research is described and analysed in detail. Next, the findings about identifying the mother tongue of unknown users, utilising five different machine learning models, are presented. Finally, the paper concludes with a discussion of the potential applications of this research and suggestions for future extensions of the study.

2. Background

The term "mother tongue" refers to the language an individual learns from birth or acquires from their family and community during their formative years. It serves as their primary mode of communication and thought, and they are typically most proficient in using this language. However, the concept of mother tongue has evolved, leading to varying interpretations.

Some experts contend that the language spoken by an individual's biological mother is the true mother tongue, while others argue that it encompasses the language of the immediate environment. A study [5] asserts that children who receive education in their mother tongue are more likely to excel academically and achieve better long-term educational outcomes. This study defines the mother tongue as "the language a child hears at home and in the community from birth," emphasising the importance of preserving and nurturing the mother tongue in bilingual education.

Similarly, another study [6] provides a comprehensive overview of bilingual education and bilingualism, defining the mother tongue as the first language a child learns, typically spoken at home. It underscores the significance of maintaining and fostering the mother tongue in bilingual education, as it can facilitate academic success and social integration.

In conclusion, the concept of mother tongue has different interpretations, ranging from the language spoken by one's biological mother to the language of the immediate environment. However, scholars such as Cummins and Baker emphasise the importance of preserving and developing the mother tongue in bilingual education, as it can positively impact academic achievement and social integration.

The concept of a mother tongue, also known as a first language (L1), has been defined in various ways by scholars from different disciplines. In linguistics, it is often defined as the language that a person learns naturally from birth or early childhood and has a high level of proficiency in [7]. In education, the mother tongue can also refer to the language used as a medium of instruction in schools and the language of instruction in multilingual contexts [8]. However, the definition and concept of the mother tongue have evolved, and there are different views and perspectives on its content. For example, some scholars have criticised the narrow focus on language proficiency in the traditional definition of the mother tongue and have emphasised the sociocultural and affective aspects of language learning and use [9]. Kamusella [10] argued that language is a political construct and that identities can be constructed and contested through language use. He also emphasised the importance of linguistic diversity and multilingualism in creating more inclusive societies. In another study, Gorter [11] posited that language use is complex and dynamic and that individuals can have multiple and changing linguistic identities based on their social context and experiences. He also highlighted the importance of acknowledging and valuing linguistic diversity in education and society.

One part of the research focused on converting one language into another, with the conversion concerning written or spoken speech. For example, a study by Fei et al. [12] dealt with the problem of incomplete semantic role labelling in low-resource languages. They converted the labels from the source language to the target language in their method. Yi et al. [13] dealt with synthesising spoken speech of various languages from text data (text-to-speech) and tried to deal with the problem of incorrect pronunciation. They proposed a triplet training scheme composed of an anchor, a positive, and a negative sample to cover unseen cases. A similar problem was dealt with by Zhou et al. [14], who tried to improve pronunciation when converting speech into another language, using cross-lingual voice conversion techniques. Finally, Vaswani et al. [15] proposed a new simple network architecture, the “Transformer”, for translating one language to another and achieved outstanding results.

Two important terms related to the mother tongue are “language loss” and “language policy”. Language loss refers to the gradual or rapid decline in the proficiency or use of an individual’s mother tongue, often due to language change or death [16]. A study looking for the effect of the use of computers and Internet use on language loss would be noteworthy. Language policy refers to decisions and practices related to language use in various sectors, such as education, government, media, and commerce, which can significantly impact the status, use, and development of the mother tongue and other languages [17].

Regarding the detection of the mother tongue, Mechti et al. [18], utilising a gated recurrent unit (GRU) network, introduced a deep learning model that can accurately identify the mother tongue of Arabic language learners, an essential aspect of language education. The primary objective of this study was to tackle the challenge of recognising the mother tongue of Arabic language learners to customise personalised language learning strategies for each learner. The learners’ written work is presented as input to the proposed model to generate writing samples. The pre-trained word embedding layer transforms the input text into a sequence of vectors, then passed to the GRU network to capture and model the input data’s long-term dependencies, given its ability to model sequential data. The model is trained on a dataset of writing samples from Arabic language learners with different mother tongues and is evaluated on a separate test set. The results show that the proposed model outperforms several baseline models and achieves high accuracy in identifying the mother tongue of Arabic language learners.

In addition, Siddhant et al. [19] investigated the use of pronunciation information for speaker and language recognition. They tested their models on conversational speech datasets in multiple languages and found that pronunciation information improves the accuracy of mother tongue recognition.

The papers presented in this discussion demonstrate the diverse range of approaches and methods that researchers have used to identify mother tongues. In addition to the

tools used in the works mentioned above, many others can also be used to find a user's mother tongue. Some of them are attention models [20,21], transformer models [22,23], and graph models [24,25]. By developing more accurate and efficient methods for mother tongue recognition, researchers can potentially improve language-related applications such as speech recognition, language teaching, and natural language processing.

One of the earliest studies on keystroke dynamics (KD) was conducted by Gaines et al. in 1980 [26], where they investigated the variation in typing patterns between individuals and found that individuals had unique typing patterns that could be used for identification purposes. Since then, several studies have focused on using KD in authentication systems. For example, Monroe et al. [27] proposed a keyboard-based authentication system that used neural networks to identify users based on their typing patterns, achieving high accuracy rates. Bergadano et al. [28] also conducted a study on using KD for biometric authentication, developing a model based on KD and evaluating its effectiveness through experiments.

Other studies have explored this using KD to detect impostors and anomalies in typing behaviours. Killourhy and Maxion [29] used a dataset from users typing a fixed text at regular intervals over several weeks to train and test anomaly detection algorithms, such as Principal Component Analysis, Mahalanobis distance, and Support Vector Machines. They found that the performance of the algorithms varied depending on the specific keystroke features being analysed. Gunetti and Picardi [30] analysed the KD of free text to investigate its feasibility as a biometric authentication mechanism for text entry, obtaining promising results with low false alarm rates and impostor pass rates.

KD has also been explored in user classification and recognition of the user's physical or mental situation. Tsimperidis and Arampatzis [31] attempted to identify characteristics of users, such as gender, age, and handedness, using KD features and a rotation forest classifier, achieving high accuracy rates in user profiling. Tsimperidis et al. [32] used keystroke durations and diagram latencies extracted from a dataset to develop a system that could accurately distinguish the age group of an unknown user. Roy et al. [33] proposed a KD-based indicator for Parkinson's disease screening at home, using ensemble learning and addressing key hypotheses related to the screening process to enhance the accuracy and effectiveness of the method.

As it became evident from the literature, on the one hand, the identification of a user's mother tongue has been attempted using various approaches, such as methods of natural language processing and exploitation of pronunciation information. On the other hand, KD has been mainly used for user authentication, recognising some inherent or acquired characteristics of users, and recognising users' mental and physical state. However, at least according to what is known, KD has not been used so far to identify users' mother tongue.

3. Methodology

The availability of datasets on KD for research purposes is limited on the Internet, with most of them being recorded by users using a fixed-text approach, wherein users enter a given text. This approach has limitations as it may not accurately reflect the natural typing behaviour of users. On the other hand, publishing free-text logging data risks exposing personal data; therefore, such datasets are rarely found in studies or surveys conducted online. As a result, in the initial phase of this research, free-text data were collected from volunteers who willingly participated in the study. In the second phase, relevant features related to KD were extracted from the collected data. Finally, in the third and last phase, the users' mother tongue was identified, and parameters of five machine learning algorithms, namely Support Vector Machine (SVM), Naïve Bayes (NB), Radial Basis Function Network (RBFN), Random Forest (RF), and Multi-Layer Perceptron (MLP), were appropriately set.

3.1. Dataset Preparation

The main objective of this study was to develop a research methodology to classify users based on their mother tongue using KD. A data collection procedure was designed

and implemented to achieve this objective, utilising a keylogger to record users' typing behaviour during their routine computer usage for approximately three and a half months (from 2 April 2022 to 19 July 2022). For this purpose, hundreds of users were approached to participate in the project. Those who agreed to be recorded were instructed to use their computer only when the keylogger was active, so the log file contained only data from them. No other instructions or restrictions were given to the volunteers, so they were free to type whatever they wanted, whatever time of day they wanted, and in whatever application, with the aim of recording as closely as possible the actual typing. In addition, the volunteers were given guarantees that their data would not be shared with third parties and would be used exclusively for the needs of this research. Finally, volunteers were allowed to review their data before handing over the log file to the researchers and deciding whether or not to do so.

The study was conducted with users representing five languages: Albanian, Bulgarian, English, Greek, and Turkish. It is worth noting that each of these languages belongs to a distinct language family, namely Germanic languages (English), Slavic languages (Bulgarian), Turkic languages (Turkish), Hellenic languages (Greek), and Albanian languages (Albanian).

After the data recording process, a total of 194 log files were collected. The number of log files collected for each language is presented in Table 1.

Table 1. Number of logfiles per mother tongue.

| Mother Tongue | Logfiles | Percentage |
|---------------|----------|------------|
| Albanian | 51 | 26.3% |
| Bulgarian | 46 | 23.7% |
| English | 17 | 8.8% |
| Greek | 55 | 28.3% |
| Turkish | 25 | 12.9% |
| Total | 194 | 100.0% |

The dataset utilised in this study comprised a variable number of logfiles for each language. Still, the asymmetry was not deemed extreme, as the class with the fewest samples contained approximately 1/3 of the class with the most samples. The study recruited volunteers of both genders in nearly equal numbers, with ages ranging from 18 to 65 years, and the distribution of ages was almost uniform.

The logfiles of the dataset consist of records, each containing information about an action on the keyboard, such as the key used, the exact time it was used, and the type of action (press or release). Also, each log file contains information about the volunteer, including their mother tongue.

3.2. Feature Set Preparation

After data collection, features were extracted, with KD accompanied by several features. However, in this study, only keystroke durations were utilised. This decision was based on the fact that keystroke durations occur more frequently than other KD features, making the method suitable for datasets with small sample sizes where other features may be infrequent.

Each keystroke duration can be easily calculated from the difference between the time instants of the release and the pressing of a key, information contained in the logfiles. However, each key appears more than once in a logfile, so the value of the corresponding feature results from the average of all keystroke durations of this key. In order to make the values of the features more representative, a threshold of 5 occurrences of each key was defined in the log file. Thus, if a key appears less than 5 times, then the value of the corresponding feature is not considered representative or taken into account.

After the feature extraction procedure ended, the feature set was created, consisting of records, each assigned to a log file. Each record consists of the feature values, separated

by commas, that correspond to the 106 keys for which there was data. At the end of each record, there is the label of the log file, which is one of the five different mother tongues, i.e., “Albanian”, “Bulgarian”, “English”, “Greek”, and “Turkish”. Finally, the feature set file was in a suitable format to be input to machine learning models.

3.3. Classifier Selection and Model Evaluation

Multiple machine learning models were employed in this research, including SVM, NB, RBFN, RF, and MLP. These models were identified as the best-performing ones in accuracy, training time (time taken to build the model), F1 score, and area under the Receiver Operating Characteristic (ROC) curve.

4. Results and Discussion

The primary objective of this study was to classify users based on their mother tongue using keystroke durations as features and machine learning models. To evaluate the performance of these models, metrics such as accuracy, time to build model, F1 score, and area under the ROC curve (AUC) were utilised.

The AUC was chosen as a relevant metric for evaluating the model’s performance, particularly when dealing with imbalanced classes. The AUC quantifies the area under the ROC curve, which illustrates the trade-off between the true and false positive rates at various classification thresholds.

Each experiment was performed with the 10-fold cross-validation method to obtain a more reliable statistical measure since it is performed ten times with a different training set and testing set each time, and their average value is calculated. This avoids the possibility of calculating an outlier as accuracy, F1 score and AUC.

Through a series of extensive experiments, the best results obtained from each classifier were documented in Table 2.

Table 2. The best results of each of SVM, NB, RBFN, RF, and MLP models, in terms of accuracy, time to build model (in s), F1 score, and area under the ROC curve.

| Model | Accuracy | Time to Build (in s) | F1 Score | AUC |
|-------|----------|----------------------|----------|-------|
| SVM | 75.8% | 0.12 | 0.757 | 0.884 |
| NB | 66.5% | 0.01 | 0.652 | 0.878 |
| RBFN | 82.5% | 0.31 | 0.823 | 0.939 |
| RF | 77.3% | 16.25 | 0.767 | 0.926 |
| MLP | 81.4% | 11.18 | 0.814 | 0.946 |

The best results of Table 2 were obtained for the SVM model having a C parameter value equal to 10, for the RBFN model having 10 clusters and a minimum standard deviation for clusters equal to 1.5, for the RF model having 1100 trees and 150 features to be considered in random feature selection, and for the MLP model having a learning rate of 0.6 and a momentum of 0.8.

Figure 1 shows the best results of the five machine learning models in searching the mother tongue of unknown users.

Based on the findings, the RBFN model demonstrated the highest accuracy of 82.5% and an F1 score of 0.823, followed by the MLP model with 81.4% accuracy and 0.814 F1 score. Conversely, MLP had the highest value of 0.946 in terms of AUC, followed by RBFN with 0.939. Notably, the NB and SVM models were observed to have shorter training times compared to other models.

Two conclusions can be drawn from these results. Firstly, the classification of users based on their mother tongue using KD data is feasible, with an accuracy of over 82% achieved in this dataset, surpassing the baseline accuracy of around 28%. Secondly, RBFN exhibited the best overall performance in accuracy and F1 score, but SVM may be more practical for real-world applications due to its shorter training time.

Finally, a closer examination of the results is presented in the confusion matrix shown in Table 3, which corresponds to the best-performing run of the experiments.

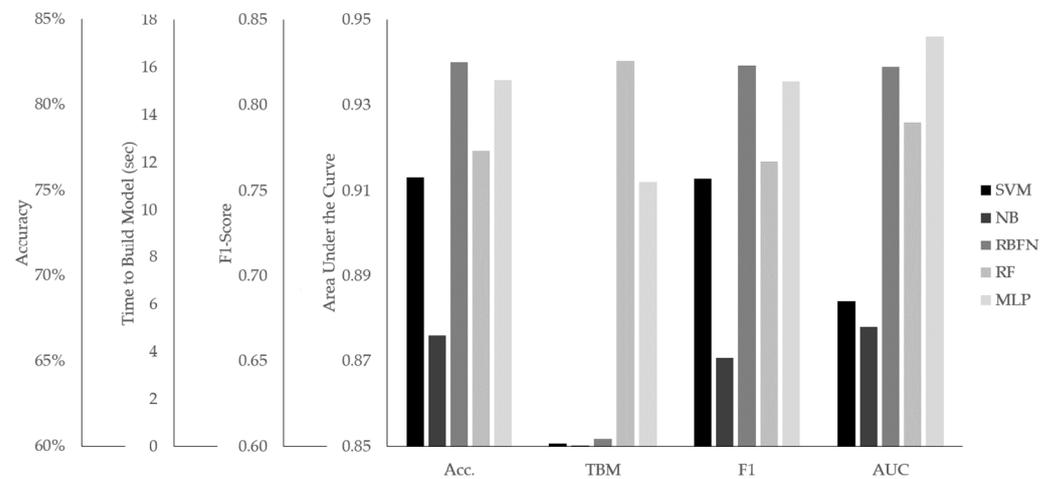


Figure 1. Models performance, in terms of accuracy (Acc., y-axis ranges from 60% to 85%), time to build model (TBM, y-axis ranges from 0 to 18 s.), F1 score (F1, y-axis ranges from 0.60 to 0.85), and area under the ROC curve (AUC, y-axis ranges from 0.85 to 0.95).

Table 3. Best run confusion matrix.

| Language | Albanian | Bulgarian | English | Greek | Turkish |
|-----------|----------|-----------|---------|-------|---------|
| Albanian | 47 | 2 | 0 | 2 | 0 |
| Bulgarian | 5 | 32 | 2 | 5 | 2 |
| English | 0 | 2 | 14 | 0 | 1 |
| Greek | 1 | 3 | 1 | 48 | 2 |
| Turkish | 2 | 1 | 0 | 3 | 19 |

Although the languages under investigation belong to distinct language families, as elucidated earlier, Table 3 presents evidence indicating a higher likelihood of misclassification of logfiles as belonging to languages spoken in geographically adjacent regions. For instance, log files attributed to the Albanian language were occasionally misclassified as Greek or Bulgarian, which are official languages of neighbouring countries. While this observation is a preliminary finding, it merits further investigation in an extended research study to better comprehend the underlying factors contributing to this phenomenon.

Since there is no other research in the literature that attempts to identify the mother tongue of users using KD, the results of this research cannot be directly compared with those of any other.

5. Conclusions

This study sheds light on the efficacy of keystroke dynamics (KD) in identifying the mother tongue of unknown Internet users. The study introduces a system that employs five widely used classifiers to analyse a dataset of KD. The results demonstrate that the system achieves a commendable accuracy rate of approximately 82% in identifying the user's mother tongue. This study contributes to the literature on KD by providing valuable insights into the utilisation of the mother tongue as a factor for user recognition. Identifying the mother tongue of a user and other characteristics, such as gender, age, educational level, handedness, etc., will help in user profiling where this is required.

The field of KD holds significant potential for various practical applications, including detecting cyber criminals in digital forensics, personalised services for Internet users, targeted advertisements, and fraud detection. Additionally, KD can serve as a complementary authentication method when continuous authentication is necessary, making it relevant for

real-time security management in unmanned computer terminals, user behaviour analysis, and timely user alerting or tagging.

There are several avenues for future research in this area. Firstly, expanding the dataset to include users from a broader range of mother tongues would help validate and further enhance the results obtained in this study. Secondly, exploring additional KD features would provide insights into which features most effectively classify users based on their mother tongue. Thirdly, perform an ablation study to test the proposed system's operation thoroughly. Fourthly, the definition of new, more specialised classes to test the impact of other user characteristics. Lastly, investigating a higher probability of misclassifying a user's mother tongue as a neighbouring country's language than other misclassifications warrants further investigation.

This study contributes to the understanding of KD in mother tongue recognition. It opens up possibilities for further research in this field with potential applications in user authentication, cybersecurity, and user behaviour analysis.

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

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The log files contain sensitive and/or personal data of the volunteers who participated in the typing recording and are therefore not available.

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

References

1. Ulker, M. The Approach of Learning a Foreign Language by Watching TV Series. *Educ. Res. Rev.* **2019**, *14*, 608–617. [\[CrossRef\]](#)
2. UNESCO. *The International Year of Indigenous Languages: Mobilizing the International Community to Preserve, Revitalize and Promote Indigenous Languages*; UNESCO Publishing: Paris, France, 2021.
3. Deng, Y.; Zhong, Y. Keystroke Dynamics User Authentication Based on Gaussian Mixture Model and Deep Belief Nets. *Int. Sch. Res. Not.* **2013**, *2013*, 565183. [\[CrossRef\]](#)
4. Roy, S.; Pradhan, J.; Kumar, A.; Adhikary, D.R.D.; Roy, U.; Sinha, D.; Pal, R.K. A Systematic Literature Review on Latest Keystroke Dynamics Based Models. *IEEE Access* **2022**, *10*, 92192–92236. [\[CrossRef\]](#)
5. Cummins, J. Bilingual children's mother tongue: Why is it important for education? *Sprogforum* **2001**, *7*, 15–20. [\[CrossRef\]](#)
6. Baker, C. *Foundations of Bilingual Education and Bilingualism*, 3rd ed.; Buffalo, N.Y., Ed.; Bilingual education and bilingualism; Multilingual Matters: Clevedon, UK, 2001.
7. Grosjean, F. *Bilingual: Life and Reality*; Harvard University Press: Cambridge, MA, USA, 2010. [\[CrossRef\]](#)
8. Petrovic, J.E.; Olmstead, S. Language, Power, and Pedagogy: Bilingual Children in the Crossfire, by J. Cummins. *Biling. Res. J.* **2001**, *25*, 405–412. [\[CrossRef\]](#)
9. Pavlenko, A.; Blackledge, A. *Negotiation of Identities in Multilingual Contexts*; Multilingual Matters: Bristol, UK, 2004. [\[CrossRef\]](#)
10. Kamusella, T. *The Politics of Language and Nationalism in Modern Central Europe*; Palgrave Macmillan UK: London, UK, 2009. [\[CrossRef\]](#)
11. Gorter, D.; Zenotz, V.; Cenoz, J. *Minority Languages and Multilingual Education: Bridging the Local and the Global*; Educational Linguistics; Springer: Dordrecht, The Netherlands, 2014; Volume 18. [\[CrossRef\]](#)
12. Fei, H.; Zhang, M.; Ji, D. Cross-Lingual Semantic Role Labeling with High-Quality Translated Training Corpus. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 7014–7026. [\[CrossRef\]](#)
13. Ye, J.; Zhou, H.; Su, Z.; He, W.; Ren, K.; Li, L.; Lu, H. Improving Cross-Lingual Speech Synthesis with Triplet Training Scheme. In Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; pp. 6072–6076. [\[CrossRef\]](#)
14. Zhou, Y.; Wu, X.; Tian, X.; Li, H. Optimization of Cross-Lingual Voice Conversion with Linguistics Losses to Reduce Foreign Accents. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2023**, *31*, 1916–1926. [\[CrossRef\]](#)
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: New York, NY, USA; pp. 5998–6008.
16. Hinton, L.; Hale, K.L. *The Green Book of Language Revitalization in Practice*; Brill: Leiden, The Netherlands; Boston, MA, USA, 2013.

17. García, O.; Baetens Beardsmore, H. *Bilingual Education in the 21st Century: A Global Perspective*; Wiley-Blackwell Pub: Malden, MA, USA; Oxford, UK, 2009.
18. Mechti, S.; Alroobaea, R.; Krichen, M.; Rubaiee, S.; Ahmed, A. Deep Learning Model for Identifying the Arabic Language Learners Based on Gated Recurrent Unit Network. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 620–627. [[CrossRef](#)]
19. Siddhant, A.; Jyothi, P.; Ganapathy, S. Leveraging Native Language Speech for Accent Identification Using Deep Siamese Networks. In Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, 16–20 December 2017; pp. 621–628. [[CrossRef](#)]
20. Fei, H.; Zhang, Y.; Ren, Y.; Ji, D. Latent emotion memory for multi-label emotion classification. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 7692–7699. [[CrossRef](#)]
21. Wu, S.; Fei, H.; Ren, Y.; Ji, D.; Li, J. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. In Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI-21), Montreal, QB, Canada, 19–26 August 2021; pp. 3957–3963. [[CrossRef](#)]
22. Thara, S.; Poornachandran, P. Transformer Based Language Identification for Malayalam-English Code-Mixed Text. *IEEE Access* **2021**, *9*, 118837–118850. [[CrossRef](#)]
23. Ranasinghe, T.; Zampieri, M. An Evaluation of Multilingual Offensive Language Identification Methods for the Languages of India. *Information* **2021**, *12*, 306. [[CrossRef](#)]
24. Huang, Y.-H.; Harryyanto, K.; Tsai, C.-W.; Pornvattanavichai, R.; Chen, Y.-S. Graph Knowledge Transfer for Offensive Language Identification with Graph Neural Networks. In Proceedings of the 23rd International Conference on Information Reuse and Integration for Data Science (IRI), San Diego, CA, USA, 9–11 August 2022; pp. 216–221. [[CrossRef](#)]
25. Mishra, P.; Tredici, M.D.; Yannakoudakis, H.; Shutova, E. Abusive Language Detection with Graph Convolutional Networks. 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), ACL, Minneapolis, MN, USA, 2–7 June 2019; pp. 2145–2150. [[CrossRef](#)]
26. Gaines, R.S.; Lisowski, W.; Press, S.J.; Shapiro, N. *Authentication by Keystroke Timing: Some Preliminary Results*; Rand Corporation; R-2526-NSF. Rand: Santa Monica, CA, USA, 1980.
27. Monroe, F.; Reiter, M.K.; Wetzel, S. Password Hardening Based on Keystroke Dynamics. In Proceedings of the 6th ACM Conference on Computer and Communications Security, Singapore, 1–4 November 1999; pp. 73–82. [[CrossRef](#)]
28. Bergadano, F.; Gunetti, D.; Picardi, C. User Authentication through Keystroke Dynamics. *ACM Trans. Inf. Syst. Secur.* **2002**, *5*, 367–397. [[CrossRef](#)]
29. Killourhy, K.S.; Maxion, R.A. Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. In Proceedings of the 2009 IEEE/IFIP International Conference on Dependable Systems & Networks, Lisbon, Portugal, 29 June–2 July 2009; pp. 125–134. [[CrossRef](#)]
30. Gunetti, D.; Picardi, C. Keystroke Analysis of Free Text. *ACM Trans. Inf. Syst. Secur.* **2005**, *8*, 312–347. [[CrossRef](#)]
31. Tsimperidis, I.; Arampatzis, A. *User Profiling Using Keystroke Dynamics and Rotation Forest: In Advances in Information Security, Privacy, and Ethics*; Lobo, V., Correia, A., Eds.; IGI Global: Hershey, PA, USA, 2022; pp. 1–24. [[CrossRef](#)]
32. Tsimperidis, I.; Yucel, C.; Katos, V. Age and Gender as Cyber Attribution Features in Keystroke Dynamic-Based User Classification Processes. *Electronics* **2021**, *10*, 835. [[CrossRef](#)]
33. Roy, S.; Roy, U.; Sinha, D.; Pal, R.K. Imbalanced Ensemble Learning in Determining Parkinson’s Disease Using Keystroke Dynamics. *Expert Syst. Appl.* **2023**, *217*, 119522. [[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.