Semantic structuring of video collections from speech: segmentation and hyperlinking

Anca Şimon

PhD advisors:

Pascale Sébillot & Guillaume Gravier

INSA de Rennes CNRS

LINKMEDIA research team

Inria INVENTEURS DU MONDE NUMÉRIQUE UMR IRISA INSA RENNES cnrs

General context

- Size of world electronic data in 2013: 4.4 ZB (IDC report)
  ~ 9 ZB in 2015

… …

General context

➢ Size of world electronic data in 2013: 4.4 ZB (IDC report)
~ 9 ZB in 2015

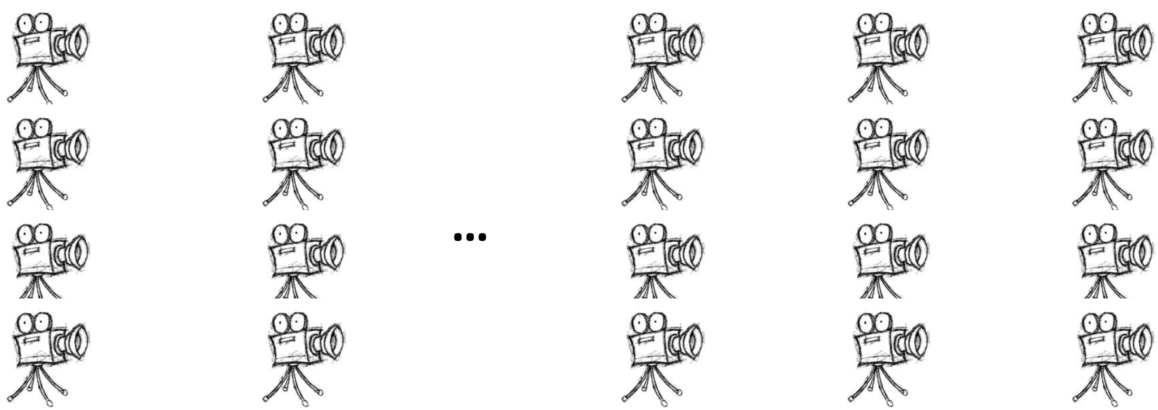
Semantic structuring of video collections from speech: segmentation and hyperlinking

Anca Şimon

PhD advisors:
Pascale Sébillot & Guillaume Gravier
INSA de Rennes CNRS
LINKMEDIA research team

2500 BC
Pyramids of Giza

2015
The day of the defense



~ 2,600 cameras filming 24/7

General context

- Size of world electronic data in 2013: 4.4 ZB (IDC report)
  ~ 9 ZB in 2015

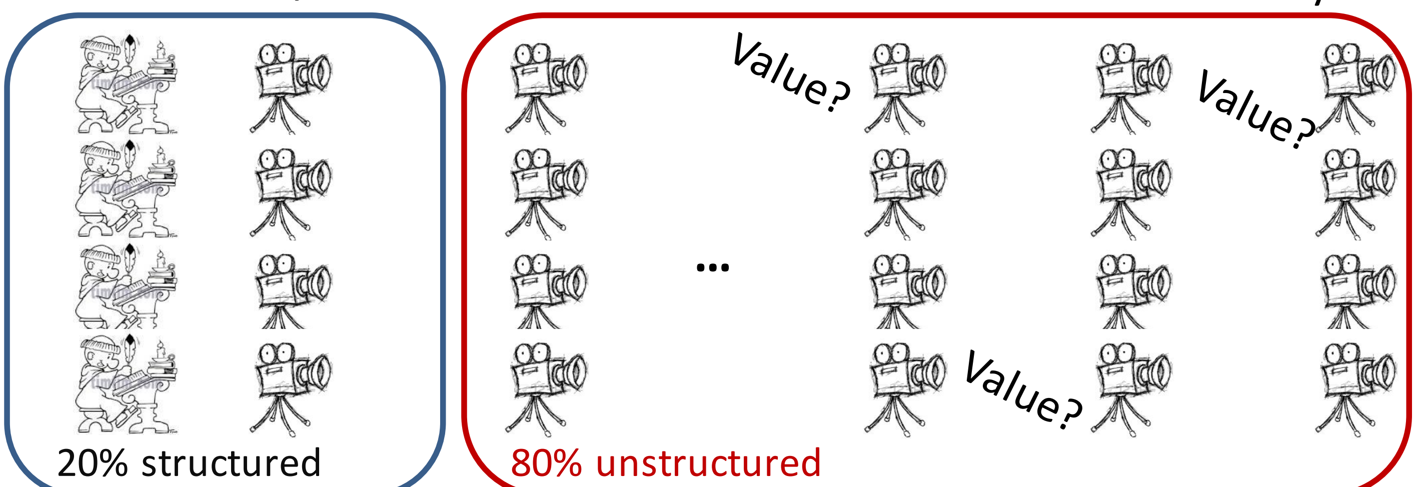

Semantic structuring of video collections from speech: segmentation and hyperlinking

Anca Şimon

PhD advisors:
Pascale Sébillot & Guillaume Gravier
INSA de Rennes CNRS
LINKMEDIA research team

2500 BC
Pyramids of Giza

2015
The day of the defense

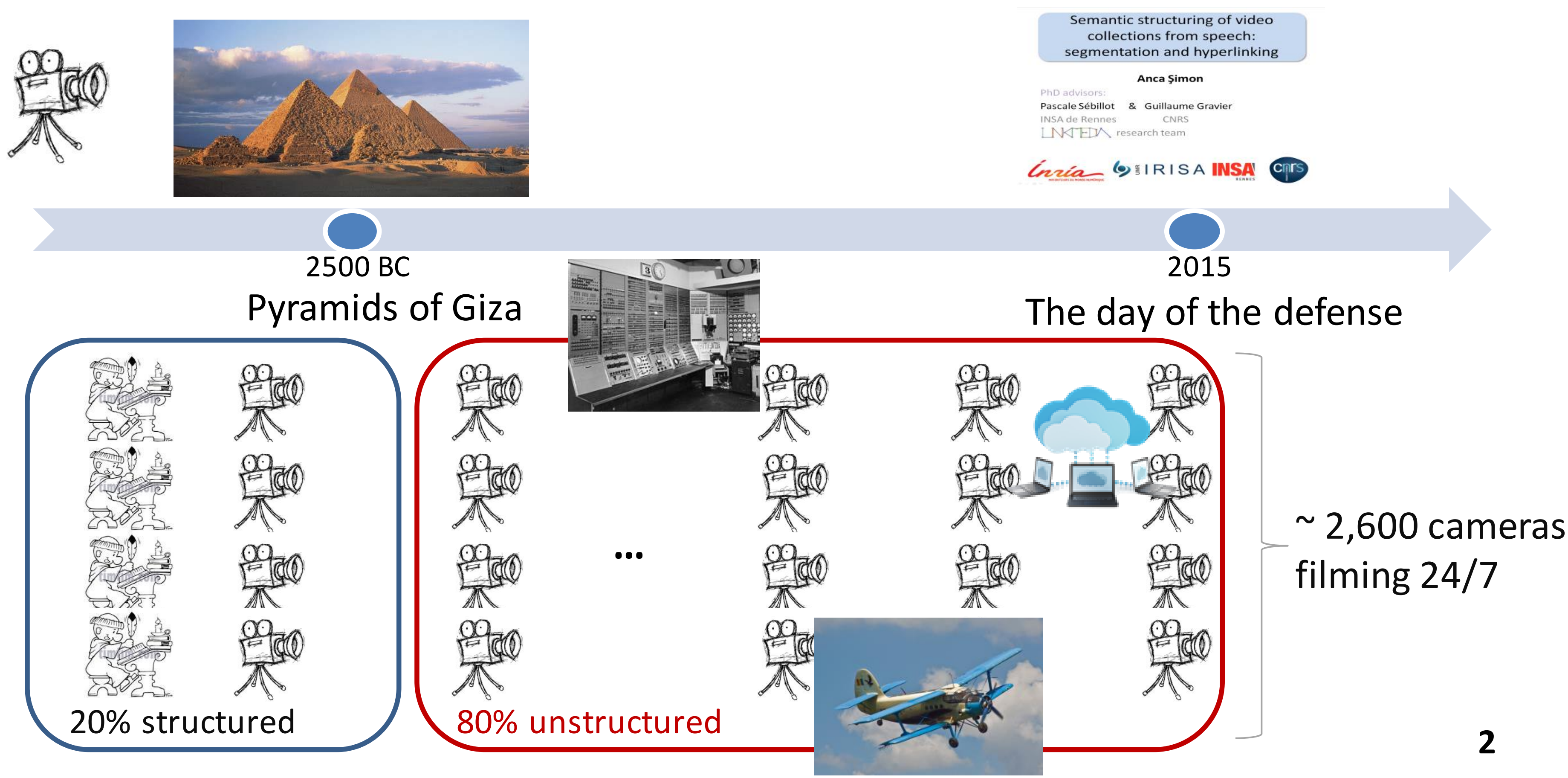


~ 2,600 cameras filming 24/7

General context

- Size of world electronic data in 2013: 4.4 ZB (IDC report)

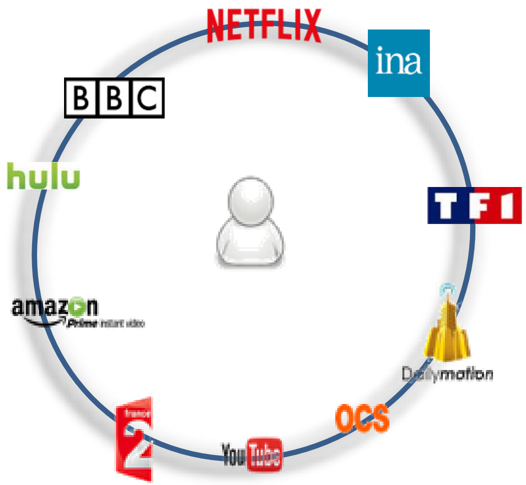
  ~ 9 ZB in 2015



~ 90% of the internet traffic is video data

DIGITAL UNIVERSE

Audiovisual landscape



... INA archive > 5 million hours of programs;

Youtube > 300 hours of videos/minute;

Netflix subscribers > 60 million;

98.3% of French households have at least 1 TV ...

Watch what we want, when we want , on whatever device we want

Challenges:

- user centric model
- unstructured data
- heterogeneous content

Motivating examples

➢ Have access to points of interest in a video

…


Semantic structuring of video collections from speech: segmentation and hyperlinking

Anca Şimon

PhD advisors:

Pascale Sébillot & Guillaume Gravier

INSA de Rennes CNRS

LINKMEDIA research team

2500 BC
Pyramids of Giza

80
Collosseum

1990
Hubble telescope

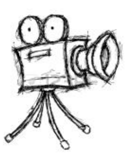
2015
The day of the defense

Motivating examples

- Study how a topic is presented by different TV shows

2500 BC
Pyramids of Giza

2015
The day of the defense

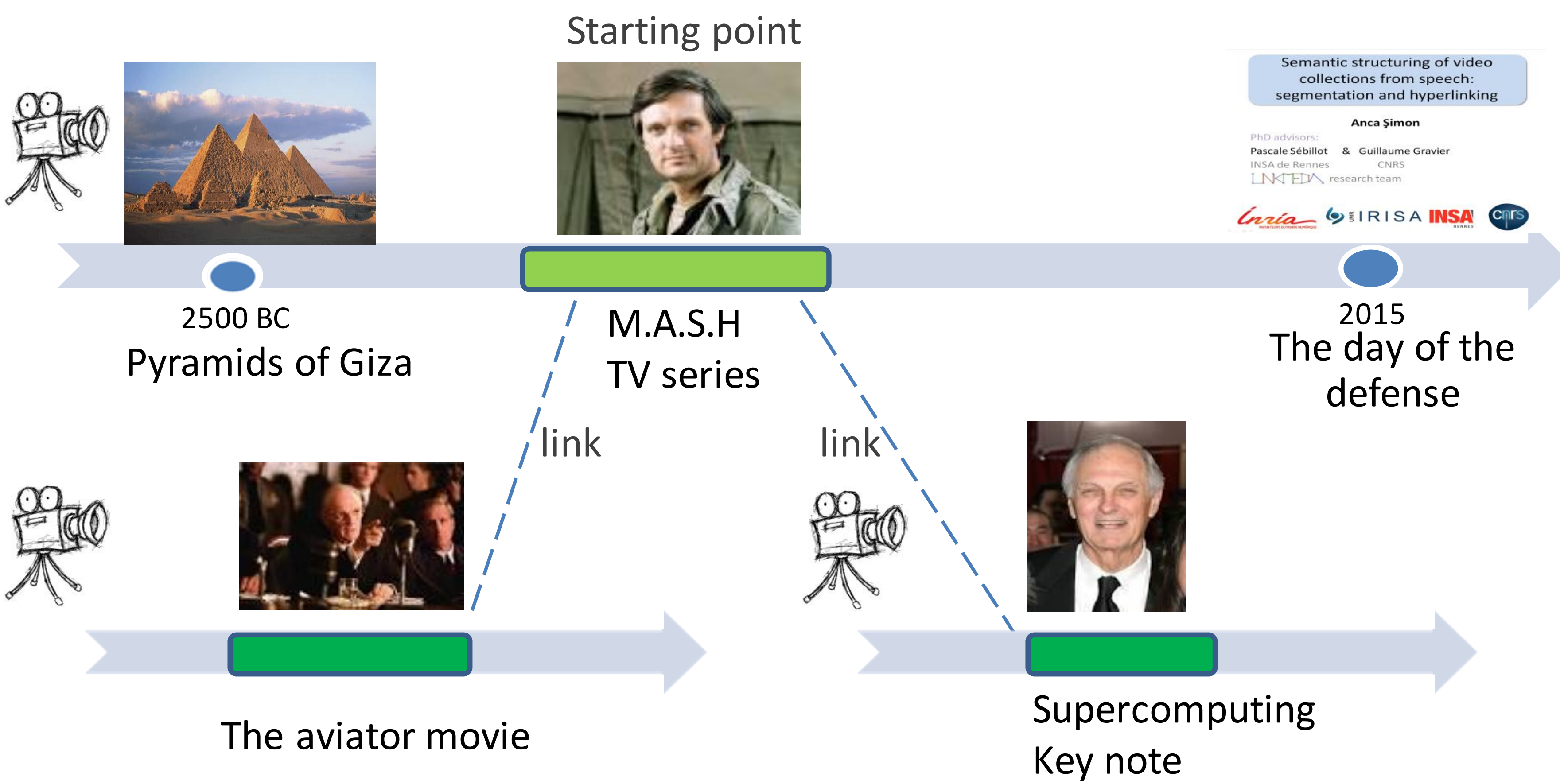


2500 BC
Pyramids of Giza

2015
The day of the defense

Motivating examples

➢ Discover *interesting* and *unexpected* information starting from a video fragment



Research questions

1. How to *structure* audiovisual content?

Research questions

1. How to *structure* audiovisual content?

Provide automatic and generic techniques for *topical structuring* of TV shows.

- challenging data: automatic TV show transcripts (ASR system)

Research questions

1. How to *structure* audiovisual content?

Provide automatic and generic techniques for *topical structuring* of TV shows.

- challenging data: automatic TV show transcripts (ASR system)

2. How to *exploit* structured content?

Research questions

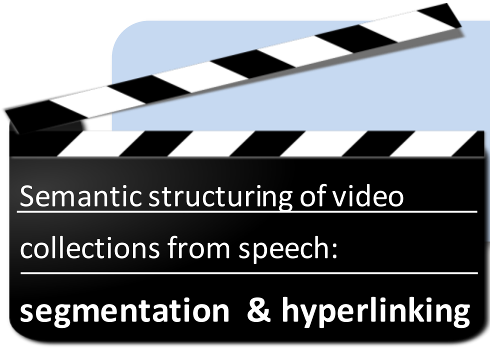
1. How to *structure* audiovisual content?

Provide automatic and generic techniques for *topical structuring* of TV shows.

- challenging data: automatic TV show transcripts (ASR system)

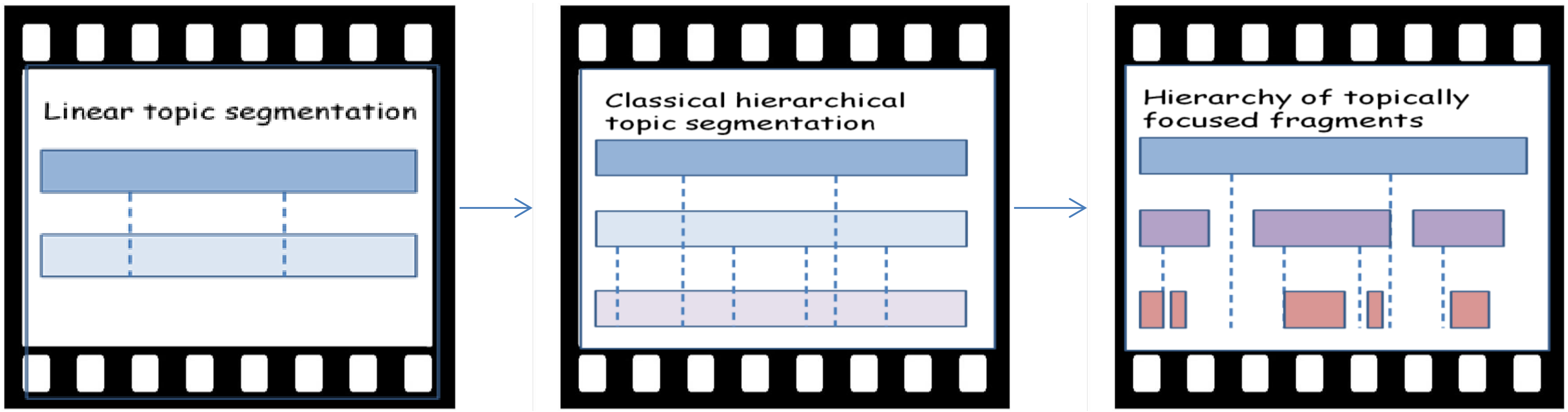
2. How to *exploit* structured content?

Study the implications of the topical structure in the context of *video hyperlinking.*

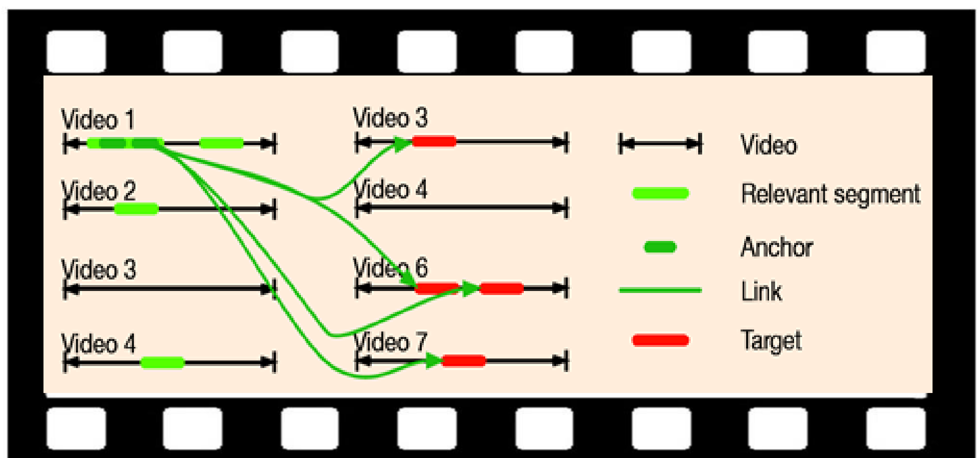


Outline

Thesis contributions in a nutshell



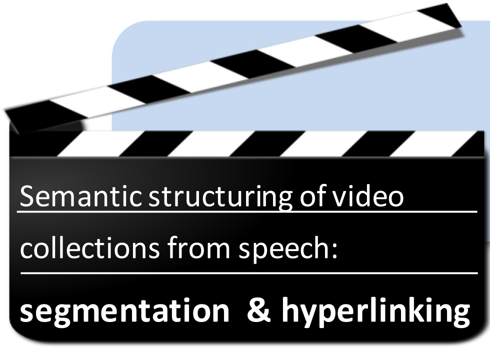
- Anchor and target generation
- Link justification & diversity control

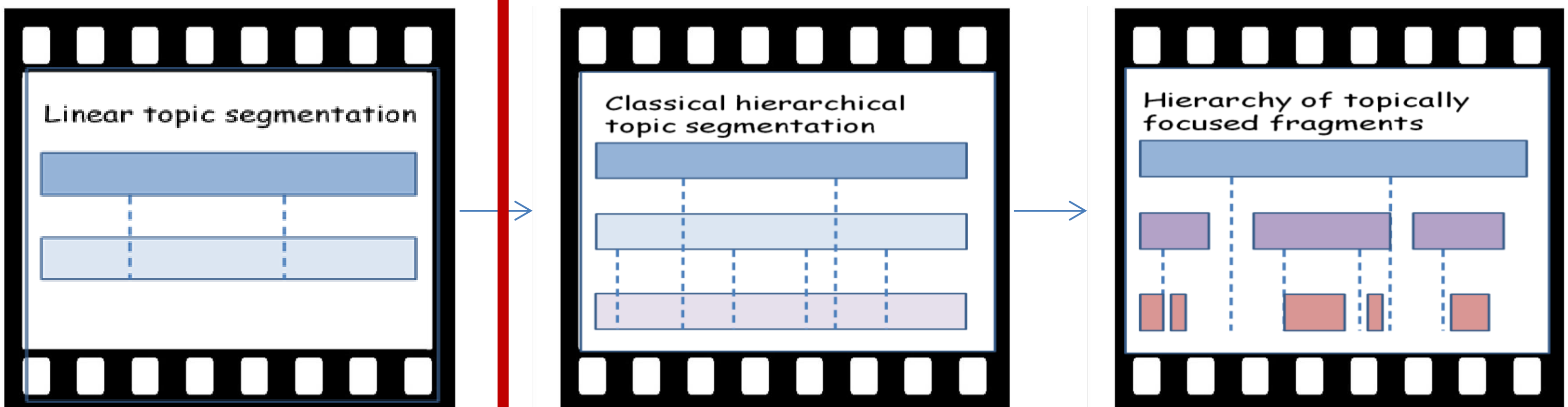


MediaEval benchmark initiative
Search & Anchoring & Hyperlinking

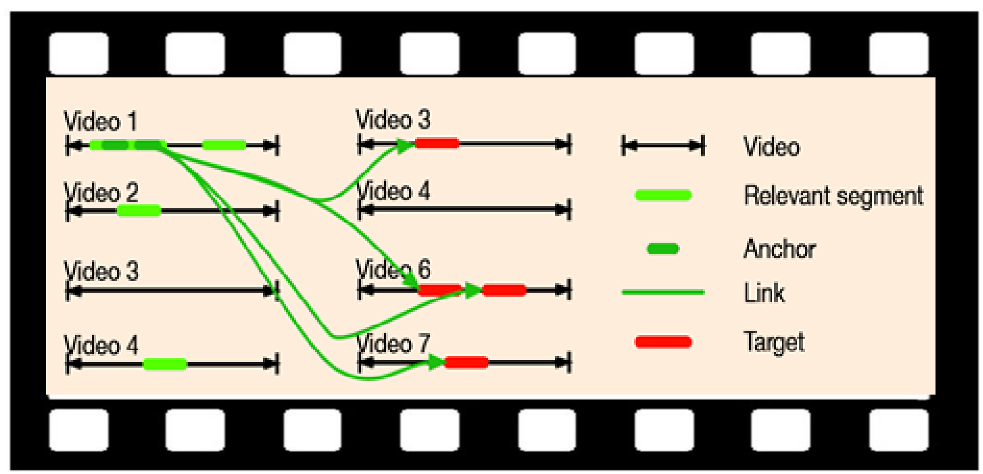
Outline

Thesis contributions in a nutshell





- Anchor and target generation
- Link justification & diversity control



MediaEval benchmark initiative

Search & Anchoring & Hyperlinking

Linear topic segmentation

Divide data into topically coherent segments.

TV show

ASR: textual transcripts of the speech

...	La France de la débrouille	Le soutien scolaire privé	...

<u>Difficulties</u>:

- automatic transcripts ≠ written text
- subjectivity of the concept of topic
- evaluation

<u>Objective</u>:

- provide a solution for topic segmentation that is:
  - \+ generic
  - \+ robust

Topic segmentation
-lexical cohesion-based techniques-

- Exploit words distributions or lexical chains (Hearst 1997, Morris and Hirst 1991)

Key notion: significant change in vocabulary → topic change

1. Local methods: locally detecting the *lexical disrupture*

(Hearst 1997, Hernandez et al. 2002, Ferret et al. 1998, Claveau et al. 2011)

- Drawbacks: selecting the window size; choosing the threshold to decide if a frontier should be placed;

2. Global methods: globally measuring the *lexical cohesion*

(Choi 2000, Reynar 1994, Utiyama et al. 2001, Eisenstein et al. 2008)

- Drawbacks: potential oversegmentation; need the number of segments a priori;

Topic segmentation
lexical cohesion-based techniques

- Exploit words distributions or lexical chains (Hearst 1997, Morris and Hirst 1991)

Key notion: significant change in vocabulary → topic change

1. Local methods: locally detecting the *lexical disrupture*

(Hearst 1997, Hernandez et al. 2002, Ferret et al. 1998, Claveau et al. 2011)

- Drawbacks: selecting the window size; choosing the threshold to decide if a frontier should be placed;

2. Global methods: globally measuring the *lexical cohesion*

(Choi 2000, Reynar 1994, Utiyama et al. 2001, Eisenstein et al. 2008)

- Drawbacks: potential oversegmentation; need the number of segments a priori;

Can they be reconciled?

Reconciling lexical cohesion & disrupture

Propose :

1. A segmentation criterion that combines both cohesion and disrupture
2. The corresponding algorithm for topic segmentation

(similar concept: Malioutov and Barzilay, 2006)

Reconciling lexical cohesion & disrupture

Propose :

1. A segmentation criterion that combines both cohesion and disrupture
2. The corresponding algorithm for topic segmentation

(similar concept: Malioutov and Barzilay, 2006)

Starting point: Utiyama and Isahara (2001) global algorithm TextSeg

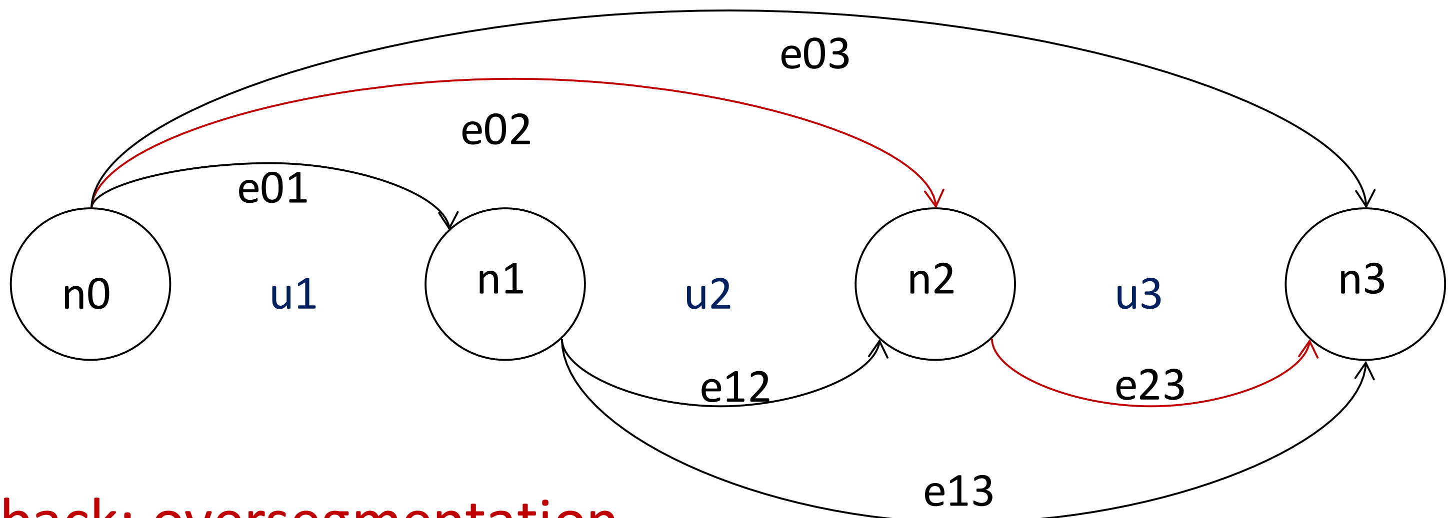
- State-of-the-art
- Domain independent
- Can deal with topical segments of highly varying lengths
- Provides an efficient graph based implementation

Statistical model TextSeg

Find the most probable segmentation among all possible ones, assuming that segments are mutually independent:

$$\hat{S} = arg \max_{S} \sum_{i=1}^{m} \ln(P[W_i|S_i]) - \alpha ln(n)$$

Probabilistic graph-based segmentation:



Drawback: oversegmentation

Introduction of the lexical disruption
MSeg

Assume a Markovian hypothesis between the segments in order to take into account, for each segment, the previous one:

$$\hat{S} = arg \max_{S} \sum_{i=1}^{m} \ln(P[W_i|S_i]) - \lambda \sum_{2}^{m} \Delta(W_i, W_{i-1}) - \alpha ln(n)$$

Disruption computation: Δ

- Cosine similarity, cross probabilities ($P[W_i \mid S_{i-1}]$ and $P[W_{i-1} \mid S_i]$)
- Weights: TF-IDF, Okapi

Experiments

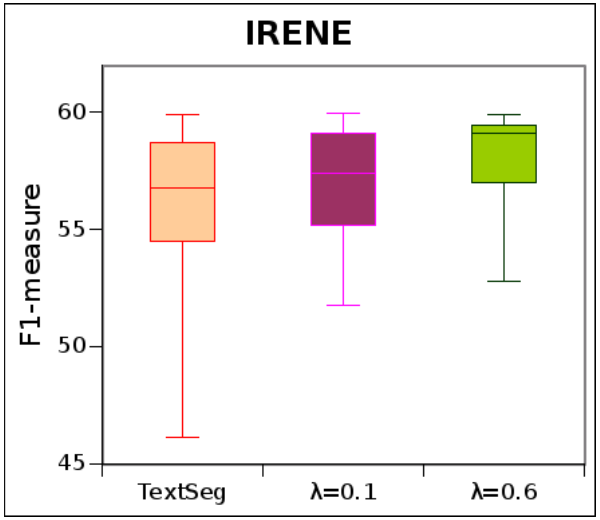
Corpora

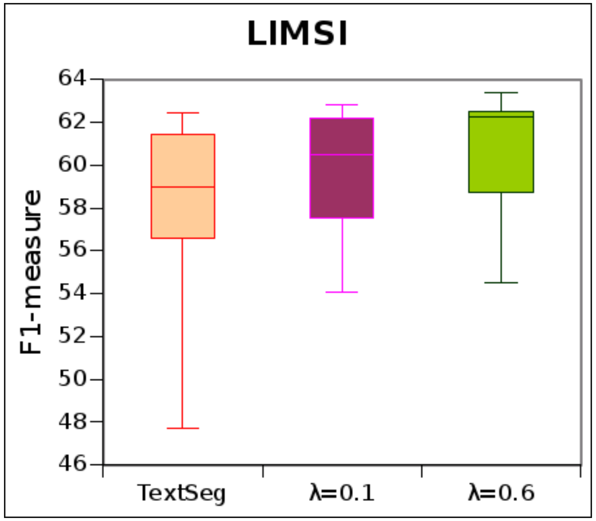
1. TV news transcripts (IRENE and LIMSI ASR systems)
    - 56 news programs (~1/2 hour each, reports duration ~ 2-3 min.)
    - Reduced number of word repetitions
    - IRENE has WER higher that that of LIMSI by ~ 6 points
    - TreeTagger: data lemmatized
    - Groundthruth: manual annotation
2. Choi's artificial data set
3. Medical textbook

Evaluation

- Recall, precision, F1-measure
- Tolerance: 10 sec.

Results: TextSeg vs. MSeg

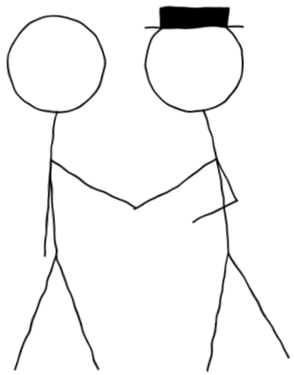




Corpus	F1 gain	Confidence interval 95% TextSeg (λ=0)	MSeg (λ≠0)
IRENE (WER 36%)	0.3	[54.4,57.6]	[56.92,59]
LIMSI (WER 30%)	**0.86**	[56.7,60.2]	[59.44,61.95]
REFERENCE (6)	**0.77**	[70.39,72.29]	[71.7,73.29]
IRENE(6)	0.2	[56.81,60.94]	[59.51,63.43]
LIMSI(6)	**0.5**	[64.27,68.64]	[67.7,71.56]

λ is the importance given to the disruption
α controls the contribution of the prior model

Lessons learned

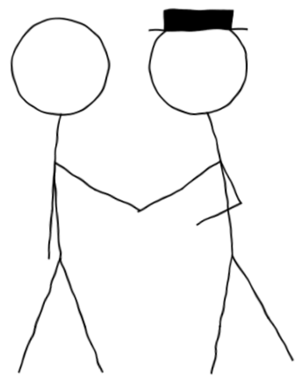


Lexical cohesion

Lexical disruption

- overcome challenges characteristic to local and global methods
- diminish the influence of the prior model
- eliminate wrong hypothesis
- impact of disruption is bigger on longer segments
- automatic transcripts ≠ written text
- automatic transcripts ≠ manual transcripts
- deal with abrupt vs. smooth topic changes
- BoW model looses semantic information

Lessons learned

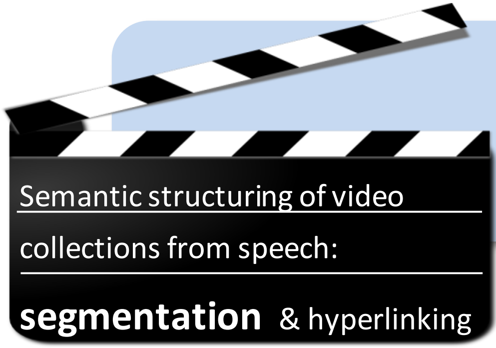


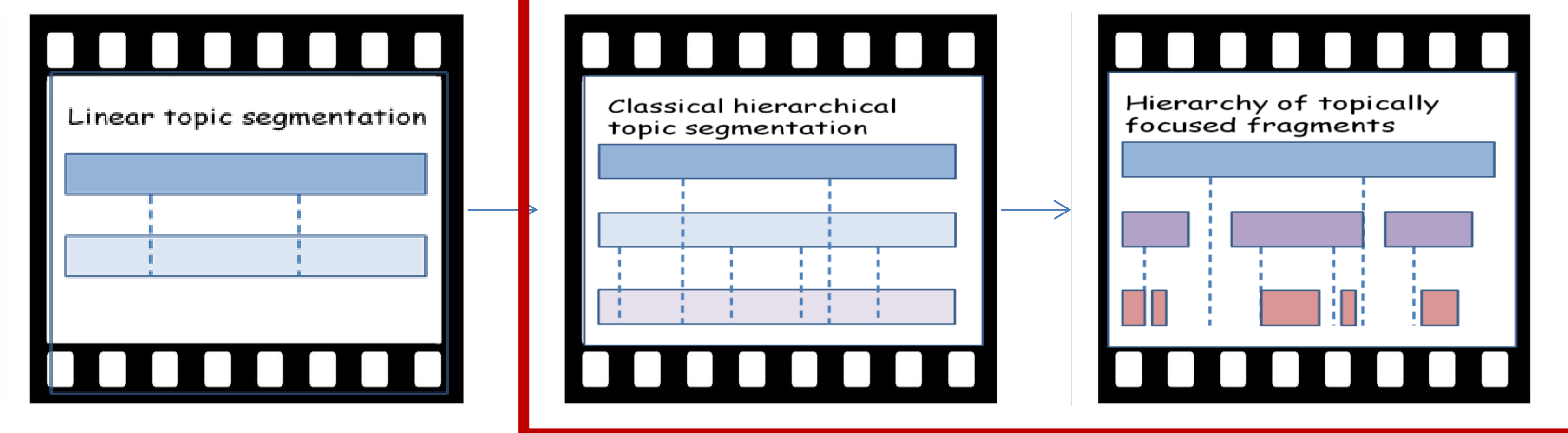
Lexical cohesion

Lexical disruption

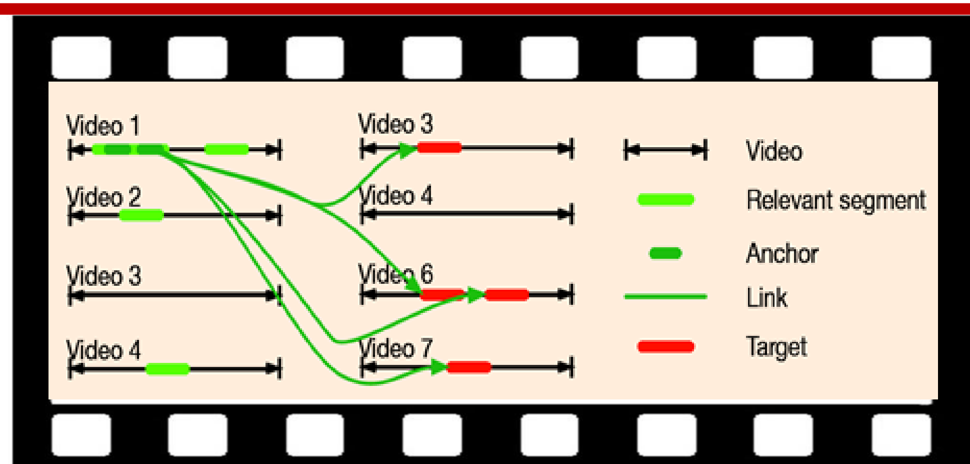
- overcome challenges characteristic to local and global methods
- diminish the influence of the prior model
- eliminate wrong hypothesis
- impact of disruption is bigger on longer segments
- automatic transcripts ≠ written text
- automatic transcripts ≠ manual transcripts
- deal with abrupt vs. smooth topic changes
- BoW model looses semantic information

Outline
-Thesis contributions in a nutshell-





- Anchor and target generation
- Link justification & diversity control

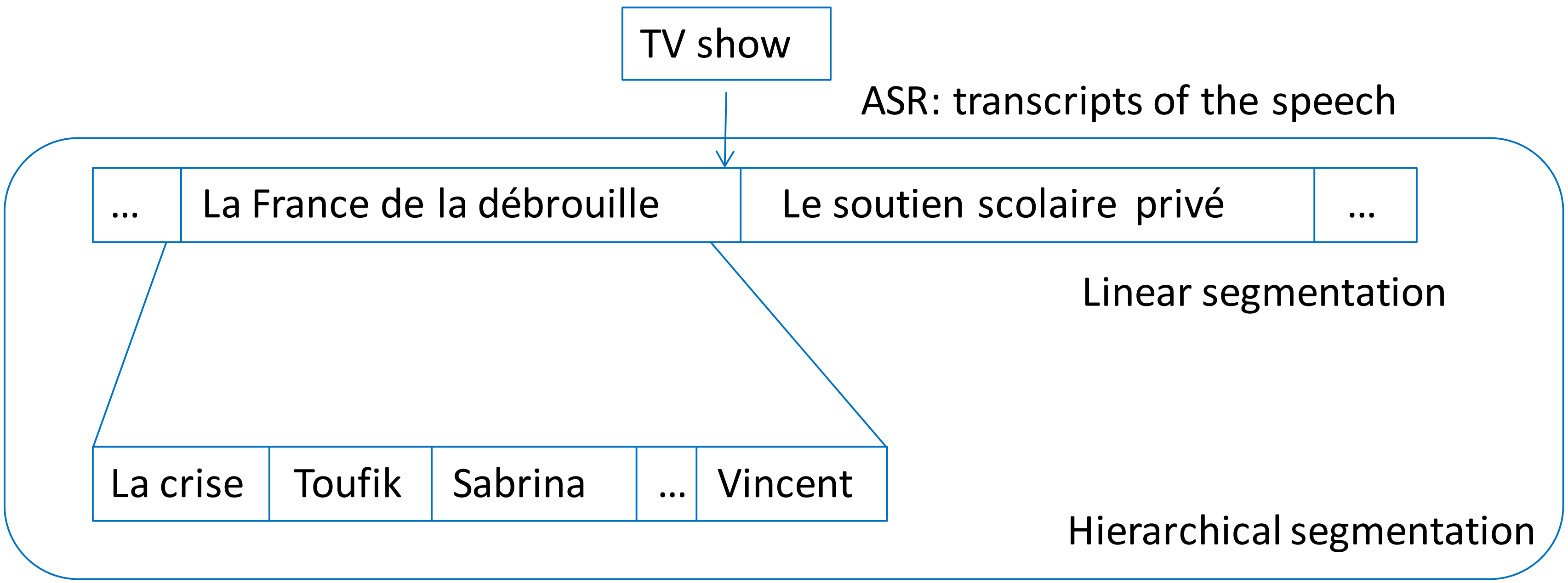


MediaEval benchmark initiative
Search & Anchoring & Hyperlinking

Hierarchical topic segmentation

Discourse structure often displays a hierarchical form

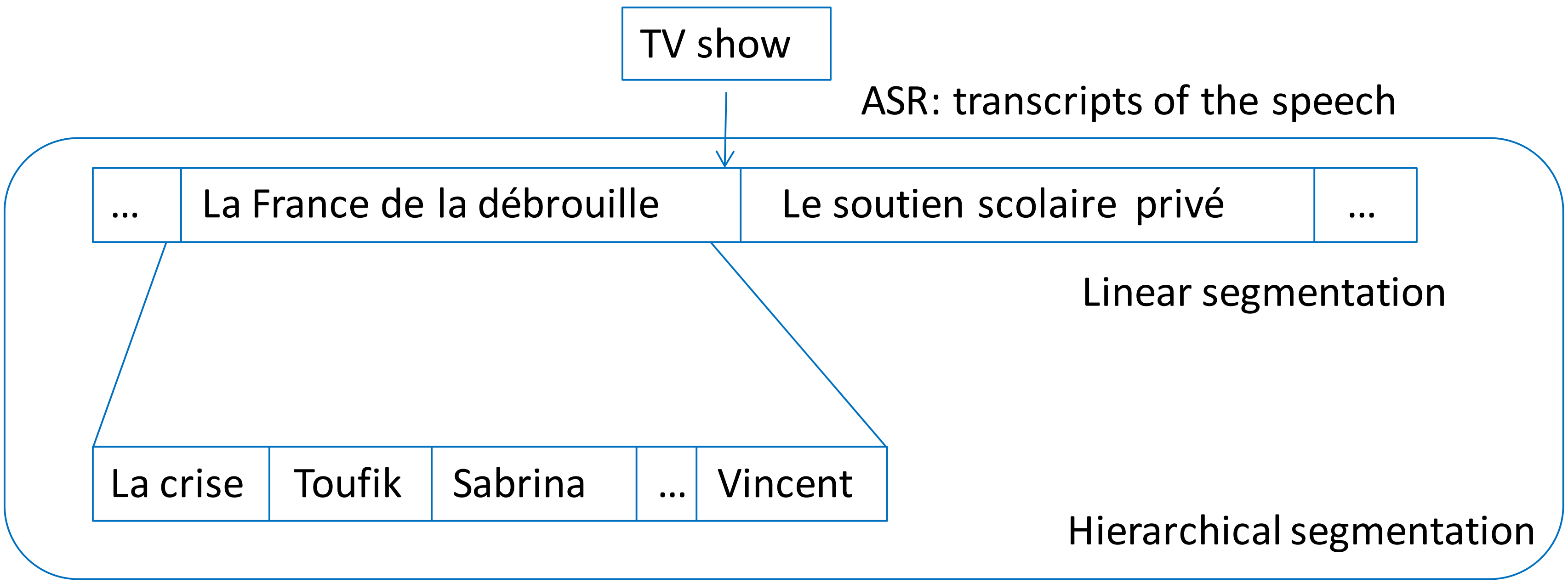
(Grosz and Sidner 1986, Eisenstein 2009, Carroll 2010, etc.)



Hierarchical topic segmentation

Discourse structure often displays a hierarchical form

(Grosz and Sidner 1986, Eisenstein 2009, Carroll 2010, etc.)



<u>Difficulties</u>:
- automatic transcripts ≠ written text
- number of words available
- subjectivity of the concept of topic and sub-topic
- evaluation

Existing solutions for hierarchical segmentation

1. Recursive application of a linear segmentation technique

(Guinaudeau 2011, Carroll 2010)

- Drawbacks: decide when to stop; errors from one level get propagated to another one

2. Obtain directly the hierarchical structure

(Moens and Busser 2001, Eisenstein 2009, Kazantseva, 2014)

- Drawbacks: need information about the granularity level; expected segment durations

Existing solutions for hierarchical segmentation

1. Recursive application of a linear segmentation technique

(Guinaudeau 2011, Carroll 2010)

- Drawbacks: decide when to stop; errors from one level get propagated to another one

2. Obtain directly the hierarchical structure

(Moens and Busser 2001, Eisenstein 2009, Kazantseva, 2014)

- Drawbacks: need information about the granularity level; expected segment durations

How well do they work?

Classical measures have limitations

Segmentation results

Method	F1-measure					
	TV shows				Wikipedia	
	Manual (4)		Automatic (7)		(66 articles)	
	coarse	fine	coarse	fine	coarse	fine
Eisenstein	100	**28.3**	100	**21.2**	18.15	**27.94**
(recursive) TextSeg	100	**30.6**	95.24	**27.11**	33.6	**37.7**
(recursive) MSeg	100	**31**	95.24	**27.47**	33.6	**40.2**

Classical measures have limitations

Segmentation results

Method	F1-measure					
	TV shows				Wikipedia	
	Manual (4)		Automatic (7)		(66 articles)	
	coarse	fine	coarse	fine	coarse	fine
Eisenstein	100	**28.3**	100	**21.2**	18.15	**27.94**
(recursive) TextSeg	100	**30.6**	95.24	**27.11**	33.6	**37.7**
(recursive) MSeg	100	**31**	95.24	**27.47**	33.6	**40.2**

Need something new…

➢leverage the **burstiness** phenomenon in word occurrences:

if a word appears once it is more likely to appear again, instead of independently

(Rasmus, 2005)

Proposed approach

1) Leverage the burstiness phenomenon in word occurrences

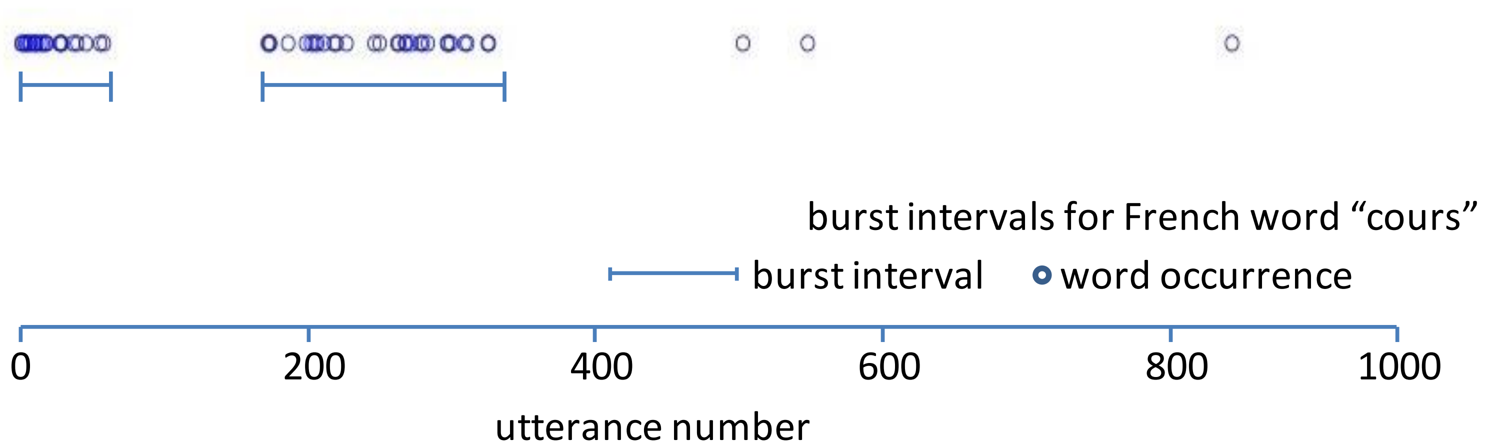
- Bursty words: characterized by long inter-arrival times followed by short inter-arrival times;
- Non-bursty words: exhibit inter-arrival times with smaller variance.

Proposed approach

1) Leverage the burstiness phenomenon in word occurrences

- Bursty words: characterized by long inter-arrival times followed by short inter-arrival times;
- Non-bursty words: exhibit inter-arrival times with smaller variance.

✓Starting point: Kleinberg's algorithm (Kleinberg, 2002)

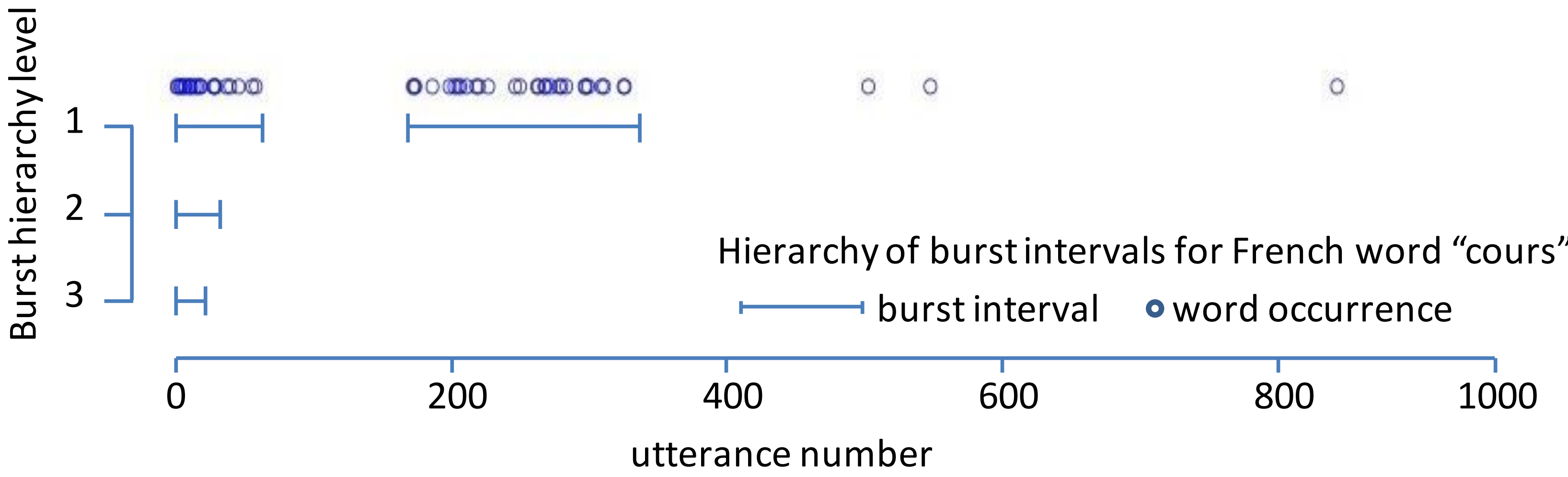


Proposed approach

1) Leverage the burstiness phenomenon in word occurrences

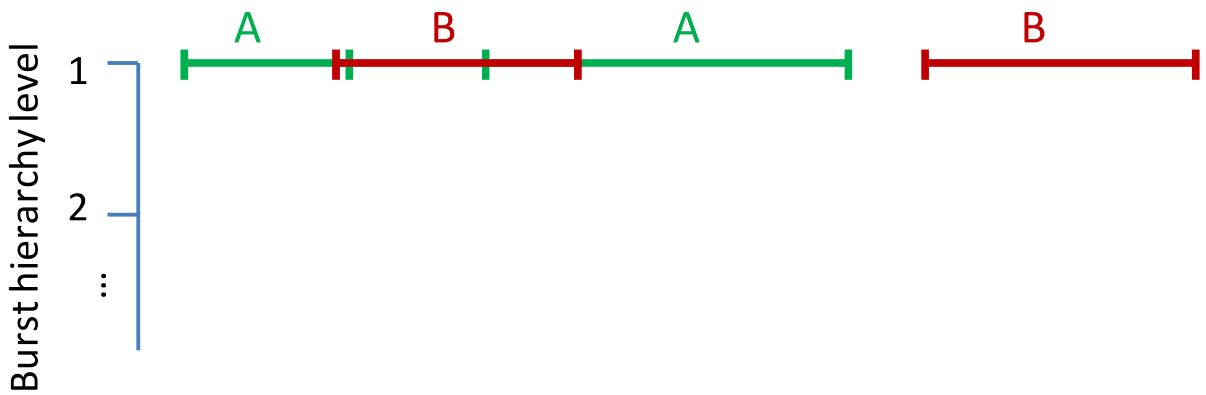
- Bursty words: characterized by long inter-arrival times followed by short inter-arrival times;
- Non-bursty words: exhibit inter-arrival times with smaller variance.

✓Starting point: Kleinberg's algorithm (Kleinberg, 2002)



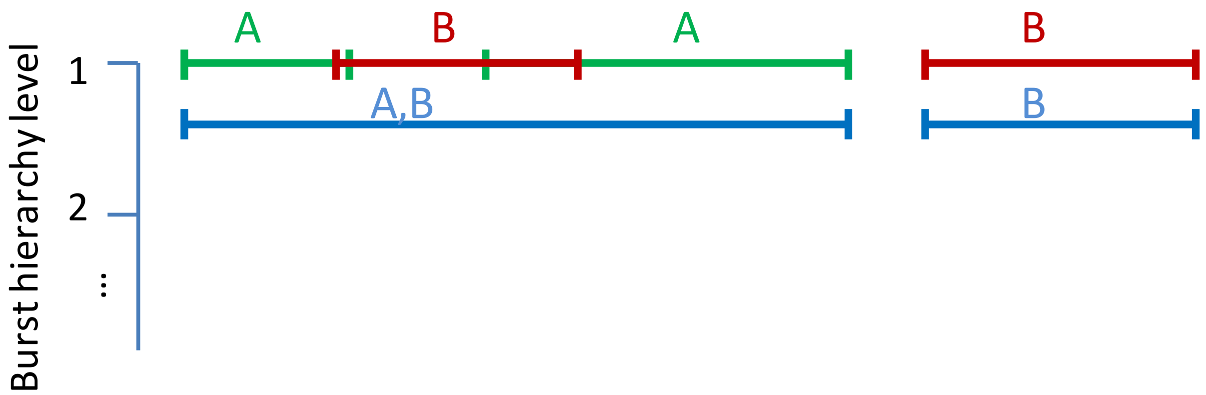
Proposed approach

2) Agglomerative clustering of burst intervals



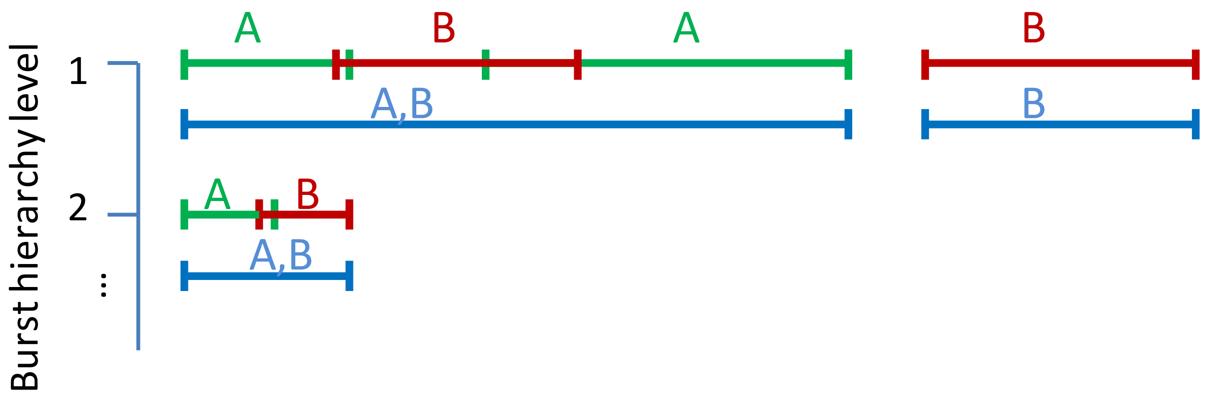
Proposed approach

2) Agglomerative clustering of burst intervals



Proposed approach

2) Agglomerative clustering of burst intervals



Proposed approach

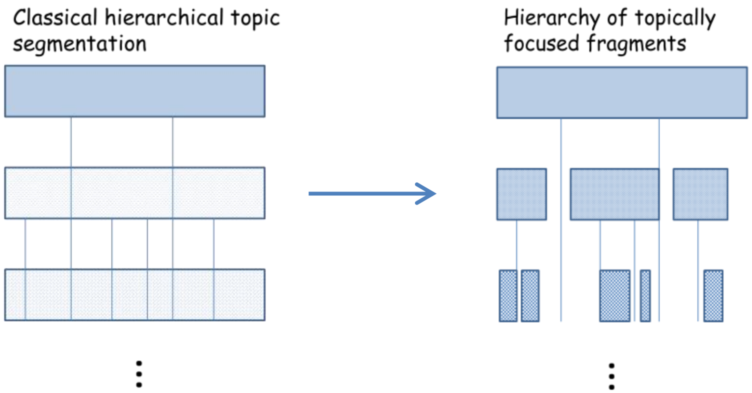
2) Agglomerative clustering of burst intervals

Result: *a hierarchy of topically focused fragments*

Proposed approach

2) Agglomerative clustering of burst intervals

Result: *a hierarchy of topically focused fragments*



Hierarchy of topically focused fragments

Automatic transcript: *Castle in the country*
[start time: 0.01 -> end time: 29.23]

⋮

0.01 -> 10.5
(place, ballroom, color, king, royal, rock, tweed, palace, etc.)

0.01- > 1.50
(royal, tweed, site, ballroom, palace, etc.)

2.29- > 3.21
(build, prosperous)

…

7.43- > 8.35
(friend, Picasso)

…

0.01 -> 0.10
(ballroom)

0.18 -> 1.50
(royal, site, tweed, palace)

1.33 -> 1.50
(royal, palace)

Experiments

Corpora

1. TV shows, manual and automatic transcripts
    - 7 episodes of a report show (Envoyé Spécial) (~2 hour each)
    - 3 levels of topic hierarchy (manual annotation)
2. Medical textbook
    - 227 chapters and 1136 sections
    - 2 levels of topic hierarchy
3. Wikipedia articles
    - 66 articles
    - 4 levels of hierarchy

Evaluation

M1: proportion of topical fragment belonging to a unique reference segment
M2: proportion of reference segments with at least one matching topical fragment

Comparison to dense segmentation

Corpus	Level	HTFF		Eisenstein (HierBayes)	
		M1	M2	M1	M2
TV shows manual transcripts	Level1 (coarse)	**0.75**	1	0.51	1
	Level2	**0.56**	0.74	0.15	1
	Level3 (fine)	**0.47**	0.17	--	--
Medical textbook	Level1 (coarse)	0.82	0.89	0.22	1
	Level2 (fine)	0.71	0.64	0.06	1
Wikipedia articles	Level1 (coarse)	**0.22**	0.97	0.29	1
	Level2	**0.62**	0.66	0.42	1
	Level3	0.69	0.29	--	--
	Level4 (fine)	0.49	0.06	--	--

HTFF: provide a better topical focus (M1);
the topic coverage at lower levels is smaller (M2)
HierBayes: segments usually do not belong to a unique topic;

Lessons learned: topic segmentation

- Question the fundamental aspects:
  - When is it worth to segment?
  - Can we actually find the segments in the groundthruth?

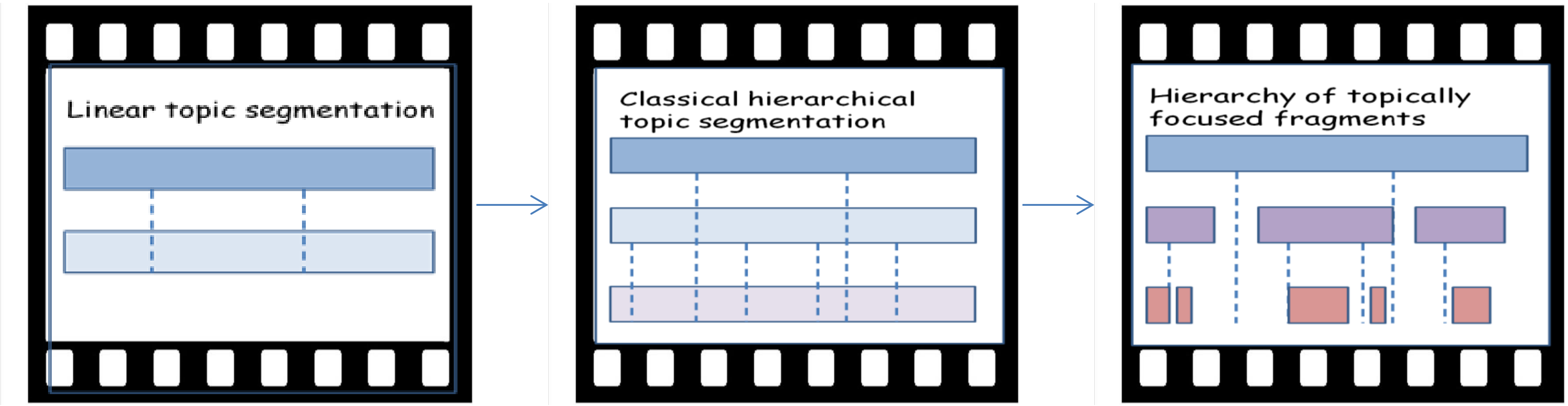
- Go in a different direction:
  - Propose something new
  - ✓ HTFF - a new representation

- Use of topic segmentation in NLP-related applications:
  - TextSeg, Mseg: target generation
  - HTFF: decide when to stop a segmentation; compression; summarization; anchor generation;

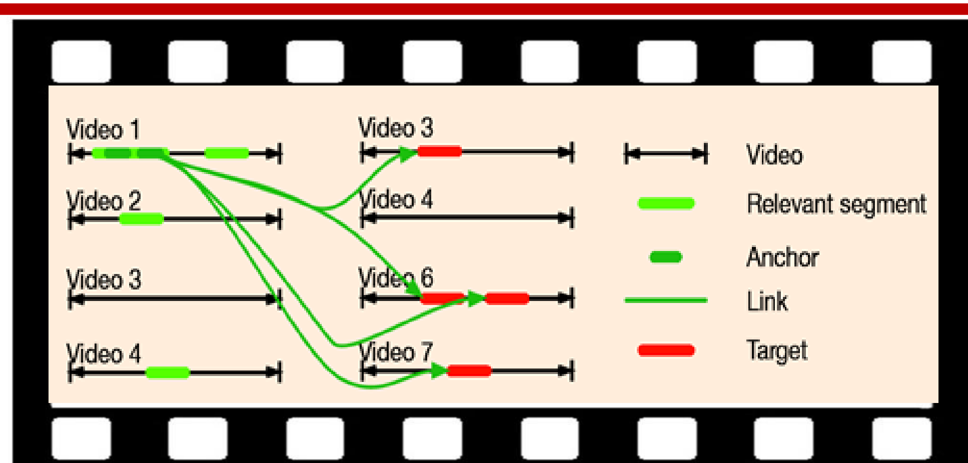
Semantic structuring of video collections from speech: segmentation & **hyperlinking**

Outline

-Thesis contributions in a nutshell-



- Anchor and target generation
- Link justification & diversity control

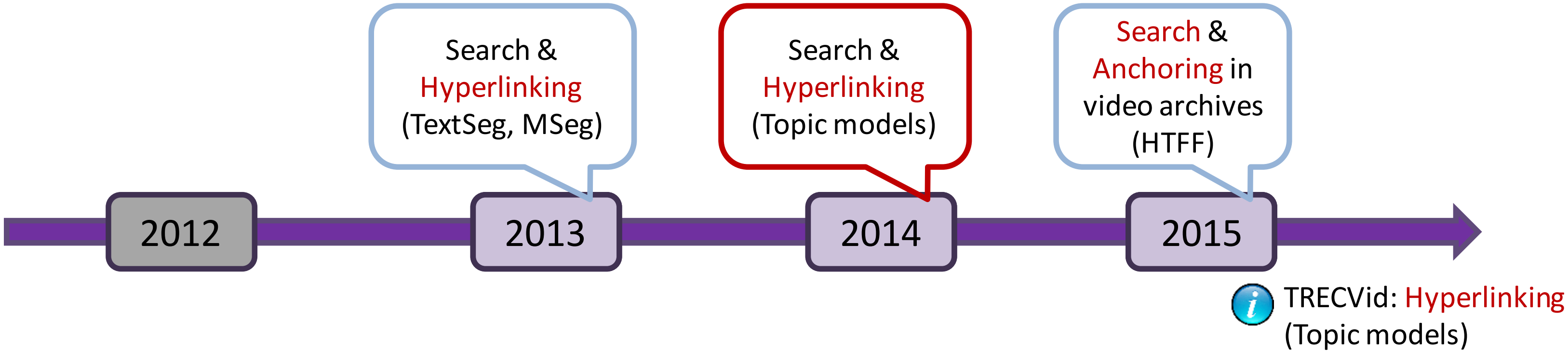


MediaEval benchmark initiative

Search & Anchoring & Hyperlinking

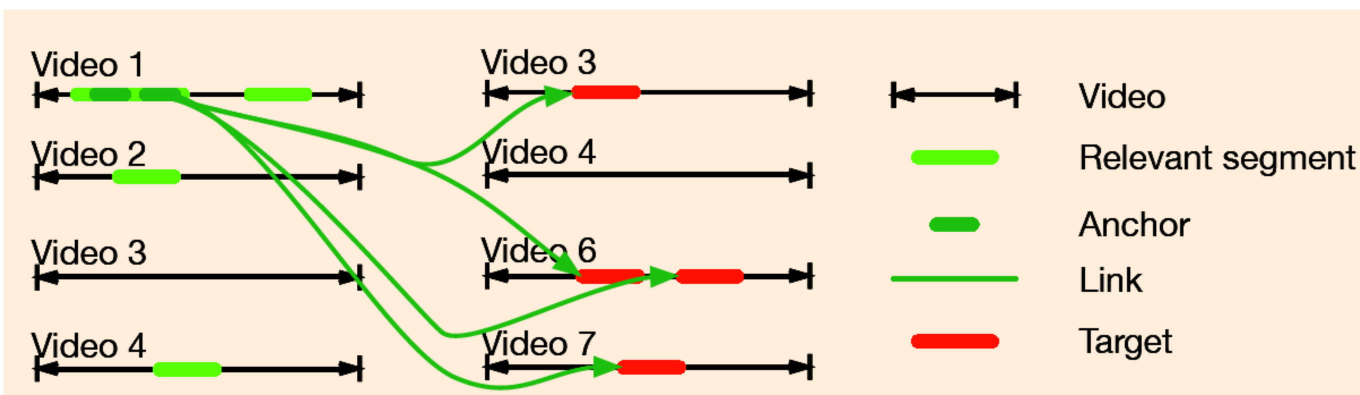
Context

MediaEval benchmarking initiative: Search and Hyperlinking task



Use case

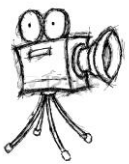
Text query
- speech cue
- visual cue



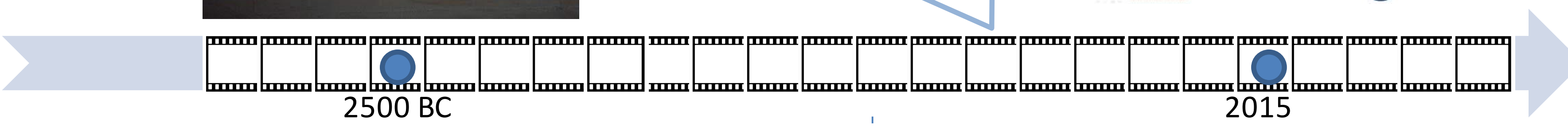
Video hyperlinking

A two-step approach:

1. Segmentation



-Fixed-length segments
-Video shots
-Topic segments
-Utterances



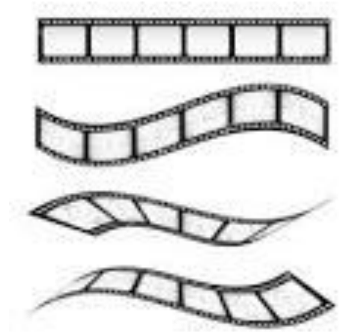
2500 BC

Pyramids of Giza

2015

The day of the defense

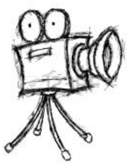
Potential targets



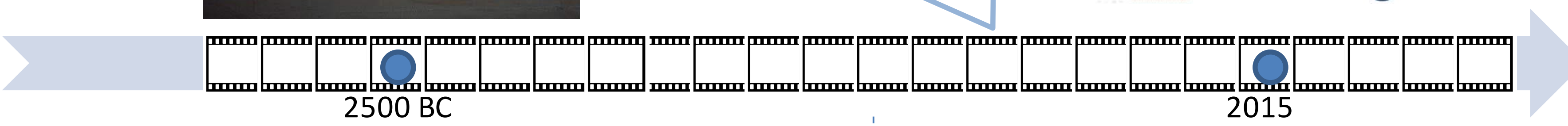
Video hyperlinking

A two-step approach:

1. Segmentation



-Fixed-length segments
-Video shots
-Topic segments
-Utterances

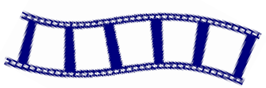


Pyramids of Giza

The day of the defense

2. Target selection

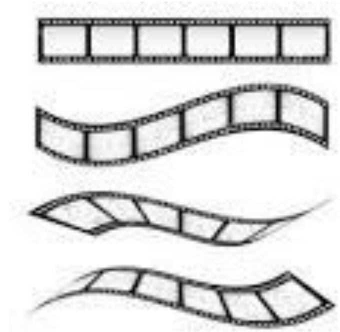
Anchor



comparison & selection

-Language via transcripts (entities, prosody)
-Visual content (concepts)
-Metadata

Potential targets



What about diversity?

Direct comparison in vector space with cosine similarity!

Targets very similar to the anchor

- near duplicates
- timeline events
- … but no diversity and no serendipity

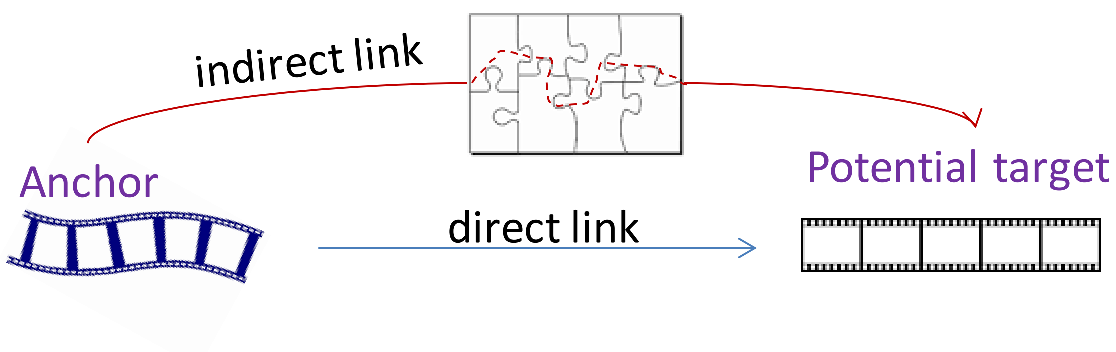
What about diversity?

Direct comparison in vector space with cosine similarity!

Targets very similar to the anchor

- near duplicates
- timeline events
- … but no diversity and no serendipity

<u>Solution</u>: Indirect comparison



+ link anchor-target pairs with few words in common

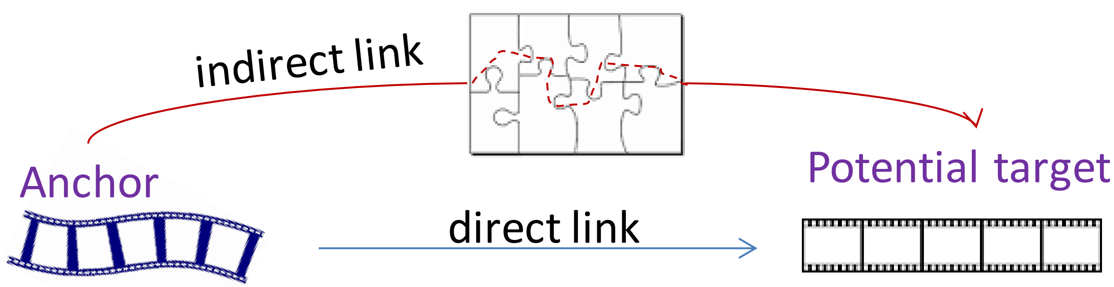
What about diversity?

Direct comparison in vector space with cosine similarity!

Targets very similar to the anchor

- near duplicates
- timeline events
- … but no diversity and no serendipity

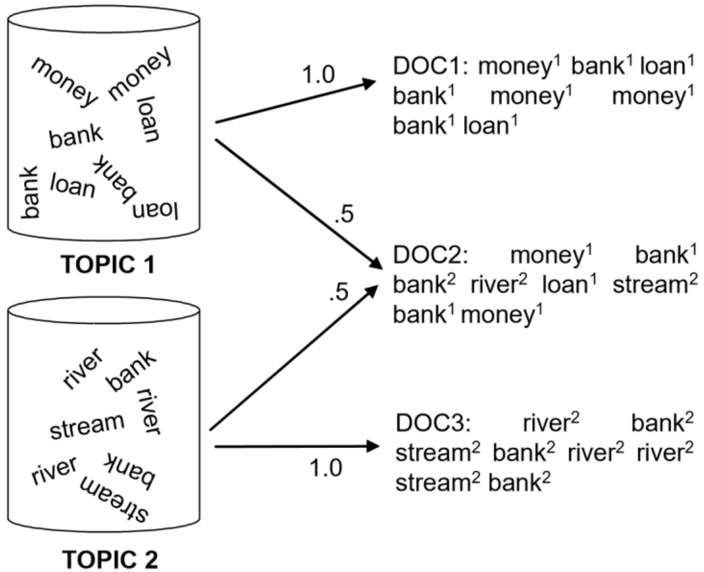
Solution: Indirect comparison via a hierarchy of topic models



+ link anchor-target pairs with few words in common
+ control diversity
+ link justification

LDA model

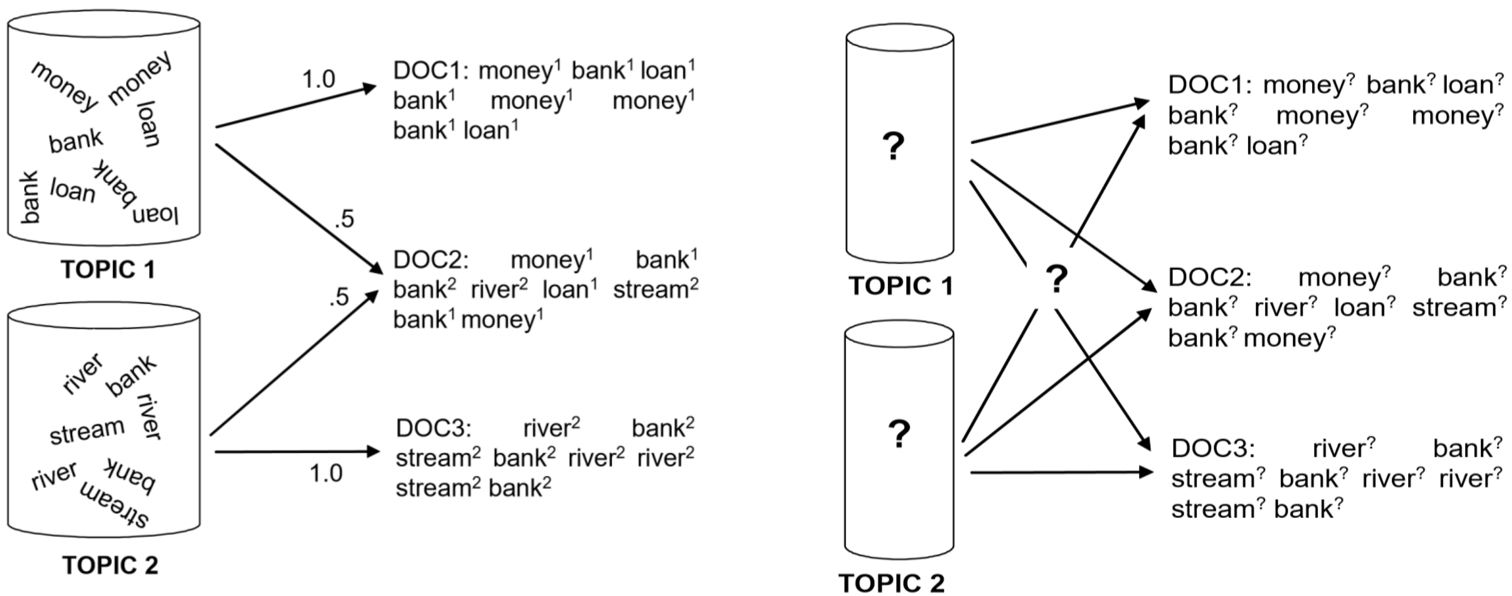
<u>Key idea:</u> there exist latent topics which uncover how words in documents have been generated



Steyvers and Griffiths, 2010

LDA model

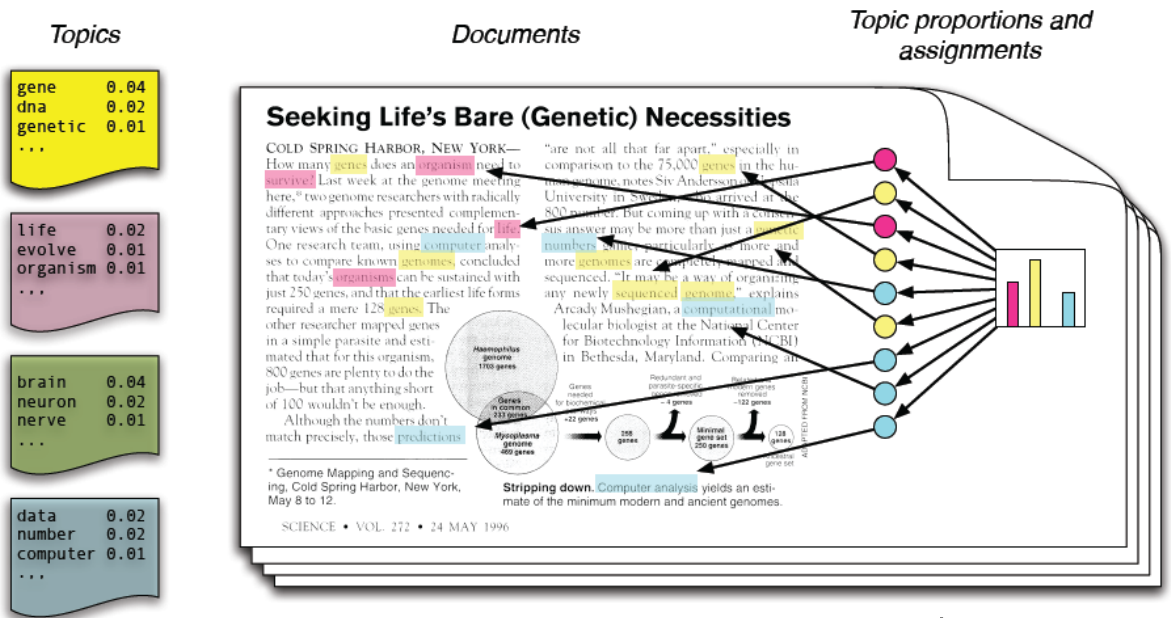
<u>Key idea:</u> there exist latent topics which uncover how words in documents have been generated



Steyvers and Griffiths, 2010

LDA model

<u>Key idea:</u> there exist latent topics which uncover how words in documents have been generated



Blei, 2012

- Each topic: a probability distribution over words
- Each document: a mixture of topics

Leverage LDA for hyperlinking

Create a hierarchy of topics:

$$K \in \{50,100,150,200,300,500,700,1000,1500,1700\}$$

- Level 1, $K_1 = 50$, broad topics $z_i^1, i \in [1, K_1]$
- Level 10, $K_{10} = 1700$, fine-grained topics $z_i^{10}, i \in [1, K_{10}]$

Leverage LDA for hyperlinking

Create a hierarchy of topics:

$$K \in \{50,100,150,200,300,500,700,1000,1500,1700\}$$

- Level 1, $K_1 = 50$, broad topics $z_i^1, i \in [1, K_1]$
- Level 10, $K_{10} = 1700$, fine-grained topics $z_i^{10}, i \in [1, K_{10}]$

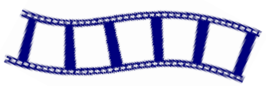
broad	fine-grained
z_3^1, K_1=50	z_{50}^{10}, K_{10}=1700
People	Referendum
Government	Minister
Tax	Scotland
Minister	Independence
Party	Alexander

$$\begin{array}{llllll}
z_1^1 & z_2^1 & z_3^1 \ldots & z_{50}^1 & & \\
z_1^2 & z_2^2 & z_3^2 \ldots & z_{50}^2 \ldots & z_{100}^2 & \\
\vdots & & & & & \\
z_1^{10} & z_2^{10} & z_3^{10} \ldots & z_{50}^{10} \ldots & z_{100}^{10} \ldots & z_{1700}^{10}
\end{array}$$

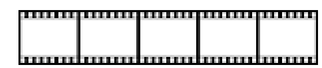
Changing the representation space

$$
\begin{array}{llllll}
z_1^1 & z_2^1 & z_3^1 \;\ldots\; z_{50}^1 & & & \\
z_1^2 & z_2^2 & z_3^2 \;\ldots\; z_{50}^2 \;\ldots\; z_{100}^2 & & & \\
\vdots & & & & & \\
z_1^{10} & z_2^{10} & z_3^{10} & \ldots\; z_{50}^{10} & \ldots\; z_{100}^{10} & \ldots\; z_{1700}^{10}
\end{array}
$$

Anchor



Potential target

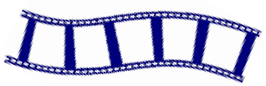


- New representation of an anchor/target segment

$$
x_l = (p(x \mid z_1^l) \ldots p(x \mid z_{K_l}^l))
$$

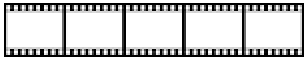
Changing the representation space

Anchor



$$\begin{array}{llllllll}
z_1^1 & z_2^1 & z_3^1 & \ldots z_{50}^1 & & & & \\
z_1^2 & z_2^2 & z_3^2 & \cdots z_{50}^2 & \ldots & z_{100}^2 & & \\
\vdots & & & & & & & \\
z_1^{10} & z_2^{10} & z_3^{10} & \ldots z_{50}^{10} & \ldots & z_{100}^{10} & \ldots & z_{1700}^{10}
\end{array}$$

Potential target



- New representation of an anchor/target segment

$$\boxed{x_l = (p(x \mid z_1^l) \ldots p(x \mid z_{K_l}^l))}$$

- 1st strategy: independent topic levels (IT)
- 2nd strategy: hard and soft links between topics

Independent levels

- Anchor segment x $\quad x_l = (p(x \mid z_1^l) \ldots p(x \mid z_{K_l}^l))$
- Target segment y $\quad y_l = (p(y \mid z_1^l) \ldots p(y \mid z_{K_l}^l))$

$$\boxed{Similarity(x, y) = \sum_l \alpha_l \log(x_l \cdot y_l)}$$

IT_k only level *k* $\quad \alpha_{\mathrm{k}} = 1, \alpha_{\mathrm{i \neq k}} = 0$

$IT_=$ equal weights $\quad \alpha_{\mathrm{k}} = 0.2, \forall k \in \{1,3,5,7,9\}$

$IT_<$ general<specific $\quad \alpha_1 = 0.1, \alpha_3 = 0.15, \alpha_5 = 0.2, \alpha_7 = 0.25, \alpha_9 = 0.3$

$IT_>$ specific<general $\quad \alpha_1 = 0.3, \alpha_3 = 0.25, \alpha_5 = 0.2, \alpha_7 = 0.15, \alpha_9 = 0.1$

Data

2013 & 2014 Search & Hyperlinking data

- BBC broadcast videos
- automatic speech transcripts (LIMSI)

Task considered: reranking targets

- Targets proposed by all the participants!
- Relevance judgments provided by turkers (AMT)

year	#hours of video	#anchors	avg. anchor duration (95% interval)	#targets (% relevant)	avg. target duration (95%interval)
2013	1,335	30	32.2 [13.4,51]	9,973 (29.9%)	83.38 sec. [82.58,84.18]
2014	2,686	30	22.9 [11.1,34.8]	12,340 (15.3%)	58.85 sec. [58.1,59.58]

Relevance assessment

- Baseline: direct cos-similarity (DirectH)
- Measures: relevance (P@10);
  tolerance to irrelevance (P@10_tol)

	2013		2014	
method	P@10	P@10_tol	P@10	P@10_tol
DirectH	0.61	0.25	0.41	0.19
IT_{50}	0.65	**0.44***	0.26	0.18
IT_{150}	0.57	**0.34***	0.37	**0.25***
IT_{300}	0.61	**0.35***	0.34	**0.26***
IT_{700}	0.64	**0.34***	0.31	0.21
IT_{1500}	0.59	**0.32***	0.32	0.24
$IT_{Comb=}$	0.66	**0.35***	0.27	0.22
$IT_{Comb<}$	0.67	**0.37***	0.27	0.21
$IT_{Comb>}$	0.65	**0.35***	0.29	0.22

* Statistical significant values (paired t-test, $p<0.05$)

Diversity assessment

Success of a hyperlinking system:
cover potential (idiosyncratic) user interest & enable serendipity

Links differ between systems

System 1	System 2	% difference 2013	2014
IT_{700}	*DirectH*	93	86
IT_{700}	$IT_{Comb>}$	82	90
IT_{700}	*Hierarchy*	98	93
$IT_{Comb=}$	*Hierarchy*	94	95

AMT evaluation scenario at MediaEval

- 1 judgement/anchor-target pair
- yes/no relevance assessment
- description of potential targets

Diversity in the links

Design a new evaluation scenario:

- At least 3 assessments per anchor-target pair
- Each participant should do 5 tests
- Test for: relevance (same topic, related topic, same show);
  unexpectedness;
  interestingness;

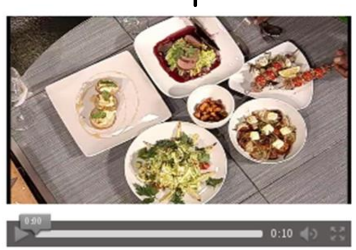
Clip A

Anchor:

Two video clips (B and C) that could be linked to video A are recommended to you that should encourage this further exploration. Please watch the two videos and answer the questions.

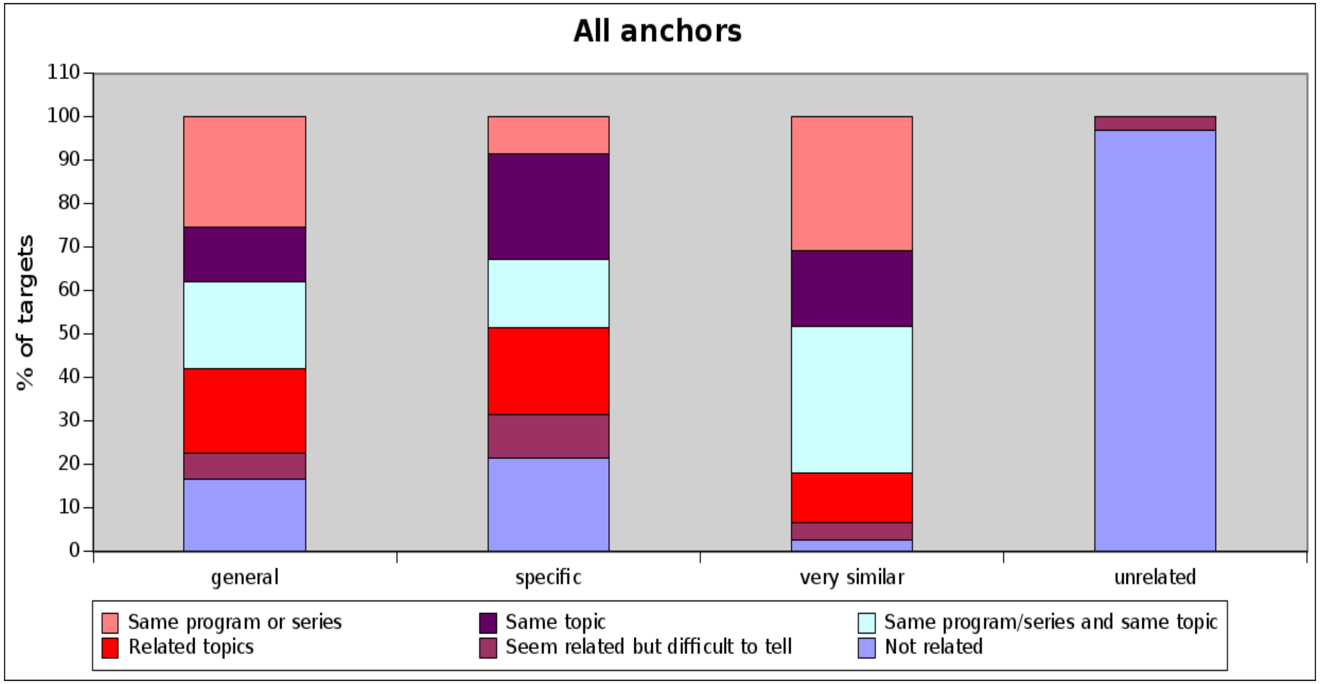
Targets:

Clip B



Clip C

Results for the new scenario

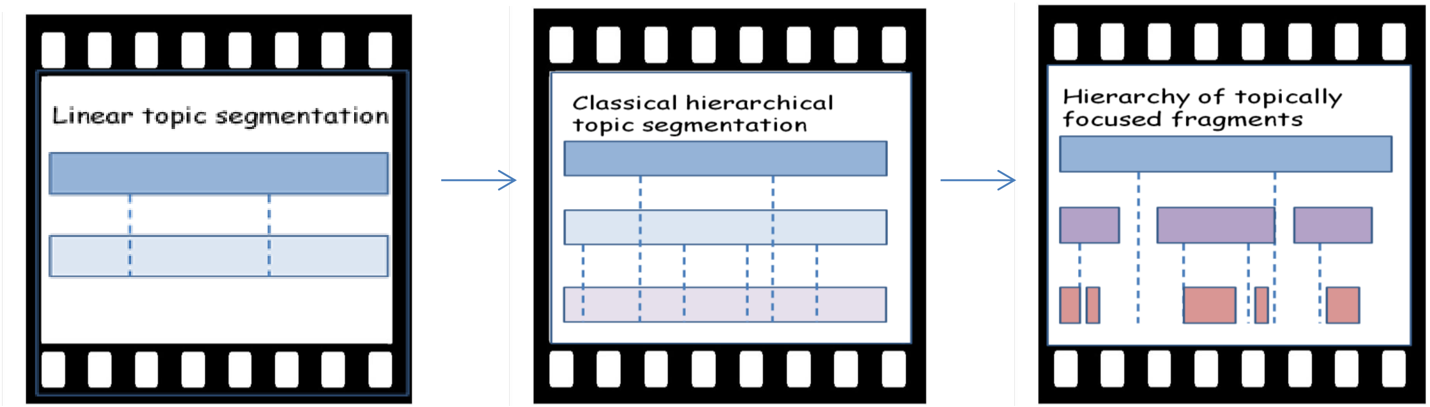


- Very similar targets:
  - same program/series and same topic (91% expected; 9% possibly)
  - most expected
- Specific topics:
  - same topic (47% expected; 53% possibly)
  - less expected

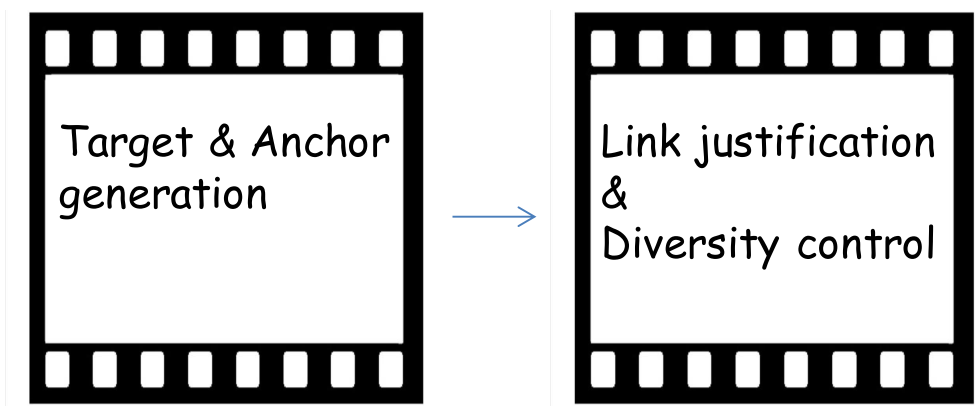
Conclusions & Perspectives

Answering the research questions

1.How to *structure* audiovisual content?



2. How to *exploit* structured content?



Answering the research questions

1.How to *structure* audiovisual content?

- ✓ EMNLP 2013
- ✓ TALN 2013
- ✓ RANLP 2015

2. How to *exploit* structured content?

✓ SLAM 2014, SLAM 2015, MediaEval 2013,2014,2015

Challenges:

✓ MediaEval(2013-2015), TRECVid 2015

➢ Collaborations: Sien Moens, Camille Guinaudeau, Rémi Bois, Ronan Sicre, Emmanuel Morin, Martha Larson

Perspectives