TVD: a reproducible and multiply aligned TV series dataset

Anindya Roy, Camille Guinaudeau, Hervé Bredin, Claude Barras

Context

Data associated with TV series is multimodal

- Speech
- Visual data
- Crowd-sourced textual content

TV series are potential source of data for various applications

- Summarization
- Rich Speech Retrieval
- Personal Identification without privacy issue

Technologies developed applicable in

- TV production context: second screen applications
- Other kinds of multimedia documents

Context

Few resources around TV series

Copyright restrictions

At most open-source version of the algorithm and pre-extracted features

Difficult to reproduce and compare results without the original dataset and associated annotations

Three contributions:

① Provide computer scripts to reproduce the corpus **from DVD**

② Parse and structure data related to TV series

③ Align units from different sources

Outline

① Corpus description

② Tracks alignment

③ How to use the TVD corpus?

④ Conclusions & future work

Corpus description

Two popular TV series

Situation comedy : ***The Big Bang Theory***

20 minutes long

5/6 main characters

Fantasy drama : ***Game of Thrones***

50 minutes long

More than 35 main characters

Large fan base:

- ✓ Manual transcripts
- ✓ Episode descriptions
- ✓ Comments and discussions on Internet Forums

Corpus description

Computer scripts to reproduce tracks from various sources

From DVD:

✓ Video tracks

✓ Multi-lingual audio tracks

✓ Multi-lingual subtitles

From websites:

✓ Manual transcripts

✓ Episode outlines

✓ Summaries

From manual annotations:

[Tapaswi et al., 2012; Bäuml et al., 2013]

✓ Shot boundaries

✓ Speech turns

✓ Face tracks

From ASR:

✓ Automatic transcripts

Corpus description

Manual transcripts from website (MTR)

Scene location: A Chinese restaurant.
Sheldon: I’m sorry, we cannot do this without Wolowitz.
Leonard: We can’t order Chinese food without Wolowitz?

Jorah Mormont: You need to drink, child. And eat.
Daenerys Targaryen: Isn’t there anything else ?
Jorah Mormont: The Dothraki have two things in abundance: grass and horses. People can't live on grass.

Corpus description

Multi-lingual subtitles from DVD (SUB)

```
00:13:21,520 -> 00:13:24,318
I’m sorry. we cannot do this without Wolowitz.
00:13:24,480 -> 00:13:27,278
We can’t order Chinese food without Wolowitz?
```

```
00:02:06,520 --> 00:02:07,953
You need to drink, child.
00:02:08,079 --> 00:02:09,990
(She sighs)
00:02:12,159 --> 00:02:14,070
And eat.
```

Corpus description

Episode outlines from website (OL)

Scene location: Hallway outside apartments
Event: Penny gives Leonard the key to her apartment.
Event: The four guys get into a discussion about Superman’s flight skills in front of Penny.

Scene location: Hallway
Event: Leonard invites Penny over.

Event: Khal Drogo’s khalasar is several days from Pentos crossing the plains known as the Flatlands...
Event: The Dothraki make camp and Daenerys is helped from her horse by Ser Jorah and her handmaidens...

Corpus description

Summaries from website (SUM)

When Leonard and Sheldon meet Penny, Leonard is immediately interested in her (saying "our babies will be smart and beautiful"), but Sheldon feels his friend is chasing a dream he’ll never catch (adding "not to mention imaginary")...

Three rangers of the Night’s Watch: Ser Waymar Royce, Will, and Gared depart from the Wall to investigate reports of wildlings in the Haunted Forest which lies to the north...

Corpus description

	# Episodes								
Tracks	TBBT	GoT	Type	Manual	Time-stamped	Multi-lingual	Identity	Location	Original source
Manual transcripts	132	5	dialogue	✔			✔	✔	WWW
Subtitles	17	10	dialogue	✔	✔	✔			DVD
Automatic transcripts	17	10	dialogue		✔	✔			On request
Episode outlines	69	17	description	✔				✔	WWW
Summaries	69	30	description	✔				✔	WWW
Speech turns	6	-	annotation	✔	✔		✔		(Tapaswi et al., 2012)
Face Tracks	6	-	annotation	✔	✔		✔		(Tapaswi et al., 2012)
Shots	6	-	annotation	✔	✔				(Bäuml et al., 2013)

To improve the usability of the dataset

→ Automatic alignment between tracks

Corpus description

	# Episodes								
Tracks	TBBT	GoT	Type	Manual	Time-stamped	Multi-lingual	Identity	Location	Original source
Manual transcripts	132	5	dialogue	✔			✔	✔	WWW
Subtitles	17	10	dialogue	✔	✔	✔			DVD
Automatic transcripts	17	10	dialogue		✔	✔			On request
Episode outlines	69	17	description	✔				✔	WWW
Summaries	69	30	description	✔				✔	WWW
Speech turns	6	-	annotation	✔	✔		✔		(Tapaswi et al., 2012)
Face Tracks	6	-	annotation	✔	✔		✔		(Tapaswi et al., 2012)
Shots	6	-	annotation	✔	✔				(Bäuml et al., 2013)

To improve the usability of the dataset

→ Automatic alignment between tracks

Corpus description

	# Episodes								
Tracks	TBBT	GoT	Type	Manual	Time-stamped	Multi-lingual	Identity	Location	Original source
Manual transcripts	132	5	dialogue	✓			✓	✓	WWW
Subtitles	17	10	dialogue	✓	✓	✓			DVD
Automatic transcripts	17	10	dialogue		✓	✓			On request
Episode outlines	69	17	description	✓				✓	WWW
Summaries	69	30	description	✓				✓	WWW
Speech turns	6	-	annotation	✓	✓		✓		(Tapaswi et al., 2012)
Face Tracks	6	-	annotation	✓	✓		✓		(Tapaswi et al., 2012)
Shots	6	-	annotation	✓	✓				(Bäuml et al., 2013)

To improve the usability of the dataset

→ Automatic alignment between tracks

Corpus description

	# Episodes								
Tracks	TBBT	GoT	Type	Manual	Time-stamped	Multi-lingual	Identity	Location	Original source
Manual transcripts	132	5	dialogue	✔			✔	✔	WWW
Subtitles	17	10	dialogue	✔	✔	✔			DVD
Automatic transcripts	17	10	dialogue		✔	✔			On request
Episode outlines	69	17	description	✔				✔	WWW
Summaries	69	30	description	✔				✔	WWW
Speech turns	6	-	annotation	✔	✔		✔		(Tapaswi et al., 2012)
Face Tracks	6	-	annotation	✔	✔		✔		(Tapaswi et al., 2012)
Shots	6	-	annotation	✔	✔				(Bäuml et al., 2013)

To improve the usability of the dataset

→ Automatic alignment between tracks

Corpus description

	# Episodes								
Tracks	TBBT	GoT	Type	Manual	Time-stamped	Multi-lingual	Identity	Location	Original source
Manual transcripts	132	5	dialogue	✔			✔	✔	WWW
Subtitles	17	10	dialogue	✔	✔	✔			DVD
Automatic transcripts	17	10	dialogue		✔	✔			On request
Episode outlines	69	17	description	✔				✔	WWW
Summaries	69	30	description	✔				✔	WWW
Speech turns	6	-	annotation	✔	✔		✔		(Tapaswi et al., 2012)
Face Tracks	6	-	annotation	✔	✔		✔		(Tapaswi et al., 2012)
Shots	6	-	annotation	✔	✔				(Bäuml et al., 2013)

To improve the usability of the dataset

→ Automatic alignment between tracks

Outline

① Corpus description

② Tracks alignment

③ How to use the TVD corpus?

④ Conclusions & future work

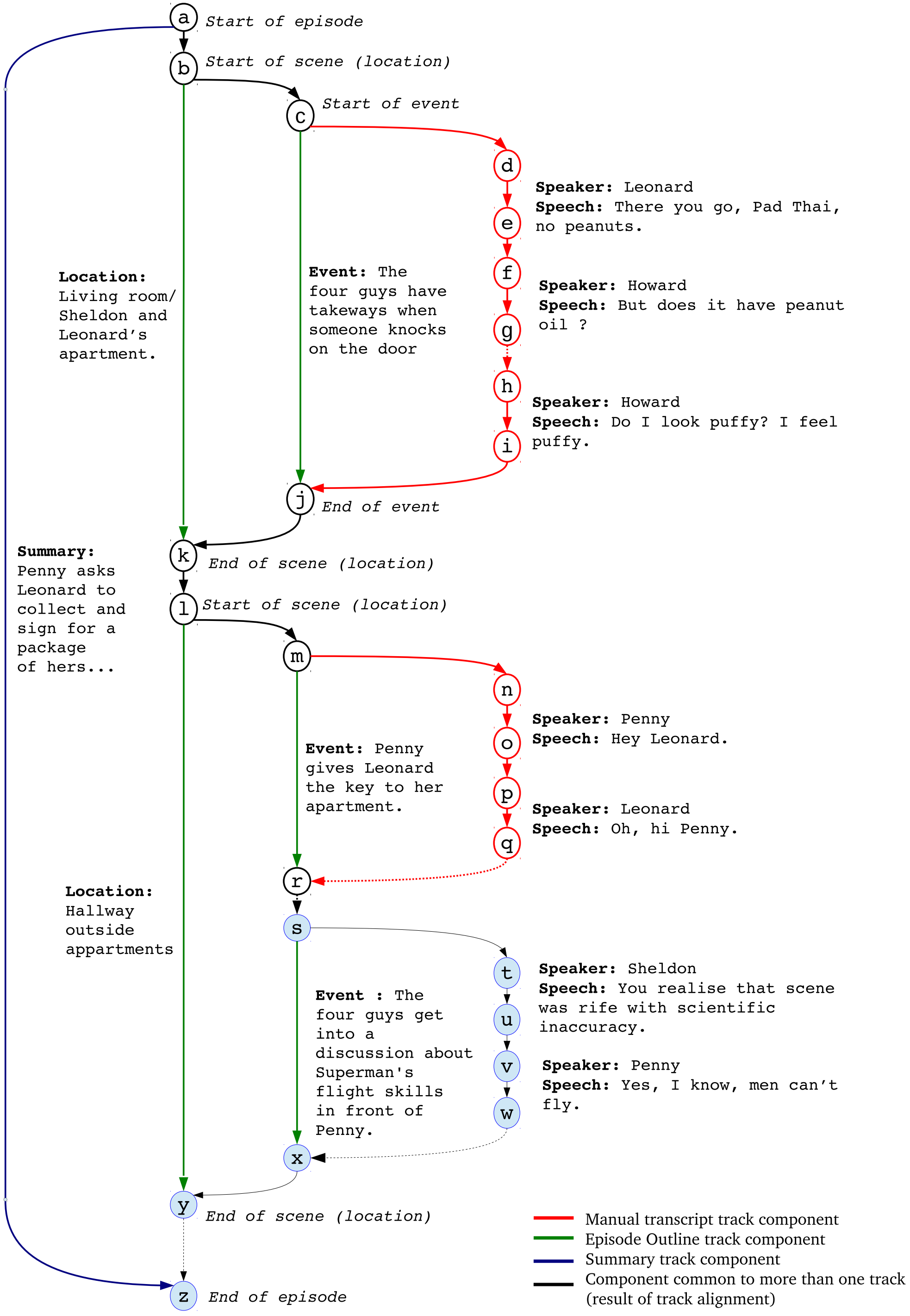
Alignment

Three pairs of track alignment to improve the dataset usability

① **Manual transcripts (MTR) ⇔ subtitles (SUB)**
merges time-stamps from SUB with exact dialogue from MTR

② **Subtitles (SUB)⇔automatic transcripts (ATR)**
enhances time resolution from *sentence*-level (SUB) to *word*-level (ATR)

③ **Episode outlines (OL) ⇔ manual transcripts (MTR)**
merges speaker and dialogue lines from MTR with event descriptions from OL



Start of episode

Start of scene (location)

Start of event

Speaker: Leonard
Speech: There you go, Pad Thai, no peanuts.

Location: Living room/ Sheldon and Leonard's apartment.

Event: The four guys have takeways when someone knocks on the door

Speaker: Howard
Speech: But does it have peanut oil ?

Speaker: Howard
Speech: Do I look puffy? I feel puffy.

End of event

Summary: Penny asks Leonard to collect and sign for a package of hers...

End of scene (location)

Start of scene (location)

Speaker: Penny
Speech: Hey Leonard.

Event: Penny gives Leonard the key to her apartment.

Speaker: Leonard
Speech: Oh, hi Penny.

Location: Hallway outside appartments

Speaker: Sheldon
Speech: You realise that scene was rife with scientific inaccuracy.

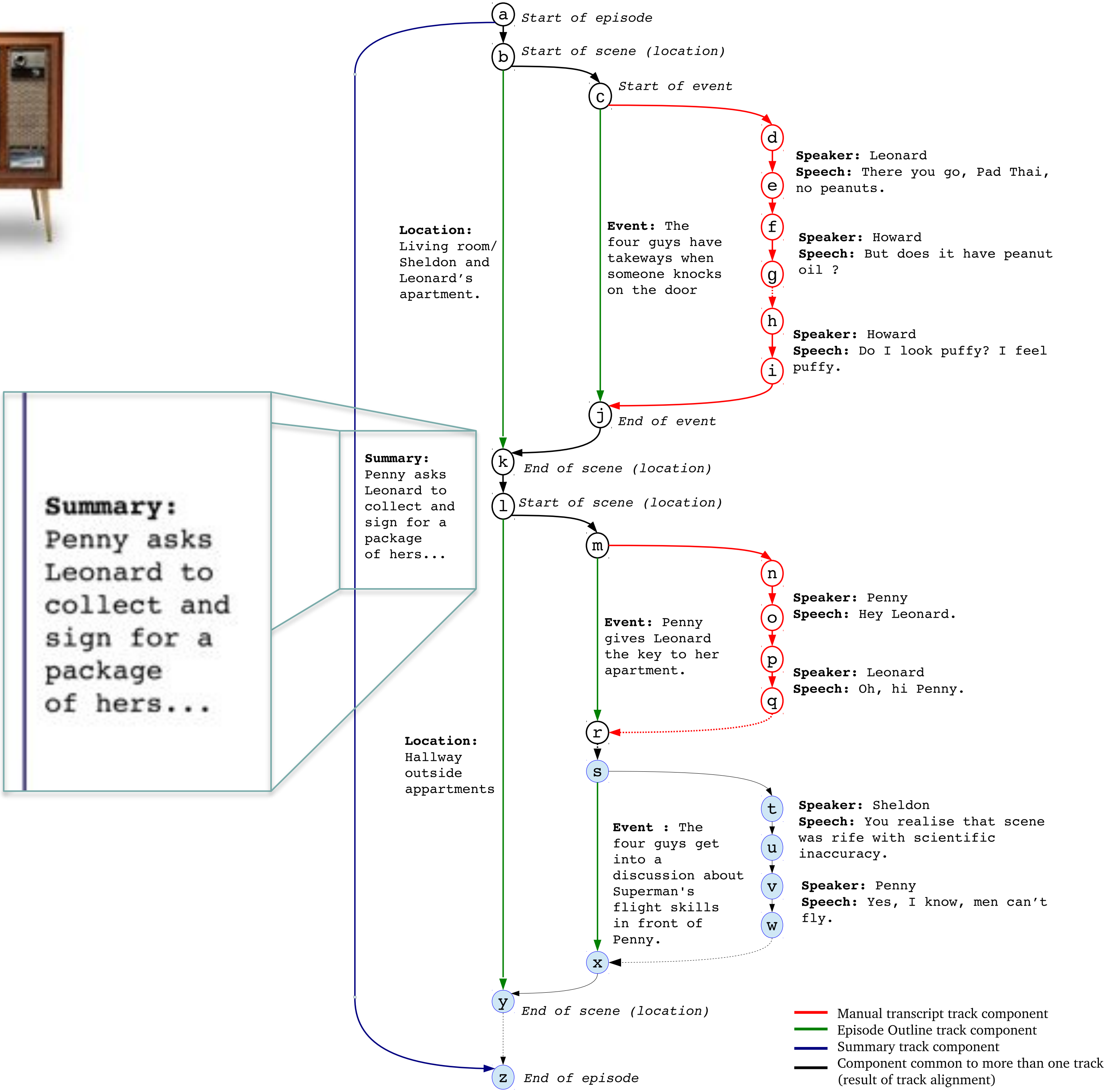
Event : The four guys get into a discussion about Superman's flight skills in front of Penny.

Speaker: Penny
Speech: Yes, I know, men can't fly.

End of scene (location)

End of episode

🟥 Manual transcript track component
🟩 Episode Outline track component
🟦 Summary track component
⬛ Component common to more than one track (result of track alignment)



Start of episode

Start of scene (location)

Start of event

Speaker: Leonard
Speech: There you go, Pad Thai, no peanuts.

Location: Living room/ Sheldon and Leonard's apartment.

Event: The four guys have takeways when someone knocks on the door

Speaker: Howard
Speech: But does it have peanut oil ?

Speaker: Howard
Speech: Do I look puffy? I feel puffy.

End of event

End of scene (location)

Start of scene (location)

Summary: Penny asks Leonard to collect and sign for a package of hers...

Summary: Penny asks Leonard to collect and sign for a package of hers...

Speaker: Penny
Speech: Hey Leonard.

Event: Penny gives Leonard the key to her apartment.

Speaker: Leonard
Speech: Oh, hi Penny.

Location: Hallway outside appartments

Speaker: Sheldon
Speech: You realise that scene was rife with scientific inaccuracy.

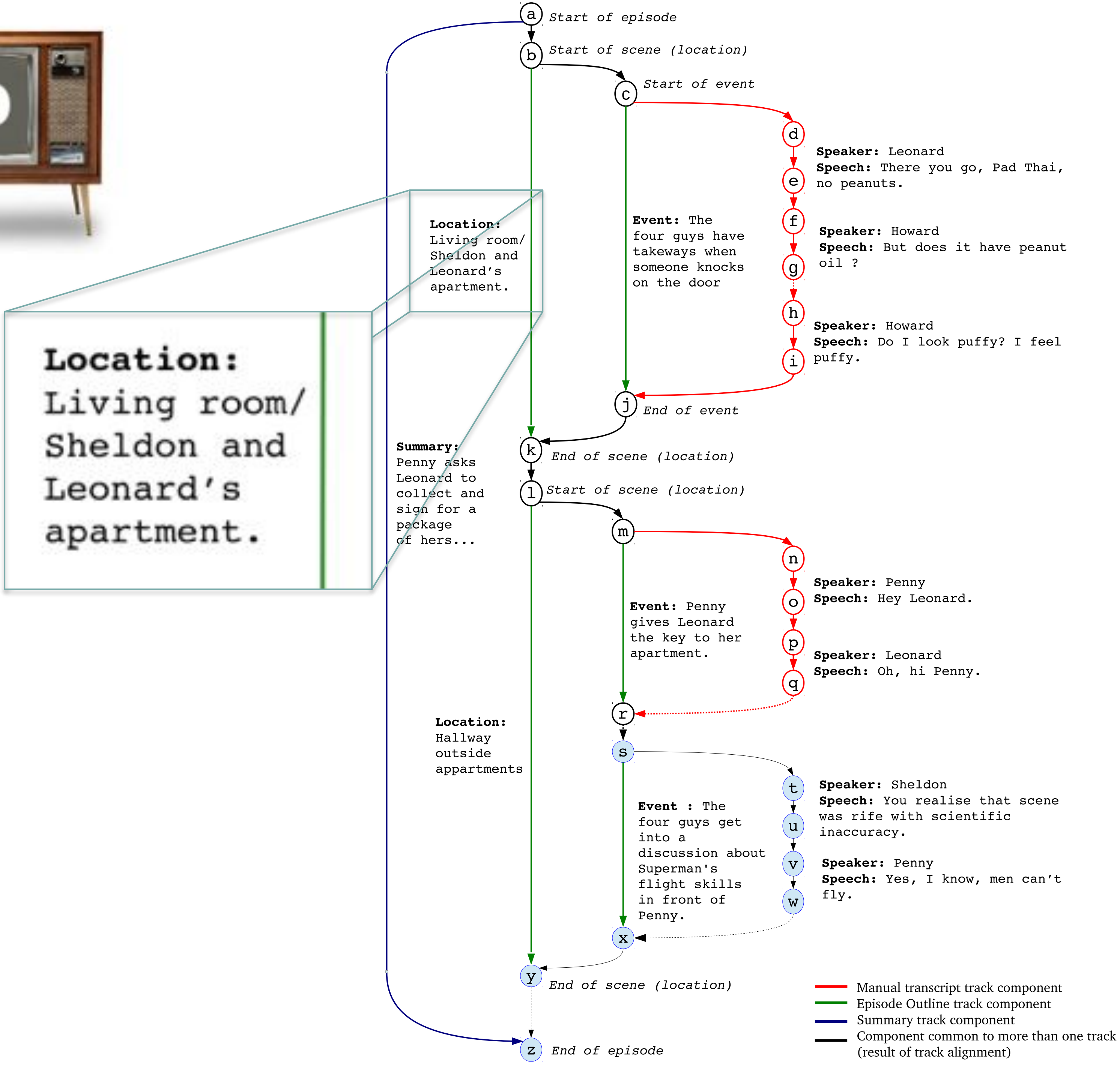
Event : The four guys get into a discussion about Superman's flight skills in front of Penny.

Speaker: Penny
Speech: Yes, I know, men can't fly.

End of scene (location)

End of episode

🟥 Manual transcript track component
🟩 Episode Outline track component
🟦 Summary track component
⬛ Component common to more than one track (result of track alignment)



a Start of episode

b Start of scene (location)

c Start of event

d

e

Speaker: Leonard
Speech: There you go, Pad Thai, no peanuts.

f

g

Speaker: Howard
Speech: But does it have peanut oil ?

h

i

Speaker: Howard
Speech: Do I look puffy? I feel puffy.

Event: The four guys have takeways when someone knocks on the door

j End of event

Location: Living room/ Sheldon and Leonard's apartment.

Location: Living room/ Sheldon and Leonard's apartment.

Summary: Penny asks Leonard to collect and sign for a package of hers...

k End of scene (location)

l Start of scene (location)

m

n

o

Speaker: Penny
Speech: Hey Leonard.

p

q

Speaker: Leonard
Speech: Oh, hi Penny.

Event: Penny gives Leonard the key to her apartment.

r

Location: Hallway outside appartments

s

t

u

Speaker: Sheldon
Speech: You realise that scene was rife with scientific inaccuracy.

v

w

Speaker: Penny
Speech: Yes, I know, men can't fly.

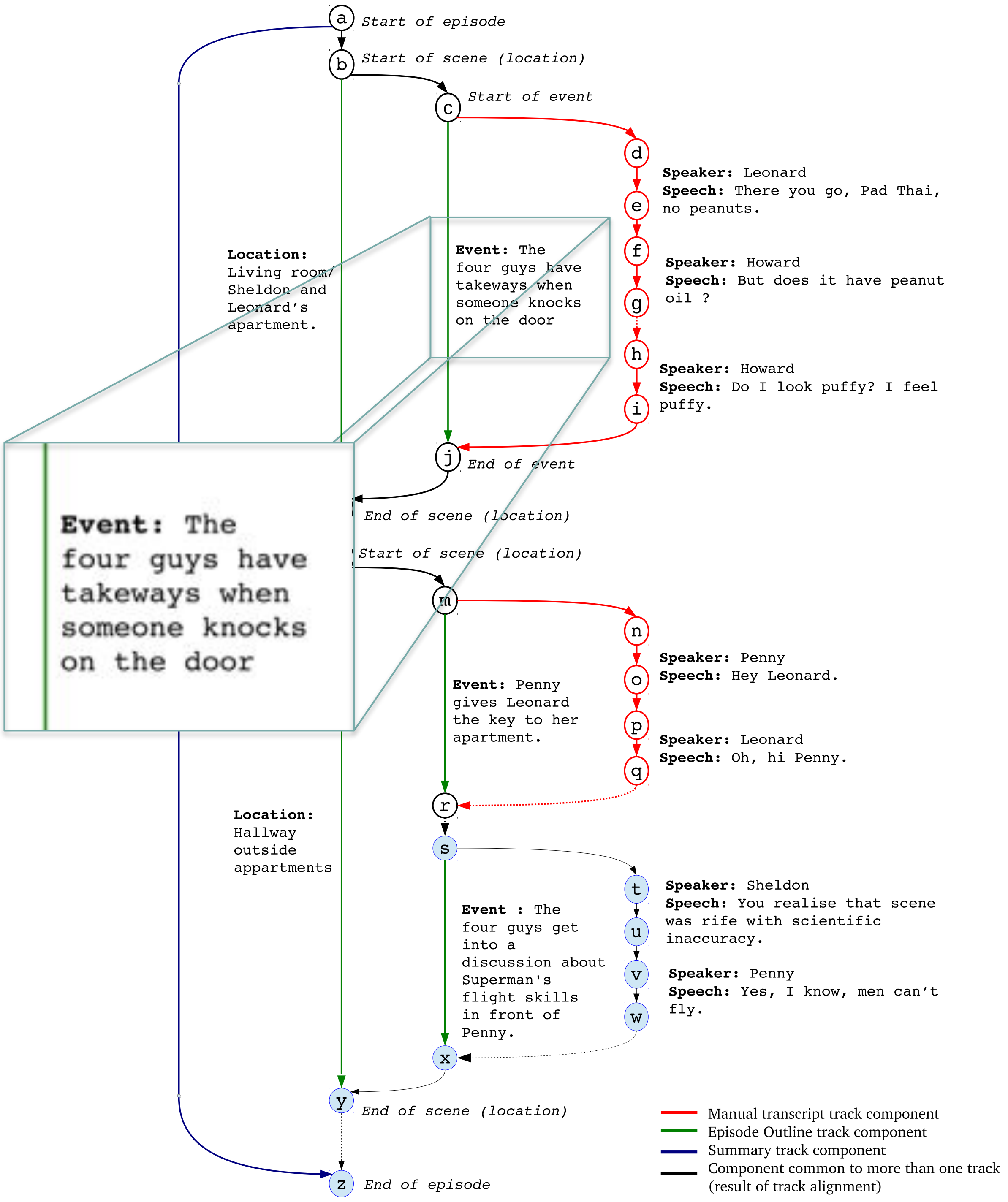
Event : The four guys get into a discussion about Superman's flight skills in front of Penny.

x

y End of scene (location)

z End of episode

🟥 Manual transcript track component
🟩 Episode Outline track component
🟦 Summary track component
⬛ Component common to more than one track (result of track alignment)



Start of episode

Start of scene (location)

Start of event

Speaker: Leonard
Speech: There you go, Pad Thai, no peanuts.

Speaker: Howard
Speech: But does it have peanut oil ?

Speaker: Howard
Speech: Do I look puffy? I feel puffy.

Location: Living room/ Sheldon and Leonard's apartment.

Event: The four guys have takeways when someone knocks on the door

Event: The four guys have takeways when someone knocks on the door

End of event

End of scene (location)

Start of scene (location)

Event: Penny gives Leonard the key to her apartment.

Speaker: Penny
Speech: Hey Leonard.

Speaker: Leonard
Speech: Oh, hi Penny.

Location: Hallway outside appartments

Event : The four guys get into a discussion about Superman's flight skills in front of Penny.

Speaker: Sheldon
Speech: You realise that scene was rife with scientific inaccuracy.

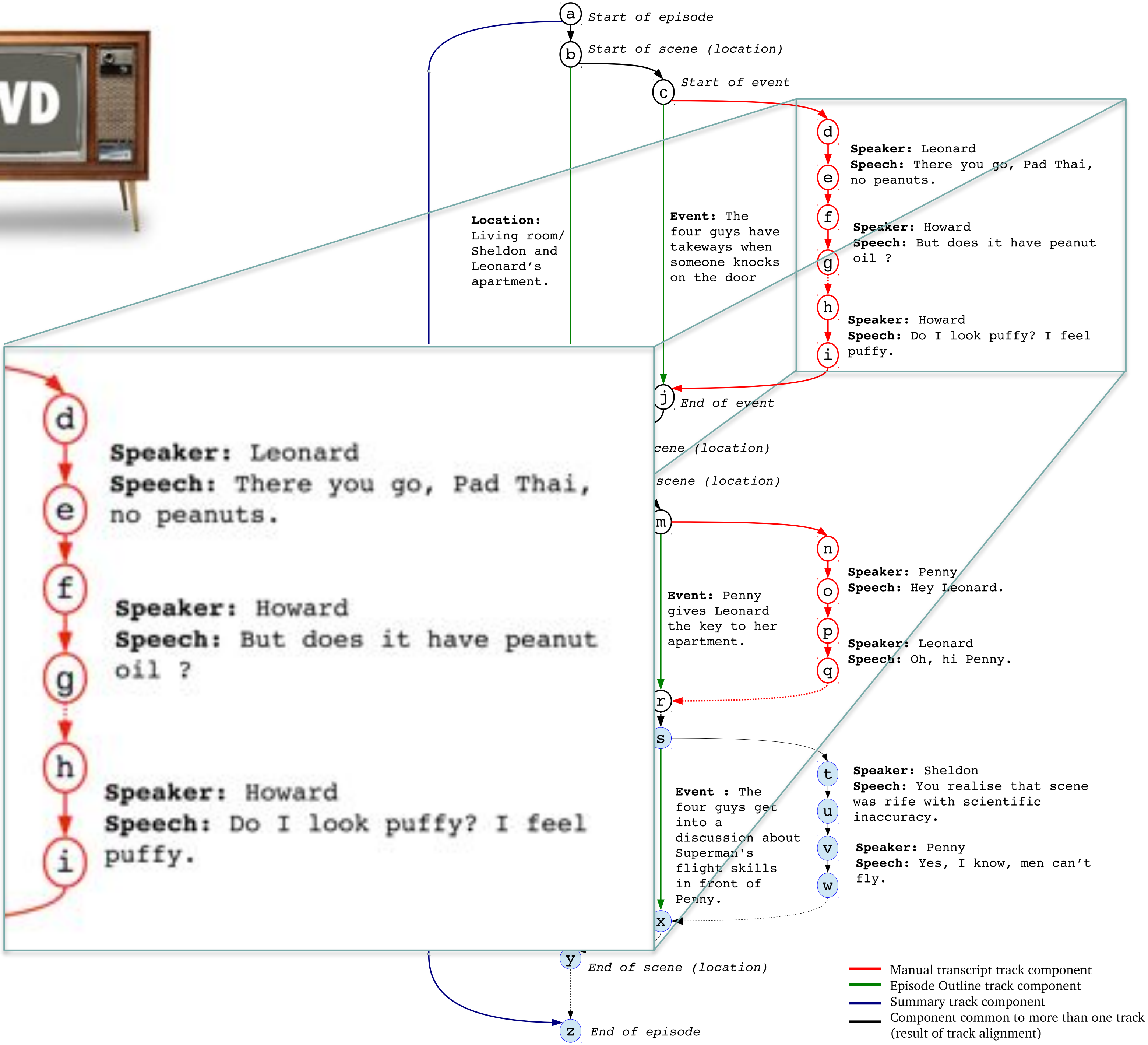
Speaker: Penny
Speech: Yes, I know, men can't fly.

End of scene (location)

End of episode

🟥 Manual transcript track component
🟩 Episode Outline track component
🟦 Summary track component
⬛ Component common to more than one track (result of track alignment)

TVD



Start of episode

Start of scene (location)

Start of event

Location: Living room/ Sheldon and Leonard's apartment.

Event: The four guys have takeways when someone knocks on the door

Speaker: Leonard
Speech: There you go, Pad Thai, no peanuts.

Speaker: Howard
Speech: But does it have peanut oil ?

Speaker: Howard
Speech: Do I look puffy? I feel puffy.

End of event

cene (location)

scene (location)

Event: Penny gives Leonard the key to her apartment.

Speaker: Penny
Speech: Hey Leonard.

Speaker: Leonard
Speech: Oh, hi Penny.

Event : The four guys get into a discussion about Superman's flight skills in front of Penny.

Speaker: Sheldon
Speech: You realise that scene was rife with scientific inaccuracy.

Speaker: Penny
Speech: Yes, I know, men can't fly.

End of scene (location)

End of episode

Speaker: Leonard
Speech: There you go, Pad Thai, no peanuts.

Speaker: Howard
Speech: But does it have peanut oil ?

Speaker: Howard
Speech: Do I look puffy? I feel puffy.

🟥 Manual transcript track component
🟩 Episode Outline track component
🟦 Summary track component
⬛ Component common to more than one track (result of track alignment)

Alignment

Dynamic Time Warping (DTW)

Two tracks: $U_{1:N} \equiv \{u_{1,1}, \ldots, u_{1,N}\}$ and $U_{2,M} \equiv \{u_{2,1}, \ldots, u_{2,M}\}$

Global alignment $\mathscr{S}(i,j)$ between $u_{1,i}$ and $u_{2,j}$ is calculated as

MTR ⇔ SUB

SUB ⇔ ATR

$$\mathscr{S}(i,j) = \max \begin{cases} \mathscr{S}(i-1,j-1) + s(i,j) \\ \mathscr{S}(i-1,j) + s(i,j) \\ \mathscr{S}(i,j-1) + s(i,j) \end{cases}$$

1 if $u_{1,i}$ and $u_{2,j}$ are equal

Best path → backtracking from $\mathscr{S}(N,M)$ to $\mathscr{S}(1,1)$.

Alignment

Words in outlines may have limited overlap with words in manual transcripts
→ abstractive summarization

Global alignment $\mathscr{S}(i,j)$ between $u_{1,i}$ and $u_{2,j}$ is calculated as

$$\mathscr{S}(i,j) = \max \begin{cases} \mathscr{S}(i-1,j-1) + s(i,j) \\ \mathscr{S}(i-1,j) + s(i,j) \\ \mathscr{S}(i,j-1) + s(i,j) \end{cases}$$

OL ⇔ MTR

Cosine similarity between TFIDF vectors
+ context
+ scene location
+ wordnet

Best path → backtracking from $\mathscr{S}(N,M)$ to $\mathscr{S}(1,1)$.

Outline

① Corpus description

② Tracks alignment

③ How to use the TVD corpus?

④ Conclusions & future work

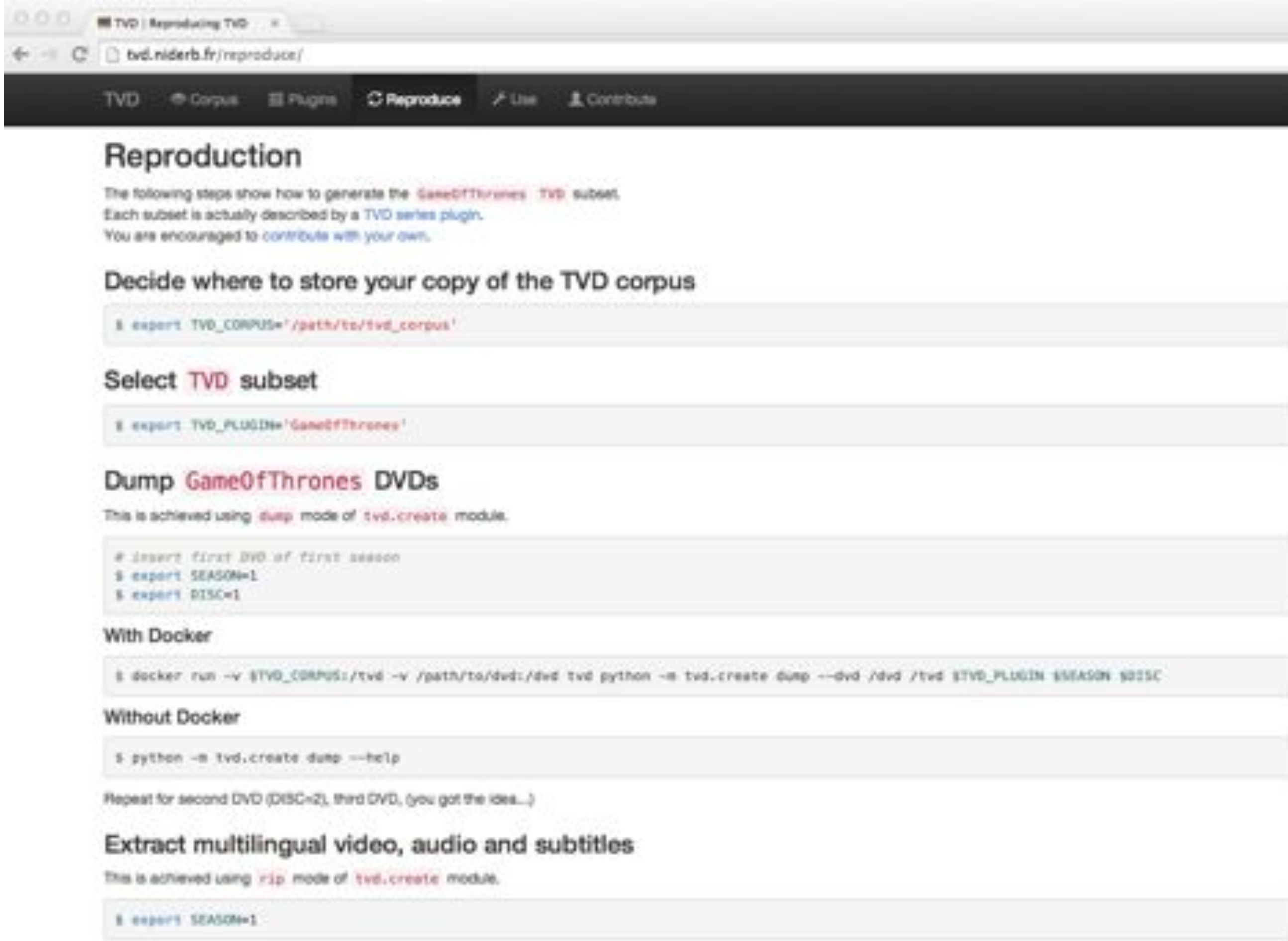
How to reproduce the TVD corpus?

tvd.niderb.fr

Contains all scripts to reproduce the TVD dataset locally

Scripts in Python

Docker image with every dependency pre-installed



Reproduction

The following steps show how to generate the `GameOfThrones` `TVD` subset.
Each subset is actually described by a TVD series plugin.
You are encouraged to contribute with your own.

Decide where to store your copy of the TVD corpus

```
$ export TVD_CORPUS='/path/to/tvd_corpus'
```

Select TVD subset

```
$ export TVD_PLUGIN='GameOfThrones'
```

Dump GameOfThrones DVDs

This is achieved using `dump` mode of `tvd.create` module.

```
# insert first DVD of first season
$ export SEASON=1
$ export DISC=1
```

With Docker

```
$ docker run -v $TVD_CORPUS:/tvd -v /path/to/dvd:/dvd tvd python -m tvd.create dump --dvd /dvd /tvd $TVD_PLUGIN $SEASON $DISC
```

Without Docker

```
$ python -m tvd.create dump --help
```

Repeat for second DVD (DISC=2), third DVD, (you got the idea...)

Extract multilingual video, audio and subtitles

This is achieved using `rip` mode of `tvd.create` module.

```
$ export SEASON=1
```

How to contribute to the TVD corpus?

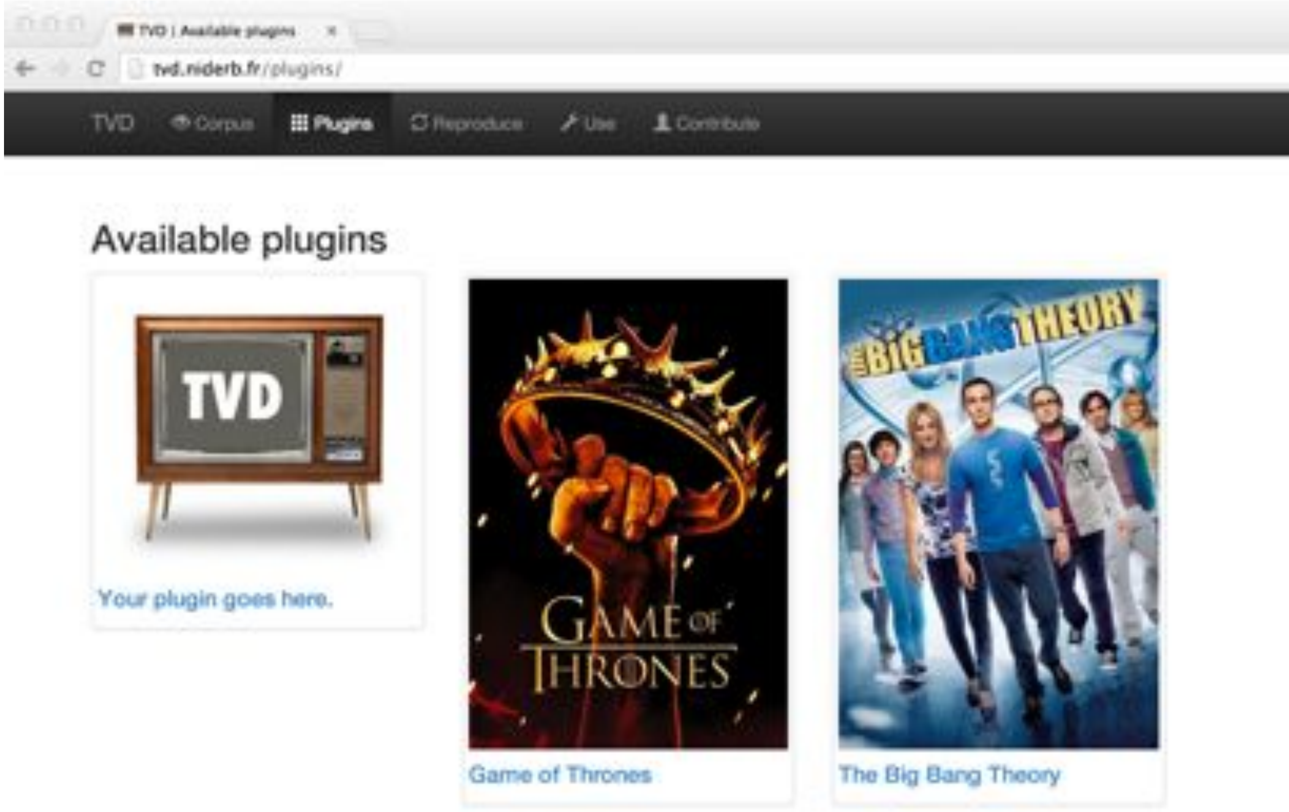
tvd.niderb.fr

Possibility to add new plugins

add new metadata to an existing plugin

Collaborative dataset

Join us !!



Available plugins

Your plugin goes here.

Game of Thrones

The Big Bang Theory

Outline

① Corpus description

② Tracks alignment

③ How to use the TVD corpus?

④ Conclusions & future work

Conclusions & future work

New multi-track TV series dataset

Computer scripts to locally regenerate the dataset from legally acquired DVD

Alignment of tracks in the dataset using:

- ✓ DTW
- ✓ context-dependent TFIDF
- ✓ scene locations
- ✓ WordNet

Conclusions & future work

Dataset that can be used for various application

✓ **Rich speech retrieval**

Rich speech acts (“X *invites* Y”, “X *tries to convince* Y”) are often explicitly mentioned in episode outlines but not in speech transcripts

→ Initial experiments based on this idea give promising results

✓ **Speaker diarization and identification**

✓ **Scene segmentation**

✓ **Automatic Summarization**

Conclusions & future work

Increasing the size of the corpus

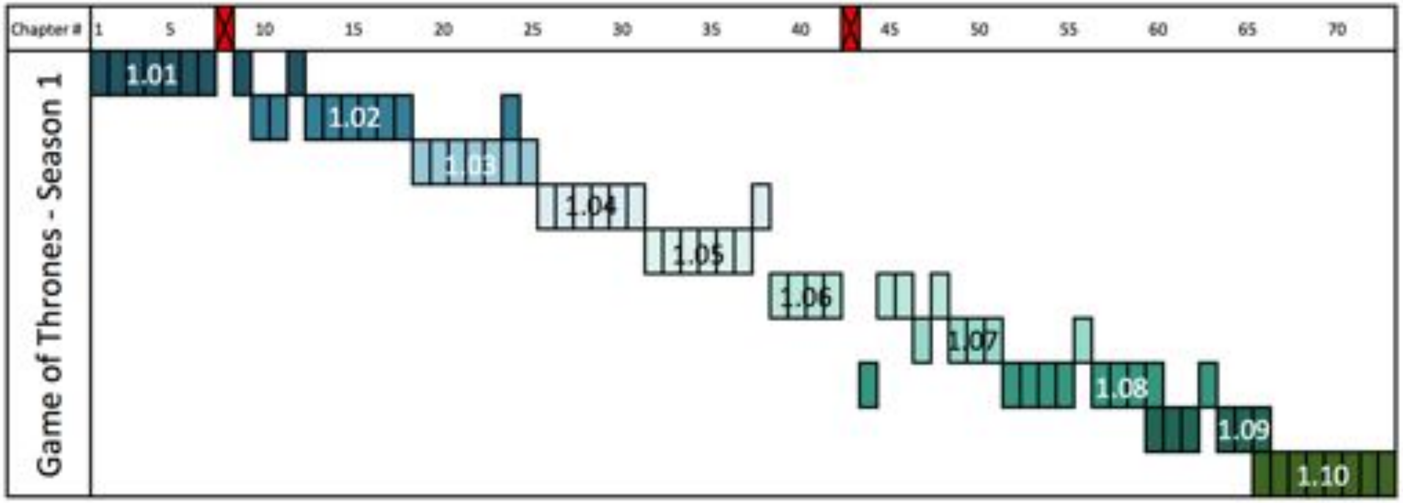
① **Adding new plugins**

✓ *Friends* – 10 seasons with summaries and multi-lingual manual transcripts

✓ *Real humans*

② **Adding information to existing plugins**

✓ Alignment between books and episodes in *Game of Thrones*



QUESTIONS?