Extractive Text-Based Summarization of Arabic videos: Issues, Approaches and Evaluations

M A Menacer, C E González-Gallardo, K Abidi, Dominique Fohr, Denis Jouvet, D Langlois, Odile Mella, F Sadat, J M Torres-Moreno, Kamel Smaïli

Extractive Text-Based Summarization of Arabic videos: Issues, Approaches and Evaluations

M.A. Menacer[1], C.E. González-Gallardo[2], K. Abidi[1], D. Fohr[1] D. Jouvet[1], D. Langlois[1], O. Mella[1], F. Sadat[3], J.M. Torres-Moreno[2,4], and K. Smaïli[1]

[1] Loria, University of Lorraine, France
{mohamed-amine.menacer, karima.abidi, fohr, jouvet, langlois, mella, smaili}@loria.fr
[2] LIA, Avignon Université, France {carlos-emiliano.gonzalez-gallardo, juan-manuel.torres}@univ-avignon.fr
[3] UQAM, Montreal (Quebec), Canada
sadat.fatiha@uqam.ca
[4] Poliyechnique Montréal, (Quebec) Canada

Abstract. In this paper, we present and evaluate a method for extractive text-based summarization of Arabic videos. The algorithm is proposed in the scope of the AMIS project that aims at helping a user to understand videos given in a foreign language (Arabic). For that, the project proposes several strategies to translate and summarize the videos. One of them consists in transcribing the Arabic videos, summarizing the transcriptions, and translating the summary. In this paper we describe the video corpus that was collected from YouTube and present and evaluate the transcription-summarization part of this strategy. Moreover, we present the Automatic Speech Recognition (ASR) system used to transcribe the videos, and show how we adapted this system to the Algerian dialect. Then, we describe how we automatically segment into sentences the sequence of words provided by the ASR system, and how we summarize the obtained sequence of sentences. We evaluate objectively and subjectively our approach. Results show that the ASR system performs well in terms of Word Error Rate on MSA, but needs to be adapted for dealing with Algerian dialect data. The subjective evaluation shows the same behaviour than ASR: transcriptions for videos containing dialectal data were better scored than videos containing only MSA data. However, summaries based on transcriptions are not as well rated, even when transcriptions are better rated. Last, the study shows that features, such as the lengths of transcriptions and summaries, and the subjective score of transcriptions, explain only 31% of the subjective score of summaries.

Keywords: Text Summarization · Video Summarization · Automatic speech recognition · Segmentation.

1 Introduction

Understanding the content of a video in a foreign language could be considered as a dream. However, research in video analysis, automatic speech recognition and machine

translation has evolved significantly and the results today can be considered encouraging. In this article, part of the Chist-Era founded AMIS[5] (Access Multilingual Information opinionS) project, we address the problem of understanding a video in a foreign language.

In the scope of this project, we consider that we understand the content of a video if we can summarize it correctly. Therefore, this project uses several research disciplines related to natural language processing, namely video analysis, automatic speech recognition, segmentation of speech transcriptions and automatic summarization. Moreover, it is essential to evaluate the performance of such a system, either to make it public or to highlight the new research challenges related to this problem. It is indeed very difficult to find an objective measure allowing to assess the whole system, since this one is the result of several technologies and models.

As part of this project we considered that the foreign language is the Arabic language, so we developed a speech recognition system for Arabic that we named ALASR [17] (Arabic Loria Automatic Speech Recognition system). We have also developed a machine translation system that translates the results of the Arabic transcript into English. We worked on real data that we crawled from TV channels broadcasting in Arabic, such as: Euronews, AlArabiya, Skynews, etc. We also collected videos from Algerian channels broadcasting in Arabic, but necessarily using sometimes the Algerian dialect.

When testing ALASR on Algerian channels data, the performance collapsed. This drove us to adapt ALASR to dialectal data, which led to better results. Regarding the global assessment, we conducted a subjective evaluation that allowed us to test not only the result of speech recognition, but also the automatic summarizing system.

The rest of this paper is organized as follows. Section 2 presents our video corpus. The Arabic ASR system is presented in Section 3, and its adaptation on Algerian dialect in Section 4. An automatic sentences segmentation module is shown in Section 5; Section 6 shows the automatic text summarizer employed in this work. Section 7 presents our results, and finally, Section 8 concludes this paper.

2 Video corpus

A project such as AMIS requires to collect videos in order to estimate the parameters of our models and to evaluate our approach. For that, French, English and Arabic videos have been collected. Videos have been selected according to a set of controversial Twitter hash-tags such as `#womenrights` or `#syria` given that one goal of the AMIS project is to compare opinions on videos in different languages that deal with the same topic; more details on the collection process can be found in [11]. The overall video corpus corresponds to more than 300 hours of video, that is about 100 hours in each of the three languages (French, English and Arabic). The video data come from various channels such as Euronews, France24, BBC and AlArabiya.

With respect to the Arabic videos, more than 1,500 videos have been collected. They come from channels such as AlArabiya, France24, SkynewsArabia, Euronews,

[5] http://deustotechlife.deusto.es/amis/

EchoroukTV, EnnaharTV, BBC, etc. The duration of the videos vary from one minute up to more than one hour.

3 Arabic automatic speech recognition

The training of the acoustic models and the recognition experiments were carried out with the ALASR system developed at LORIA laboratory. ALASR is based on the Kaldi toolkit [20]. For the acoustic parameters, 13-dimensional Mel-Frequency Cepstral Coefficients (MFCC) augmented with their first and second order derivatives were computed. 37 acoustic models were trained: 34 phone models, one model for silence, one for respiration and one for noise. A Deep Neural Network (DNN) was used to produce posterior probabilities for the context dependent phone densities of the Hidden Markov Models (DNN-HMM models). The DNN consists of 6 layers with 2,048 hidden neurons each. For the input layer, 11 frames were concatenated, and the output layer has 4,264 output neurons, corresponding to the 4,264 senones (contextual phone densities). A total of 30 millions parameters were estimated using 54 hours of Arabic Broadcast News Speech Corpus. 5 hours of spoken data were used for tuning (Dev) and 5 other hours for evaluating the performance of ALASR system (Test).

Linguistic knowledge is required to capture the properties of the language. For this reason, we trained two 4-gram language models one on the Gigaword corpus and the other on the train transcripts of the acoustic data. Since these two corpora are unbalanced, the two language models were combined linearly by optimizing the weights of the linear interpolation on the transcripts of the acoustic Dev set. Due to memory constraints, we decided to prune the full 4-gram language model by minimizing the relative entropy between the full and the pruned model [22]. This led to a total number of 4M n-grams in the pruned language model compared to 983M n-grams in the full language model. This later model will be used for rescoring the lattice produced by the system.

The pronunciation lexicon makes the link between the language model and the acoustic model. The absence of the short vowels (diacritics) in written texts brings issues in the pronunciation modelling. In fact, for each Arabic grapheme-based form, the ASR system has to consider all the pronunciation possibilities. There are two approaches to deal with this issue:

Grapheme-based approach This approach considers for Arabic that the pronunciation of each word is simply its grapheme decomposition, and therefore, graphemes represent the basic units for the acoustic model. While this approach is the simplest way to build a lexicon, it will not provide an explicit representation of short vowels, which might lead to recognition errors.

Phoneme-based approach Unlike the previous approach where short vowels are implicitly modeled with the surrounding consonants in the acoustic modelling, this approach provides an explicit representation of short vowels in the pronunciation modelling. This approach is adopted in this work.

In order to create the phoneme-based model, we selected the 109k most frequent words from the Gigaword corpus (1 billion word occurrences) plus the words that appear more than 3 times in the transcripts of the acoustic Train set. Afterwards, only

words for which pronunciation variants exist in an external lexicon [2] were kept. This process produces a lexicon having 95k unique grapheme-based words and 485k pronunciation variants, that is an average of 5.07 pronunciations per word. This lexicon is referred in the following as MSA_{lex}. Table 1 illustrates the evaluation of ALASR system on the Test corpus.

Table 1. Performance of ALASR before and after rescoring the lattice (WER: Word Error Rate, OOV: Out-Of-Vocabulary.).

System	**WER (%)**	**OOV (%)**
ALASR	15.32	2.5
ALASR+Rescoring	14.02	

Using a pruned language model accelerates the decoding process but it affects the performance of the system. By rescoring the produced lattice, new hypotheses are generated based on the probabilities of the full 4-gram language model, which leads to an absolute improvement of 1.3%.

4 Adaptation of the Automatic Speech Recognition system to the Algerian dialect

Most of Arab people do not use MSA in their daily conversations, since their mother tongue is an Arabic dialect that is mainly derived from MSA. The Arabic dialect varies from one country to another and sometimes more than one dialect can be found within a country. These variants are mainly influenced by the history of the region itself [15].

The Algerian dialect is one of the Maghrebi dialects spoken in the western Arab countries. It is one of the hardest dialect to be recognized by an ASR system. This is due to the fact that this variant of Arabic language uses many borrowed words (mainly French) and alters the pronunciation of many words of MSA [9, 10]. Furthermore, the borrowed words could be used such as in the original language, or they could be altered in order to respect the morphological structure of the Arabic language.

Building a robust speech recognition system requires feeding the training models with spoken and written data of the targeted language. Unfortunately, these kinds of data does not exist for the Algerian dialect since it is mainly spoken and there is no standards nor rules to write it. Our approach to recognize the Algerian dialect is to explore data sharing between the languages that impact the dialect, namely MSA and French. The main idea is to extend a small spoken corpus of the Algerian dialect with speech data from the MSA and the French languages, for training the acoustic models.

The aligned dialectal spoken corpus was created by having native Algerian people reading 4.6k sentences extracted from PADIC [14, 16] and CALYOU [1] corpora. Statistics about the resulted corpus, named ADIA (Algerian Dialect) in the following, are presented in Table 2. It should be noted that the speakers of the test data are different from those of the training and development data.

The same architecture used to build the ALASR system is used to train an initial acoustic model for the Algerian dialect based on the Train part of the ADIA corpus.

Table 2. Some figures of ADIA corpus.

Subset	Duration	Number of speakers		
		Female	Male	Total
Train	240 min	1	3	4
Dev	40 min	1	1	2
Test	75 min	1	2	3

This Train corpus was increased, afterwards, gradually by using acoustic data extracted from those used in ALASR system (MSA corpora) and with data extracted from ESTER (a French corpus) [5]. The optimal amount of acoustic data of each language to include in the training data was determined by minimizing the WER on the ADIA Dev corpus. We found that using a too large amount of MSA and French spoken data has a negative impact on the system performance. The optimal WER was obtained by adding 12 hours of MSA data and 12 hours of French data to the ADIA Train corpus.

The language model we propose, is a linear combination of four bigram models. Two of them were trained on MSA textual data: Gigaword and transcripts of the MSA acoustic Train set. The two others were trained on dialectal data: PADIC and CALYOU. The weights of the linear interpolation are estimated on a development corpus composed by a mixture of MSA and dialect data.

The initial MSA lexicon (MSA_{lex}) was extended by the most frequent words extracted from dialectal textual data (PADIC and CALYOU), which led to a lexicon of size of 125k words. The pronunciation variants of these dialectal words were produced by adapting the G2P approach proposed in [8].

In the first experimental phase, we want to evaluate how ALASR system performs on dialectal spoken data. Afterwards, we report the system performance by combining data from the three languages (dialect, MSA and French) to recognise the ADIA Test corpus. Table 3 summarizes the obtained results.

Table 3. Performance of the ASR systems on ADIA Test corpus.

System	Training Acoustic Data	WER(%)	OOV (%)
ALASR	MSA	78.5	33.6
S_1	ADIA	40.0	6.8
S_2	ADIA+MSA+Fr	**37.7**	

Since the Algerian dialect does not share many words with MSA (this is indicated by the high percentage of the OOV rate), ALASR system collapses completely when it was applied on the Test ADIA corpus. On the other side, with only 4 hours of dialectal training data (S_1 system), a WER of 40% was obtained. Moreover, by increasing this limited training corpus with data that come from MSA and French corpora, an absolute improvement of 2.3% is achieved. This shows the possibility to use data covering several languages to improve the recognition of a specific language.

5 Sentence Boundary Detection

Automatic speech recognition (ASR) systems aim to transform spoken data into a textual representation which may be used on further NLP tasks including POS tagging, semantic parsing, question answering, machine translation and automatic text summarization, [4, 12]. The vast majority of ASR systems focus on generating the correct sequence of transcribed words without taking into account the structure of the transcribed document, thus producing transcripts that lack of syntactic information like sentence boundaries. [7, 26]. However, optimal sentence boundary segmentation over ASR transcripts has shown to be crucial over further NLP tasks like entity and relation extraction, topic detection and automatic summarization [13, 18, 21].

Sentence Boundary Detection (SBD) aims to automatically split into sentences an unpunctuated text; nevertheless in spoken language the notion of sentence is not as well defined as in formal written sources. Separating into speaker utterances is a straightforward solution in spoken language, but in a standard conversation, utterances may be very long thus producing very long segments. In addition, disfluencies like repetitions, restarts, revisions, hesitations and interruptions make the definition of a sentence unclear. The concept of Semantic Unit (SU), introduced by the Linguistic Data Consortium on the SimpleMDE V5.0 guideline, is considered to be an atomic element of the transcript that achieves to express a complete idea [23]. A SU may correspond to the equivalent of a sentence in written text, a phrase or a single word. It seems to be an inclusive conception of a segment and is flexible enough to deal with the majority of spoken language troubles.

We implemented the SBD system based on character embeddings and Convolutional Neural Networks (CNN) described in [6] to segment the automatic transcripts into SUs. In this architecture, the CNN classifies the middle word of a 5-word window into *boundary* or *not boundary*. Character embeddings are word embedding representations where each word is expressed as the sum of their *n*-gram character vectors. This type of embedding representation is very useful for morphology rich languages like Arabic. To conduct our experiments we opted for the FastText character embedding [3] pre-trained vectors [6], which consist of 300 dimensions 610,977 vectors. The input layer of the CNN architecture proposed in [6] is represented by a 5×300 matrix representing the relation between a window of 5 words and their 300 dimension FastText vectors. The hidden architecture of the CNN consist of an arrange of convolutional, pooling and fully connected layers blocks followed by three fully connected layers. Finally, the output layer is composed of two neurons corresponding to two the possible output classes.

We performed the CNN training with a 70M words subset (Asharq Al-Awsat news wire) from the Arabic Gigaword[7] dataset. Table 4 shows the performance of the system in terms of the F1-score[8] for both classes over an evaluation set of 10.5M samples Detailed explanation of the CNN architecture and extended performance evaluation are available in [6].

[6] https://fasttext.cc/docs/en/crawl-vectors.html
[7] https://catalog.ldc.upenn.edu/LDC2011T11
[8] Harmonic mean combining Precision and Recall.

Table 4. Performance of the CNN based SBD system for the classes *boundary* and *no boundary*.

Class	**F1-score**
boundary	0.684
no boundary	0.980

6 Automatic Text Summarization

An automatic summary is a text generated by a software, that is coherent and contains a significant amount of relevant information from the source text. Usually, the compression rate ρ of the summary is less than a third of the length of the original document [25]. Automatic Text Summarization (ATS) systems aim to produce summaries from a source document. In general, the ATS algorithms work well if the source contains well-written documents like news, books, chapters, etc. In these kinds of documents, the sentences are reasonably well delimited: the borders of sentences are the final point, and the markers ? and !. In our case, where source documents correspond to transcripts from an ASR system, the deal is very different. Punctuation marks are non-existent and no phrase delimitation are available, thus the SBD system described in section 5 is applied before any summarization process is performed and segments salience is computed.

An extract is the assembly of fragments that have been extracted from a source text. The aim of an extract is to give a quick overview of the original document content. Extraction is an efficient topic and genre independent ATS method [25]. Surface-level methods do not delve into the linguistic depths of a document; rather they use some linguistic elements in order to identify the relevant segments of a document. Used in several studies on summarization, surface-level techniques use the occurrences of words to weight sentences.

In order to produce extractive text-based summaries, we opted for the Artex algorithm [24, 25]. This method is very simple, fast and efficient. The main idea is to map the source document (P sentences, n types terms, in a suitable space representation of a matrix $S_{[P\times n]}$. Each term is weighted by a classical *TF.IDF*, without stop-words and punctuation. All terms are stemmed using a Porter algorithm [19]. The original Artex version is able to process English, French and Spanish [24], but we adapted the prepossessing modules in order to process Arabic language. In the matrix space, Artex searches to compute a weight for each sentence i, using a scalar product between the main topic, the sentence i and the main type "word". The main topic is computed as the sum of P vector sentences. The main type "word" is computed as the sum of n vector words. The sentences close to the main topic and using several terms ad hoc the topic, are retained to generate the summary following a ρ ratio.

7 Experiments

To evaluate the results of the automatic summarization system, we decided to conduct a subjective evaluation. The evaluators are asked to give a score between 1 and 5 for both ALASR system and the automatic summarization system according to the ranking

assessment of Tables 5 and 6. It is necessary to evaluate the automatic speech recognition system because the automatic summarization system depends on it. In Table 7, we give some details about the evaluation of 27 videos. Each of them was summarized 3 times depending on several percentage (ρ ratio) of the original video. The Arabic videos concerned by the evaluation are those extracted in the framework of the project AMIS and concern the following channels: Euronews, AlArabiya and Skynews. Three native Arabic speakers evaluated the videos. The smallest transcribed video is composed of 52 words and the longest one of 394 words.

Table 5. Rating scale for the ALASR system assessment.

1	Incomprehensible transcription
2	Only certain segments of the video are understandable
3	A substantial proportion of the transcription is understandable
4	The transcription is very understandable
5	The transcription is not only understandable, but it is fluid and does not seem to involve linguistic errors (syntactic or semantic).

Table 6. Rating scale for the automatic summarization system assessment.

1	Incomprehensible summary
2	Only some events of the original video are found in the summary and overall the text is incomprehensible
3	A substantial proportion of the events in the original video are in the summary and overall the text is understandable
4	Very good summary and the text is very correct
5	Excellent summary

Table 7. Some figures concerning the subjective evaluation.

Count	**Value**
Videos	27
Summary per Video	3
Channel TV	3
Evaluators	3
Size of the shortest summary (in words)	52
Size of the longest summary (in words)	394

In Fig 1, we draw the Box plot of the results of the subjective evaluation of the ALASR system and of the automatic summary system. The latter system depends obviously on the result of the ASR system. That is why we report them in the same diagram. Half of the population of the ASR evaluation received an evaluation between 3 and 4 and the upper Quartile is equal to 4 which means that 25% of the transcriptions have

received the highest score. These results indicate that the developed ALASR system performs very well.

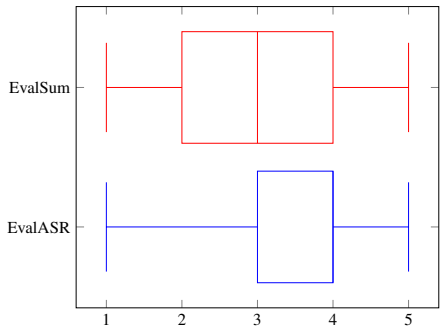

Fig. 1. The Box plot corresponding to the subjective evaluation of the Arabic ASR and the automatic summarization systems on MSA data.

In Fig 2, we analyse the lengths, in terms of words, of the transcriptions and the summaries in order to attempt to find a relationship between the size of the summary, and the performance of the automatic summarization system. The Quartile Q_1 is equal to 81, that means that 25% of the summaries have a length smaller than 81 words knowing that the longest transcription is composed of 394 words. Also, 25% of the population has a length greater than 140 words, which correspond to 35% of the longest video.

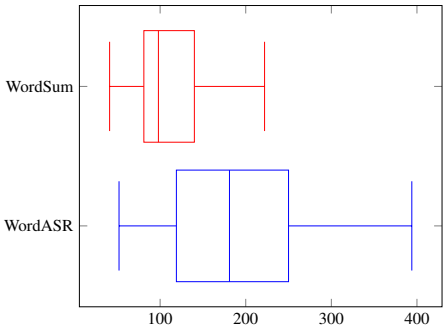

Fig. 2. The Box plot corresponding to the number of words of the Arabic ASR and the automatic summarization systems.

The same evaluators conducted another assessment, it concerns the evaluation of the Arabic ASR and the automatic summarization systems in which some video sequences are in Algerian dialect. To do so, 6 videos from Algerian TV, namely Echorouk and Ennahar were recognized by ALASR and by the system we adapted to better recognize the Algerian dialect. Fig 3 shows the number of the examples that receive scores between 1 and 5. We can remark that no video received a rating of 5 and consequently no more summary received this score. Only 6 videos have been ranked 4, but unfortunately no

summary was ranked 4. 12 evaluations on the Arabic ASR are considered as not understandable and 15 among the population have a bad summary (score = 1). These bad results were expected with an Arabic ASR not adapted to the Algerian dialect.

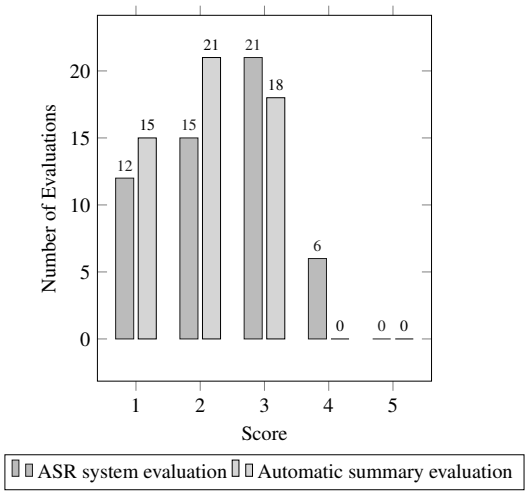

Fig. 3. The number of responses for each score of the subjective assessment of dialectal data with ALASR system.

By transcribing the videos with the adapted Arabic ASR system (Fig 4) for Algerian dialect, no improvement on high score ratings and especially for score 5 was found, on the other hand 12 examples of the population were ranked 4 and this led to 2 summaries with a score 4.

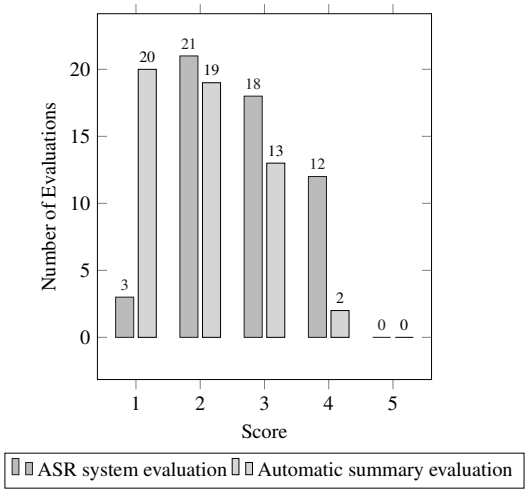

Fig. 4. The number of responses for each score of the subjective assessment of dialectal data with the adapted ASR system.

In order to study the relationship between the scores of the summary and the other parameters such as: the number of words (*ASRWord*) of the original video, the score of the ASR system (*ASRScore*) and the number of words of the summary (*SumWord*), we decided to use the multiple linear regression that has the objective to model the linear relationship between the explanatory independent variables mentioned above and the dependent response variable (*EvalSum*). We use the statistical metric (R^2), named coefficient of determination to measure how much of the variation in outcome can be explained by the variation in the independent variables. It measures the adequacy between a model resulting from a multiple linear regression and the observed data which made it possible to establish the relationship. On our dataset of 243 examples, $R^2 = 0.310$, this indicates that 31% of the dispersion is explained by the regression model. This is not high value, but it is not completely null. If we consider the null hypothesis as $H_0 : a_1 = a_2 = a_3 = 0$ and the alternative hypothesis as at least one of the a_i is different from 0. The model F depending on R^2 is calculated as follows:

$$F = \frac{\frac{R^2}{p}}{\frac{1-R^2}{n-p-1}} \tag{1}$$

Where n is the size of the sample and p is the number of degrees of freedom. The calculated value of F is equal to 35.899. F follows a Fisher law at $(p, n-p-1)$ degrees of freedom. The theoretical $F(2,240)$ is equal to 3.239. In conclusion, the critical region of the test is therefore: rejection of H_0 because $F > F_{0.95}(2,240)$. The hypothesis that there is a relationship between the explanatory variables and the score of the automatic summarization system can not be ruled out.

8 Conclusion

In this paper, we present and evaluate an extractive text-based summarization method for Arabic videos, which is proposed in the scope of AMIS project. AMIS aims at helping a user to understand videos given in a foreign language (Arabic in this study and research), by translating and summarizing the videos through several strategies. One strategy consists in transcribing the Arabic videos and summarizing the transcriptions. The evaluations of summaries were objective and also subjective.

The objective evaluation of the ASR system showed the necessity to include dialectal material in the training data when the Algerian dialect is used in the videos. This result was confirmed by the subjective evaluation of ASR outputs: when dialectal data is used for training, transcriptions of Algerian dialect videos are better evaluated. However, the automatic summaries obtained from the transcriptions do not lead to the same conclusion: with dialectal data in training, the summaries are judged less good. In order to better understand these contrasting results, we tried to measure which features of summaries influence the judgement. This study showed that original lengths of videos, lengths of summaries and ASR performance explain only 31% of the subjective scores. Furthermore, a statistical analysis shows that a relationship between these features and the scores given to summaries can not be ruled out.

This research shows the difficulty to evaluate results for complex projects such as AMIS as the summarization task requires a high degree of cognitive effort during the evaluation. So the question is how to automatically predict the quality of summaries? To answer to this question, in future work, we would like to more deeply explore which features influence the quality of summaries. For that, it will be necessary to increase the number of evaluated videos.

Acknowledgment

We acknowledge the support of Chist-Era for funding this research through the AMIS (Access Multilingual Information opinionS) project.

References

1. Abidi, k., Menacer, M.a., Smaili, K.: CALYOU: A comparable spoken algerian corpus harvested from youtube. In: 18th Annual Conference of the International Communication Association (Interspeech) (2017)
2. Ali, A., Zhang, Y., Cardinal, P., Dahak, N., Vogel, S., Glass, J.: A complete kaldi recipe for building arabic speech recognition systems. In: Spoken Language Technology Workshop (SLT), 2014 IEEE. pp. 525–529 (Dec 2014). https://doi.org/10.1109/SLT.2014.7078629
3. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics **5**, 135–146 (2017)
4. Che, X., Wang, C., Yang, H., Meinel, C.: Punctuation prediction for unsegmented transcript based on word vector. In: LREC (2016)
5. Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J.F., Gravier, G.: The ester phase ii evaluation campaign for the rich transcription of french broadcast news. In: Ninth European Conference on Speech Communication and Technology (2005)
6. González-Gallardo, C.E., Pontes, E.L., Sadat, F., Torres-Moreno, J.M.: Automated sentence boundary detection in modern standard arabic transcripts using deep neural networks. Procedia Computer Science **142**, 339–346 (2018)
7. Gotoh, Y., Renals, S.: Sentence boundary detection in broadcast speech transcripts. In: ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW) (2000)
8. Harrat, S., Meftouh, K., Abbas, M., Smaïli, K.: Grapheme to phoneme conversion - an Arabic dialect case. In: Spoken Language Technologies for Under-resourced Languages (2014)
9. Harrat, S., Meftouh, K., Smaïli, K.: Creating Parallel Arabic Dialect Corpus: Pitfalls to Avoid. In: 18th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING). Budapest, Hungary (Apr 2017), `https://hal.archives-ouvertes.fr/hal-01557405`
10. Harrat, S., Meftouh, K., Smaïli, K.: Maghrebi Arabic dialect processing: an overview. Journal of International Science and General Applications **1** (2018), `https://hal.archives-ouvertes.fr/hal-01873779`
11. Leszczuk, M., Grega, M., Koźbiał, A., Gliwski, J., Wasieczko, K., Smaïli, K.: Video summarization framework for newscasts and reports–work in progress. In: International Conference on Multimedia Communications, Services and Security. pp. 86–97. Springer (2017)
12. Linhares Pontes, E., González-Gallardo, C.E., Torres-Moreno, J.M., Huet, S.: Cross-lingual speech-to-text summarization. In: Choroś, K., Kopel, M., Kukla, E., Siemiński, A. (eds.) Multimedia and Network Information Systems. pp. 385–395. Springer International Publishing, Cham (2019)

13. Makhoul, J., Baron, A., Bulyko, I., Nguyen, L., Ramshaw, L., Stallard, D., Schwartz, R., Xiang, B.: The effects of speech recognition and punctuation on information extraction performance. In: Ninth European Conference on Speech Communication and Technology (2005)
14. Meftouh, K., Harrat, S., Smaïli, K.: PADIC: extension and new experiments. In: 7th International Conference on Advanced Technologies ICAT. Antalya, Turkey (Apr 2018), https://hal.archives-ouvertes.fr/hal-01718858
15. Meftouh, K., Bouchemal, N., Smaïli, K.: A Study of a Non-Resourced Language: The Case of one of the Algerian Dialects. In: The third International Workshop on Spoken Languages Technologies for Under-resourced Languages - SLTU'12. pp. 1–7. Cape-town, South Africa (May 2012), https://hal.archives-ouvertes.fr/hal-00727042
16. Meftouh, K., Harrat, S., Jamoussi, S., Abbas, M., Smaili, K.: Machine translation experiments on PADIC: A parallel arabic dialect corpus. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. pp. 26–34 (2015)
17. Menacer, M.A., Mella, O., Fohr, D., Jouvet, D., Langlois, D., Smaïli, K.: Development of the Arabic Loria Automatic Speech Recognition system (ALASR) and its evaluation for Algerian dialect. In: ACLing 2017 - 3rd International Conference on Arabic Computational Linguistics. pp. 1–8. Dubai, United Arab Emirates (Nov 2017), https://hal.archives-ouvertes.fr/hal-01583842
18. Mrozinski, J., Whittaker, E.W., Chatain, P., Furui, S.: Automatic sentence segmentation of speech for automatic summarization. In: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. vol. 1, pp. I–I. IEEE (2006)
19. Porter, M.F.: An algorithm for suffix stripping. Program **14**(3), 130–137 (1980)
20. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The kaldi speech recognition toolkit. In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. IEEE Signal Processing Society (2011), iEEE Catalog No.: CFP11SRW-USB
21. Shriberg, E., Stolcke, A., Hakkani-Tür, D., Tür, G.: Prosody-based automatic segmentation of speech into sentences and topics. Speech communication **32**(1-2), 127–154 (2000)
22. Stolcke, A.: Entropy-based pruning of backoff language models. arXiv preprint cs/0006025 (2000)
23. Strassel, S.: Simple metadata annotation specification v5 (01 2003), http://www.ldc.upenn.edu/Projects/MDE/Guidelines/SimpleMDE
24. Torres-Moreno, J.M.: Artex is another text summarizer. arXiv preprint arXiv:1210.3312 (2012)
25. Torres-Moreno, J.M.: Automatic Text Summarization. Wiley (2014)
26. Yu, D., Deng, L.: Automatic speech recognition. Springer (2016)