NoCap Captions: Providing Subtitle for Content Creators

Husna Shafirah Binti Helmirizal, Zuraidah Binti Mohd Sulaiman & Afifah Binti Fadhlullah

Universiti Teknologi MARA, Malaysia
Email: 2022964043@isiswa.uitm.edu.my

Abstract

The growing global interest in personalised media content via streaming platforms have contributed to the rising number of content creators on platforms such as YouTube all around the world. Along with that, subtitles have been the medium that bridge the users of various demographic background with the media content that are in languages that are considered foreign to them. In Malaysia however, the majority of the content creators did not include accurate English subtitles in their content due to their incapability to navigate both English and Bahasa Melayu properly, which result in low-quality subtitle inclusion (from auto-generated YouTube CC). Due to that, we believe that locally based YouTube channels struggle to make their viewers grow locally and globally. A case study in 2013 on Digital Discovery Channel on YouTube on the other hand has proven the benefit of YouTube closed captioning where the channel recorded an increase in viewer engagement by 7.32% after the use of closed captions (3Play Media, 2021). To further test this finding, in this study, two videos from a local-based YouTube channel are chosen as the subject of observation. The researchers translated the contents and produced English closed captions that are suitable for the YouTube audience for both videos. Then, the view counts for 20 days before and 20 days after English close captions were added were recorded and compared. The primary data was gained from the YouTube Analytic system and the result reflected an increase in view counts after English closed captioning at 28% and 24% respectively. The findings indicate that the insertion of the English closed captions does influence the view count on a YouTube video and overall reflects the significance of proper YouTube subtitle translation and closed captioning in the industry of content creation.

Keywords: Close Caption (CC), Subtitle Translation, Viewer Engagement, Content Creation

Introduction

YouTube has remained as the most popular global video sharing platforms since its launching in 2005. Having users from diverse demographic, interlingual subtitle or closed captions (CC) has become the tool to bridge the language barrier within the platform and enable contents of various language accessible to the global audience. As of 2022, YouTube closed captions comprises of two types of closed captions: autogenerated CC and uploader's CC. The autogenerated CC can generate closed caption from the audio of the video according to the chosen language but it can only work as the intralingual subtitle. Besides, the autogenerated

CC have low rate of language precision when generating CC from foreign languages audio such as Bahasa Melayu. Therefore, to produce the accurate interlingual subtitle, Malaysian content creators must upload their own CC to their videos. This limitation is what encourages me to plan the NoCap Captions project.

The name NoCap Captions originates from the slang "no cap" which means "no lie" and is also a pun for the phrase "no caption". This project centres around the subject of Translation. For this project, the researcher has approached several YouTube channels which content is mainly in Malay language to provide unsolicited interlingual subtitle service (Bahasa Melayu to English) through the YouTube's closed caption feature. The scope of the video content that is chosen for this translation project are tutorial videos with hobby-based contents which are close to the heart of the general audience. The significant behind choosing tutorial videos for this translation project is the fact that tutorial video have the main objective to explain and guide the audience on a specific process or procedure. Thus, it is very important for audience to understand the language used in the video. They also consist of only one speaker in each video, which is suitable for my level of skill in subtitling. Although the initial focus was to have more than one channels to expose the researchers to the translation of content in various language registers, only one channel agreed to be a part of this project. However, this matter is taken positively as by reducing the number of channel and language register involved in this translation project, the researcher can focus more on the quality of subtitle in the specific register.

The channel which collaborated with the researcher for this project is called the Kebun Bandar channel and it is fully operated by Mr Mohamad Izham Bin Mohd Alias who is also an Agro-technology lecturer in a polytechnic college in Kelantan. The channel focuses on urban gardening tips contents in the Southeast Asian climate. Majority of the audiences from the channels are Malaysians but there are also international audiences who are shows interest on the videos. However, the most common comments left by international audiences are that the video would be better if there are English subtitle inserted. Thus, the two videos that have been chosen in this project are videos with demands of English Subtitles and a video with comments in English which complains on the poor presentation of information in the video. This is as I believed, through the insertion of English subtitle on the videos, the international audience that will stumbled upon the video in the future will have better comprehension on the video content and will hopefully leave more positive comments on the video.

Methods

The first method done is the project planning which included literary research on subtitling, choosing the target channel and videos, and communication with the owner of Kebun Bandar Channel. The next method is the production of the subtitles. Firstly, transcription process required the extraction of verbal content in the original language (Malay) to be transformed into written content. Then, the written content was translated according to suitable translation strategies into the target language (English). Timing was done to synchronise the written subtitle with the verbal content in the video. Then, formatting process transformed the text files into SubRip Subtitles (SRT) files. Next, errors in time stamp, format, and content were screened and edited. After that the language aspects were proofread. All production processes were completed via Notepad++ software. The third method is the subtitle

publication where the channel owner uploaded the subtitles into the videos through the YouTube Studio (closed caption feature) and the subtitles were set for public view. Next, the subtitle undergoes a quality evaluation process by a subtitle translator and a subtitle editor. The fifth method, the data collection, is where the view counts and traffic source from 17 December 2021 to 25 January 2022 were collected from the channel owner via YouTube Analytics. The primary data received were then analysed and compared. The difference between view counts for 20 days before and 20 days after subtitling was calculated and the view count increase percentage was generated. Finally, all data obtained are represented in tables and graphs for comparison and the conclusion was drawn.

Methodology

This study aims to answer the question of whether closed- captioning is significant for the viewer engagement of YouTube videos. The data obtained from this study is a qualitative data which is obtained through observational method (forms) and derivational data collection (YouTube analytics). Firstly, the researcher has produced two subtitles for two chosen video which is already on YouTube. Then, the subtitles were uploaded and published through the Closed-caption function of the video. The view count for both video 20 days before and 20 days after close captioning is then recorded and compared. The increase percentage in the view after subtitling will indicate the impact of the closed-captioning process on the viewer engagement of the YouTube videos.

Data Analysis

For this study, two primary qualitative data are obtained. The first data is the score given by expert evaluators to rate the quality of the subtitles which is collected via google form. The score is produced through likert scale evaluation from the scale of 1 (very poor), to the scale of 5 (very good). This data highlights the technical and the translation quality of the subtitle along with the competitiveness of the subtitle. Next the second primary data is obtained through the owner of the channel. The view counts from 17 December 2021 to 25 January 2022 is obtained via the YouTube analytics system in the form of google sheet. To ensure the quality of the primary data, the traffic source of the view is also tracked, and majority of the views originated from the YouTube suggestion while the remaining percentage originated from external link and direct search result. To compare the view counts before and after the captioning process, the data is presented in a line graph (x axis: day 1 until day 20 before and after subtitling, y axis: the view counts recorded on the video). To clearly compare the performance of the video the difference of total views for each video before and after subtitling is calculated (Total view count 20 days after subtitling - Total view count 20 days before subtitling). Next the percentage of increase in view is calculated (Difference of views before and after subtitling/ Total view counts 20 days before subtitling x 100%).

Results

Quality Evaluation for Subtitle (Likert Scale 1 - 5)								
Qualities	Technical Quality			Translation Quality			Overall Attributes	
	Readability	Speed	Synchronisation	Adaptation	Tone	Consistency	Competitiveness	Suitability to Target Audience
Evaluator A	3/5	3/5	3/5	3/5	4/5	4/5	3/5	3/5
Evaluator B	3/5	3/5	3/5	3/5	4/5	4/5	3/5	4/5

Table 1.0

Reference:
1 – Very Poor
2 – Poor
3 – Average
4 – Good
5 – Very Good

For the evaluation process, I managed to contact two expert evaluators; Ms Siti Zaharah Abdul Rahim who is a freelance subtitle translator and Ms Hidayah Ibrahim who is a subtitle editor under the Iyuno-SDI Group which is a leading translation service provider in Malaysia. Both evaluation forms were sent via Email and Linkedin and the full response can be seen in the Appendix. The result of the evaluation is presented in Table 1.0. Overall, both evaluators rated the technical quality of the subtitle 3/5 for readability, delivery speed, and synchronisation. As for the translation quality, both rated the language adaptation at 3/5 and 4/5 for language tone, and language consistency. When asked regarding the suitability of the subtitles to the general audience on YouTube, Ms Hidayah chose 4/5 (agree) while Ms Siti Zaharah chose 3/5. Finally, both evaluators chose 3/5 on the question whether they agree that the subtitle is competitive enough to be in the subtitle production industry.

View counts for 40 days (17 December 2021 - 25 January 2022) Video 1: How to Grow Hydroponic Lettuce																					
View/ Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Σ
Before Subtitling	31	36	30	21	23	17	33	32	42	40	33	33	22	27	24	30	24	23	34	23	578
After Subtitling	44	31	28	20	35	44	27	25	32	44	43	58	45	36	42	37	30	32	39	47	739
Difference	(Σ After subtitling - Σ Before subtitling)																				161
View Increase	(Difference/ Σ Before subtitling x 100%)																				28%
View counts for 40 days (17 December 2021 - 25 January 2022) Video 2: Making Lactic Acid Bacillus (LAB) Using Milk and Rice Wash																					
View/ Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Σ
Before Subtitling	29	27	27	30	26	17	34	16	15	14	30	21	16	18	19	23	24	21	24	21	452
After Subtitling	22	20	27	29	31	25	27	23	28	30	22	26	26	37	34	24	37	31	39	23	561
Difference	(Σ After subtitling - Σ Before subtitling)																				109
View Increase	(Difference/ Σ Before subtitling x 100%)																				24%

Table 2.0

How to Grow Hydroponic Lettuces
Comparison of View Counts:
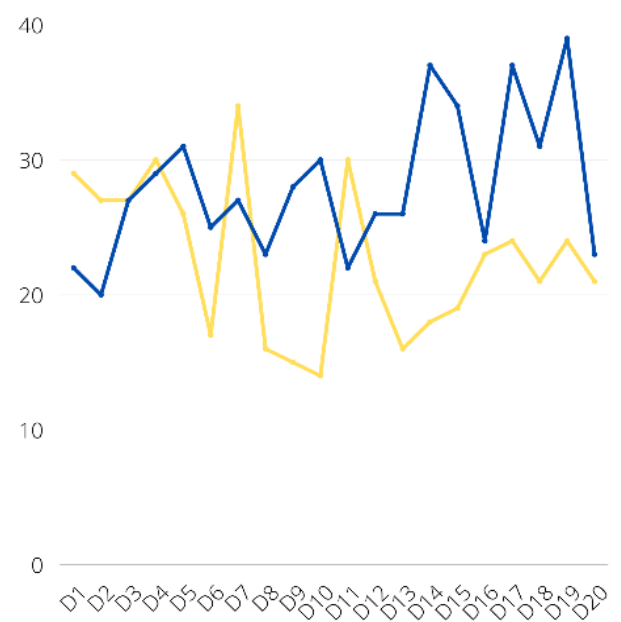20 Days Before Subtitling and 20 Days After Subtitling
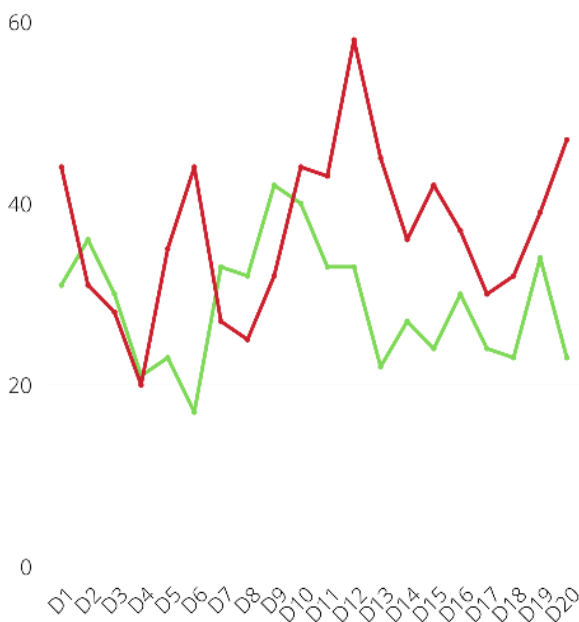
Legend:
V1 Before Subtitling
V1 After Subtitling
Source:
YouTube Analytics Data. Kebun Bandar Channel,
17 December 2021 - 25 January 2022

Figure 1.0

Making Lactic Acid Bacillus (LAB) Using Milk and Rice Wash.
Comparison of View Counts:
20 Days Before Subtitling and 20 Days After Subtitling

Legend:
V2 Before Subtitling
V2 After Subtitling
Source:
YouTube Analytics Data. Kebun Bandar Channel,
17 December 2021 - 25 January 2022

Figure 2.0

As agreed per the initial discussion with the channel owner, after 20 days of publication of subtitle in the two videos, Mr Izham had shared the data from YouTube Analytics on both video with me (Figure 1.0 & 2.0). To see the impact of the subtitle insertion on the audience engagement, I compared the view counts twenty days before subtitling was done and twenty days after subtitling on both videos. For the first video, "How to Grow Hydroponic Lettuces" (refer to Figure 1.0), the view count spiked up to 44 views right on the first day of subtitle insertion (6 January 2022) compared to 31 views the day before (5 January 2022). On day two, three and four, the view count after subtitling gradually decreased before increasing again to 44 views on day six after subtitling. The pattern of decrease and increase in view after subtitling repeated in a pattern in which the view will decrease for four days and increase on the fifth repeatedly. However, starting from day ten until day twenty (after subtitling), the view counts remained higher than the view counts before subtitling despite the ups and downs. The highest view counts recorded after subtitling is 58 views which is on day 12 of subtitling. For the second video (refer to Figure 2.0), unlike the first video, the view on the first day of subtitle insertion decreased to 22 views from 29 views on 5 January. The view count after subtitling only exceeded the view count before subtitling on the fifth day after subtitling and from D6 to D9 of subtitling. However, starting from day twelve until day twenty after subtitling the view counts after subtitling were significantly higher than the view counts before subtitling with its highest view count at 39 views on 24 January 2022. The summary of the finding is recorded in table 2.0. Altogether, the difference between the total view counts of twenty days before and after subtitling for the first video is 161 views which is while for second video, the difference is 109 views. The increase of views after subtitling is recorded at 28% and 24% respectively.

Discussions
Firstly, the completion of this project highlights the possibility of gaining subtitle writing skills via self-learning through the successful production of the published subtitles. However, through the quality evaluation it is prominent that the subtitles produced through this basic subtitle writing skills are overall at an average level especially in term of technical quality and there a lot of room for improvement. Next, the findings from the view count of the two videos after subtitling have proven the significance of closed captioning in increasing the viewer's engagement for YouTube videos. Despite the increase and decrease in pattern of viewing, the study managed to record a clear difference in term of view count where the number of views after closed captioning is higher than the number of views before closed captioning.

Conclusion
Wrapping up this project, I have gained precious knowledge in regards of the importance of applied linguistics practices like especially translation towards the industry of media and digital content. To improve better engagement with the audience online, content creator not only have to consider on the videography of their content, but also on the quality of language usage in their content. This project highlights the value that a language expertise can bring to the world of content creation. In this sense, this project showed how the insertion of subtitle into the video not only help clarify the information delivery in the video, but also assist the algorithm of YouTube to identify the content of video to reach wider audience. High quality subtitle also helps the wider audiences to learn various knowledge across culture and country more efficiently. This is as a subtitle with good quality of translation, grammar, sentence

structure, is proven to be more engaging compared to inaccurate auto-generated closed captions on YouTube.

Although the technology for auto-generated closed captions by YouTube will continue to improve, the nature of the language itself which are spoken by many different individuals with different unique nuances will always be the challenge that is hard to be tackled by technology. Even if there will be a day where artificial technology can help translate audio directly into a chosen target language, the human touch from a linguist will always be the one assisting it. Thus, I believe more translation expert should consider venturing into digital platforms such as YouTube to provide their expertise in improving the environment of digital content. More language experts in the field of subtitle translation must be produced to assist the development of digital media which brings many benefits to the society.

Overall, I managed to fulfill my project 90% objectives in which I have develop subtitle writing skills which is suitable for YouTube closed captions and produced subtitles for two video which have fair potential to compete in the market of subtitling service. I am also glad that my subtitle has an impact in increasing the audience engagement of the two YouTube videos involved. Through this project I have applied all the language skills which I learnt throughout my five semesters in English Professional in Communication. My attempt to venture into the field of subtitling and content translation has given me many insights in terms of the way to communicate professionally with professional connection in workplace setting, applying and adapting with new set of knowledge and information, data management, entrepreneurial opportunities valuation, and discipline in managing individual project. Through this project, I gained new experience and inspiration of my future prospect in career and education pathway for my self-development.

Limitations and Future Studies
Firstly, the biggest limitation for this project would be the communication challenges which is faced when working with a third-party channel. As the YouTube channel owner is not a full-time content creator, he had to prioritise his work commitments, aside from the project. Thus, there are times where the schedule of project had to be delayed and altered, resulting the inability to perform the subtitling according to the initial planned quantity. Next, since the researcher does not have full control on the process of subtitle insertion through YouTube Studio, I could not fix the synchronisation of the subtitle to video. This issue is reflected through the expert evaluation rating especially on the technical quality of the subtitles. Therefore, it is better to observe a YouTube video which have close rapport with the researcher to enable more flexibility of access in YouTube analytics and YouTube Studios features. For future studies, firstly, in terms of subtitle writing skills, constant learning progress is required to produce a higher quality subtitle. To produce a subtitle with high-quality technical properties, the technology like automation function should be utilized fully to assist the process of subtitle synchronisation. This included the use of software such as Aegisub Advanced Subtitle Editor which have automation function and enable easier monitoring of subtitle quality. Next, for translation aspect, as suggested by both evaluators I should improve in producing a more concise translation especially when translating for subtitle as there are limited screen space for better readability. In term of language usage, Evaluator A pointed out some mistranslation which I made in the video including the translation of filler words which I omitted. I also must improve on my punctuation and

sentence structure of subtitle by revising the subtitle individually without comparing it to the original audio or transcription. That way, I can reduce my tendency to translate the subtitle too literally which caused structure confusion when translating.

References
3Play Media. (2021). *Case Studies for Closed Captioning Company*. Retrieved October 12, 2021, from https://www.3playmedia.com/why-3play/case-studies/

Cambridge Dictionary | English Dictionary, Translations & Thesaurus. (2022, February 2). Cambridge Dictionary. Retrieved 2021, from https://dictionary.cambridge.org/

Dewan Bahasa dan Pustaka. (2022). *Pusat Rujukan Persuratan Melayu @PRPM*. Pusat Rujukan Persuratan Melayu. Retrieved 2021, from https://prpm.dbp.gov.my/

Jean-Marc, J. (2021). *How to Do Good Subtitles: the Golden Rules Guidelines*. Checksub. Retrieved October 13, 2021, from https://www.checksub.com/subtitle/do-good-subtitles-golden-rules/

Kebun Bandar. (2020). *Cara Membuat Lactobacillus (EM-1) Sendiri Guna Air Basuhan Beras dan Susu / Homemade Yakult Vitagen* [Video]. YouTube. https://www.youtube.com/watch?v=hCL3562LcB0&list=PL8mTN5wRtZl82yFZBdQDrhs8NHM9SlwP0&index=3

Kebun Bandar. (2019). *Ini Cara Saya Tanam Salad Hidroponik Hanya Guna Bekas RM2* [Video]. YouTube. https://www.youtube.com/watch?v=Tq2AdbemJck&list=PL8mTN5wRtZl82yFZBdQDrhs8NHM9SlwP0&index=7

KomalSrivastava@TWC. (2021). *How to create SRT subtitles in Windows 11/10*. The Windows Club. Retrieved December 11, 2021, from https://www.thewindowsclub.com/how-to-create-srt-subtitles-in-windows

Laurence, H. (2021). *Convert Subtitle: Excel formatting to SRT formatting (Excel to SRT)* [Video]. YouTube. https://www.youtube.com/watch?v=rhZdv80tV6s&list=PL8mTN5wRtZl82yFZBdQDrhs8NHM9SlwP0&index=1

Microsoft. (2022). *How to use SRT files for displaying subtitles during video playback*. Retrieved December 11, 2021, from https://support.microsoft.com/en-us/topic/how-to-use-srt-files-for-displaying-subtitles-during-video-playback-ea2ae7f6-e49d-bd06-446a-77d2daa00615