SegChain: Towards a generic automatic video segmentation framework, based on lexical chains of audio transcriptions

Adrian-Gabriel CHIFU
LSIS UMR 7296 CNRS
Université Aix-Marseille
Marseille, France
adrian.chifu@lsis.org

Sébastien FOURNIER
LSIS UMR 7296 CNRS
Université Aix-Marseille
Marseille, France
sebastien.fournier@lsis.org

ABSTRACT

With the advances in multimedia broadcasting through a rich variety of channels and with the vulgarization of video production, it becomes essential to be able to provide reliable means of retrieving information within videos, not only the videos themselves. Research in this area has been widely focused on the context of TV news broadcasts, for which the structure itself provides clues for story segmentation. The systematic employment of these clues would lead to thematically driven systems that would not be easily adaptable in the case of videos of other types. The systems are therefore dependent on the type of videos for which they have been designed. In this paper we aim at introducing SegChain, a generic unsupervised framework for story segmentation, based on lexical chains from transcriptions. SegChain takes into account the topic changes by perceiving the fluctuations of the most frequent terms throughout the video.

CCS Concepts

•**Information systems** → **Recommender systems; Video search;** *Web searching and information discovery;*

Keywords

Video retrieval; Story segmentation; Lexical chains; Transcriptions

1. INTRODUCTION

With the multimedia broadcasting development, via increasingly diversified channels and with the democratization of video production, like the case of textual content production, it becomes essential to provide effective means for retrieving the information contained within videos and not only the videos themselves. Furthermore, the very use of audiovisual content is not linear anymore, but becoming an "on demand" service. Users themselves choose the content they wish to watch. However, the large amount of existing videos requires assistance, so that users may pertinently choose what they desire to watch and eventually offer them not a video but rather a series of videos, or going even further, propose a sequence of video segments. To achieve this result, it is essential to have a description of the video and, if we want to go further, a description of the segments or semantically homogeneous chapters within the videos. This is the case of the huge collection of Internet hosted videos that challenge both video content providers and users when it comes to video information retrieval. Thus, one of the challenges is to automatically build such segments and to provide data regarding these segments with the purpose of facilitating their research. In order to automatically build these video segments, according to various studies, one of the most effective solutions is to use the inherent video multimodality. To better exploit multi-modality, while still remaining as generic as possible, it is particularly necessary to understand the interlacing of these various modalities. Indeed, the majority of research is essentially focused so far on the construction and discovery of semantically homogeneous segments within videos coming from the field of television news. Yet, it is a relatively small sub-type and it has a well defined structure (with invariants easy to operate, like the presenter, or some repeating key phrases), facilitating the discovery of these semantically homogeneous segments. Therefore, to cover a wider range of videos, our work is part of a more generic framework. SegChain, the framework that we want to develop, will use all the existing terms within a video, while trying to extract specific elements present in such systems. Therefore, we want at first to study the contribution of each of these terms on border detection between two chapters or segments. Our work in this article focuses on the use of video transcripts to identify the semantically homogeneous segments. For this, we propose in this paper a method based on lexical chains [17], coupled with similarity measures in order to follow the topic variations reflected in the perception of the term fluctuation throughout the video.

The educational video (MOOC) has lately known a popularity boost. Many educational and online course platforms have been developed, thus the large amount of video content should be treated such that the information could be optimally retrieved and exploited. We therefore consider, in addition to TV news broadcasts, MOOC videos in our evaluation (see Section 5), since MOOC platforms could benefit from our framework. SegChain could be integrated in multimodal treatment of MOOC videos, being useful for story

WIMS'16 June 13–15, 2016, Nîmes, France

segmentation, keyword identification, content recommendation, linkage with semantically homogeneous segments from other videos and so on.

The novelty of SegChain consists in the way the lexical chains for the most frequent terms are considered. We introduce a chain compactness measure to filter out the terms that are too dispersed along the transcription. Another distinctive trait is the two-step boundary detection when SegChain seeks first the similarity variation between short segments of transcription text in order to suggest and then checks once again the similarity between the suggestions, in order to aggregate them in larger zones that are semantically homogeneous. SegChain is also language independent, requiring only a stop-word list for the considered language.

This article is organized as follows. The second section presents the related work both on story segmentation and on lexical chains. The third section presents our video segmentation lexical chains-based method that exploits changes in topics through the perception of the fluctuation of these terms throughout the video. The fourth section presents an example of application on a video of a French television newscast, followed by a preliminary evaluation in the fifth section. Finally, the article is concluded and the future work is presented.

2. RELATED WORK

2.1 Story segmentation

Story unit (chapter) detection for videos has been studied with focus on TV news videos and more particularly during TRECVID competitions in 2003 and 2004. In order to achieve such a task, as mentioned in [2] and [10], multimodal use of available resources (transcript, video and audio dialogue) yielded the best results.

Thus, Dumont *et al.* [5], through machine learning techniques (random forest), combine features such as silence, speaker detection and descriptors obtained from image processing such as visual activity or logo detection to discover the transitions between chapters. In [15], several modalities are used, such as the text with keywords or named entities, the video with transitions between shots, the audio with the use of silences and the internal structure of TV news.

However, a key element to detect chapters in TV news is the main presenter (anchor) detection and its consequently employment as the starting point of chapters. Thus, O'hare *et al.* [14] use the presence of the presenter coupled with techniques based on shot classification. From work which is also based on the presence of a presenter, we can also mention the research of [4], [13] and [19].

Colace [4] use Bayesian networks to identify the presenter and get begining of chapters. Misra [13] also rely on the presence of main presenter to detect chapters, employing afterwards the color similarity of adjacent shots. In [19] anchor identification is also employed while building a true signature using the points of interest identification in the image, also including colorimetry. To build this signature, the authors also use the internal structure of the program.

Other work focuses on audio transcript analysis and detects transitions chapters by extracting key terms like "welcome" or "no transition".

Approaches from the literature achieve excellent results, sometimes exceeding 90% or 95% of precision and recall as in the work of [19]. However, even if newscasts have slightly different structures, they are still often created by following simple rules that provide important clues for automatic content analysis [7]. Moreover, the systematic taking into account of these clues leads to a strong specialization of systems making them difficult to adapt to other types of videos. These systems are highly dependent on the video program for which they have been designed.

2.2 Lexical chains

Various approaches have been applied to speech or text based topic segmentation. One of such approach is lexical chains. For instance, in [11] the authors rely on text representation as weighted lexical chains. The research from [9], [16] and [1] yields close approaches, using similarities but use different way to represent the text. Another approach, presented by [18], uses hidden Markov models.

In [6] and in [8], the authors extended previous research and try to address some of their drawback by integrating semantically related terms to the segmentation model in order to extend the description of the possible segments.

More recently, [3] used an approach that can be seen as an improved variant of text-tiling using new segmentation technique inspired from the image analysis field and relying on a new way to compute similarities between candidate segments called vectorization.

3. STORY SEGMENTATION BASED ON LEXICAL CHAINS

To have an independent and generic approach, we apply the lexical chains using a different text representation and we define additional pretreatment operations on these chains. The proposed method is based on the following hypothesis: within a video segment, there is a homogeneous distribution of the most frequent terms. Obviously, the topic shift involves changes in this distribution. The detection of these changes establishes the borders between two homogeneous segments.

We work in this context with the textual information contained in video transcripts, which is represented here by a subtitle file. This file is composed by multiple subtitle units. A subtitle unit is the text chunk uniquely identified in time by the start and end timestamps. One story unit lasts for 3-4 seconds on average.

Let S be the set of subtitle units such that:

$$S = \{st_i\}$$

Let T be the set of the most frequent terms such that:

$$T = \{t_i\}$$

Definition 1. **(Apparition function)** The apparition function determines if a term t_i occurs in a subtitle unit st_j. The apparition function (denoted *app*) is defined as follows:

$$app\,(t_i, st_j) = \begin{cases} 1, & \text{if } t_i \in st_j, \\ 0, & \text{otherwise} \end{cases}, i \in [0, |T|] \text{ et } j \in [0, |S|) \quad (1)$$

Definition 2. **(The set of positions of initial terms)** In our method, since we have sparse vectors, for each term t_i, we shall store only the positions of subtitle units st_j for

which the function *app* has a value different than zero. This set is called *n0app* and it is defined as follows:

$$n0app\left(t_i\right) = \left\{j \mid app\left(t_i, st_j\right) \neq 0\right\}, i \in [0, |T|) \text{ and } j \in [0, |S|) \tag{2}$$

As shown by Sitbon and Bellot [17], there are multiple ways to compute the value of a hiatus. We consider the following definition.

Definition 3. **(Hiatus)** A hiatus represents the average of distances between two occurrences divided by the total length of this term's repartition along the entire subtitle set. Thus, the hiatus is defined as follows:

$$hiat\left(t_i\right) = \frac{\sum_{k=0}^{|n0app(t_i)|-2} \left[n0app\left(t_i\right)_{k+1} - n0app\left(t_i\right)_k\right]}{|n0app\left(t_i\right)| - 1}, \quad i \in [0, N) \tag{3}$$

Definition 4. **(Lexical chains)** For two consecutive occurrences k and $k+1$ of a term t_i, a chain is created between k and $k+1$ if the distance between these occurrence positions is inferior or equal to the hiatus value computed for the term t_i. For $i \in [0, |T|)$ et $j \in [0, |S|)$, the belonging of a term position to a chain is computed by:

$$appC\left(t_i, st_j\right) = \begin{cases} 1, & \text{if } j \in n0app\left(t_i\right) \\ 1, & \text{if } n0app\left(t_i\right)_k < j < n0app\left(t_i\right)_{k+1} \text{ and} \\ & n0app\left(t_i\right)_{k+1} - n0app\left(t_i\right)_k < hiat\left(t_i\right) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Thereby, a lexical chain is the maximal set of consecutive not null positions in *appC*, as in equation 5.

$$chain(t_i) = \left\{ (d, f) \,\middle|\, \forall k, d \leq k \leq f, appC\left(t_i, st_k\right) = 1 \text{ and } \nexists \left(d', f'\right), \forall k', d' \leq k' \leq f', appC\left(t_i, st'_k\right) = 1 \right\} \tag{5}$$

The lexical chain length represents the difference between the first and the last position in the chain, as follows:

$$lenCh\left(t_i\right) = \left\{f - d \mid (d, f) \in chain\left(t_i\right)\right\} \tag{6}$$

Definition 5. **(Chain compactness)** The chain compactness defines a chain's discriminant ability. More a chain is compact, more it will be discriminating in relation to the topic change in the video. The chain compactness measure takes into account the first and the last occurrence of a term, the total number of subtitle units, the total number of a term's occurrences, the number of chains in which the term occurs and the maximal length of all the chains of all the terms. It is defined as follows:

$$comp\left(t_i\right) = \frac{\left[n0app\left(t_i\right)_{|n0app(t_i)|-1} - n0app\left(t_i\right)_0\right]}{|S|} \times \frac{|n0app\left(t_i\right)| \times |chain\left(t_i\right)|}{\max\limits_{j \in [0,N]} \left(\max\left(lenCh\left(t_j\right)\right)\right)} \tag{7}$$

Next, the terms are ordered descending descending with respect to their *comp* score. These terms are filtered afterwards by their score computed in equation 7. Terms with *comp* value inferior to a threshold are discarded. The threshold is chosen such that a percentage of the terms are filtered out (from 25% to 40% in our experiments). The new set of terms T' is defined as follows:

$$T' = T - \left\{t_i \mid comp\left(t_i\right) < \alpha\right\} \tag{8}$$

Definition 6. **(Apparition by subtitles)** Each subtitle unit is represented by the set of *appC* for each. We therefore obtain a term apparition set, denoted *cSt*, with respect to each subtitle unit and defined as follows:

$$cSt\left(st_i\right) = \left\{appC\left(t_j, st_i\right) \mid j \in \left[0, |T'|\right)\right\} \tag{9}$$

Definition 7. **(Subtitle unit similarity)** The similarity between two subtitle units corresponds to the cosine similarity between two consecutive subtitle units represented by *cSt* (equation 10). The similarity is defined as follows:

$$sim\left(k\right) = cos\left(cSt\left(st_k\right), cSt\left(st_{k+1}\right)\right), k \in \left[0, |T'| - 1\right) \tag{10}$$

With the purpose of identifying topic shifts, we first identify the subtitle units that are candidates for borders between two topics. These border positions are the local minima of the similarity values computed by equation 10.

$$minima = \left\{i \mid sim\left(i\right) \leq sim\left(x\right), \forall x \in [i, i + x]\right\} \tag{11}$$

Definition 8. **(Subtitle unit segment)** A subtitle unit segment represents the sum of *cSt* values between two consecutive local minima positions.

$$segment = \left\{ \sum_{k=i}^{i+1} cSt\left(st_k\right) \,\middle|\, i \in minima \right\} \tag{12}$$

Definition 9. **(Semantically homogeneous segment border)** In order to compute the definitive border positions that separate every two consecutive and semantically homogeneous segments, we compute the similarity between the subtitle unit segments from the *segment* set. This similarity measure is computed by the function called *sim2* which also employs the cosine similarity measure. The similarity value is then compared to a threshold (0.6 in our experiments).

$$sim2\left(k\right) = cos\left(segment\left(k\right), segment\left(k+1\right)\right), \quad k \in [0, |segment| - 1) \tag{13}$$

The border positions for the semantically homogeneous segments are thus the positions that verify the following equation:

$$front = \left\{k \mid sim2\left(k\right) \geq \beta\right\} \tag{14}$$

To obtain the final results, that is to say the segmentation in semantically homogeneous areas within the video and also to enable the capability of information retrieval inside the video, we propose the algorithm (Algorithm 1) that brings together the elements composing SegChain.

Algorithm 1 SegChain: The story segmentation method

Requires: A subtitle file in TRS or SRT format.
Ensures: Subtitle unit IDs corresponding to the segment borders, as well as the most frequent terms for each segment.

if TRS file format **then**
    convert file into SRT format
extract the raw text from the subtitle file
remove punctuation
tokenize text into words
remove stop words
compute term frequency
$T \leftarrow$ the most frequent N terms
build S
for all $t_i \in T$ **do**
    compute the $app\,(t_i, st_j)$ vector
    compute $hiat\,(t_i)$
    compute $chain\,(t_i)$
    compute $comp\,(t_i)$
compute T'
for all $St_j \in S$ **do**
    compute $sim\,(st)$
compute $minima$
for all $St_j \in S$ **do**
    compute $segment$
compute $front$

4. EXAMPLE

We have ran SegChain on a video in order to test it. This video has a length of approximately 40 minutes and it is a TV news broadcast from France TV. In Figure 1 we display the terms considered after the removal of the less discriminant terms, judged by equation 7. In this figure, the terms are ordered by their compactness values of their lexical chain. For instance, the terms "dette" and "pense" are considered as the terms that have the most compact lexical chains. The horizontal lines represent the lexical chains identified for each term. This figure also shows the candidate positions for borders, marked with red vertical lines, as well as the definitive borders (black vertical lines) that split the semantically homogeneous segments. The horizontal axis represents the subtitle units, while the vertical axis represents the terms.

The *sim* similarity values are displayed in Figure 2 to show the term similarity variation between neighbor subtitle units. From the local minima of this curve the candidate positions for segmentation are selected.

5. PRELIMINARY EVALUATION

To conduct a preliminary evaluation we considered several aspects: employment of effectiveness measures, evaluation on TV news, evaluation on MOOC videos, comparison with state-of-the-art.

Regarding the evaluation measures, three measures have been employed: *Precision*, *Recall* and *F_measure*. These measures are well known in information retrieval and there are widely used in story segmentation evaluation [12], [13].

Definition 10. **(Precision)** The segmentation precision measures how many good frontiers have been detected among all the detections and it is defined as follows:

$$Precision = \frac{|good\_front|}{|all\_detected\_front|}, \tag{15}$$

where $|good\_front|$ represents the number of correctly detected frontiers, with an established degree of confidence, while the total number of detected frontiers is represented by $|all\_detected\_front|$. The degree of confidence's value ($conf$) represents the number of subtitle units, at left and at right from the detected frontier ($detected\_front$), for which the following condition is verified:

$$true\_front \in [detected\_front - conf, detected\_front + conf],$$

where $true\_front$ represents a real frontier, from the ground truth.

If the condition is true, the $detected\_front$ will be considered as $good\_front$.

We need to mention that our notion of $good\_front$ is very restrictive, since we set the degree of confidence $conf$ with the value 1. This means that, a $detected\_front$ is a $good\_front$ only if it is itself a $true\_front$, or at least one of it's left or right neighbors are a $true\_front$.

Definition 11. **(Recall)** The segmentation recall measures how many good frontiers have been detected among all the true frontiers and it is defined as follows:

$$Recall = \frac{|good\_front|}{|all\_true\_front|}, \tag{16}$$

where $|all\_true\_front|$ represents the number of all true frontiers, from the ground truth.

Definition 12. **(F_measure)** The $F\_measure$ is the harmonic mean of $Precision$ and $Recall$ values:

$$F\_measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

We evaluated SegChain over the TV news video in French (40 minutes) from Section 4, which has been manually annotated with true frontiers. In order to test the generality of the framework, we also tested the performance on a MOOC video in English (41 minutes). The method is language independent, the only requirement being a stop-word list for the respective language. The ground truth in the MOOC case is automatically generated as follows: several small MOOC videos are gathered together in a single larger video and the ground truth frontiers represent the limits of each small video. We also compare our method's effectiveness with the effectiveness of TextTiling [9], a state-of-the-art method.

The performances of the two methods are displayed in Table 1 and Table 2, in a comparative manner.

Table 1: Performance measures for TextTiling and SegChain, on the TV news video

TV news video	TextTiling	SegChain
Precision	0.1071	**0.3333**
Recall	**0.4285**	0.2143
F_measure	0.1714	**0.2609**

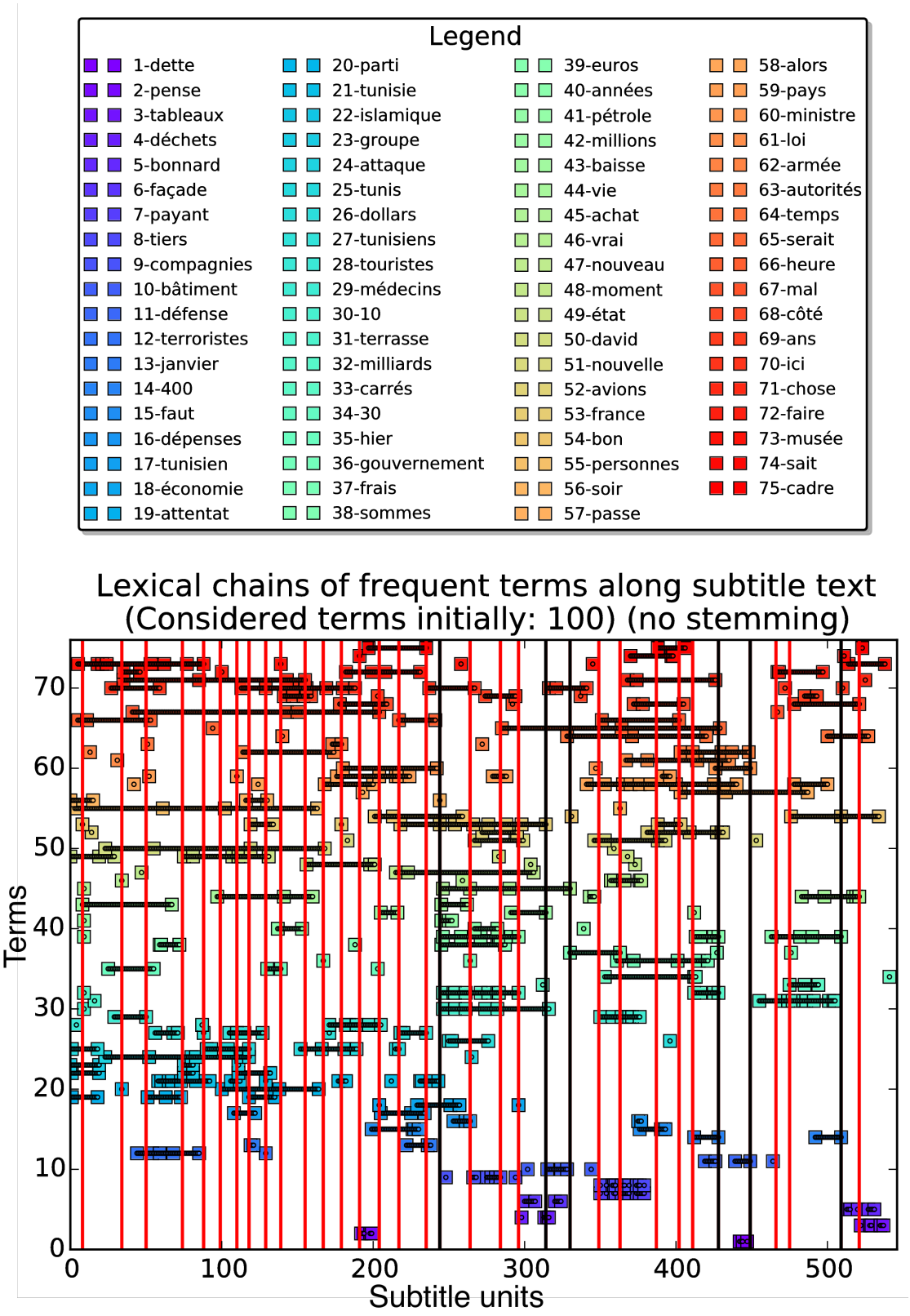

Figure 1: Segment borders for the lexical chains of the most frequent terms

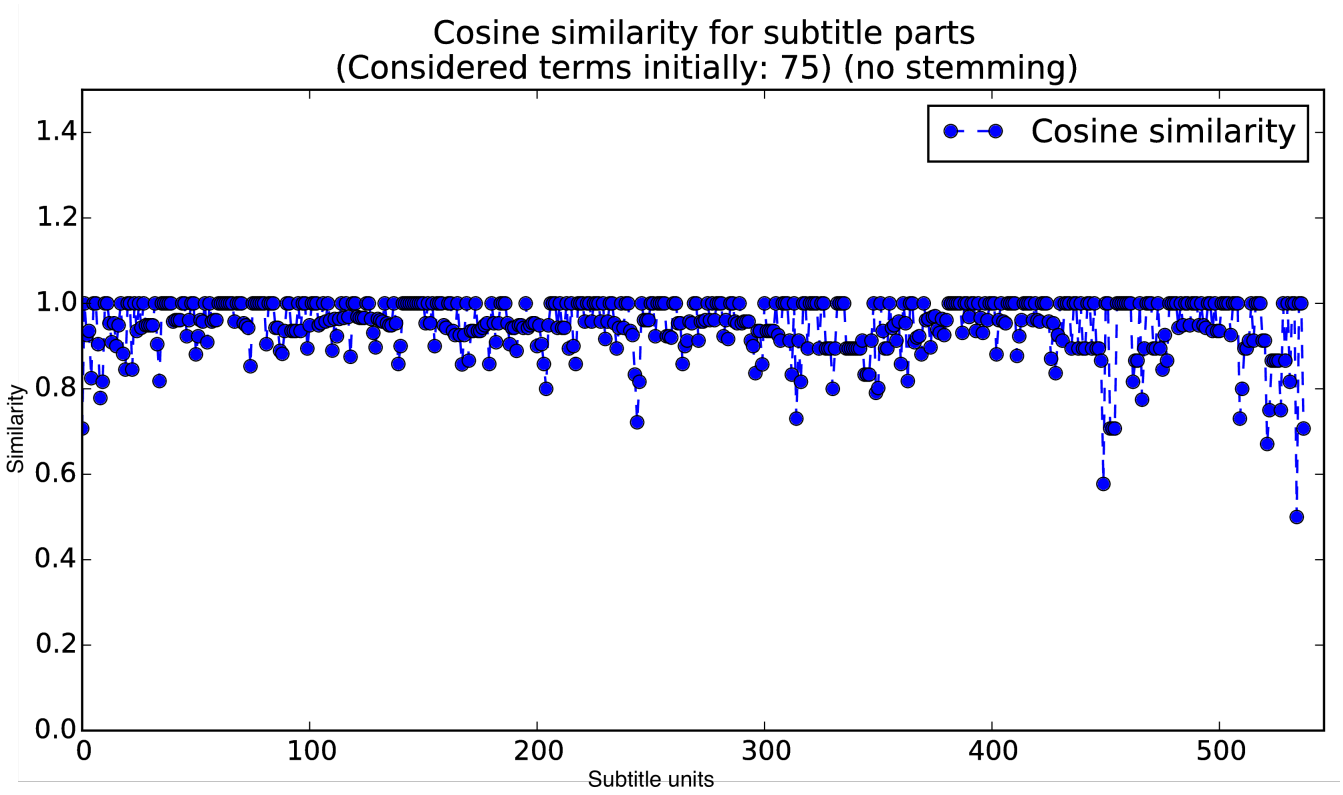

Figure 2: Cosine similarity between neighbor subtitle units

One can notice that the TextTiling has a better recall in the case of the TV news video, while SegChain yields a better precision and a better *F_measure* value. Having a better precision is a good advantage since our objective is to segment as precisely as possible.

Table 2: Performance measures for TextTiling and SegChain, on the MOOC video

MOOC video	TextTiling	SegChain
Precision	0.0408	**0.1429**
Recall	0.6667	0.6667
F_measure	0.0769	**0.2353**

On the other hand, on the MOOC video we obtain the same recall value (which is also a good one), and a much better precision, also implying a better *F_measure* value. This is encouraging, having in mind the difficulties of MOOC video segmentation, mentioned below.

The MOOC videos represent a challenge compared to TV news videos, since the structure is different. In most cases, within a MOOC the vocabulary remains the same. On the other hand, in TV news videos the topics are very well segregated, there are hints such as the anchor talking and various keywords. Moreover, from the multi-modal perspective of video image treatment, the shot boundaries from MOOC videos are more difficult to be correctly extracted. For example, in the case of TV news videos the shot changes are more clear (going from studio to field broadcast), while in the case of MOOC videos, the changes are more subtle (the slide background colors remain basically the same when slides are changing). Therefore, the success of a textual segmentation approach is crucial when multi-modal treatment fails.

We also list here several limitations of TextTiling. It requires a context window that impacts severely on the number of segments. For instance, for a window of 30 words (default value), 57 segments have been found, while for a window of 100 words 14 segments have been found. In addition, TextTiling is designed to work with paragraphs and well formatted sentences, and it is not the case of automatically generated audio transcriptions. Moreover, TextTiling outputs segments that basically have the same length, which is unrealistic since the topics can vary in time length in an important manner (from less than a minute up to a few dozens of minutes). The generic framework proposed in this paper does not have these limitations.

6. CONCLUSION

The research from this article introduced SegChain, an approach that aims at defining a generic framework for video segmentation. In order to achieve this objective, our work is firstly based on audio transcription processing to identify semantically homogeneous segments. For this, we proposed in this article a lexical chain-based method [17], combined with cosine-based similarity measures in order to detect topic variations through the perception of term fluctuation within chains, all along the video. We provided several definitions that are employed in a semantically homogeneous segment detection algorithm. We have presented an application example over a 40 minute long video.

We have conducted a preliminary evaluation of SegChain on the French TV news broadcast taken as example and on an English MOOC video. The performance of SegChain was compared to the performance of the TextTiling method [9] through three effectiveness measures: *Precision*, *Recall* and *F_measure*. SegChain has obtained better performance in terms of *Precision* and *F_measure* in both cases, but a lower *Recall*. In the case of the English MOOC the *Recall* values are the same as TextTiling.

On a short term perspective, we wish to validate our method on the TRECVID 2003-4 that are most commonly employed for this task in order to be able to compare our results with the state of the art.

However, we want to insist on assessing the SegChain's robustness on other video types than news TV, such as various types of MOOCs (of different length, subject, etc.) or videos that have a structure less clearer than news TV, since we believe we would obtain better results on videos with a more "free" structure than methods specifically built for news TV.

7. ACKNOWLEDGMENTS

This work has been carried out thanks to the support of the A*MIDEX project (n° ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR).

8. REFERENCES

[1] F. Y. Y. Choi. Advances in domain independent linear text segmentation. In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*, NAACL 2000, pages 26–33, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics.

[2] T.-S. Chua, S.-F. Chang, L. Chaisorn, and W. Hsu. Story boundary detection in large broadcast news video archives. In *Proceedings of the 12th annual ACM international conference on Multimedia - MULTIMEDIA '04*, page 656, New York, New York, USA, oct 2004. ACM Press.

[3] V. Claveau and S. Lefèvre. Topic segmentation of TV-streams by watershed transform and vectorization. *Computer Speech & Language*, 29(1):63–80, jan 2015.

[4] F. Colace, P. Foggia, and G. Percannella. A Probabilistic Framework for TV-News Stories Detection and Classification. In *2005 IEEE International Conference on Multimedia and Expo*, pages 1350–1353. IEEE, 2005.

[5] É. Dumont and G. Quénot. Automatic Story Segmentation for TV News Video Using Multiple Modalities. *International Journal of Digital Multimedia Broadcasting*, 2012:1–11, 2012.

[6] O. Ferret. Improving Text Segmentation by Combining Endogenous and Exogenous Methods. *Proceedings of the International Conference RANLP-2009*, pages 88–93, 2009.

[7] A. Goyal, P. Punitha, F. Hopfgartner, and J. M. Jose. Split and Merge Based Story Segmentation in News Videos. In *Proceeding ECIR '09 Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, pages 766–770, 2009.

[8] C. Guinaudeau, G. Gravier, and P. Sébillot. Enhancing lexical cohesion measure with confidence measures, semantic relations and language model interpolation for multimedia spoken content topic segmentation. *Computer Speech & Language*, 26(2):90–104, apr 2012.

[9] M. A. Hearst. TextTiling: segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64, mar 1997.

[10] W. H. Hsu, L. S. Kennedy, S.-f. Chang, M. Franz, and J. R. Smith. COLUMBIA-IBM NEWS VIDEO STORY SEGMENTATION IN TRECVID 2004 Dept . of Electrical Engineering , Columbia University , New York { winston , lyndon , sfchang } @ ee . columbia . edu IBM T . J . Watson Research Center , New York Columbia University ADVENT Tech. Technical report, Dept. of Electrical Engineering, Columbia University, New York, 2005.

[11] M.-Y. Kan, J. L. Klavans, and K. R. McKeown. Linear Segmentation and Segment Significance. page 9, sep 1998.

[12] M. Lin. for Lecture Videos : A Linguistics-Based Approach. *International Journal*, 1(June):27–45, 2005.

[13] H. Misra, F. Hopfgartner, A. Goyal, P. Punitha, and J. M. Jose. TV news story segmentation based on semantic coherence and content similarity. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5916 LNCS:347–357, 2009.

[14] N. O'hare, a.F. Smeaton, C. Czirjek, N. O'Connor, and N. Murphy. A generic news story segmentation system and its evaluation. *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3:iii–1028–31, 2004.

[15] G. J. Poulisse, M. F. Moens, T. Dekens, and K. Deschacht. News story segmentation in multiple modalities. *Multimedia Tools and Applications*, 48(1):3–22, 2010.

[16] J. C. Reynar and M. P. Marcus. *Topic Segmentation: Algorithms and Applications*. PhD thesis, University of Pennsylvania, US, 1998.

[17] L. Sitbon and P. Bellot. Topic segmentation using weighted lexical links (wll). In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 737–738, New York, NY, USA, 2007. ACM.

[18] M. Utiyama and H. Isahara. A statistical model for domain-independent text segmentation. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics - ACL '01*, pages 499–506, Morristown, NJ, USA, jul 2001. Association for Computational Linguistics.

[19] T. Zlitni, B. Bouaziz, and W. Mahdi. Automatic topics segmentation for TV news video using prior knowledge. *Multimedia Tools and Applications*, 2015.