



RESEARCH PAPER

A Computational Discourse Analysis of Macmillan ELT Language Tree Textbooks of Grades 3 and 4: Using NIP Stanford Core's Named Entity Recognition Method

¹Momina Khattak, ²Dr.Muhammad Adnan Maqbool and ³Dr. Fasiha Altaf

1. Visiting Lecturer, University of Education Lahore, Attock Campus, , Punjab, Pakistan
2. Assistant Professor (Education), University of Education, Lahore, Punjab, Pakistan
3. Assistant Professor, Department of Education, GC Women University Sialkot, Punjab, Pakistan

Corresponding Author: fasiha.altaf@gcwus.edu.pk

ABSTRACT

This study is conducted to examine the Macmillan ELT textbooks of language tree grades 3 and 4 of Pakistan which teach at private sector schools. The pre-eminent publishing houses in Pakistan are like "Cambridge University press, Peak publisher, Macmillan publishers and Oxford University. This research also investigates the usage of foreign cultural words in ELT textbooks which alienates learners, from their own culture and they are getting confused and cannot resonate with these cultural items in their own. This study is based on descriptive quantitative analysis and analytical observation methods and to further categorize manually NLP identified entities while using one of its feature NER in ICC categorization. The data is afterwards analyzed by statistical methods, including percentage based line graphs and frequency counts. Consequently, this study framework is based on manifest content based quantitative analysis. Which provides assistance to the educators and policymakers towards more culturally responsive textbook selection.

KEYWORDS: ELT, NLP Stanford core (v. 4.5.9), NER, ICC, Macmillan textbooks grades 3 and 4

Introduction

Academic discourse is the formation of meaningful words, such as fiction, classroom conversation, reliable materials, and books used in language teaching classrooms. According to Nawaz et al. (2015), "Textbooks are considered a fundamental source for English language teaching in Pakistan". Cambridge University press, Peak publisher, Macmillan publishers and Oxford University Press are considered the pre-eminent publishers within private sector schools in Pakistan." The selection of these textbooks takes into account the aptitude, behavior, and culture of the specific target groups, schools, and society. However, these authors are often unaware of the cultural and religious background of Pakistani learners' society, leading to content that fails to fulfill the social, cultural, and religious requirements of English second language learners in Pakistan" (Shah, 2014).

The majority of learners remain confused during English second and foreign language classrooms in Pakistan due to the difference in social identity, beliefs, behavior, social and political institutes, socialization, and national history and geography. The content in these ELT textbooks reflects the cultural deviation, and it alienates the apprentices from their native culture. This distinction between the learners' culture and the target culture engenders a novel and earnest nuisance in the academic careers of apprentices. It is imperative to design ELT textbooks according to the learners' culture to bolster their confidence, relatability, connectivity, and association with the content of textbooks. The innateness of the target culture is integral to the survival of learners in the globalized world. Due to the proliferation of dependency on international economic institutions, the English language has become imperative to learn by all nations of the world, but learners in English as a second and foreign language classrooms confront cultural confusion with academic travails.

Natural Language Processing (NLP) has proven to be a strong computational model for processing and analyzing human language so that significant information can be derived from large amounts of text. NLP is based on a variety of tools such as Named Entity Recognition (NER), part-of-speech tagging, sentiment analysis, and syntactic parsing that enable computers to process and analyze nuances of language. NER, for example, allows the identification of such as names, places, and institutions in a text and provide information. Afterward, statistical analysis to be count manually count the frequency of each item or entity and visualize the result using a graphs. This makes NLP a desirable way of studying the contents of text in a systematic manner, particularly when criticizing cultural representation.

So, the focal point of this research is a corpus based analysis using NLP tools particularly, named entity recognition (NER) while using Stanford Core NLP (v. 4.5.9). It was employed to examine the cultural content entities such as, schools, places, events and institutions from the 3 randomly different units of Macmillan publisher language tree grades 3 and 4 ELT textbooks. Afterward, these entities have been manually extracted and then analyzed by through statistical method in which each entity frequency has been recorded. This is been access by Byram's (1994) components of intercultural communicative competence, attitudes, knowledge, interpretative competence, and critical cultural awareness. By integrating Byram's model and NLP, this study investigates to what degree English Language Teaching (ELT) textbooks in Pakistan were culturally sensitive and whether they were able to encourage intercultural competence within the Pakistani learner context.

Literature Review

Language plays a vital role in human communication, connect different individuals to one another and promoting societal development. It is deeply intertwined with culture, beliefs, and values, providing a way for interaction over gestures. Yule (2014), describes "language as a human system that is vital in articulating arbitrary vocal system used in communication". Wardhaugh (2014), states that "language is an orderly system that structures sound patterns into meaningful expressions, once again reinforcing its contribution in deciding the way people communicate and exchange the ideas and that changes in language often mirror shift in society". Trudgill (2021), further emphasizes that "language is a means of communication that shape human society, and transmit beliefs, norms and values it also plays an important role in the construction in the construction of society and togetherness".

As Morris (2023), states that "language is a system of communication it has proper system of sounds, words and grammar through which they convey emotions and brings people together in society and holds them in common terms. So, the evolution of language is parallel to human civilization, from rudimentary symbolic expression to increasingly complex forms of communication".

Similarly Culture, is an attribute and awareness of a specific circle of people including language, food, festivals, rituals beliefs, customs, values, norms and traditions. It is a collection of attitudes and association that induced by socialization. Rocher (2004), states that "values, feelings, and practices towards culture is the core of group identity". Brown (2007), also sketches that "culture decides how individuals think and behave". Whereas, Porter, R. E., & McDaniel, E. R. (2015) propose that "culture conveys both linguistic and social behavior patterns which determines that environment and experience influence communication norms and personal attitudes".

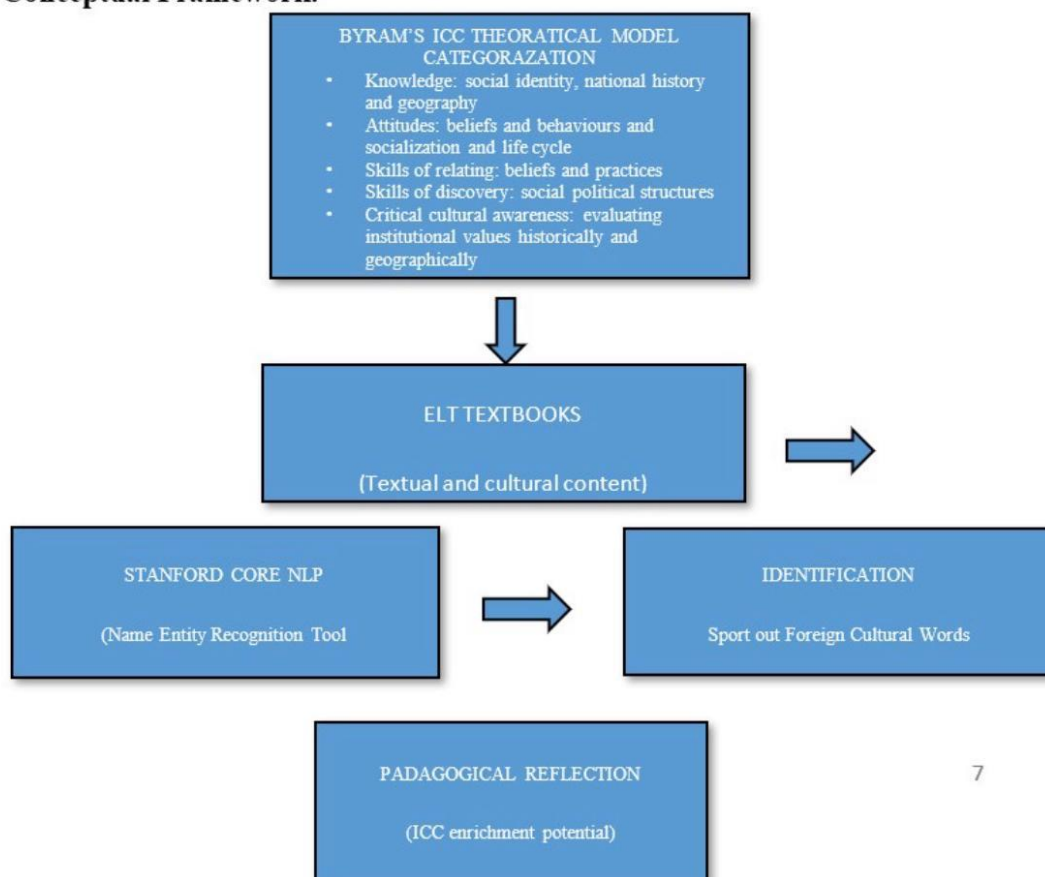
In educational context, particularly in English Language Teaching (ELT), language is not a subject but a vehicle through which one gains linguistic competence as well as cross-cultural consciousness. Therefore, ELT has the twin functions one of the function is teaching a globalized language and the other function is transmitting the cultural values embedding

with it. The close relationship between language and culture requires to integrate cultural material deliberately into ELT classrooms. Baker (2015), states “culture and language teaching are intertwined with each other, one without another is incomplete”. However, this heightened emphasis on global and foreign cultures in ELT materials in nations such as Pakistan has made learners familiar with more about Western culture and values than their own, creating a state of cultural self-ignorance. When English grew as a lingua franca, the curricula for language shifted towards more global and multicultural material. This shift focused on intercultural competence, so language students needed to interact with various cultural points of view. English language textbooks used in Pakistani private schools contained contemporary content, integrating these wider cultural narratives into academic life.

Furthermore, NLP has emerged a powerful tool in ELT, offering innovative approaches in English language teaching to enhance personal learning and cultural relevance. According to Mehrotra, (2019), Pakistan being a multinational country where over 70 languages has spoken is bridging the linguistic gap and cultural inclusivity between ELT and NLP. NLP tools like sentiment analysis, named entity analysis and topic modelling, have been utilized to improve the cultural content of ELT textbooks to ensure the balance diversity of relevant culture.

In contrast, studies by the Government of Pakistan (2006) and Majeed (2012) illustrate classroom challenges whereby, textbooks are prioritized in the learning content, barring the utilization of additional material such as visual aids or library usage. Textbooks should supplement, and not dominate, learning. Brown (2004) states that “ELT materials may be inadequate to facilitate effective cultural understanding.” The outcome is that students from private sector, especially in the case of Pakistan can become more and more disconnected from their native culture, more and more accustomed to foreign cultures, yet disconnecting themselves from local customs and ethics. Language and culture being closely related to each other, cultural content must be incorporated into foreign language classes in order to make real communication feasible and further enhance the students' knowledge about different communities. In situations like those in Pakistan, where resources could be limited, it is of the greatest significance that teachers are meticulous about assessing and choosing not only materials improving linguistic ability but also increasing awareness of culture. An essential research gap lies under this investigation how to emerge a balanced cultural proportion between native and foreign ELT course which enables the learners to support their linguistic development and cultural identity. Although, the application of Natural Language Processing tools as a problem-solving framework has not been explored in analyzing textual data. Natural Language Processing is unable to identify independently culturally biased or prejudiced content. Instead of that, these tools provide preparatory categorization which is needed for manual interoperation to accurately differentiate between native and foreign cultural references. The study of NLP in the context of ELT culturally adaptive material as an adjunct analytical tool is not been fully explored. When it incorporated with the manual analysis, NLP has the potential to enhance more culturally responsive language by helping to adapt content, check for cultural appropriateness, and identify likely cultural imbalance within textbooks.

3. Conceptual Framework:



BYRAM'S ICC THEORITICAL CATEGORI

The conceptual framework of this research integrates computational linguistics and intercultural education which examine the cultural representation of English language teaching textbooks (ELT) at private sector. This study employ Stanford Core NLP (v. 4.5.9) software to extract relevant lexical items from ELT textbooks which are than analyzed by theoretical lens of Byram's intercultural communicative competence (ICC) model. The framework adopt quantitative frequency analysis. The research started with the selection and digitalization of Macmillan publishers ELT textbooks commonly used to teach at different private sectors primary level in Pakistan. After the process of ELT textbooks digitalization the text are pre-processed for clarity and analyzed by Stanford Core NLP (v. 4.5.9) software tool Name Entity Recognition (NER) which used to filter or extract lexical items from the textbooks like; food, rituals, holidays, social values, religion, national identity these words extracted by NLP using its NER tool match the categories of Byram's ICC model, which includes attitudes, knowledge, interpreting and relating skills, discovery and interaction skills, and critical cultural awareness then it incorporated with the manual analysis. Afterwards, these textual and cultural words analysis involves quantifying the frequency and distribution of cultural items from the textual data. These frequencies are provided in tables or graphs for determining which of the ICC elements are highlighted or omitted. This help to determine the extent and balance of cultural content among grades 3 or 4 level textbooks. The above methods provide a comprehensive analysis that how much these ELT materials are embedded with intercultural components, guiding the educators and policymakers towards more culturally responsive textbooks selection.

Material and Methods

This study employs a quantitative descriptive type while using analytical observation method to categorize manually NLP identify entities while using one of his feature NER into ICC categorizes. The data afterwards analyzed by statistical methods, including percentage based line graphs and frequency count.

Results and Discussion

This part of the study represent the findings of quantitative data, which explored the cultural imbalance within a Pakistani private sector ELT textbooks. The data has been gathered by using analytical observation method to categorize by manually annotations through NLP identify entities while using one of its feature NER into ICC categorizes furthermore, a screen short of these annotative entities is provided to accompany this categorization process. The data afterwards analyzed by statistical methods, including percentage based line graphs and frequency count. This analysis is organized by manifest content based research objectives, focusing on the English Language Teaching (ELT) textbooks in Pakistan are culturally sensitive and whether they are able to encourage intercultural competence within the Pakistani learner context. Here are some screen shorts of NLP Stanford core identify entities while using its NER feature the tool highlights different categorizing like person, country, title and location which are later categorized by manually annotation process.



Figure 1(a) NER annotation grade 3 language tree ELT story book

Figure 1 (a) is taken from Stanford core NLP while using one of its prominent feature NER named entity recognition, which clearly categorized different entities from unit 1”The Swing” such as, person, time and cause of death.

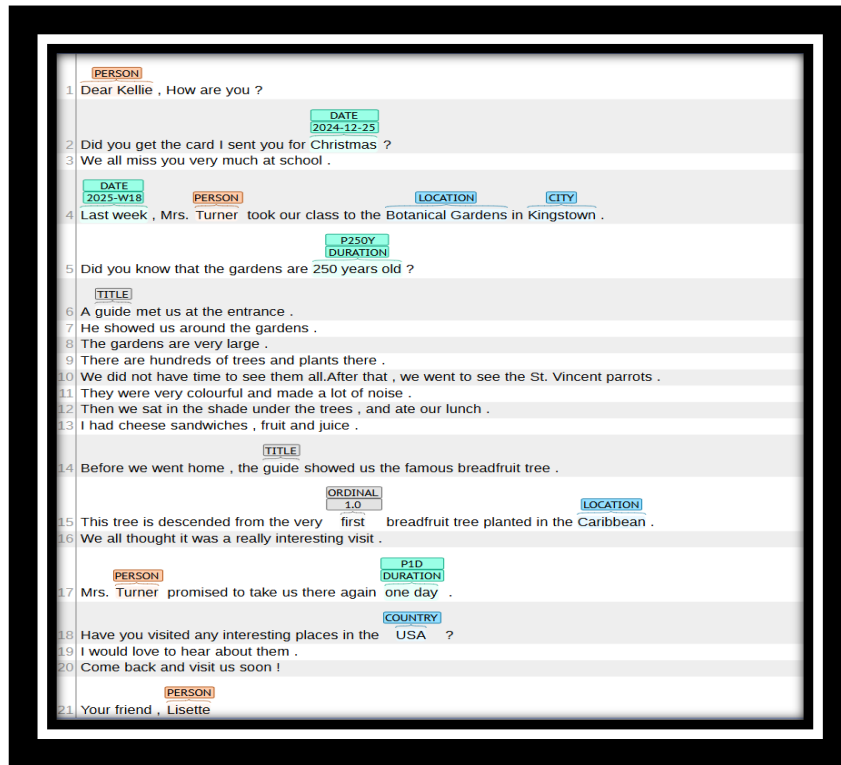


Figure 1(b) NER annotation grade 3 language tree ELT story book

Figure 1 (b) is also taken from Stanford core NLP while using one of its prominent feature NER named entity recognition, which shows two more different categorizes of entities from unit 1. In unit 8”The Class Visit” shows entities like, person, duration, location and title.

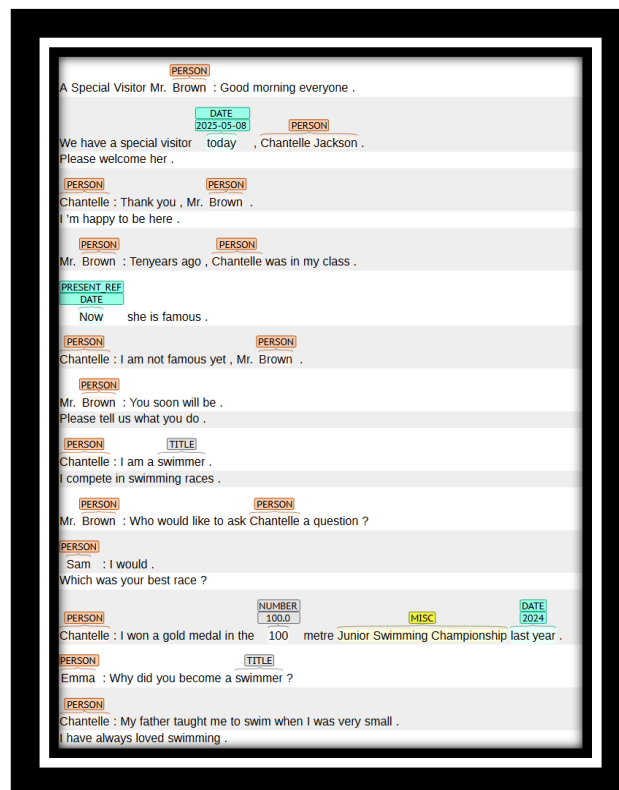


Figure 1(c) NER annotation grade 3 language tree ELT story book

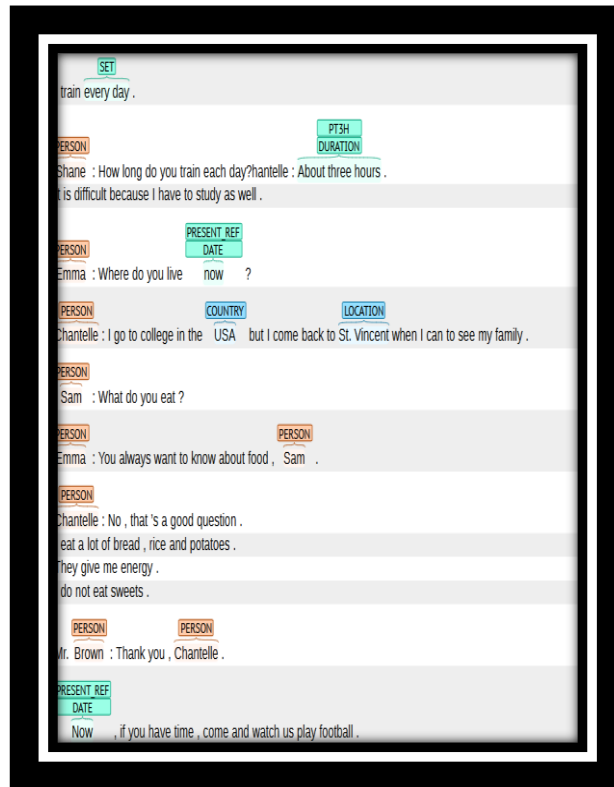


Figure 1 (c) is also taken from Stanford core NLP while using one of its prominent feature NER named entity recognition, which shows more different categorizes of entities from unit 1 and 8. In unit 2 "A Special Visitor" shows entities like, person, duration, location, title, date and number.

Later, in this study we manually apply it to Byram's ICC theoretical categorization model which bases these categorizations on the components like knowledge, attitude, skills of discovery, skills of relating and critical cultural awareness, which are additionally connected with different sub sets, such as social-political situations, social identity, history and geography and personal identity. Here is the statistical or graphical analysis which shows how many foreign cultural words in these chapters exist:

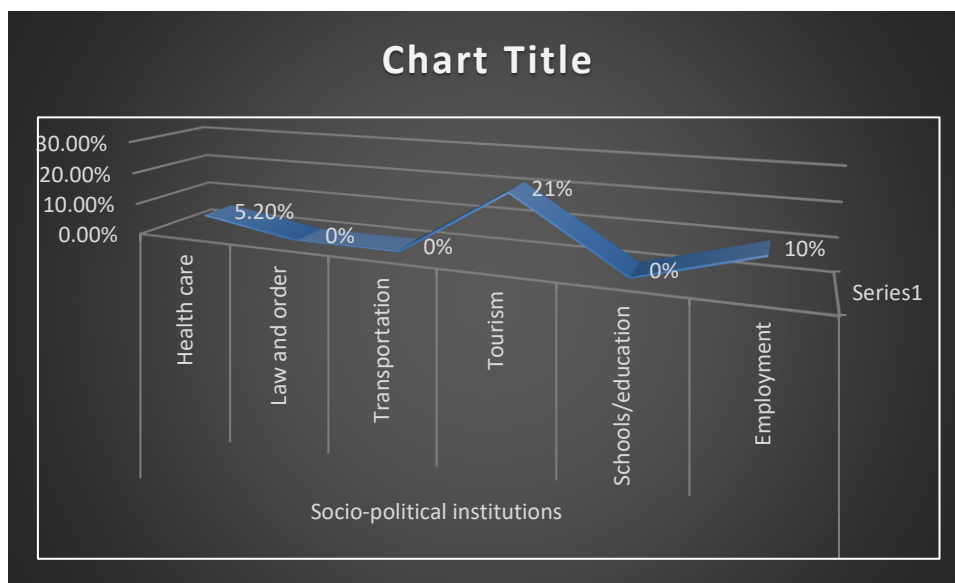


Figure 2 (a) Statistical line graph of the social-political institutions category of Byram's ICC model Mammalian language tree grade 3 ELT storybook

Figure 2 (a) of this line graph clearly shows the percentage of foreign cultural social-political institutions from the Mammalian language tree grade 3 textbook. These percentages are from three different units which are unit 1 “The Swing” unit 2 “A Special Visitor” and unit 8 “A Class Visit”. This graph shows the foreign cultural words in the field of health care are used 5.20% whereas, law and order, transportation and schools or educational based cultural words are 0% in these units. The percentage line goes higher at the tourism component which is 21% and after that employment is another component on which the line goes up.

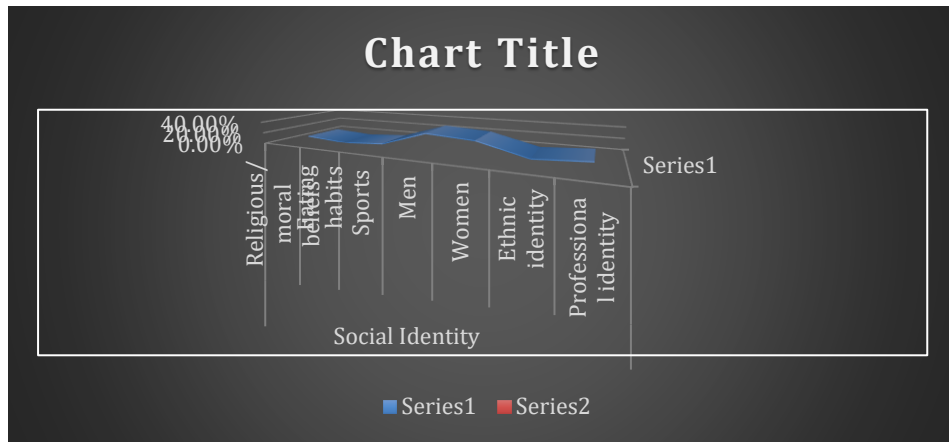


Figure 2 (b) Statistical line graph of the social-identity category of Byram’s ICC model Mammalian language tree grade 3 ELT storybook

Figure 2 (b) of this line graph shows different foreign cultural elements of social-identity taken from different units which are unit 1 “The Swing” unit 2 “A Special Visitor” and unit 8 “A Class Visit” from Macmillan language tree grade 3 textbooks these components are religious and moral beliefs, eating habits, sports, men, women, ethnic identity and professional identity. The highest percentage of cultural words in this graph is men and women which is 31% and 26% which clearly shows that in these units no cultural associated names of men and women has been found in these units. Similarly different cultures ethnic identity and professional identities are also found which are 5.20% and 10.20%. Whereas, we also have found international sports elements which are 5.20% but the last component which do not found in these units are eating habits which is why it is 0%.

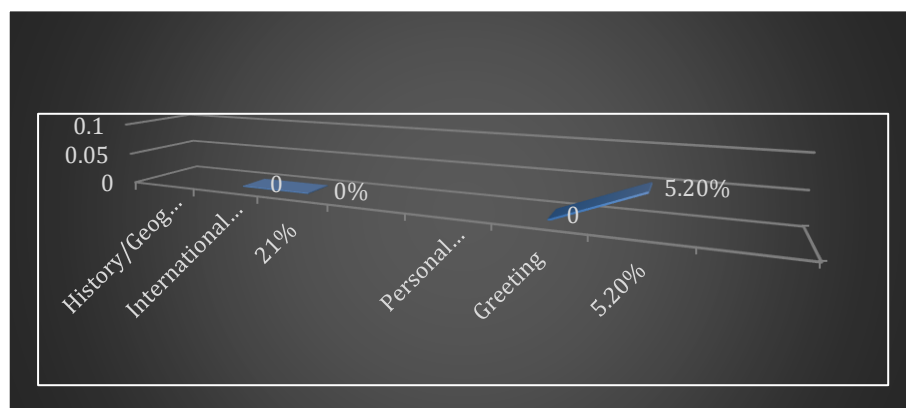


Figure 2 (c) Statistical line graph of the personal-identity and history or geography category of Byram’s ICC model Mammalian language tree grade 3 ELT storybook

Figure 2 (c) shows the two categorizes of byram’s model of history and geography and personal identity which is further divided into international history or geography and rituals and greeting. The line graph clearly shows that cultural elements from Mammalian publishers’ language tree textbook of grade 3 unit 1 “The Swing” unit 2 “A Special Visitor” and

unit 8 “A Class Visit” has geographical elements higher which is 21% whereas, historical elements are 0%. On the other hand personal identity has rituals and greetings in which greeting elements are 0% therefore, rituals have been found 5.20%.

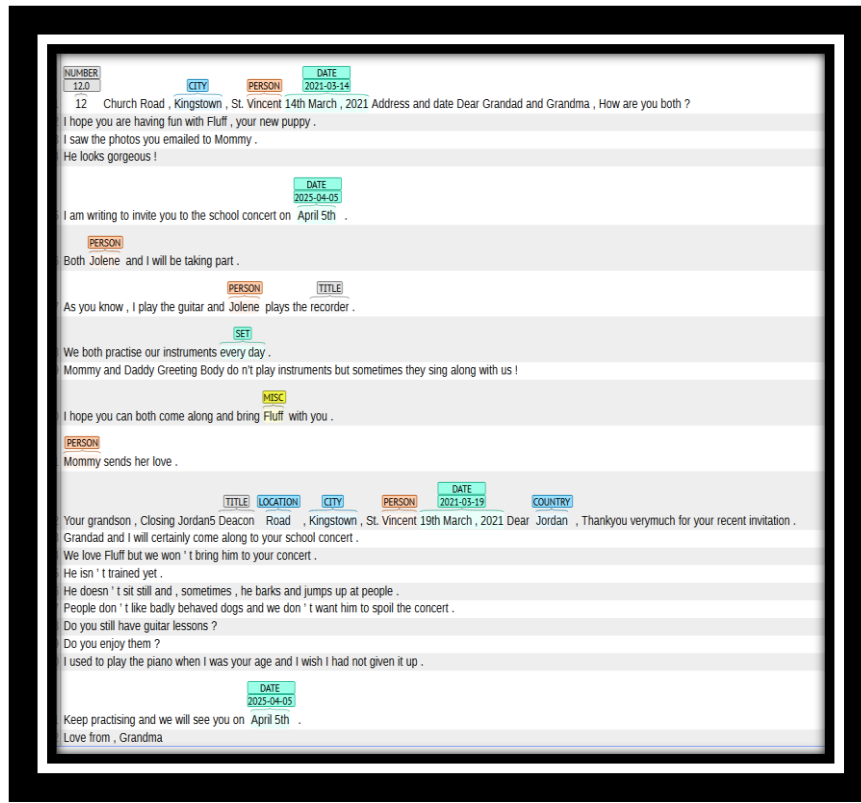


Figure 3 (a) NER annotation grade 4 language tree ELT story book

Figure 3 (a) is taken from Stanford core NLP while using one of its prominent feature NER named entity recognition, which clearly categorized different entities from unit 1 “An Invitation” such as, person, city, country, location, title, number and set.



Figure 3 (b) NER annotation grade 4 language tree ELT story book

Figure 3 (b) is also taken from Stanford core NLP while using one of its prominent feature NER named entity recognition, which shows different categorizes of entities from unit 1. In unit 13” The New Park” shows the entities like, person, duration, location, title and organization. Which are not seen in previous screenshots.

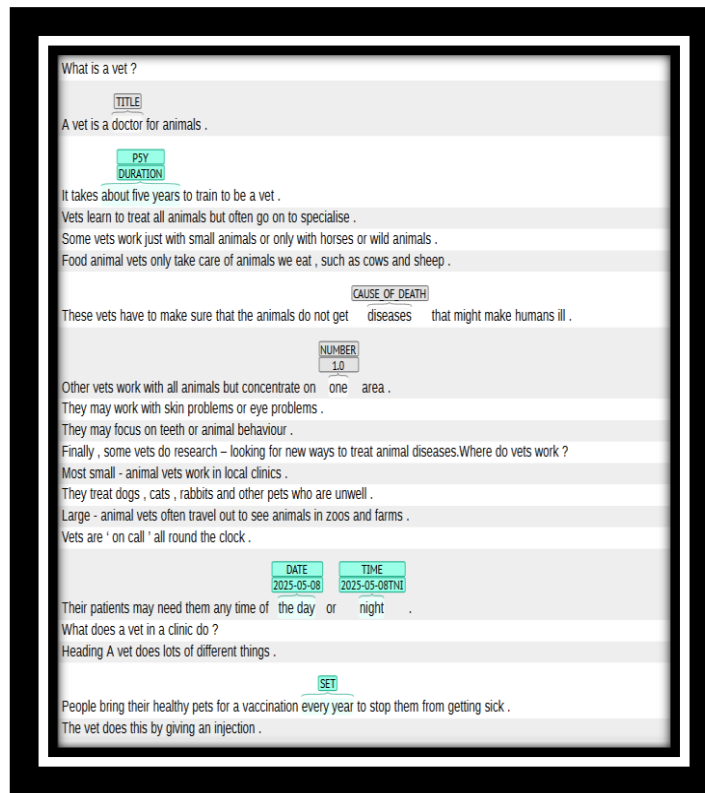


Figure 3 (c) NER annotation grade 4 language tree ELT story book

Figure 3 (c) is also taken from Stanford core NLP while using one of its prominent feature NER named entity recognition, which shows same categorizes of entities from unit 1 and 13. In unit 2”A vet” shows entities like, person, title, set, date and number. Again, in this study we manually apply it to Byram’s ICC theoretical categorization model which bases these categorizations on the components like knowledge, attitude, skills of discovery, skills of relating and critical cultural awareness, which are additionally connected with different sub sets, such as social-political situations, social identity, history and geography and personal identity. Here is the statistical or graphical analysis of grade 4 language tree textbooks which shows how many foreign cultural words in the above mention units exist:

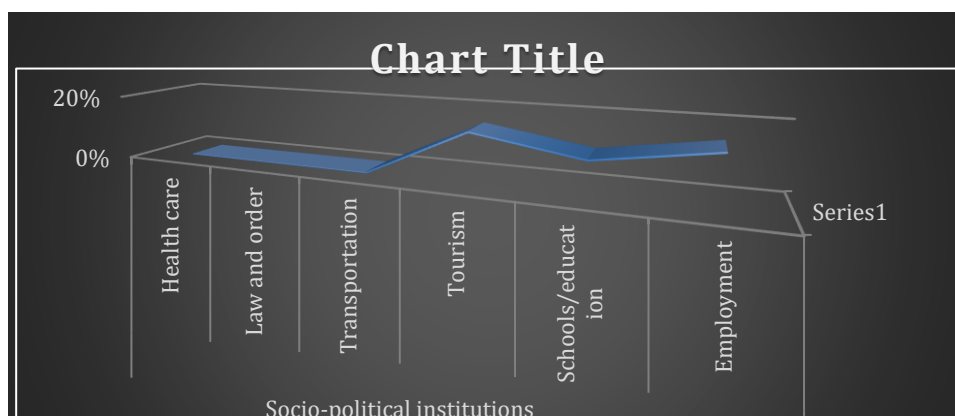


Figure 4 (a) Statistical line graph of the social-political institutions category of Byram’s ICC model Mammalian language tree grade 4 ELT storybook

Statistical figure 4 (a) from Macmillan publisher grader 4 story book shows almost the same analysis which is seen above in the grade 3 storybook the foreign cultural elements like healthcare, law and order and transportation has the same percentage which is 0% in these grade 4 units “A Vet”, A New Park and An Invitation. Whereas, the differences can see the other element which are in tourism foreign cultural words are used 15% similarly in education and school the line graph is a bit higher which is 10% and lastly, employment percentage in this graph is 15%.

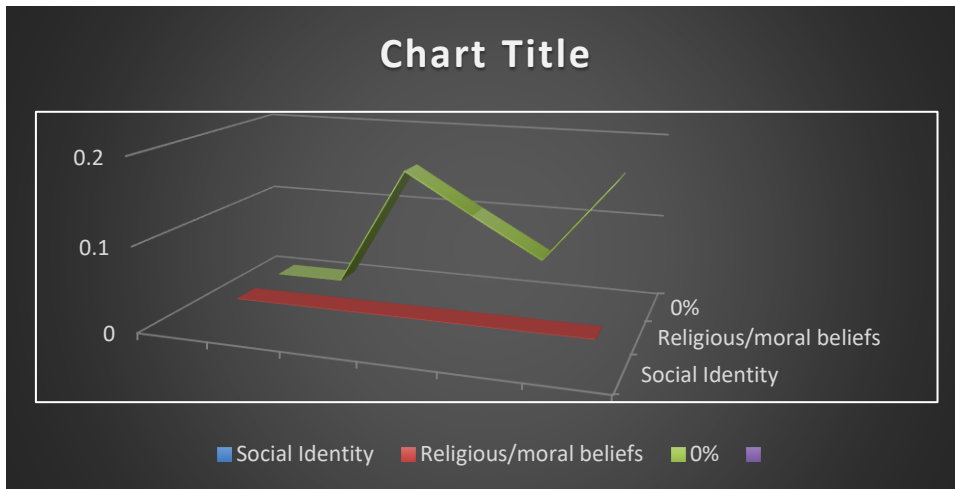


Figure 4 (b) Statistical line graph of the social-political institutions category of Byram’s ICC model Mammalian language tree grade 4 ELT storybook

Figure 4 (b) some results are quite similar to figure 3 (b) in that graph social identity elements such as, men and women have a higher line of foreign cultural names likewise, we see in this graph men are 15% and women are 10%. Other than that ethnic and professional identities are 5.2% and 15.7% which are quite near number to the previous graph 3 (b) result. Similarly another element eating habit result is same as above figure 3 (b) which is 0%. The differentiation comes into other two elements which are sports and religious in previous units of grade 3 we have seen these elements but here in grade 4 units we don’t see such elements. So there percentages are 0%.

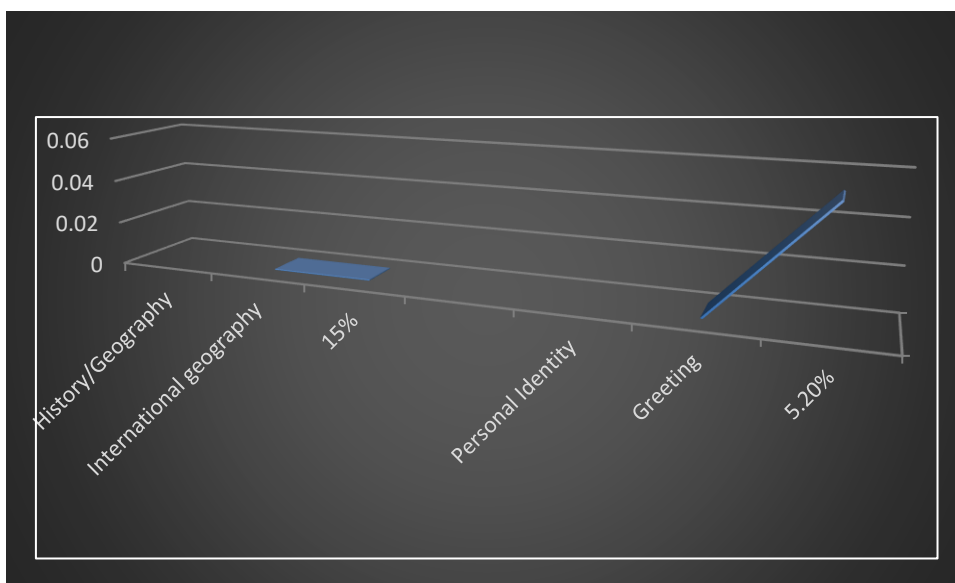


Figure 4 (c) Statistical line graph of the social-political institutions category of Byram’s ICC model Mammalian language tree grade 4 ELT storybook

Figure 4 (c) results are similar almost at each elements international geography is the highest in previous line graph number 3(c) in this graph you can see it is on top which is 15% international history is 0% in these units. Whereas, the other two elements ritual and greetings have the same percentage which is 5.20%.

The analysis of both ELT textbooks units show that these books which we taught at private sector are embedded with foreign cultural words. Native culture or local cultural entities are not found in this study which alienates the students with their own culture as we can see now-a- days. So, this research is guiding the educators and policymakers towards more culturally responsive textbooks selection.

Conclusion

In conclusion, this study finds a gap in ELT textbooks which are mostly taught by Pakistani private sectors schools, where students are not getting any affiliation with their own culture instead of that, they are leaning about western and Caribbean culture that is not relevant to their culture. This thing also create identity crisis in young learners where they get confused between their native culture and foreign culture. This study is held with two ELT Macmillan textbooks of grade 3 and 4 from which we randomly choose 3 chapters each and then put that text into NLP Stanford core where its NER feature it categorized all its entities which further manually characterized into Byram's ICC model and lastly, does frequency analysis for the clear result. This study is quite useful in guiding the educators and policymakers towards more culturally responsive textbooks selection.

References

- Baker, W. (2015). *Culture and identity through English as a lingua franca: Rethinking concepts and goals in intercultural communication*. De Gruyter Mouton
- Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). Pearson Education: New York.
- Majeed, A. (2012). *The representation of target culture in the ELT textbooks in Pakistan*.
- Mehrotra, S. (2019). Harnessing artificial intelligence for dynamic landscape: Re-envisioning English language teaching in Pakistan. *Journal of Language and Linguistic Studies*, 7(2), 257-270. <https://jll.uoch.edu.pk/index.php/jll/article/download/304/255>
- Morris, M. (2023). *An introduction to the philosophy of language*. Cambridge University Press
- Nawaz. (2015). Representation of English Culture in Oxford English Textbooks of Class *International Journal of Academic Research and Reflection*, 3, 89-94. Retrieved from <https://www.idpublications.org/wp-content/uploads/2014/10/>
- Rocher, F. (2004). *Culture, civilisation et idéologie*. Presses de l'Université de Montréal.
- Samovar, L. A., Porter, R. E., & McDaniel, E. R. (2015). *Communication between cultures* (9th ed.). Cengage Learning
- Shah. (2014). Representation of Target Culture in the ELT Textbooks in Pakistan: Evaluation of Oxford Progressive English for Cultural Relevance. *Journal of Education and Practice*.13, 89-101. Retrieved from <https://www.iiste.org/Journals/index.php/JEP/article/view/12716>
- Trudgill, P. (2021). *European language matters: English in its European context*. Cambridge University Press.
- Wardhaugh, R., & Fuller, J. M. (2015). *An introduction to sociolinguistics* (7th ed.). Wiley-Blackwell.
- Yale University. (2014). *what is language?* Yale Linguistics. <https://ling.yale.edu/undergraduate/what-language>