Evaluation of Abstractive Summarisation Models with Machine Translation in Deliberative Processes

Anonymous EMNLP submission

Abstract

In this project, we conduct an extensive evaluation of a wide range of abstractive summarisation models in combination with an off-the-shelf machine translation model. The evaluation includes different model architectures and is performed in a deliberation dataset. Unlike commonly studied datasets, such as news articles, the dataset evaluated here presents the difficulty of combining multiple narratives in a single text, mostly of poor grammatical quality. We obtain promising results regarding the fluency, consistency, and relevance of the summaries produced by using a system that is easy to implement for production purposes.

1 Introduction

The processes of deliberation and collective intelligence production have evolved radically thanks to the possibility of carrying them out digitally. However, this often results in large amounts of generated content in the deliberations, causing information overload that prevents their potential from being fully realised (Arana-Catania et al., 2021; Davies and Procter, 2020). To address this, we evaluate the potential value of abstractive summarisation models when combined together with a machine translation system in synthesising and filtering information collected through such processes. Whereas the current technology of language models is mostly limited to a few languages, preventing the majority of the population from using them, our approach can be deployed for many languages just changing the translation model without the need to generate new, ad-hoc corpora for the task or costly retraining for each language. The current evaluation is done in a Spanish deliberation dataset.

We have carried out an evaluation with 6 abstractive summarisation models: BART (Lewis et al., 2019), T5 (Raffel et al., 2019), BERT (PreSumm – BertSumExtAbs: Liu and Lapata, 2019), PG (Pointer-Generator with Coverage Penalty) (See et al., 2017), CopyTransformer (Gehrmann et al., 2018), and FastAbsRL (Chen and Bansal, 2018). Those models are applied in combination with the machine translation system MarianMT (Junczys-Dowmunt et al., 2018) using the Opus-MT models (Tiedemann and Thottingal, 2020). We have evaluated the quality of the summaries for each model and their comparison.

Early research on the problem of text summarisation in low resourced languages (although not focused on deliberation) (Orăsan and Chiorean, 2008) demonstrated the limitation of machine translation systems at the time. Recently, Ouyang et al. (2019) revisited the problem of low quality translations in low resourced languages and successfully demonstrated the possibility of using abstractive summarisation by retraining their model on corpora that have gone through the same machine translation process. In this study, we complete the cycle, translating from the original language to English, summarising, and translating back to the original language, thus avoiding the retraining.

Using other approaches, Yao et al. (2015) conducted a study of English-to-Chinese summarisation in which they combined an extractive approach with a process of sentence compression that effectively abstracts the results. Duan et al. (2019), following on from the work of Shen et al. (2018), exploited the capability of a resource-rich language summariser in a teacher-student framework that connects it to the target language summariser.

2 Dataset

The evaluation has been carried out with a dataset of deliberative processes in Spanish, translated into English to carry out the summarisation and the generated summaries have been translated back into Spanish for their evaluation. Thus, the evaluators have evaluated Spanish summaries of Spanish texts, without knowing the intermediate English versions.

The dataset is available in the Madrid City Council 'Datos Abiertos' repository[1], called 'Comentarios'. It contains public deliberations in relation to citizen proposals made in the participation platform of the city council. The dataset has been selected due to the great success of the participation platform, which has led to $26,400$ proposals and $125,135$ comments being submitted. This is one of the most successful cases of digital participation in the world and is therefore a perfect case study for evaluating the information overload problem in deliberation (Arana-Catania et al., 2021).

Each of the citizen proposals presents a debate space where the public comments can be found. 40 debates were selected covering the different deliberation scenarios of the dataset. These represent three main cases: 20 debates with ($n = 10$) comments, representing the most common situation of debates with few comments; 15 debates with ($20 \leq n \leq 30$) comments, for the medium case; and 5 debates with ($60 \leq n \leq 70$) comments, for the case with a large number of comments.

The debates in each category were also selected to cover the different comment scenarios, from very short to very lengthy comments. In the first case from $1,000$ to $5,000$ total characters; in the medium case from $3,000$ to $13,000$; and in the large case from $10,000$ to $18,000$ characters. The text to summarise for each debate was created concatenating its comments in a single document.

By using debates from all scenarios regarding number of comments and comments length we ensure that the selection is not biased to a specific case of deliberation that could skew our results.

Examples of the debates can be found in the Appendix, where it can be seen the combination of multiple narratives through the different comments and the poor grammatical quality of the texts.

3 Abstractive Summarisation Methodology

Different models have been selected, covering some of the best available summarisers, but also exploring different model architectures:

- BART (Lewis et al., 2019)[2]. This combines a bidirectional transformer as an encoder, similar to the following T5 and BERT cases, with a left-to-right autoregressive decoder similar as GPT (Radford et al., 2018). The 'large-cnn' pre-trained model has been used here.

- T5 (Raffel et al., 2019)[2]. This uses an encoder-decoder transformer architecture, trained in the Colossal Clean Crawled Corpus. The 'small' pre-trained model has been used.

- BERT (PreSumm – BertSumExtAbs: Liu and Lapata, 2019)[3]. This uses BERT (Devlin et al., 2018) as the encoder and a randomly initialized transformer as a decoder, fine-tuning first the encoder as an extractive summariser and then as an abstractive one. The BertSumExtAbs pre-trained model has been used.

- PG (Pointer-Generator with Coverage Penalty) (See et al., 2017)[4]. This uses a 1-layer bidirectional LSTM encoder and a 1-layer unidirectional LSTM decoder with attention, with the possibility of switching between copying words or generating them (Pointer-Generator) and including a coverage mechanism adding up attention distributions of previous steps to minimise repetitions.

- CopyTransformer (Gehrmann et al., 2018)[5]. This uses the transformer architecture, but one attention head defines the copy distribution.

- FastAbsRL (Chen and Bansal, 2018)[6]. An extractor agent is used to select sentences (using LSTM layers to represent and copy sentences) and an abstractor network is used to compress and paraphrase the selected sentences. Both are trained separately and then the full model is trained with reinforcement learning by using A2C (Mnih et al., 2016).

The Rouge scores of these models (Lin, 2004) reported by the authors are shown in Table 1.

In an initial exploration, additional models were evaluated: Adversarial Reinforce GAN (Wang and Lee, 2018), using Generative Adversarial Networks; Contextual Matching (Zhou and Rush, 2019), using ELMo (Peters et al., 2018) in combination with a domain fluency model using LSTM

[1] https://datos.madrid.es

[2] Implementation by HuggingFace https://github.com/huggingface/transformers

[3] Implementation by the authors https://github.com/nlpyang/PreSumm

[4] Implementation by OpenNMT https://opennmt.net/OpenNMT-py/Summarization.html

[5] OpenNMT implementation thanks to https://github.com/sebastianGehrmann/bottom-up-summary

[6] Implementation by the authors https://github.com/ChenRocks/fast_abs_rl

Model	R1	R2	RL
BART - large-cnn	44.16	21.28	40.90
T5 - small	41.12	19.56	38.35
BERT - BertSumExtAbs	42.13	19.60	39.18
PG - OpenNMT – BRNN	39.12	17.35	36.12
CopyT - OpenNMT	39.25	17.54	36.45
FastAbsRL	40.88	17.80	38.54

Table 1: Rouge scores reported on the CNN/DailyMail dataset (Hermann et al., 2015).

layers; PoDA (Wang et al., 2019), using a transformer encoder-decoder structure combined with a pointer-generator attention layer, and training it as a denoising autoencoder; and GenParse (Song et al., 2018), combining sequential word generation with tree-based parsing. However, after an initial qualitative evaluation, we found that none of these models produced sufficiently competitive results in comparison with the selected models. We should note that several of these models work at the sentence level, which may impact their relevance in our deliberative case, where the paragraphs are composed of multiple authors comments.

The machine translation system used was MarianMT (Junczys-Dowmunt et al., 2018) using its HuggingFace implementation, with Opus-MT models[7] developed by the Helsinki-NLP group.

Machine translation was applied to the original text of the deliberations before applying the summarisers, and a second time to the summary generated to convert back to the original language (see Appendix). Thus, even when the summarisation models are trained with English datasets, the full system can be used in deliberations of any language supported by the machine translation system. The Opus-MT models used in this work count currently with pre-trained models for 1738 language pairs. It is left for future work to evaluate the effect of the translation model, and to apply it to other languages to determine the quality obtained in other cases. The models used here show a good performance (see BLEU scores in OpusMTen; OpusMTes) for the used languages.

4 Evaluation Design

We developed a protocol for the human evaluation of the summaries generated by the different models, following designs of previous studies (Amplayo and Lapata, 2020; Liu and Lapata, 2019; Narayan et al., 2018; Paulus et al., 2017; Yoon et al., 2020; Song et al., 2018). First, the different models were compared regarding their relative overall quality using the Best-Worst scaling (Louviere et al., 2015), shown to be more accurate than a generic individual scoring model, and simultaneously reducing the number of assessments required (Kiritchenko and Mohammad, 2017).

For each debate, 6 different summaries were generated, one for each of the models to be evaluated. These summaries were organised in 9 tuples of 4 elements each, where each summary appeared in 6 of the tuples in random order and without the possibility of the evaluator of identifying which model was used in each case. In total, considering all the debates, 360 tuples were produced. Each of these tuples was evaluated by 5 independent evaluators (native Spanish speakers with a minimum education level of a Bachelor's degree), producing a total of $1,800$ evaluations. The score for each summary consisted of the percentage of times it was evaluated as Best, minus the percentage of times it was evaluated as Worst.

In addition, a second evaluation was carried out for two summaries in each debate. The models were selected randomly in each case while ensuring that each model had the same number of evaluations. In this case we were interested in whether the models produce results of sufficient quality to be useful to the participants in the debate. Thus, we have not used a relative score but an absolute score, asking the evaluators to rate the following aspects (the definitions below were shared with the evaluators) on a Likert scale from 1 (Strongly disagree) to 4 (Strongly agree):

- Informativeness/Relevance. The summary contains the most relevant ideas and positions of the debate.

- Fluency/Readability/Grammaticality. The summary sentences are grammatically correct, easy to read and understand (considering as a baseline the fluency of the original debate).

- Consistency/Faithfulness. The ideas or facts contained in the summary appear in the original debate.

- Creativity. The summary has been written with its own words and sentences (instead of copying sentences directly from the debate).

[7] https://github.com/Helsinki-NLP/Opus-MT

Model	comp	σ	comp$_{[0,100]}$	$\sigma_{[0,100]}$
BART	33.08	11	66.54	5
BERT	23.33	10	61.67	5
PG	6.25	13	53.13	6
T5	-16.08	14	41.96	7
CopyT	-16.42	5	41.79	2
FastAbsRL	-30.17	10	34.92	5

Table 2: Comparison scores using the Best-Worst scaling (and thus in the range $[-100, 100]$) with its standard deviation, and normalised to the $[0, 100]$ range.

Model	Informative	σ	Fluent	σ
BART	2.58	0.8	2.85	0.8
BERT	2.53	0.8	2.65	0.9
PG	2.33	0.7	2.28	0.8
T5	2.50	0.8	2.30	0.8
CopyT	2.14	0.6	2.02	0.8
FastAbsRL	2.02	0.7	1.73	0.6
	Consistent	σ	**Creative**	σ
BART	2.88	0.8	2.08	0.7
BERT	2.72	0.9	1.98	0.6
PG	2.67	0.8	2.02	0.6
T5	2.63	0.9	1.97	0.6
CopyT	2.46	0.9	1.81	0.6
FastAbsRL	2.13	0.7	1.82	0.7

Table 3: Rating and standard deviation for each model in a scale from 1 (Strongly disagree) to 4 (Strongly agree).

5 Evaluation Results

The results obtained for the overall comparison between models are shown in Table 2, which reports the average scores of all the evaluators.

Paired Student's t-tests were performed between all pairs of models to confirm that the difference was statistically significant. This is not the case for the BERT and BART models ($p = 0.09$), showing very close results. There is also a clear overlap between T5 and CopyTransformer. All the other combination pairs are found within a difference statistically significant ($p < 0.05$).

These results are in line with the previous evaluation results on English datasets that BART and BERT are the top two summarisers (Lewis et al., 2019; Liu and Lapata, 2019). However, in the present evaluation a state-of-the-art model (T5) falls below a much older model (PG).

The results for the evaluation of the qualitative aspects of each summariser are shown in Table 3. It is important to note that in this case the standard deviation is larger compared to the first case, which is due to the smaller number of evaluations, and thus the following comments should be understood considering their statistical significance.

In this individual evaluation of each model, it can be seen again how BART obtains the best ratings in all four categories evaluated. BERT is the second best for the categories of '*Informativeness*', '*Fluency*' and '*Consistency*', while PG jumps to the second position for '*Creativity*'. T5 is in the third position for the categories '*Informativeness*' and '*Fluency*' and PG is the third best for '*Consistency*'.

This confirms the best results of BART and BERT, and a close result for T5 for generating informative summaries, but a poorer result for fluency. This may be the reason why the T5 model performed worse in the general overall comparison.

BART and BERT perform well in terms of '*consistency*', with scores close to 3. They perform a bit worse for '*fluency*' and '*informativeness*', around the middle of the possible rating 2.5. Regarding '*creativity*', the models have a poor performance, with a score of around 2, meaning that they tend to copy instead of paraphrase.

6 Conclusions

In this study we have evaluated the application of state-of-the-art abstractive summarisation models to deliberative processes in Spanish using an off-the-shelf machine translation model. Although we focused on a Spanish evaluation, our proposed pipeline can be easily deployed without additional complication to multiple other languages, offering an important value for production applications (especially cases dealing with wide ranges of languages) rarely present in other approaches. However, the evaluation of the quality for other languages is left for future work. We have done a comparative evaluation of the overall quality of the models, and an evaluation of each model with respect to different qualitative aspects: *informativeness, fluency, consistency*, and *creativity*.

As a general conclusion, from the models evaluated BART and BERT have the best results, and satisfactory results are obtained in the proposed pipeline for the summaries quality. With regard to the most important aspects, the models show a good result in the categories of *fluency* and *consistency*, and an average result regarding the *informativeness*. These results are especially promising considering the complexity and low grammatical fluency and consistency involved in deliberation. BART and BERT are the only models that score over the middle score in the three categories, and thus from our point of view are satisfactory enough to be used.

References

Reinald Kim Amplayo and Mirella Lapata. 2020. Unsupervised opinion summarization with noising and denoising. *arXiv preprint arXiv:2004.10150*.

Miguel Arana-Catania, FA Van Lier, Rob Procter, Nataliya Tkachenko, Yulan He, Arkaitz Zubiaga, and Maria Liakata. 2021. Citizen participation and machine learning for a better democracy. *arXiv preprint arXiv:2103.00508*.

Yen-Chun Chen and Mohit Bansal. 2018. Fast abstractive summarization with reinforce-selected sentence rewriting. *arXiv preprint arXiv:1805.11080*.

Jonathan Davies and Rob Procter. 2020. Online platforms of public participation: a deliberative democracy or a delusion? In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*, pages 746–753.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Xiangyu Duan, Mingming Yin, Min Zhang, Boxing Chen, and Weihua Luo. 2019. Zero-shot cross-lingual abstractive sentence summarization through teaching generation and attention. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3162–3172.

Sebastian Gehrmann, Yuntian Deng, and Alexander M Rush. 2018. Bottom-up abstractive summarization. *arXiv preprint arXiv:1808.10792*.

Karl Moritz Hermann, Tomáš Kočiskỳ, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. *arXiv preprint arXiv:1506.03340*.

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, et al. 2018. Marian: Fast neural machine translation in c++. *arXiv preprint arXiv:1804.00344*.

Svetlana Kiritchenko and Saif M Mohammad. 2017. Best-worst scaling more reliable than rating scales: A case study on sentiment intensity annotation. *arXiv preprint arXiv:1712.01765*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345*.

Jordan J Louviere, Terry N Flynn, and Anthony Alfred John Marley. 2015. *Best-worst scaling: Theory, methods and applications*. Cambridge University Press.

Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937. PMLR.

Shashi Narayan, Shay B Cohen, and Mirella Lapata. 2018. Ranking sentences for extractive summarization with reinforcement learning. *arXiv preprint arXiv:1802.08636*.

OpusMTen. 2020. https://huggingface.co/helsinki-nlp/opus-mt-en-es.

OpusMTes. 2020. https://huggingface.co/helsinki-nlp/opus-mt-es-en.

Constantin Orăsan and Oana Andreea Chiorean. 2008. Evaluation of a cross-lingual romanian-english multi-document summariser. *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008.*

Jessica Ouyang, Boya Song, and Kathleen McKeown. 2019. A robust abstractive system for cross-lingual summarization. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2025–2031.

Romain Paulus, Caiming Xiong, and Richard Socher. 2017. A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304*.

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training (2018).

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*.

Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*.

Shi-qi Shen, Yun Chen, Cheng Yang, Zhi-yuan Liu, Mao-song Sun, et al. 2018. Zero-shot cross-lingual neural headline generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(12):2319–2327.

Kaiqiang Song, Lin Zhao, and Fei Liu. 2018. Structure-infused copy mechanisms for abstractive summarization. *arXiv preprint arXiv:1806.05658*.

Jörg Tiedemann and Santhosh Thottingal. 2020. Opus-mt–building open translation services for the world. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480.

Liang Wang, Wei Zhao, Ruoyu Jia, Sujian Li, and Jingming Liu. 2019. Denoising based sequence-to-sequence pre-training for text generation. *arXiv preprint arXiv:1908.08206*.

Yau-Shian Wang and Hung-Yi Lee. 2018. Learning to encode text as human-readable summaries using generative adversarial networks. *arXiv preprint arXiv:1810.02851*.

Jin-ge Yao, Xiaojun Wan, and Jianguo Xiao. 2015. Phrase-based compressive cross-language summarization. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 118–127.

Wonjin Yoon, Yoon Sun Yeo, Minbyul Jeong, Bong-Jun Yi, and Jaewoo Kang. 2020. Learning by semantic similarity makes abstractive summarization better. *arXiv preprint arXiv:2002.07767*.

Jiawei Zhou and Alexander M Rush. 2019. Simple unsupervised summarization by contextual matching. *arXiv preprint arXiv:1907.13337*.

A Appendix

We present below an example of a debate used in the evaluation in Spanish and its machine translation to English. Following them we present the summaries generated using T5, FastAbsRL, BART, and BERT. Finally, we include the translations of these summaries.

The texts are presented in the same order used in the project. We start with a debate in Spanish, which is translated into English. This translated debate is summarised, and finally the summary is translated back into Spanish. The evaluators analysed only the original debate in Spanish and the final summaries in Spanish.

A.1 Original Spanish debate

- ademas proponemos tranvía.
- el casco no es obligatorio para mayores de 15 años mientras circulan en ciudad. lo dice la dgt.por lo demás, te doy la razón. deben cumplir la normativa de circulación. pero, eh!... los conductores de coches y motos también. hay demasiados que no respetan a los ciclistas... ¿sabias que en ciudad, un ciclista debe ocupar 1 carril de circulación... y no ir por el borde?.
- se deberían sancionar las bicis que van por las aceras o fuera de los carriles bicis.
- si las bicis van por las aceras es porque es muy peligroso ir por los carriles de los coches aunque estén marcados. no existe concienciación todavía por parte de los usuarios conductores. por otro lado, el hecho en sí de ir por la acera no es peligroso, siempre que se vaya "a paso de peatón". lo que no se puede es ir rápido.para mí el verdadero peligro es en las horas nocturnas, en que muchos ciclistas van sin luz alguna y no se ven hasta que estás prácticamente encima de ellos... eso en amsterdam está rigurosamente prohibido y se multa. aquí he visto a la policía municipal pasar de todo al verlos....
- obviamente quien dice eso no ha cogido una bici en su vida, el casco en bici no salva vidas, es un hecho, salva vidas el conductor respetuoso.
- nunca,pero nunca jamás he visto parar un ciclista en un semáforo rojo,o se suben a la acera para cruzar sorteando a los peatones o directamente se lo saltan,en un paso de peatones menos se paran.¿qué pasa,que las norma no son para todos por igual? si un coche se salta un semáforo,la multa es bestial! un poco más de respeto,sobre todo cuando circulan por la acera a la velocidad que les da la gana,con el peligro que conlleva.se creen que todo vale y la calle es suya.
- se puede circular por la calzada, aunque haya carril bici vecin@.
- no me lo creo....nunda digas nunca!.
- ¿no cree que está generalizando demasiado? no todos van con auriculares, no todos se saltan los semáforos, y los coches se tienen que aconstumbrar a la presencia de las bicis....es un medio de transporte más, y se merece respeto.
- la obligación del casco desincentiva el uso d ela bicicleta, que en el caso de mardid está mejorando la movilidad sin aumentar la contaminación

A.2 Machine translated debate

- and we're proposing a tram.
- the helmet is not mandatory for more than 15 years as they travel in the city. says dgt. otherwise, I give you the reason. they must comply with the traffic regulations. but, uh!... the drivers of cars and motorcycles also. there are too many that do not respect cyclists... did you know that in the town, a cyclist must occupy 1 lane of traffic... and not go by the edge?.
- bikes that go along the sidewalks or off the bike lanes should be sanctioned.
- if the bikes go through the sidewalks it is because it is very dangerous to go down the lanes of the cars even if they are marked. there is no awareness yet on the part of the users drivers. On the other hand, the fact itself of going down the sidewalk is not dangerous, as long as it goes "by foot".What you can not do is go fast.For me the real danger is in the night hours, in which many cyclists go without any light and you don't see until you are practically above them... that in Amsterdam is rigorously forbidden and is fined. here I have seen

the municipal police pass everything when you see them....

- obviously whoever says that hasn't taken a bike in his life, the bike helmet doesn't save lives, it's a fact, it saves lives the respectful driver.

- never, but I've never seen a cyclist stop at a red light, or get on the sidewalk to cross by shooting pedestrians or directly jump him, at a pace of pedestrians less stop.What happens, that the rules are not for everyone alike? if a car jumps a light, the ticket is best! a little more respect, especially when they circulate along the sidewalk at the speed that gives them the desire, with the danger that it carries. they believe that everything is good and the street is theirs.

- you can drive along the road, even if there is a nearby bicycle lane.

- I don't believe it... never say it!

- Don't you think it's generalizing too much? Not everyone goes with headphones, not everyone jumps the traffic lights, and cars have to settle for the presence of bikes....it's one more means of transport, and it deserves respect.

- the obligation of the helmet to discourage the use of bicycles, which in the case of Tuesday is improving mobility without increasing pollution

A.3 Generated summaries

- **T5**. the rules are not for everyone alike. not everyone jumps the traffic lights, not everyone goes with headphones, and not everybody jumps traffic lights. a little more respect, especially when they circulate along the street at the speed that gives them the desire, with the danger that it carries. I don't believe it... never say it! Don't you think it's generalizing too much? Don'

- **FastAbsRL**. the helmet is not mandatory for more than 15 years ." the real danger is in which many cyclists go without any light and you don't see until you are practically above them... that in amsterdam is rigorously forbidden . otherwise, i give you the reason. . they must comply with the traffic

- **BART**. Bikes that go along the sidewalks or off the bike lanes should be sanctioned. The real danger is in the night hours, in which many cyclists go without any light. Not everyone goes with headphones, not everyone jumps the traffic lights, and cars have to settle for the presence of bikes. It's one more means of transport, and it deserves respect.

- **BERT**. the helmet is not mandatory for more than 15 years as they travel in the city it is because it is very dangerous to go down the lanes of the cars even if they are marked there is no awareness yet on the part of the users drivers drivers

A.4 Machine translated summaries

- **T5**. las reglas no son para todos por igual. no todos saltan los semáforos, no todos van con auriculares, y no todo el mundo salta semáforos. un poco más de respeto, especialmente cuando circulan por la calle a la velocidad que les da el deseo, con el peligro que conlleva. No lo creo... nunca lo digo! ¿No crees que está generalizando demasiado?

- **FastAbsRL**. el casco no es obligatorio durante más de 15 años. " el verdadero peligro es en el que muchos ciclistas van sin ninguna luz y no se ve hasta que usted está prácticamente por encima de ellos... que en amsterdam está rigurosamente prohibido. Si no, te doy la razón. deben cumplir con el tráfico.

- **BART**. Las bicicletas que van por las aceras o fuera de los carriles bici deben ser sancionadas. El verdadero peligro es en las horas de la noche, en las que muchos ciclistas van sin ninguna luz. No todos van con auriculares, no todos saltan los semáforos, y los coches tienen que conformarse con la presencia de bicicletas. Es un medio de transporte más, y merece respeto.

- **BERT**. el casco no es obligatorio por más de 15 años ya que viajan por la ciudad es porque es muy peligroso ir por los carriles de los coches, incluso si están marcados todavía no hay conciencia por parte de los conductores de los usuarios