# A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi D. Fohr D. Jouvet D. Langlois O.Mella K. Smaïli

16 Oct 2019

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

troduction

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis







#### Introduction

AMIS (Access to Multilingual Information and Opinions) **Objective 1**: Understanding the main idea of a video by summarizing.

- Input: A video in Arabic or French.
- Output: A summary of the input video subtitled in English.

Objective 2: Cross-lingual opinion Analysis

- Input: An Arabic video will be compared to French or English video.
- Output: A review concerning the degree of divergence between the two videos.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

Introduction

MIS corpus

Identifying comparable videos

Multilingual
Fine-granularity
sentiment analys

onclusion

4 ロ ト 4 同 ト 4 目 ト 4 同 ト 4 回 ト

#### **AMIS** corpus

A large corpus of videos is collected by the AMIS consortium.

Table: The number of videos per language

Language	Number of videos
English	1874
Arabic	1503
French	2046

- Several controversial Hashtags are used to retrieve videos (#Syria, #Animal-rights, #Trump, #Women's-rights, #Homosexual-marriage, ...etc).
- All the videos have been transcribed by our Arabic and English ASR systems depending on the language of the videos

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

# Identify comparable videos

- Dictionary-based approach.
- Word embedding approach.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

.....

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

#### Dictionary-based method

► This method consists in looking-up into a dictionary if the translation of the words of the source video  $V_s$  exist in the target one  $V_t$  and vice versa.

$$LG(V_s, V_t) = \frac{\sum_{w \in \{I_s \cap D_s\}} \sigma(w, I_t) + \sum_{w \in \{I_t \cap D_t\}} \sigma(w, I_s)}{|I_s \cap D_s| + |I_t \cap D_t|}$$
(1)

- Where D<sub>s</sub> is the source part (English) of the bilingual dictionary, D<sub>t</sub> is the target part (Arabic) of the dictionary.
- I<sub>s</sub> and I<sub>t</sub> are respectively the list of words of the source and the target video.

$$\sigma(\mathbf{w}, l_{s}) = \begin{cases} 1 & \text{if } T(\mathbf{w}) \cap l_{s} \neq \emptyset \\ 0 & \text{else} \end{cases}$$
 (2)

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

...........

AMIS corpus

Identifying comparable videos

Fine-granularity sentiment analysis

Conclusion

◆□▶◆□▶◆豆▶◆豆> 豆 めるぐ

# Dictionary-based method

- ► OMWN (Open Multilingual WordNet)<sup>1</sup> that contains 17,785 Arabic and English pairs.
- We used translation table of 297,176 pairs of Arabic and English entries.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

IIII Oddollon

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

<sup>1</sup>http://compling.hss.ntu.edu.sg/omw/ ←□ → ←□ → ← ■ → ← ■ → ● ■ → へ ≪

## Word embedding method

➤ The idea is to use the semantic information encoded by **words embedding** approach to retrieve the words semantically close to each other in two bilingual documents.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

Introduction

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

#### Word embedding method

The CBOW method is trained over a large parallel corpus (9 million sentences in English and Arabic).

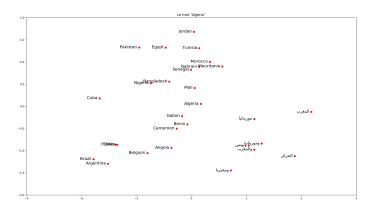


Figure: The correlated words with the entry Algeria.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

MIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analys

conclusion

#### Word embedding method

➤ To estimate the comparability between an Arabic and English videos.

$$CBOW(V_s, V_t) = \frac{\sum_{w \in \{l_s\}} \sigma(w, l_t) + \sum_{w \in \{l_t\}} \sigma(w, l_s)}{|l_s| + |l_t|}$$
(3)

In this case the function  $\sigma$  that returns 1 if the correlated words of w exist in list of a word of the source video.

$$\sigma(w, l_s) = \begin{cases} 1 & \text{if } CBOW(w) \cap l_s \neq \emptyset \\ 0 & \text{else} \end{cases}$$
 (4)

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

AMIS corpus

Identifying comparable videos

Fine-granularity sentiment analysis

#### Experimentation

- ► Test corpus composed of 123 pairs of comparable videos extracted from the Euronews web site.
- ► The classical measures in information retrieval topic: Recall (R@1, R@5 and R@10).

$$R@1 = \frac{\text{Number of the videos correctly identified in the position 1}}{\text{Total number of videos}}$$
(5)

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

Hilouuciion

MIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

onclusion

#### Experimentation

Table: The performance of different comparability methods in terms of R@1, R@5, and R@10 on a test corpus.

	R@1	R@5	<i>R</i> @10
DicMA	43	65	76
DicTT	70	90	92
CBOW	39	62	75

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analy

# Multilingual Fine-granularity sentiment analysis

- ► The methods devoted to sentiment analysis concern the polarity of text (positive, negative).
- ▶ In some other methods use fine-grained categories:
  - By using emotions (anger, disgust, fear, joy, sadness, and surprise)
  - By adopting a linguistic theory such as appraisal

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

troduction

AIS cornus

Identifying

Multilingual Fine-granularity sentiment analysis



#### The appraisal theory

13/31

The appraisal approach has been developed by White and Martin <sup>2</sup>.

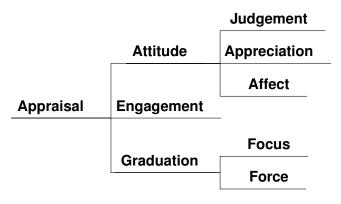


Figure: Appraisal taxonomy

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

muoduction

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

<sup>&</sup>lt;sup>2</sup>Martin, J. R. and White, P. R. R., "The language of evaluation: appraisal in English", 2005

#### Attitude category

The category Attitude gives the type of appraisal being expressed. It has multiple subcategories:

- ▶ Affect. describes the emotional reactions (happy, miserable, angry, etc.).
- Appreciation. It concerns the opinion that a person has about the inner or outer qualities of an object (beautiful, innovative, amazing, etc.).
- ▶ Judgment. describes the behaviour of somebody in a social context (lucky, brave, famous, etc.).

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

MIS corpus

Identifying

Multilingual Fine-granularity sentiment analysis

onclusion

#### Graduation category

► The Graduation describes the intensity of appraisal. It has two sub-category: focus and force.

The graduation is globally expressed via modifiers.

Word: " intelligent " Polarity: positive

Attitude category: judgement

Force: neutral

Text: " Very+ intelligent "

Polarity: positive

Attitude category: judgement

Force: Intensify

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

AMIS corpus

Identifying comparable video

Multilingual Fine-granularity sentiment analysis



#### Building appraisal lexicon

- ➤ To build the appraisal lexicon, we started from an English opinion lexicon composed of 4,913 negative words and 2,718 positive words developed by Minqing Hu and Bing Liu.
- We created a list of 363 (MW363) words with their appraisal categories inspired from the examples of Martin and White's book.

Table: Few examples of words with appraisal and polarity opinions

Word	Attitude sub-category	Polarity
Lucky	Judgment	Positive
Obscure	Judgment	Negative
Confident	Affect	Positive
Love	Affect	Positive
Helpful	Appreciation	Positive

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

AMIS cornus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis



## Building appraisal lexicon

- ➤ We assigned for each entry of the Minqing Hu and Bing Liu lexicon the corresponding appraisal Attitude category (word embedding, MW363).
- ► For each word X from Minqing Hu and Bing Liu lexicon, we find its top-n closest words to MW363.
- ► Then, we assign to X the sub-category which is dominant in this latter list.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

iiiii oddolloii

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

#### Example

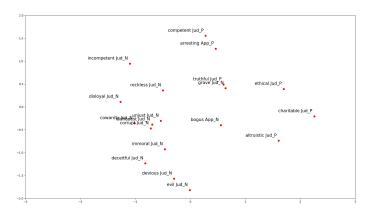


Figure: The correlated appraisal words of the MW363 with the entry *criminal*.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

MIS corpus

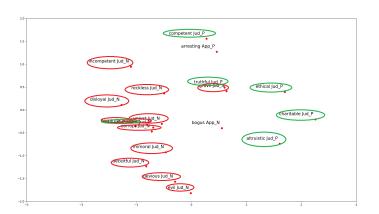
Identifying comparable videos

Multilingual Fine-granularity sentiment analysis



#### Example

► This word is close to: 11 judgment negative words, 6 positive judgment words, one positive appreciation and one negative appreciation.



A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application Arabic and English Videos

D. Jouvet, D. Langlois, O.Mella, K. Smaïli

K. Abidi, D. Fohr,

ntroduction

AIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

conclusion

Figure: The correlated appraisal words of the MW363 with the entry *criminal*.

## Building appraisal lexicon

$$S_{App}(X) = \frac{1}{d_n} \sum_{i=1}^{d_n} cosine(X, W_i) * P_{W_i}$$
 (6)

#### Where:

- d<sub>n</sub>: The number of words in the dominant sub-category in the list of the n closest words with X.
- W<sub>i</sub>: A word belonging to the list of the dominant attitude sub-category.
- ➤ X : A word of Bing Liu's sentiment lexicon.
- $P_{W_i} = \begin{cases} +1 & \text{if } W_i \text{ is positive} \\ -1 & \text{otherwise.} \end{cases}$

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

IIIIOddction

MIS corpus

Identifying

Multilingual Fine-granularity sentiment analysis



# Building appraisal lexicon

Table: Few examples of BingApp.

English word	Arabic trans	App-cat /Sub-cat	S <sub>App</sub>	Pol
Criminal	مجرم	Attitude/Judgment	-0.45	N
Attentive	منتبه	Attitude/Judgment	0.41	Р
Worried	قلق	Attitude/Affect	-0.45	N
Satisfied	زاض	Attitude/Affect	0.24	Р
Harmonious	متناغم	Attitude/Appreciation	0.63	Р

In order to work with the same material in Arabic and in English, we translated BingApp into Arabic.

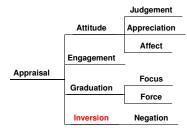
A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet. D. Langlois, O.Mella, K. Smaïli

Multilingual Fine-granularity sentiment analysis

# Dealing with the Negation

We added to BingApp, the negation words (Not, No, Neither, Nor, etc.) and assigned them to the Inversion category.



During the analysis step, if the *Inversion* category is identified in an utterance, then the polarity of the word following the negation item is inverted. A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

MIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis



## Dealing with the Force

- We added to the dictionary BingApp several modifier words that we assigned to the sub-category Force of the category Graduation.
- We shared these modifiers through 4 classes. Each class indicates the intensity of the modifier and it is assigned a score proportional to its capacity to intensify a word.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

Hiloduction

MIS corpus

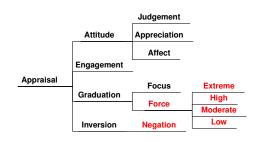
Identifying comparable video

Multilingual Fine-granularity sentiment analysis

#### Frame Title

Table: The four classes of the intensity modifiers.

Force classes	Modifiers	
Extreme	hardly, scarcely, barely, very, greatly, etc.	
High	large, less, distant, more, etc.	
Moderate	somewhat, relatively, rather, reasonably, many, etc.	
Low	slightly, least, small, etc.	



A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

troduction

Coorpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis



#### Experimentation

We used the publicly available collection of movie reviews which consists of 1,000 positive and 1,000 negative reviews.

Table: Comparison of Bing Liu's lexicon and BingApp on a test Review corpus

Method	Recall	Precision
Bing Liu's	69.0	68.8
BingApp	70.9	71.0

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

itioddddioii

/IS corpus

ldentifying comparable vide

Multilingual Fine-granularity sentiment analysis

#### **Evaluation on AMIS videos**

- Quantitative evaluation
- Qualitative evaluation

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

introduction

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis



#### Qualitative evaluation

A qualitative evaluation in which we summarize the expressed opinion in the video by using a template of the opinion review. A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

The sentiment of the video is positive with a score  $[X_p]$  and negative with a score  $[X_n]$ .  $X_{aff}$ % of the video concerns emotional reactions.  $X_{Jug}$ % of the video concerns the human behaviour according to social norms and 1 –  $(X_{aff} + X_{Jug})$ % of the video is about the appreciation of no human being entities.

ntroduction

. . . .

....

comparable video

Multilingual Fine-granularity sentiment analysis

Conclusion

Figure: The template used to generate the qualitative evaluation

#### Example

28/31\_\_.



این الاطلاب الاطلاق الفرود و المطالب المساولة الم القال أما القال الما الما القال المرافق الدرا في الدرا في الدرا الما الما الدرا في الدر في

#### Score positif = 1,56, Score négatif = -2,15

The sentiment of the video is positive with a score 1.461 and negative with a score -2.15. 50% of the video concerns the human behaviour according to social norms and 50% of the video is about the appreciation of no human being entities.



#### What will Trump's prosidency mean for the Iran nuclear deal?

the what is will dentify tray a precluding years for the rise marked with he a compact the many and what he are and several plush prevers on a visit by page are as foreign measure and the expected the new advantages to accept the agreement in on his broad and a second of the contraction of the agreement in the first proprieted as a part of the bragan has the white that proper price is a constantive growth on the first proprieted and part of the bragan has the white that proper price is a constantive generate pand for everybody that so are printenee doors to treat them a proper price is a constantive generate pand for everybody that so are printenee doors to treat them a proper price is a constantive generate price of everybody that so are printenee doors to treat them as expected after the first price are segmented by my decided to advantage to so self reviet the agreement advantage is as for size and the treat that it is correct in a page and any lay of the great price product in adultation as the contraction of the contraction of the second price of the second price price of the second p

#### Score positif = 1,42, Score négatif = -3,275

The sentiment of the video is positive with a score **1.425** and negative with a score **-3.275**. **7.14**% of the video concerns emotional reactions. **78.57**% of the video concerns the human behaviour according to social norms and **14.29**% of the video is about the appreciation of no human being entities.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

ntroduction

AMIS corpus

comparable videos

Multilingual Fine-granularity sentiment analysis

Conclusion

#### Conclusion

In this work the objective was twofold:

- Identify comparable videos :
  - We tested three methods and compared them.
  - The best one has been used to align the whole database of AMIS.

#### Multilingual Fine-granularity sentiment analysis :

- We used a new method based on the appraisal approach allowing to have a fine-grained opinion analysis.
- We created a new lexicon which contains words with their appraisal category.
- This lexicon served to evaluate quantitatively and qualitatively the content of videos.

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet. D. Langlois, O.Mella, K. Smaïli



# Thank You

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet, D. Langlois, O.Mella, K. Smaïli

AMIS corpus

Identifying comparable videos

Multilingual Fine-granularity sentiment analysis

#### Quantitative evaluation

$$S = \sum_{i=1}^{N} \alpha(\mathbf{w}_{i-k}^{i-1}) * S_{App}(\mathbf{w}_i)$$
 (7)

- Where N is the size of the video in terms of number of words.
- $ightharpoonup \alpha$  is a weight depending on the *Inverted* or the *Force* sub-category of the k words preceding the word  $w_i$ (k is set to 2)

A Fine-grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos

K. Abidi, D. Fohr, D. Jouvet. D. Langlois, O.Mella, K. Smaïli