MDPI

Article

English Language Learning via YouTube: An NLP-Based Analysis of Users' Comments

Husam M. Alawadh [1], Amerah Alabrah [2], Talha Meraj [3,*] and Hafiz Tayyab Rauf [4]

1 Department of English Language and Translation, College of Languages and Translation, King Saud University, Riyadh 11451, Saudi Arabia
2 Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia
3 Department of Computer Science, COMSATS University Islamabad—Wah Campus, Wah Cantt 47040, Pakistan
4 Centre for Smart Systems, AI and Cybersecurity, Staffordshire University, Stoke-on-Trent ST4 2DE, UK
* Correspondence: talhameraj32@gmail.com

Abstract: Online teaching and learning has been beneficial in facilitating learning of English as a foreign language (EFL). In online EFL learning, YouTube is one of the most utilized information and communication technology (ICT) tools because of its inherent features that make it a unique environment for learners and educators. Many interesting aspects of YouTube-based learning can be beneficial in supplementing conventional classroom methods, and, therefore, such aspects must be identified. Previous scholarly work aimed at improving YouTube learning environment was predominantly conducted manually by gathering learners' impressions through interviews and questionnaires to analyze the differences between YouTube- and classroom-based EFL learning. However, such methods are tedious and time-consuming and can lead to results that are of less generalizable implications. User comments on YouTube channels are useful in identifying such aspects, as they present a wealth of information related to the quality of the content provided, challenges the targeted audience faces, and areas of potential improvement. Therefore, in our current study, YouTube API is used to collect the comments of three randomly selected and popular YouTube channels. Following a data cleaning process, people's sentiments about EFL learning were first identified via a TextBlob method. Second, the automated latent semantic analysis (LSA) method of topic finding was used to collect global and open-ended topics of discussion on YouTube-based EFL learning. Users' sentiments on the most popular topics of discussion are discussed in this paper. Further, based on the results, hypothetical findings on YouTube EFL learning are provided as recommendation for future content, including more variety of the content covered, introduction of the meanings and punctuation following words, the design of the course such that it addresses a multinational audience of any age, and targeted teaching of each variety of English, such as British and American. We also make suggestions for learners of English who wish to utilize online and offline learning, which include finding the course of interest first based on one's needs which can be discussed with a tutor or any English teacher to optimize the learning experience, participating in fearless educator–learner interaction and engagement, and asking other EFL learners for their previous experiences with learning online in order for the learner to maximize benefit.

Keywords: English learning; online education; sentiment analysis; topic modeling; YouTube learning

1. Introduction

English is considered to be the international language of communication and is termed English as a foreign language (EFL) for its learners. Therefore, proficiency in speaking and listening to English has become necessary in communicating about and understanding worldwide matters [1]. Moreover, there is a growing need for EFL learners' proficiency levels to be reliably assessed for a variety of reasons ranging from enrolling in higher

education to migrating to an English speaking country for professional reasons. This is determined by certain international organizations [2] which categorize EFL learners into different groups based on their English proficiency levels. The age of technology has resulted in technology becoming popular in each aspect of current modern life, with teaching and learning being in the forefront of this transition. Many people are studying various topics using Internet sources such as YouTube videos, webpages, etc., [3].

Similarly, many people are currently learning English from YouTube videos. Moreover, traveling and enrolling in educational institutes require good command of English, and with the advances of and availability of different ICTs, this process has now become more flexible [3]. As a result of using ICT, learning has become more flexible, creative, and enriched in quality [4]. Learning in any domain using ICT results in benefits in various areas; according to a report published by the National Institute of Multimedia Education in Japan, ICT-technology-based learning is more effective compared with traditional study in classrooms. Further, it is reported that ICT-based learners have higher presentation skills, creativity, and understanding [5] of topics.

The use of electronic devices allows the provision of more visually appealing illustrations of topics via audio and video description, graphical representation, animated description, etc. Although these ICT technologies are also being utilized in classroom-based studies to increase student interactivity, remote access to YouTube-like platforms makes learning more accessible for anyone [6]. Smart choices for a more flexible learning environment are being abundantly adopted, including in the form of mobile learning, social interaction, personal grooming and growth, global networking and engagements, etc. [7–9]. Personal growth is one of the soft critical skills that any individual needs for development and involves autonomous learning of an individual at their own pace and environment to cater to those specific needs they desire.

Further, linking and structuring elements of any specific topic is also more beneficial to the learner to gain broader and stronger conceptualization [10]. Interactivity and socialization are other vital factors in learning environments. Individuals interacting with others in academic life helps others to learn from them and explore more ideas [11]. Inclusiveness is another factor in learning environments, and it highlights the barriers that cause issues when teaching students. In this way, the students, regardless of social, national and religious status, can receive equal opportunities in their studies [12].

All these above mentioned qualities of a learning environment are available on a single ICT platform, YouTube, for learning and teaching the English language. Further, a greater variety of authentic teachers [13] from different native languages are available, providing a more specialized experience for EFL learners. Pun et al. [14] state that the content delivered on online multimedia is more functional and has more up-to-date data than textbooks. Further, learning English from YouTube can lead to more up-to-date learning compared with regular institutional studies. At present, many professors in universities and colleges have started their own YouTube channels in which they upload their lectures as a complete course, which helps many students in revising the lectures with flexibility with reference to time, and absent students can also benefit from this through remote learning.

However, many recent studies on YouTube-based English language learning have focused on conducting interviews and questionnaires of EFL learners that seek English from YouTube videos. This highlights the significance of YouTube-based learning and raises specific research questions. The use of manual interpretation of YouTube-based EFL learning can be misguided and biased, as discussed in reports on previous studies of YouTube-based English learning. Therefore, analysis of the real-time collected comments, likes, and responses to those comments by other users can give practical information regarding YouTube-based English language learning. Further, examination of the YouTube comments provided by citizens of different nations can provide greater insight into YouTube-based English language learning. Our current study solves the following challenges regarding YouTube-based EFL learning:

- An automated solution was established to collect multinational data about YouTube-based EFL learning and analyze the most discussed topics and the sentiments of learners regarding these topics.
- How and why does English learning from YouTube videos influence students compared with conventional EFL learning? Is YouTube helping or misleading online consumers?
- What are the challenges and difficulties of learning from internet sources?
- Suggestions and hypotheses were derived based on LSA-based topic modeling of user feedback in the form of YouTube video comments.

To answer these questions, in our study, the data of comments and other information from several popular YouTube channels were collected. These include the benefits of learning from YouTube videos and the gaps that some viewers lack. The remaining sections of this work are Section 2: Related Work; Section 3: Material and Methods; Section 4: Findings and Discussions; Section 5: Hypothetical Findings and Suggestions; and Section 6: Conclusion, which includes future suggestions.

2. Related Work

Many of the studies that investigated YouTube-based English language learning highlight the benefits of learning from YouTube. They mostly relied on manual collection of datasets via conducting interviews and the use of questionnaires. This targeted specific users, while it also did not involve real-time comments or data. However, we highlight a few studies which were experimental in nature here.

A study conducted by a Taiwanese university reported certain facts about YouTube-based English learning and included 20 students in the hypothetical experimentation, and answers to certain predefined questions were obtained from these students. The three major areas of concern were highlighted in terms of three questions asked by the interviewers and classified into multiple options based on the EFL learners. Essentially, classroom- and YouTube-based EFL learning were differentiated. Further, the study explored whether learning is self-regulatory or not using the responses of EFL learners. The study was based on SRL models and focused on six goals. The inter-rater agreements were calculated based on EFL responses and the SRL models, and the responses were correspondingly classified. Lastly, the differences and similarities between EFL and regular classroom learners were highlighted [15]. In another example, YouTube-based language learning research was used as the basis for a conceptual study that marked the advancements and limitations of YouTube-video-based language learning environments. It reported that classroom-based language learning could be more effective if YouTube-based assistance is added to guide the students. Further, due to the availability of authentic content, the learners can be more involved in the selection of the instructor in the YouTube videos [16]. The study was conducted in a school where English was studied as a subject, and for data collection, 14 randomly selected students were enrolled. The students had both online and offline experiences, and the conducted study based its conclusions on student reviews.

In another study, three main questions were analyzed based on the responses of students to certain questions on (1) the differences in offline and online studies for EFL learners, (2) user interaction regarding study using ICT tools or YouTube as a learning environment, interaction, and engagement to multimedia, and (3) the learner's personal experience of learning and speaking English using ICT tools. For the quantitative exploration of the results, the Pearson correlation test and regression analysis were used. The time range of YouTube channel videos was analyzed, wherein the exposure between these watching times was analyzed. For qualitative analysis, the improvement in writing skills and the affordance of YouTube videos were reported by asking all learners simple questions. The study also concluded that negative aspects were related to personal growth and the accuracy in the fluency of learning and writing English [17].

Wahyuni et al. conducted a study on YouTube-based English language learning in an Indonesian setting, and a class of 40 students was observed. It was reported that

the learning environment of any learning is highly significant. Therefore, the English language learning-based study was conducted by applying a qualitative approach in the experimental setup. Mainly English speaking was analyzed, where four steps were repeatedly applied to answer the questions asked in the questionnaire reported in that study. The statistical measures of mean, standard deviation, maximum, and minimum were applied to the questions and their answers. It was concluded, based on the statistical analysis, that YouTube provides a good environment for learning English speaking [18].

The author of another study claimed that there are two modes of learning: audio and video. The study recommends video illustration over audio for English speaking. The cognitive reasoning regarding the use of video and audio media is described in the methodology section, where the multiple styles of learning and effective reasons are further described. It was reported that when using video-based EFL learning, there is less cognitive load on the learners. It was further stated that YouTube is not only useful for entertainment but can also be beneficial for language learning [19].

Putri et al. [20] reported in their study that YouTube is a more flexible, entertaining, and fun environment to learn the English language. The study included experimental data based on 12 students. Qualitative analysis was performed by asking questions on different cultural, social, meta-cognitive, resource, effectiveness, and goal commitment aspects. The analysis was based on the different factors, concerned with their effect on how swiftly the English language can be learned.

Another study based on making and uploading English-speaking videos on YouTube was conducted based on 10 participants and qualitative analysis of their answers to questions was performed. The questionnaire in use comprised five questions with an 80% response rate, which showed that before making English-speaking assignment videos and uploading them on YouTube, extensive practice was conducted on correcting their English pronunciation. Therefore, to ensure correct English pronunciation, English speaking could be improved by uploading English-speaking videos. The different suggestions to improve English pronunciation made by all learners were also provided [21].

To analyze YouTube-based EFL learners, the authors of [22] stated, in their study, that this platform contributes to developing individual participation as well as critical thinking and social engagement. Authentic English teaching material is available on YouTube. Additionally, a few channels also provide supplementary materials for English language learning. Including all these aspects results in YouTube being an effective source of learning for EFL learners. Another study on YouTube-based EFL learning included 32 students [23]. The pros and cons of the learning environment were discussed by presenting students with open-ended questions. The report concluded that YouTube is a very familiar, cheap, and flexible platform for EFL learning. However, participants encountered difficulties that could be resolved by using different YouTube-video-watching features and by discussing problems with teachers and friends.

A study in an Indonesian university included students participating in EFL learning via YouTube [24]. This involved the use of the Web 2.0 program of the YouTube platform, which provides many other benefits over traditional EFL learning. This research uncovered the benefits of YouTube-based learning over classroom learning. Certain practices were also suggested for improving teaching in the future to ensure students felt more at ease in that environment.

There are many areas of EFL learning in which different people are weak and wish only to improve in that specific aspect. To investigate this, a study was conducted on YouTube-based EFL learning for improving pronunciation [25]. The benefits of using YouTube were revealed through such remarks as videos are free to watch, different speakers from different nations even with good British and American English could be found to improve pronunciation, in particular, and the possibility to watch videos anywhere and also on different devices. The recommendations given were to learn pronunciation while watching YouTube videos, where important points were also given before and after watching videos.

In another study, student reviews were collected using Google Forms, and based on the questionnaire answers, it was concluded that YouTube-based EFL learning occurs most specifically according to the content given in the course with the addition of attractive and flexible features [26].

The different recent studies on EFL learning discussed above highlight the use of YouTube in English language learning for improving writing, speaking, grammar, and pronunciation.

However, the data that such studies utilized were only manually collected, and qualitative and quantitative analysis was mostly performed on only certain and specific questions. The research questions in those studies were built upon personal impressions and analyzed with some statistical approaches. Therefore, to bridge a clear research gap related to identifying wider challenges of YouTube-based English language learning, an automated solution to collect real-time responses from channel comments is needed. Moreover, qualitative and quantitative discussion of EFL learner responses is also required. In this way, the problems and benefits can be highlighted. Further, inter-comparison of the different YouTube-based channels could allow pedagogical implications to be formulated and shared with content makers to facilitate more positive responses from their viewers.

3. Materials and Methods

In the current study, user comments on three popular YouTube English teaching channels were collected via an API provided by Google developer services. The data were cleaned after scraping from the API, and playlist-wise topic modeling was performed, where each channel playlist topic was compared to check the user response to the topic. The primary purpose of this was to increase the audience and enhance their interest by proposing suggestions for new and already existing and popular YouTube channels. The results obtained form the current study may also carry important implications which can pour into the improvement of traditional EFL classroom environment. All primary steps of the current study are shown in Figure 1.
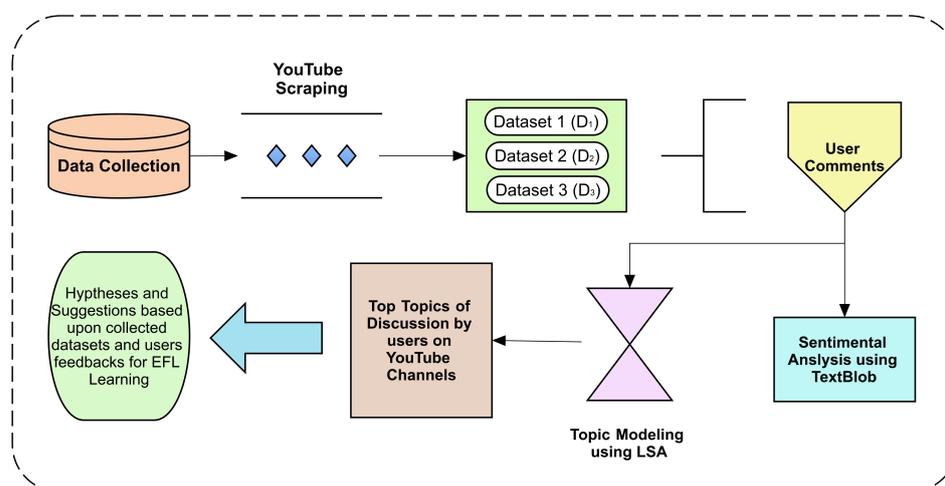
Figure 1. Flowchart containing all steps included in the current study to analyze YouTube-based EFL learning using computer-assisted methods.

Data acquired from YouTube were then subjected to cleaning. The user response on each topic of each channel is important. Therefore, at first, the sentiment of the user was checked using the TextBlob method of sentiment analysis, and, secondly, their topics of discussion in the comments box were discussed following topic modeling via the LSA method. The most popular topics and their associated used words were discussed, and suggestions were proposed by determining people's sentiments on these topics.

3.1. Data Collection and Cleaning

In previous studies, authors mostly selected interested students from different semesters and assigned them to watch videos and share narratives/notes describing their impressions. After that, interviews and questionnaires were conducted and completed to analyze the topics and behavior. The predefined questions asked mostly answer the following major and basic questions: Is YouTube a good source of learning for EFL learners or not? Similarly, are the sources on YouTube trustworthy and authentic or not?

However, in all previously used approaches, automated and real-time collection from a multinational audience was lacking. In this way, the known issues and benefits of using YouTube-based EFL learning would also be lacking. In our study, we manually selected the three of the most viewed channels on EFL learning on YouTube and scraped people's comments on them together with their likes and responses on each playlist. A YouTube developer account was created and the API key was used, and the collection of user responses according to channel and playlist was performed. The selected channel names are (1) Learn English with Englishclass101.com (D_1) [27]; (2) Learn English with Jessica (D_2) [28]; (3) Speak English with Vanessa (D_3) [29].

The textual data of comments on these channels contained stop words, URLs, and other redundant information that was removed using the regular expression library of Python. After collecting and cleaning data on all playlists, a topic wise analysis was performed. We highlight the main findings from this step below.

3.2. Dataset Topics

Different topics of discussion were found on the three different channels. The user response and interest in all topics were analyzed and included negative, positive, and neutral responses. First, the video topics are discussed here, and the findings related to these are discussed in Section 4. Table 1 describes the datasets and their discussed topics.

Table 1. Topics of videos in the three datasets used in this study.

Datasets	Topic ID	Topics of Discussion in Videos
D_1	D_1T_1	"Learn basic English vocabulary"
	D_1T_2	"English conversational phrases"
	D_1T_3	"Fun and easy"
	D_1T_4	"American English intermediate reading practice"
	D_1T_5	"American English advanced reading practice"
	D_1T_6	"British English advanced reading practice"
	D_1T_7	"English listening comprehension for absolute beginners"
	D_1T_8	"English listening comprehension for beginners"
	D_1T_9	"English listening comprehension for intermediate learners"
	D_1T_{10}	"English listening comprehension for advanced learners"
D_2	D_2T_1	"Practice English speaking conversations"
	D_2T_2	"Learn English speaking using daily life conversations"
D_3	D_3T_1	"How to learn English"
	D_3T_2	"American vs. British English"
	D_3T_3	"English listening lessons"
	D_3T_4	"English vocabulary"
	D_3T_5	"Business English"
	D_3T_6	"Travel English"
	D_3T_7	"English pronunciation"
	D_3T_8	"English grammar"
	D_3T_9	"Phrasal verbs"
	D_3T_{10}	"Live English lessons"
	D_3T_{11}	"Skype tests"
	D_3T_{12}	"Fearless discussion"
	D_3T_{13}	"Interviews in English"

We can see in Table 1 that there are several different topics discussed in the videos. The first dataset contains the topics of English vocabulary, conversational phrases, listening, and speaking at beginner and advanced levels in both British and American contexts. The second dataset involves creating animated videos containing anonymous objects. Most of the videos in this dataset contain the regular use of English regarding speaking.

The third dataset contains interesting topics compared with both of the above-discussed datasets. It contains the regular content for EFL learners as well as live interviews and the fearless discussions of its users/subscribers through Skype and other live interactions. It also contains different levels of the syllabus for EFL learners according to their usages, such as business and travel English. The user response on all these datasets and their topics is more interesting and is discussed in the following section of this manuscript.

3.3. User Sentiment in Dataset Topics via Sentiment Analysis

In previous studies, statistical analysis was performed on answers to predefined questions in questionnaires asking about YouTube videos. The analysis concerned user sentiment and suggestions to watch YouTube videos for EFL learning. Similarly, there was an automated collection of user sentiment based on open-ended comments, sentiment analysis of all comments on different topics in the videos was performed for each dataset, and the sentiments on each topic are shown in Section 4. TextBlob was used to perform sentiment analysis.

3.4. User-Response-Based Topic Modeling via Latent Semantic Analysis (LSA)

In previous studies, most of the questions and topics were user-defined and collected and provided by authors. In our study, we examined user comments on different YouTube channels in which there were different topics of discussion. To analyze the topics of discussion in comments, topic modeling based on LSA [30] was performed.

The cleaned textual data were first provided to the LDA model [31], and the most frequently occurring unique topics from each playlist of each channel were analyzed. The mathematical representation of applied steps on documents or input data are shown in Equations (1)–(3).

$$d_n(i) = p_n(i) \tag{1}$$

In Equation (1), D_n represents the three datasets, where n varies from 1 to 3, and i represents a certain comment or text document that is passed from the preprocessing step to clean the document from stop words, URLs, etc. The cleaned data are stored in D_n. These cleaned data are then used to create the count vector document matrix. The mathematical representation is shown in Equation (2).

$$d_c = d_n[count] \tag{2}$$

D_c represents the document count against the cleaned data matrix (D_n). The count is the term count in each document or text comment in this case study. To obtain multi-matrix decomposition as acquired by the LSA model, the singular value decomposition (SVD) method is used, where topics on each channel's comments data are manually specified here. SVD is a matrix term that is used to reduce the matrix dimensionality [32]. The mathematical representation is shown in Equation (3).

$$T_m = SVD(m, D_c) \tag{3}$$

In Equation (3), two input parameters are specified while making the SVD matrix that will generate the unique topics of discussion. Topics are represented by T_m, where m is the number of topics, and the output is calculated by using specific topics (m) and the cleaned count vectorizer of each document. After obtaining user topics of discussion via each dataset, sentiment analysis is performed to determine user sentiment as being either positive, negative, or neutral.

4. Findings and Discussion

There are two types of topics that need to be discussed in this section about the EFL learning topics and the user responses to them. First, let us explore the collected datasets and the sentiments of people about them.

4.1. Sentiment Analysis of Collected Datasets

There are three datasets collected where the basis of the collection was their different playlist topics. To identify people's sentiments on each topic of channels, the three basic sentiments of people were collected based on the polarity of their comments. The sentiments for each topic are shown in Table 2.

Table 2. User sentiments on topics of three YouTube channels about EFL learning.

Topics of Discussion in Videos	Sentiments (%)		
	Positive	Negative	Neutral
$D_1 T_1$	75	5	20
$D_1 T_2$	75	15	10
$D_1 T_3$	72.7	9.1	18.2
$D_1 T_4$	27.3	9.1	63.6
$D_1 T_5$	58.3	0	41.7
$D_1 T_6$	75	10	15
$D_2 T_1$	75	5	20
$D_2 T_2$	80	0	20
$D_3 T_1$	92.5	2.5	5.0
$D_3 T_2$	92.5	2.5	5.0
$D_3 T_3$	96.7	0	3.3
$D_3 T_4$	92.5	2.5	5.0
$D_3 T_5$	95.0	2.5	2.5
$D_3 T_6$	85.0	2.5	12.5
$D_3 T_7$	93.8	2.5	3.8
$D_3 T_8$	94.0	1.0	5.0
$D_3 T_9$	95.0	0	5
$D_3 T_{10}$	100	0	0
$D_3 T_{11}$	90	0	10
$D_3 T_{12}$	100	0	0
$D_3 T_{13}$	77.5	10	12.5

There are many topics, and the different sentiments regarding these can be seen in Table 2. Regarding our first dataset topics and their sentiments, we can see that there are six unique topics for which the user sentiment analysis was performed. The sentiments regarding basic vocabulary remain less negative compared with those on the discussion of conversational phrases.

Fun and ease, which is the third topic of sentiment analysis in the table, is also associated with 9% negative sentiment, and this could be due to the critical thinking about this topic of discussion. The fourth topic is associated with far less positive sentiment and is American English intermediate, and the next topic $D_1 T_5$ is more negative and is advanced-level American English reading practice.

The British English reading-based comments received more positive feedback compared with American readings, as can be seen in $D_1 T_6$. This could be attributed to higher quality exposure of and practice in the British variety of English compared with the American context, though this needs to be formally established in a future study utilizing validation and triangulation techniques.

Regarding the second dataset, which specifically utilized animated videos to teach spoken English, the self-made objects were made and used to practice spoken English. Both playlists were associated with positive sentiments and some (5%) negative sentiments. Therefore, it can be said that increasing excitement by employing non-traditional and

innovative techniques such as making animated videos for object-based English practice and learning can lead to positive impressions amongst the target audience.

4.2. LSA-Based Topic Modeling of User Comments

Topic modeling based on user comments was performed, where the most popular topics are also identified with the use of the most significant words in them. The count-vector-based feature vector is provided to the SVD matrix maker and the LSA model is then applied.

4.2.1. Dataset 1 (D_1) Topics Discussed by Users

The topics of user comments from each playlist of datasets were collected, and those for dataset D_1 are shown in Table 3.

Table 3. User topics of discussion in YouTube videos in D_1.

Topic IDs	Topics Discussed by Users
D_1T_1	"English lesson like learn thanks speaking teacher thank lot good" "Words described entire learn language kashmir just jammu imagine good"
D_1T_2	"Like engineering faculty lesson helpfuli petroleum continue awesome student studying useful" "Maximum respect appreciate coz daily improving madam really aliciai saudi great alisha" "Thank language improving saudi arabia learn english helpfuli great going faculty" "Videos going marathon yamaximum department anyone hi helpfuli"
D_1T_3	"Important jobs sends bookstores branch brands details rights restaurants"
D_1T_4 and D_1T_5	"Live pay afford bills guys bangladesh" "Cost transportation afghanistan mount education semester lot living city like home" "english help wash mim dificil fokor hi hmmm improv want" "alisha 00am practice im enjoy like btw boring beautiful simples" "Thanks video nice"
D_1T_6	"World known billgates channel hello 2nd position portuguese nice love" "Good uploading content thanks achi baat billgates bot british caption channel" "english sure british world achi baat billgates bot caption channel" "portuguese caption world english achi baat billgates bot british channel" "Speak nice world english achi baat billgates bot british caption" "Thank love world english achi baat billgates bot british caption"

The table of topics in the dataset (D_1) shows the major discussed topics based on videos of each topic. The first topic (D_1T_1) shows the type of topics that are more likely to be associated with positive sentiments and even show that the user Jammu Kashmir belongs to and is inspired by the English teaching of this channel.

Secondly, we can see in D_1T_2 that people from Saudi Arabia and Bangladesh follow this channel and strongly recommend the language learning style used here for conversational phrases videos. In the third topic of this channel, the user mentions that it is a fun and easy way to learn English. It includes user-interaction-based activity to comment in English for a given phrase. Therefore, the user comments are almost the same, talking about bills, electricity, education, etc., again and again.

Only three topics of discussion are listed in the table where other similar topic related comments are discarded. The fourth and fifth topics, which are on American English learning, already showed slightly negative sentiments of people compared with British English on this channel. In the user comments, the user is saying that the video is boring but simple and has given suggestions for where improvement is needed.

The next topic on British English learning received positive comments, and it also came from a multinational audience learning British English from this channel that positively recommended taking British English lessons through this YouTube channel.

4.2.2. Dataset 2 (D_2) Topics Discussed by Users

The users of the second channel gave mostly positive responses on data, where their discussed topics are discussed in Table 4.

Table 4. Users topic of discussion on topics of YouTube videos on D_2.

Topic IDs	Topics Discussed by Users
D_2T_1	"Really English know video dear thank wish improved love maker" "Heart efforts thank wish god know improved happy great good" "Want does practice good love know improved heart happy great" "Video thanks good wish god know improved heart happy great" "Wish god thank know improved heart happy great good fluently" "Mam super god know improved heart happy great good fluently" "Waiting to wish god know improved heart happy great good" "Pronunciation great wish maker know improved heart happy good god" "Congratulations efforts good love know improved heart happy great god" "Nice wish maker knows improved heart happy great good god"
D_2T_2	"Absolutely time listen" "English Channel story"

In the second dataset, channel comments where animated videos were given received many positive comments and best wishes, whereas comments about the video maker and story maker being nice were given. Now, it can be noted that by making objects and animated videos, plays, and by making small stories in an innovative fashion, we can increase the efficacy of EFL teaching utilizing ICT. Again, this is an area of potential future research focusing on innovative content creation in ICT and how it can increase its utility and impact in the learning process. Certain areas of language learning such as expressions, pronunciation, and vocabulary can be easier to deliver using innovative content as compared to other areas such speaking.

4.2.3. Dataset 3 (D_3) Topics Discussed by Users

The third channel includes interesting topics, compared with the two other channels, that essentially use the same strategy to teach English, which also works in regular classrooms and courses for EFL learners. The users on this channel and its topics also showed a higher proportion of comments and sharing compared with the other two channels. To examine the differences and reasons for more interaction of users of this channel, the most discussed topics are shown in Table 5.

Table 5. Topics of discussion of users of YouTube videos on D_3.

Topic IDs	Topics Discussed by Users
D_3T_1	"Day english like learn vanessa" "English doing vanessa really way"
D_3T_2	"English just vanessa interesting time use really native video expressions" "Piano asked said practice plan shoot daughter scene ready discussion" "Lot helped videos thanks dear vanessa fluency write glad friends" "Vanessa lovely write love educational exactly excellent expressions fabulous mood" "Great lesson like shoot know family breeze bless understand thank"
D_3T_3	"Vanessa thanks videos accent learning make new excited English rock" "Just years bless speaking growing teach past confidence important following" "Know useful ex situations really easy demise make love certain" "Videos learned helpful routine did thank daily things new watch" "Podcast listen amazing app beautiful, best videos vanessa teacher hope" "Lesson lot life English thanks just learning wow amazing learned"

Table 5. *Cont.*

Topic IDs	Topics Discussed by Users
D_3T_4	"Learning years hard English hope simple day daily comment good" "Regard appreciate enjoyed watching warm Indonesia video pet vanessa really" "YouTube channel really subscriber helped new relaxing love watch australia" "Like listening positive speech thank useful lot clear videos watching" "Dislike video people thanks loved vanessa class wonderful important indonesia" "Vanessa lessons dear love thank youtube hard helped hope important" "Enjoyed fun talking session learning learn just intensive inspiration indonesia"
D_3T_5	"Thank great english vanessa don lessons hello expression lesson love"
D_3T_6	"Lesson marek looking subscriber nice realy regards rent response new" "Vanessa thanks going thank english glad teacher help learning kisses" "Want bother lessons grammar don excuse world hotel helpful hey"
D_3T_7	"Love videos 70 help thanks thank enjoy teacher teach good" "Summer ll headed think wish instead lesson place portugal great" "Vanessa thanks mood nice smiling instead help helps hi hope" "Clear muchlove sweet helps pronunciation improves english india vanessa hello"
D_3T_8	"Lesson great words job beautiful vanessa thanks english hi helpful" "Thank vanessa chinese varied advanced creating media mainland love lessons"
D_3T_9	"Learn english videos different symbols fluently smile difficult" "Learn accidentaly fun channel clear success speech dear" "Lesson ve youtube easy simple sooooooo did practical" "Lessons like enjoyed struggling teach love brought vanessa" "Make isn possible plans vacations vanessa years amazing like lessons"
D_3T_{10}	"Happiness wish following tips month people relationship great" "Lots of useful love lesson thanks expressions lot channel"
D_3T_{11}	"Understand movies teacher" "youtube videos speakers"
D_3T_{12}	"Really like natural shock reverse remembered face speaking culture country" "Understand like conversations problem try speak advanced 90 listening hours"
D_3T_{13}	"Good say experienced identified inmersion interacte knowledge deep" "Learnt lot good debate work different identified hello"

As previously discussed, there are certain unique topics of discussion that are distinguished and described by the LSA model. For example, the users of dataset (D_3) showed more interest compared with those of datasets 1 and 2, and there are many aspects that seem to engage the audience. At first, the very basic start of this channel poses the question "How to learn English?", a very basic question in itself to which there are many different answers, and even people of different ages and language backgrounds could all learn English. People's sentiments and comments all remained positive and they appreciated the way to start EFL learning.

Secondly, there is a global issue of whether to learn either British or American English. The differences are highlighted when people discuss this topic, and excellent differences are highlighted with acquired expressions to speak the English of both varieties. The third topic D_3T_3, to take English lessons again, engages many audiences, the teacher's accent and expressions are, again, much appreciated, and the lesson is also much appreciated by the audience and viewers.

Acquiring and retaining Vocabulary can be a challenge for many EFL learners. Users who visited the YouTube channels under investigation appear to agree with this. On this topic, D_3T_4, many people from different countries, such as from Australia, say that Indonesian people watching a video on this topic feel more relaxed and inspired after watching these videos. The next topics of the video are related to English on business and travel

demonstrating a narrower focus on a less targeted EFL learning area especially in a free online resource.

However, business English with topic ID D_3T_5 showed all positive responses with, again, expressions-based acknowledgment. The English grammar topics and the phrasal verbs are much appreciated, and people immensely enjoyed these difficult topics. Some users found the channel to be unique and rare with interesting topics and expressions based on EFL learning. In D_3T_8, the people appreciated the advancement of this course of EFL learning and showed immense appreciation to this teacher. For this topic, it was also mentioned that a practical and easy way of learning was introduced, and teachers are increasingly struggling to teach lessons in an interesting way.

The people commenting on topic D_3T_9 also mention that they accidentally found an amazing and fluent English-speaking channel. The last topics are all practice- and interaction-based and lead to different, but fully positive, sentiments. The live sessions with students, users, and regular viewers are a unique and fruitful activity to teach English. The people themselves give positive remarks such as that they can never forget this fluent way of speaking English full of expressions and that they are more likely to understand it based on the positive expressions of the teacher.

4.3. User Difficulties Found in the Collected Datasets

In the collected datasets, although the sentiments mostly remained positive, there are a few concerns highlighted in all topic-based data from the scraping of datasets. The remarks of interest are represented as a word cloud in Figure 2, which summarizes the negative topics and remarks of dataset 1.

Figure 2. Word cloud of negative words in dataset 1 (D_1).

We can see from the word cloud that the topics of discussion are narrow in words such as that the user "wants" something; for example, "speak" loudly, "boring", "exams", "slowly", etc., are the top negative words used in dataset D_1 channel comments. The negative comments can be seen in detail in Table 6, which shows user negative remarks more deeply forthe difficulty level.

Table 6. Top negative commented topics in dataset (D_1).

Serial Number	Topics
1	"im sorry this video is very boring"
2	"Why train stations i mean over and over again"
3	"Could you speak more slowly please It's very difficult to learn with that speed"
4	"exam is not easy exam is difficult"

Users' negative comments mainly revolved around a few areas which we discuss here. First, word repetition made videos according to users boring to watch. Speaking loudly is also a topic causing negativity in users' discussion, and speaking slowly is also another issue. In addition, exam difficulty, where applicable,is another source of negative sentiments that can be exhibited as a user's personal problem with the site.

To determine the difficult topics in dataset D_3, the top words of interest are also shown as a word cloud in Figure 3. The top words show, again, the positive remarks, and why a comment is marked as negative should be discussed. The negative topics are listed in Table 7.

Figure 3. Word cloud of negative words in dataset 3 (D_3).

There is a mix of positive and negative words in negative marked comments in dataset 3. Words such as "thanks", "British", and "phrasal" need to be examined in more depth when looking at topics that are listed in Table 7.

Table 7. Top negative commented topics in dataset (D_3).

Serial Number	Topics
1	"Hi, this question has been reverberating in my mind for the past few days Is there any difference between American and British English phrasal verbs Does American use British phrasal verbs"
2	"Don't you think that is strange how they are always smiling"
3	"She speaks very fast Please someone needs tell him SLOW DOWN YOUR SPEED"

The topics with the highest keywords found in comments categorized as negative from TextBlob methods are shown in Figure 3, and the filtered topics are shown in Table 7. Most of the comments were expressing thanks to the content creator Vanessa, whereas the negative or questioning comments are showing that some users are asking whether British and American phrasal verbs are the same or different, and what the differences are. The

other topic is the same as in dataset (D_3): "Slow down the speed of speaking where you speak very fast". The other topic is that the "user feels strange that teacher always keep smiling".

The topics of discussion with negative sentiment are quite the same and show that the user feels difficulty when they are learning. However, slowing the speed, which was found to be a common difficulty in users of different datasets, could be solved using the YouTube feature to slow down the video playback. The other remarks regarding always smiling is a single person's opinion, where most of the users that commented on dataset 3 videos highlighted the importance and happiness toward the expression of its teachers. Word repetition in video creation could be an irritating point while learning English. The created video does not need to be long, and if long, it should not be boring, ensuring that interesting facts and activities are added to avoid boredom.

4.4. Key Features of YouTube-Based EFL Learning

The most prominent key feature of YouTube-based EFL learning is its remote access and control. Restrictions such as those regarding time and speed are not relevant. YouTube also provides subtitles and options for speed increase and decrease when playing videos that could be more beneficial for listening and understanding video topics according to user interest. Punctuation and pronunciation problems can be addressed in either British or American Englishes, as both can be found on YouTube. The live streaming feature is now also of interest to engage students globally to ask questions and to give suggestions for acquiring a different set of skills within the EFL learning environment. We saw reviews from people that are mostly positive regarding online or YouTube-based EFL learning. This positivity relies on many key factors. The key points or features of YouTube-based EFL learning are video sharing among friends or classmates by teachers, which reduces the reproduction of the same lecture by the instructor. It also assists absent students to catch up on certain topics. The other feature could be its larger and more extraordinary syllabus or content compared with regular classroom-based textbook learning. Global viewership is another important feature which can be considered a strength of YouTube-based EFL learning. People from different nationalities watch these YouTube-based EFL learning channels and it would be interesting to gather more information of this global viewership (e.g. where the user posting the comment is actually from). The people that respond in comments are visually illustrated according to the country in Figure 4.
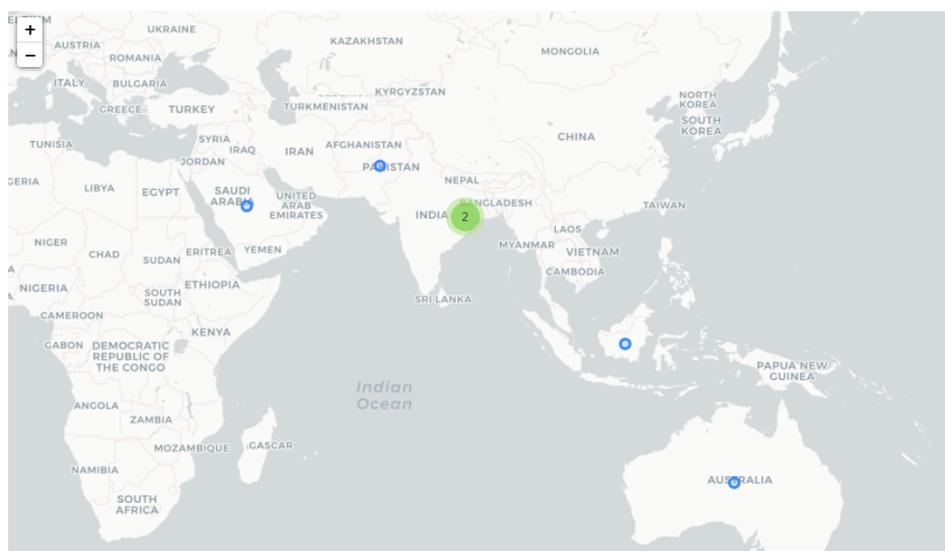

Figure 4. Users engaged in collected YouTube datasets from different regions of the world.

We can see that people from mostly the Asia Pacific region engaged in collected datasets viewership, whereas a few of them also belong to Saudi Arabia, Australia, and In-

donesia. Few users are found in the United Kingdom, United States, and other European countries. Therefore, we could emphasize here that regions that people are mostly from may need more English learning, and those from other countries may also need EFL learning but not as much as people from Asia.

5. Hypothetical Findings and Suggestions

The topics of three different YouTube channels and discussions of their users were investigated in this study. Many new findings were uncovered, compared with previous studies, and the summarized hypothetical findings are discussed here and shown in Table 8. In H_1, classroom-based textbook learning has a traditional old syllabus, whereas it is found that YouTube channels include a larger syllabus and new content compared with textbooks. An issue for consideration, as stated in H_2, is determining how to learn English vocabulary and pronunciation, which is a basic topic discussed by one of the investigated YouTube channels. Therefore, it is found to be an important factor to be included in learning. It is also found that expression-based EFL learning is important as it includes the necessary accent, expression, etc., and is described in H_3. Fearless discussions between teachers and students are also an important factor in EFL learning that increases the confidence of learners and is shown in H_4. This will boost EFL learning. EFL learning varies according to people's interests, such as regarding British, American, and business English. Therefore, an acquired syllabus and content are needed in EFL learning, as mentioned in hypothesis H_5. A general and specific age-based understanding of EFL learning is required when designing an EFL course. H_6 shows this general requirement that needs to be added in courses.

In S_1, user-based suggestions are specifically given to learn English according to your interests. In S_2, trainer authentication needs to be cross-checked to determine whether the user is learning correctly or not. In S_3, each type of question asked in comments or live interaction should be answered to engage an increasingly larger audience. In S_4, ask the EFL learner what new things, improvements, and suggestions for things need to be added in future videos or courses. In S_5, one of the channels included animated videos and stories in which topic-wise discussion was added for listening and viewing by users. There are certain problems, challenges upon exploration, and discussion of datasets that are found listed in tabular format. There are two types of categories of findings from the YouTube channel datasets, and the user discussions on them were explored and used as a basis for deriving certain hypotheses and suggestions. From H1 to H5, how things are happening on different YouTube channels based on EFL learning is discussed.

The positive and dominant factors compared with regular classroom-based textual studies are discussed. These hypotheses play major roles in YouTube-based EFL learning compared with regular classroom EFL learning. However, a few suggestions made on user responses in comments were collected that could improve both classroom- and YouTube-based EFL learning.

Table 8. Hypothetical findings and suggestions (for better convergence) based upon dataset and user response upon watching YouTube videos.

Hypothetical Finding IDs	Hypothesis
H_1	Classroom-based EFL learning is confined to textbook-based old syllabus, whereas online YouTube-based EFL learning contains new, updated, and a variety of topics in EFL learning.
H_2	Suggestion to add a basic start topic in EFL learning such as how to learn English and how to learn vocabulary, conversation phrases, etc.
H_3	Real-time expressions while teaching English are much appreciated by users and are needed and play an important role in EFL learning.

Table 8. *Cont.*

Hypothetical Finding IDs	Hypothesis
H_4	For fearless discussion and student–teacher interaction, it is suggested that online YouTube-based EFL learning is added to regular practice in classrooms.
H_5	Each topic of EFL learning needs to be categorized and separately taught, such as British and American English, business and traveling English, etc.
H_6	Make and design courses in such a way that people of all ages can easily learn and understand English.
S_1	For online EFL learning, reach out to and search for a channel particular to your interest.
S_2	Authenticate the teacher/trainer from your classroom and instructor so that you do not learn incorrectly.
S_3	Respond to subscribed users to answer their silly and important questions to increase the audience and to ensure a better teaching style.
S_4	Ask users specifically what kind of EFL learning they want and what kinds of improvements could be made to their online courses.
S_5	Animated and object-based story creation can also be used to improve EFL learning, and live interaction is essential as it engages more users in topic-based discussion according to the findings of datasets.

6. Conclusions

The current study scraped the comment data of three different YouTube channels, using the YouTube developer-provided API service. The scraped data were collected and cleaned over a textual preprocessing phase and then assessed to determine user sentiments on topics discussed in them. Then, the user comments based on top and unique topics were collected to obtain user reviews of these. To obtain topics of user discussion, the LSA method was used, which works on the sentiments of a given document. In previous studies, manual inspection of YouTube-based EFL learning was conducted on limited datasets of mostly less than 100 people. Thus, they do not highlight the benefits and challenges of global users. However, this study used an automated method for data collection as well as to find topics in user comments. This resulted in making hypotheses regarding the importance and uniqueness of YouTube-based EFL learning. The benefits of this learning are also discussed, such as having a larger syllabus on YouTube as compared to textbook learning, expressions and punctuation following words, course design that covers a multinational audience and is suitable for learning by a student of any age, and allowing separate learning of each variety of English, such as British and American. Further, classroom-based EFL learning could also be improved by implementing the following suggestions: find the course of interest first based on one's needs which can be discussed with a tutor or any English teacher to optimize the learning experience, participate in fearless educator–learner interaction and engagement, and ask other EFL learners for their previous experiences with learning online in order for the learner to maximize benefit.

Author Contributions: Conceptualization, T.M. and H.M.A.; Methodology, T.M.; Software, T.M.; Validation, A.A. and H.T.R.; Formal analysis, H.M.A. and T.M.; Investigation, A.A.; Writing–review and editing, H.T.R. and H.M.A.; Supervision, H.M.A.; Project administration, H.M.A. All authors have read and agreed to the published version of the manuscript.

References

1. Getie, A.S. Factors affecting the attitudes of students towards learning English as a foreign language. *Cogent Educ.* **2020**, *7*, 1738184. [CrossRef]
2. Macaro, E.; Jiménez Muñoz, A.; Lasagabaster, D. The importance of certification of English medium instruction teachers in higher education in Spain. *Porta Linguarum* **2019**, *32*, 103–118. [CrossRef]
3. Tahat, K.M.; Al-Sarayrah, W.; Salloum, S.A.; Habes, M.; Ali, S. The influence of YouTube videos on the learning experience of disabled people during the COVID-19 outbreak. In *Advances in Data Science and Intelligent Data Communication Technologies for COVID-19*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 239–252.
4. European Commission (EC). *Making a European Area of Lifelong Learning a Reality*; European Commission: Brussels, Belgium, 2001.
5. Aoki, K. The use of ICT and e-learning in higher education in Japan. *Int. J. Educ. Pedagog. Sci.* **2010**, *4*, 986–990.
6. Reinders, H.; Benson, P. Research agenda: Language learning beyond the classroom. *Lang. Teach.* **2017**, *50*, 561–578. [CrossRef]
7. Buchem, I.; Pérez-Sanagustín, M. Personal learning environments in smart cities: Current approaches and future scenarios. In Proceedings of the PLE Conference 2013, Berlin, Germany, 10–12 July 2013; pp. 136–154.
8. Mikuleckỳ, P. Smart environments for smart learning. In Proceedings of the DIVAI 2012 9th International Scientific Conference on Distance Learning in Applied Informatics, Sturovo, Slovakia, 2–4 May 2012; pp. 213–222.
9. Libbrecht, P.; Müller, W.; Rebholz, S. Smart learner support through semi-automatic feedback. In *Smart Learning Environments*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 129–157.
10. Vaughan, T. *Multimedia: Making It Work*; McGraw-Hill, Inc.: New York, NY, USA, 2006.
11. Al-Mukhaini, E.M.; Al-Qayoudhi, W.S.; Al-Badi, A.H. Adoption of social networking in education: A study of the use of social networks by higher education students in Oman. *J. Int. Educ. Res.* **2014**, *10*, 143–154. [CrossRef]
12. Rice, D. Appendix 1: Use of ICTs for Inclusive Education. 2020. Available online: http://inova.snv.jussieu.fr/evenements/colloques/colloques/article.php?c=70&l=en&a=361 (accessed on 15 May 2022).
13. Grzeszczyk, K.B. Using multimedia in the English language classroom. *World Sci. News* **2016**, *43*, 104–157.
14. Pun, M. The use of multimedia technology in English language teaching: A global perspective. *Crossing Bord. Int. J. Interdiscip. Stud.* **2013**, *1*, 29–38. [CrossRef]
15. Wang, H.c.; Chen, C.W.y. Learning English from YouTubers: English L2 learners' self-regulated language learning on YouTube. *Innov. Lang. Learn. Teach.* **2020**, *14*, 333–346. [CrossRef]
16. Dizon, G. YouTube for Second Language Learning: What Does the Research Tell Us? *Aust. J. Appl. Linguist.* **2022**, *5*, 19–26.
17. Alobaid, A. Smart multimedia learning of ICT: Role and impact on language learners' writing fluency—YouTube online English learning resources as an example. *Smart Learn. Environ.* **2020**, *7*, 1–30. [CrossRef]
18. Wahyuni, A.; Utami, A.R.; Education, E. the Use of Youtube Video in Encouraging Speaking Skill. *Pustakailmu. Id* **2021**, *7*, 1–9.
19. Nasution, A.K.R. YouTube as a media in English language teaching (ELT) context: Teaching procedure text. *Utamax J. Ultim. Res. Trends Educ.* **2019**, *1*, 29–33. [CrossRef]
20. Putri, F.H. Youtube for self-regulated language learning: An EFL perspective. *Engl. Educ. J. Tadris Bhs. Ingg.* **2019**, *12*, 42–57.
21. Rahayu, S.P.; Putri, W.S. Uploading speaking assignment to YouTube channel as an effort in increasing student's pronunciation skill. *EnJourMe (Engl. J. Merdeka) Cult. Lang. Teach. Engl.* **2018**, *3*, 35–45. [CrossRef]
22. Aprianto, D. To what extent does youtube contents-based language learning promote An English proficiency? *J. Engl. Lang. Teach. Lit.* **2020**, *3*, 108–126. [CrossRef]
23. Khoiroh, S.A. Using YouTube for speaking in online learning: EFL students' perception and difficulties. *Res. Engl. Lang. Teach. Indones* **2021**, *9*, 202–211.
24. Novawan, A.; Alvarez-Tosalem, S.M.; Ismailia, T.; Wicaksono, J.A.; Setiarini, R.B. Students' Experiences of Online English Language Learning by Using YouTube. In Proceedings of the The First International Conference on Social Science, Humanity, and Public Health (ICOSHIP 2020), Online, 7–8 November 2020; Atlantis Press: Dordrecht, The Netherlands, 2021; pp. 220–226.
25. Al-Jarf, R. YouTube videos as a resource for self-regulated pronunciation practice in EFL distance learning environments. *J. Engl. Lang. Teach. Appl. Linguist.* **2022**, *4*, 44–52. [CrossRef]
26. Simanjuntak, U.S.; Silalahi, D.E.; Sihombing, P.S.; Purba, L. Students'perceptions of using youtube as English online learning media during COVID-19 pandemic. *J. Lang. Lang. Teach.* **2021**, *9*, 150–159. [CrossRef]
27. Learn English with EnglishClass101.com—youtube.com. Available online: https://www.youtube.com/c/EnglishClass101 (accessed on 3 November 2022).

28. Learn English with Jessica—youtube.com. Available online: https://www.youtube.com/c/LearnEnglishwithJessica (accessed on 3 November 2022).
29. Speak English with Vanessa—youtube.com. Available online: https://www.youtube.com/c/TeacherVanessa (accessed on 3 November 2022).
30. Landauer, T.K. LSA as a theory of meaning. In *Handbook of Latent Semantic Analysis*; Psychology Press: London, UK, 2007; pp. 15–46.
31. Martinez, A.M.; Kak, A.C. Pca versus lda. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 228–233. [CrossRef]
32. Wall, M.E.; Rechtsteiner, A.; Rocha, L.M. Singular value decomposition and principal component analysis. In *A Practical Approach to Microarray Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 91–109.