
Translation and the Media

Alexandra Birch



Challenges facing the Media

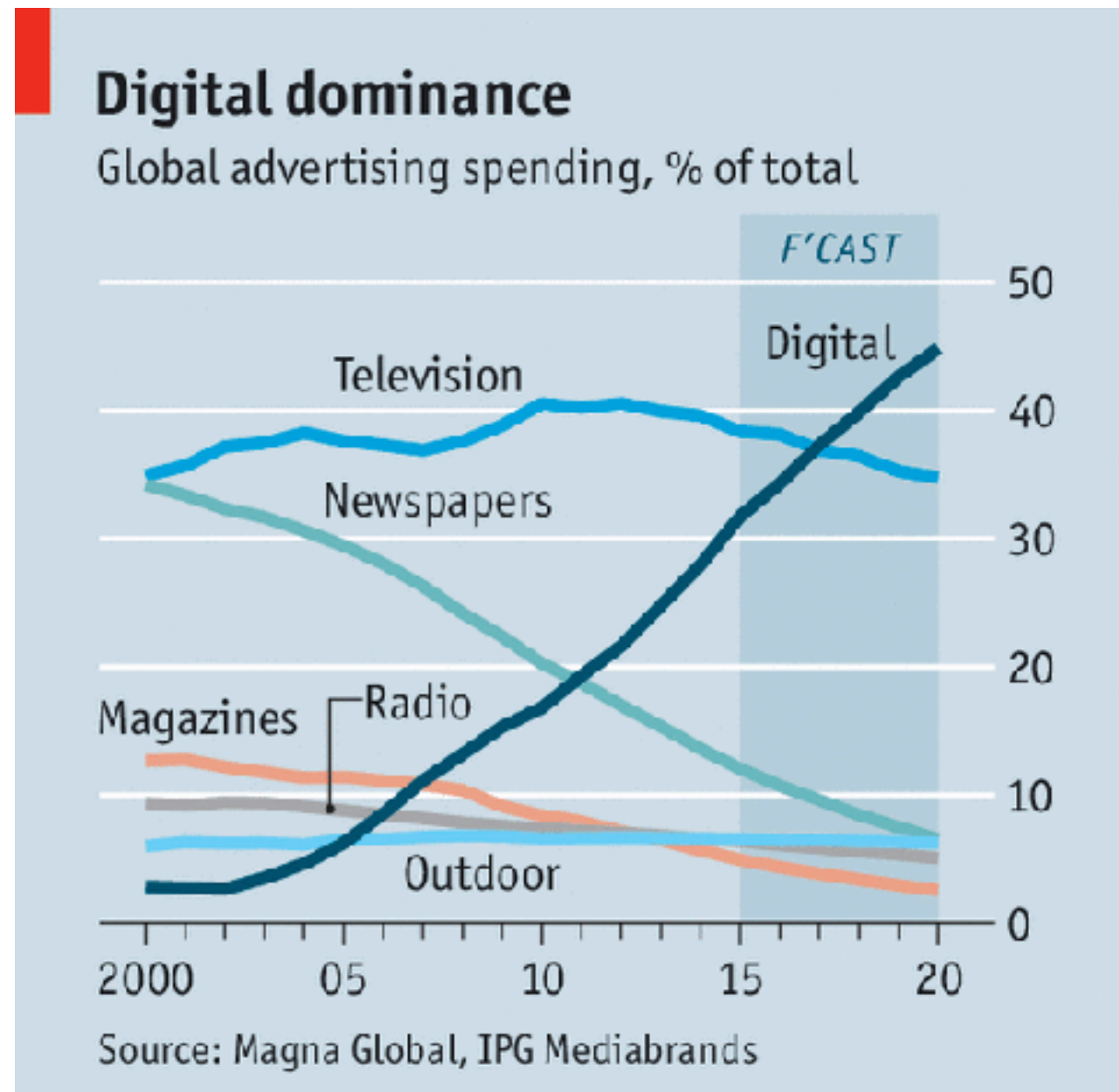


Jamal Khashoggi

Challenges facing the Media



Challenges facing the Media



Economist.com

Source: Economist

Types of Media Sources



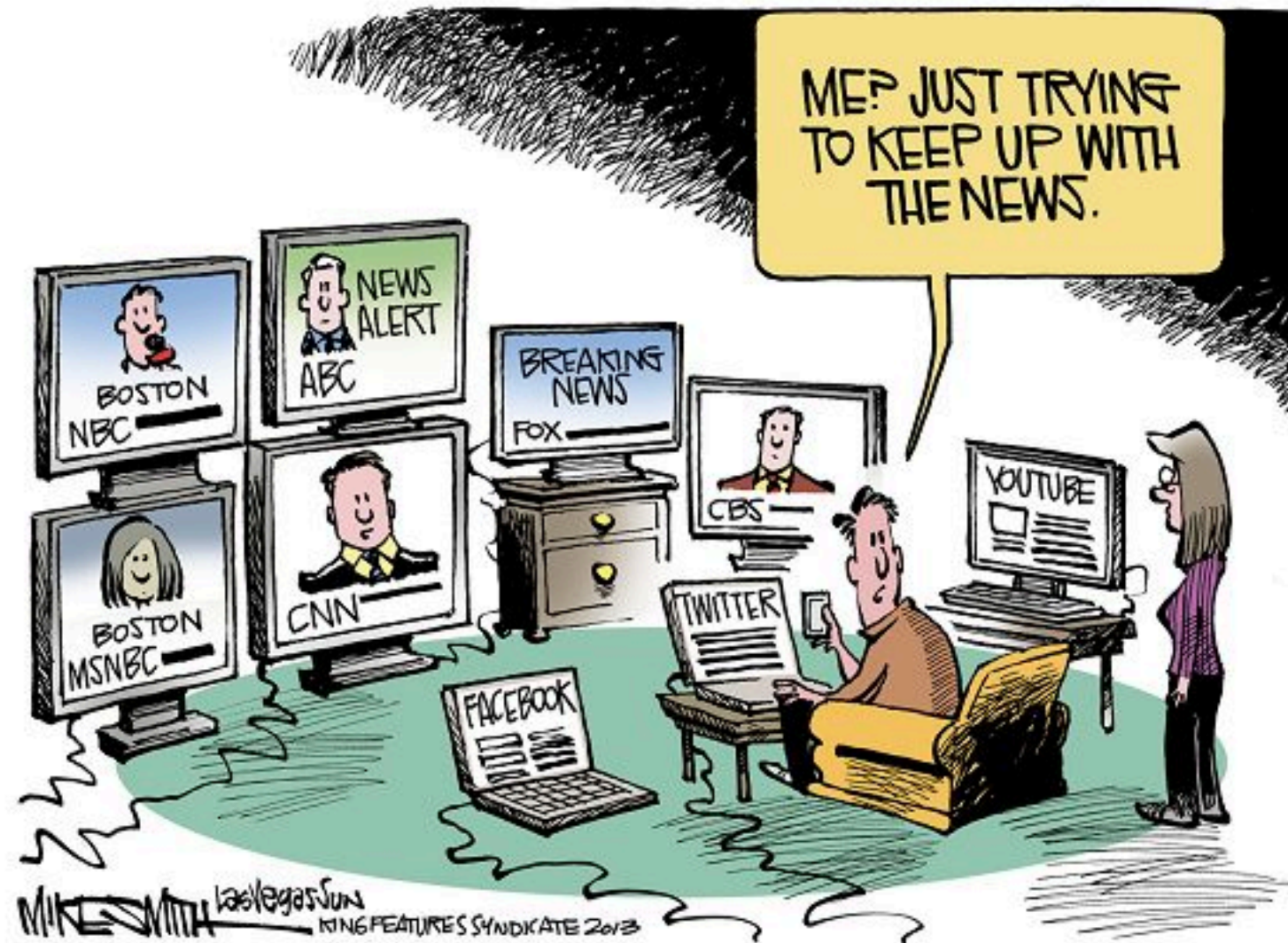
Speech Recognition, text normalisation

Diversity of Languages



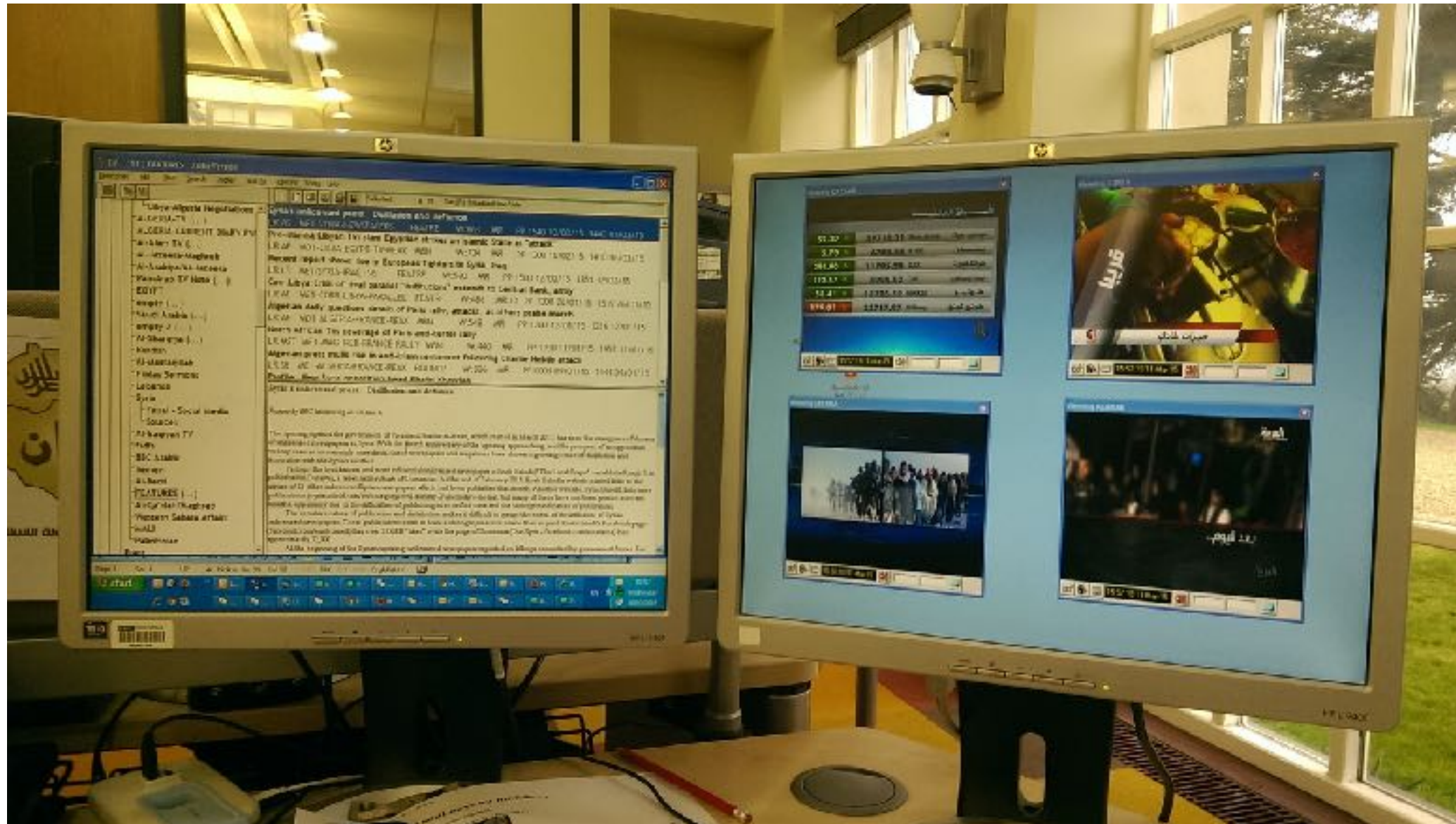
Machine Translation

Overload of Information



Natural Language Processing

BBC Media Monitoring



BBC Media Monitoring

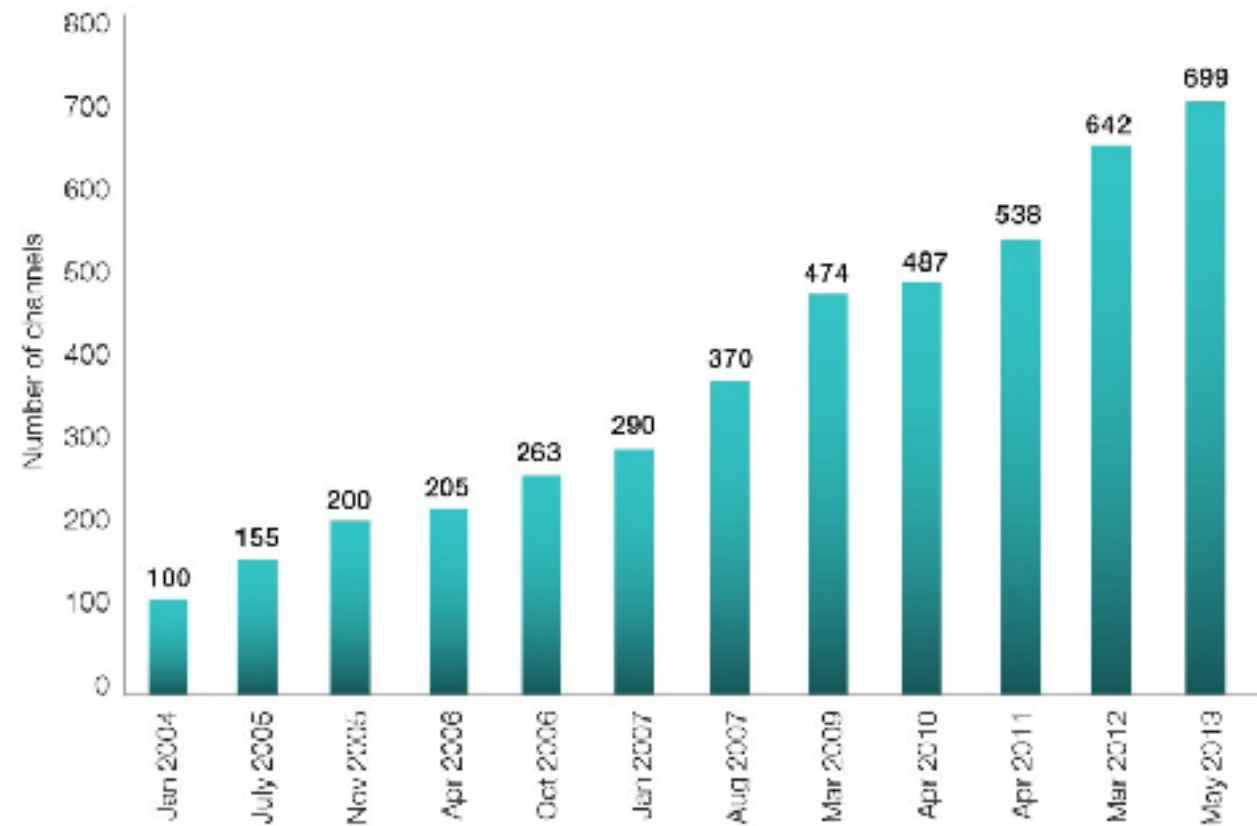


Scale of Media Sources

BBC Monitoring 2015

- 13,500 sources
- 1,500 Television broadcasters
- 1,350 Radio stations

Free To Air Channels in Arabic Countries



78% in Arabic

Source: Arab Advisors Group research



Questions:

- **How can NLP help the Media?**
 - ✦ Unlocking more content from different media types/languages
 - ✦ Making sense of deluge of data
 - ✦ Reach a wider audience
- **How can we deal with multilingual content?**
 - ✦ Translation
 - ✦ Multilingual NLP
- **Context of EU Projects:**
 - ✦ **SUMMA**: Automated Media Monitoring Platform
 - ✦ **GoURMET**: Low-resource Machine Translation



Scalable Understanding of Multilingual Media



H2020: 1 February 2016 – 31 January 2019



BBC Monitoring Use Case

BBCM Digital Toolkit



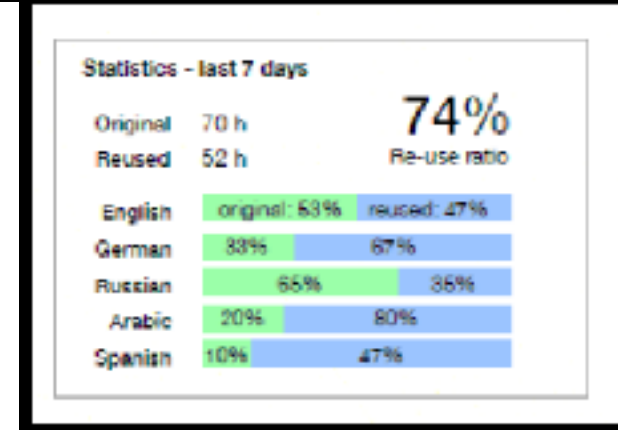
Aims
Robo eyes and ears
Find stories quicker
Alerts
Data journalism

Deutsche Welle Internal Monitoring Use Case



Aims
Monitor news
Compare/extract info
Reuse content
Analyse

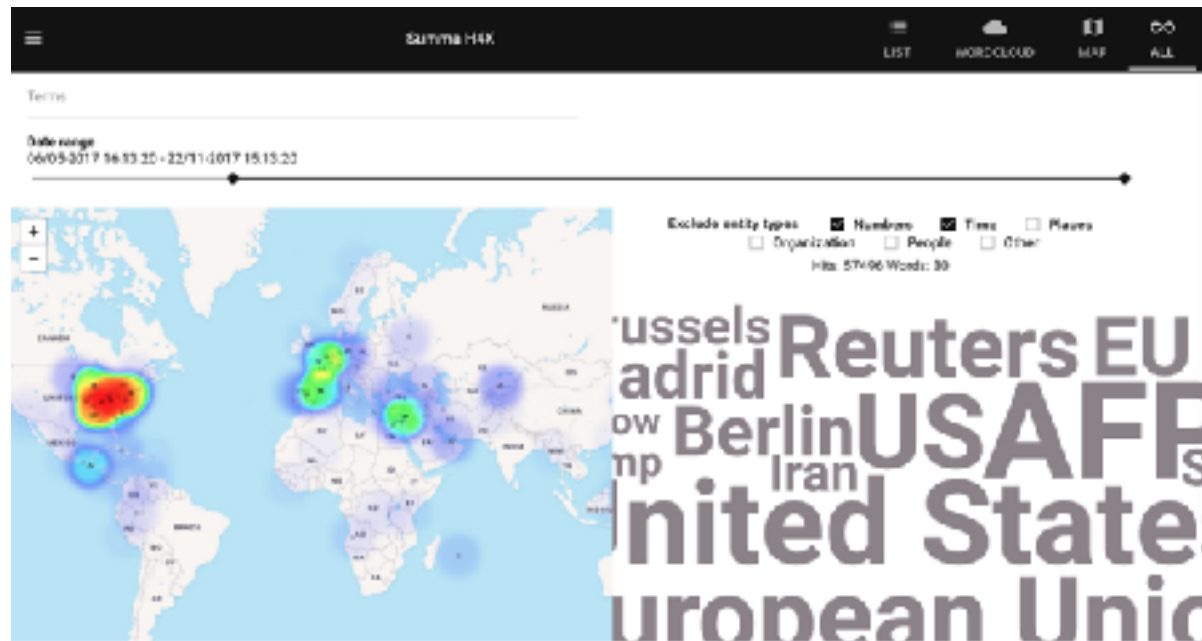
Monitor DW sources
8 SUMMA languages
Video, audio, text



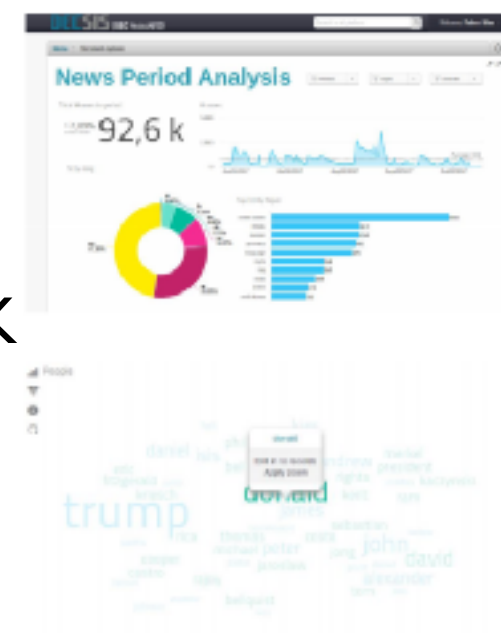
Data Journalism Use Case: NewsHacks

Paul Bradshaw:

Nose for news + scale and range of digital data



NRK



Decsis

Fact Extraction and VERification on Wikipedia

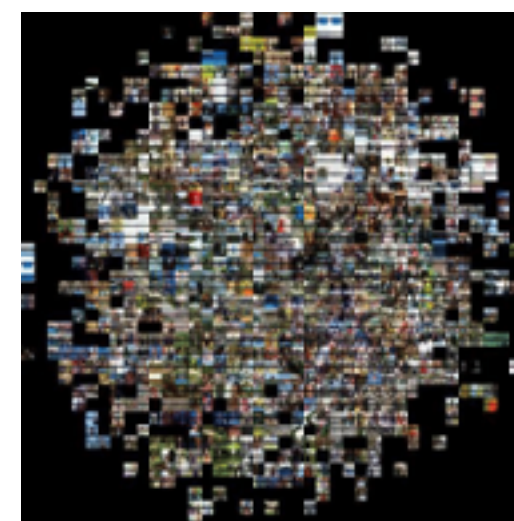
Claim:
The Reelkey King riots took place in the most populous county in the USA.

Evidence:
[wiki/Los Angeles Riots]: The 1962 Los Angeles riots, also known as the Rodney King riots, were a series of riots, looting, arson, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.
[wiki/Los Angeles County]: Los Angeles County, officially the County of Los Angeles, is the most populous county in the United States.

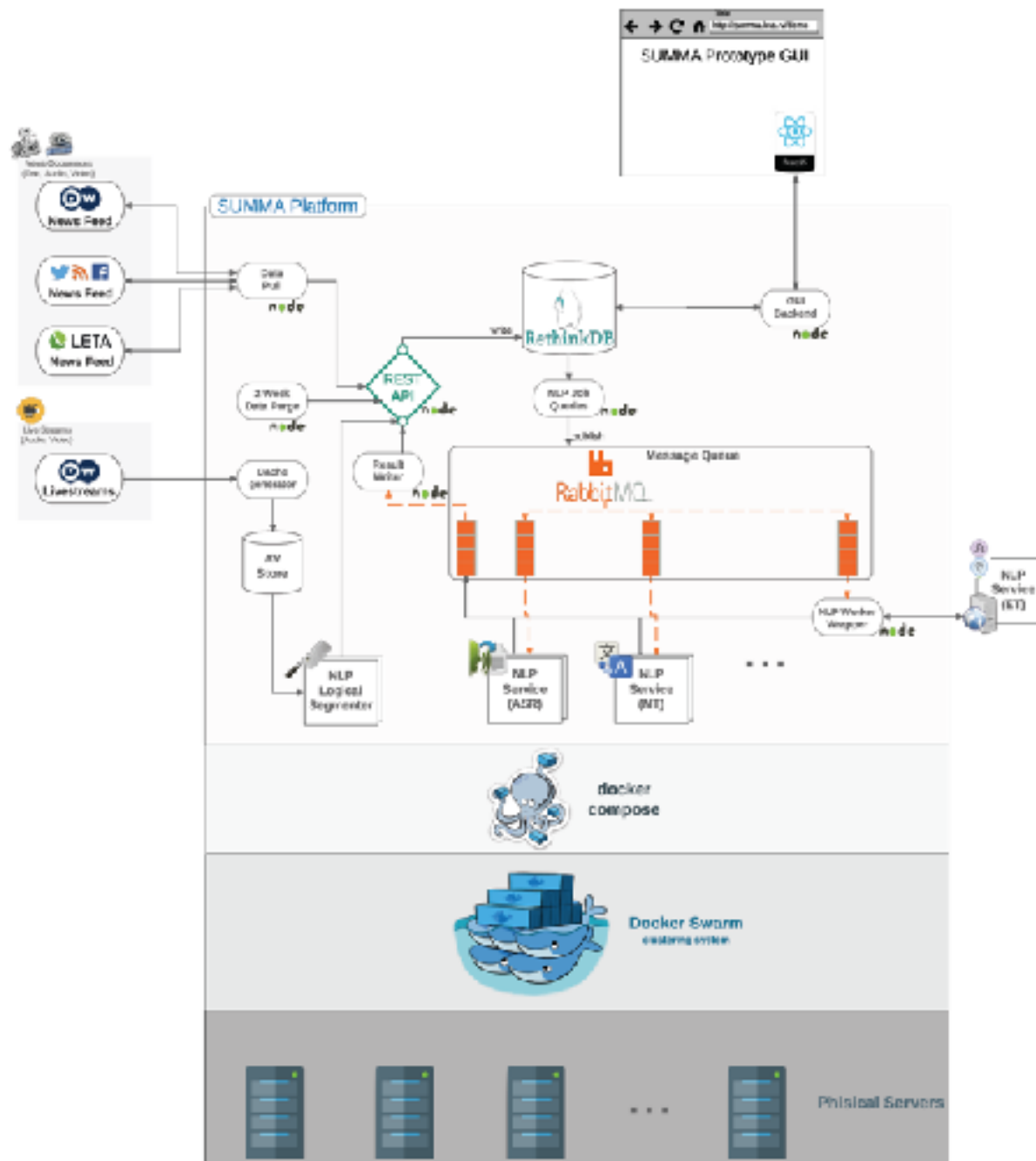
SUPPORTED

Given a claim, return whether it is SUPPORTED/REFUTED/UNVERIFIED according to Wikipedia. 2 stage process:

1. Find relevant passages for the claim
2. Decide the label for the claim given these passages.

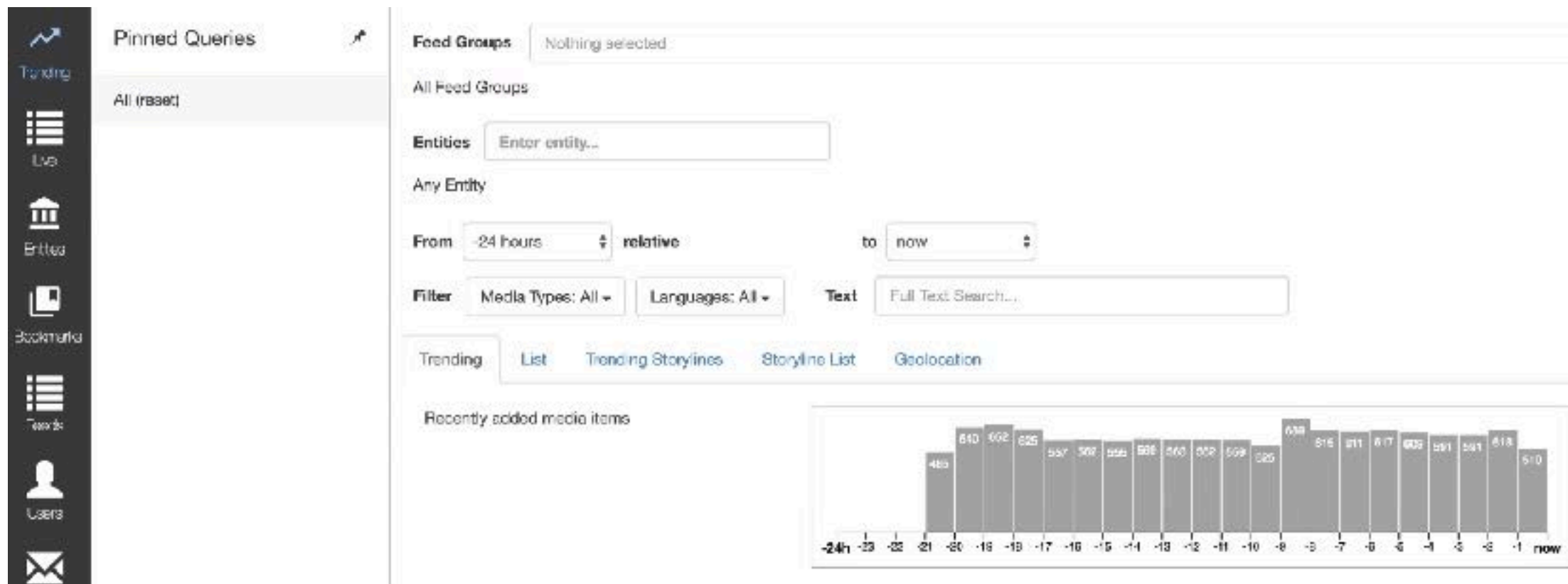


SUMMA Platform



- Microservices architecture
- All services are Docker containers
- Scalability is achieved by launching as many Docker container instances per task as required
- Open source platform and open source NLP Modules
<https://github.com/summa-platform>

SUMMA Platform: Scalability test



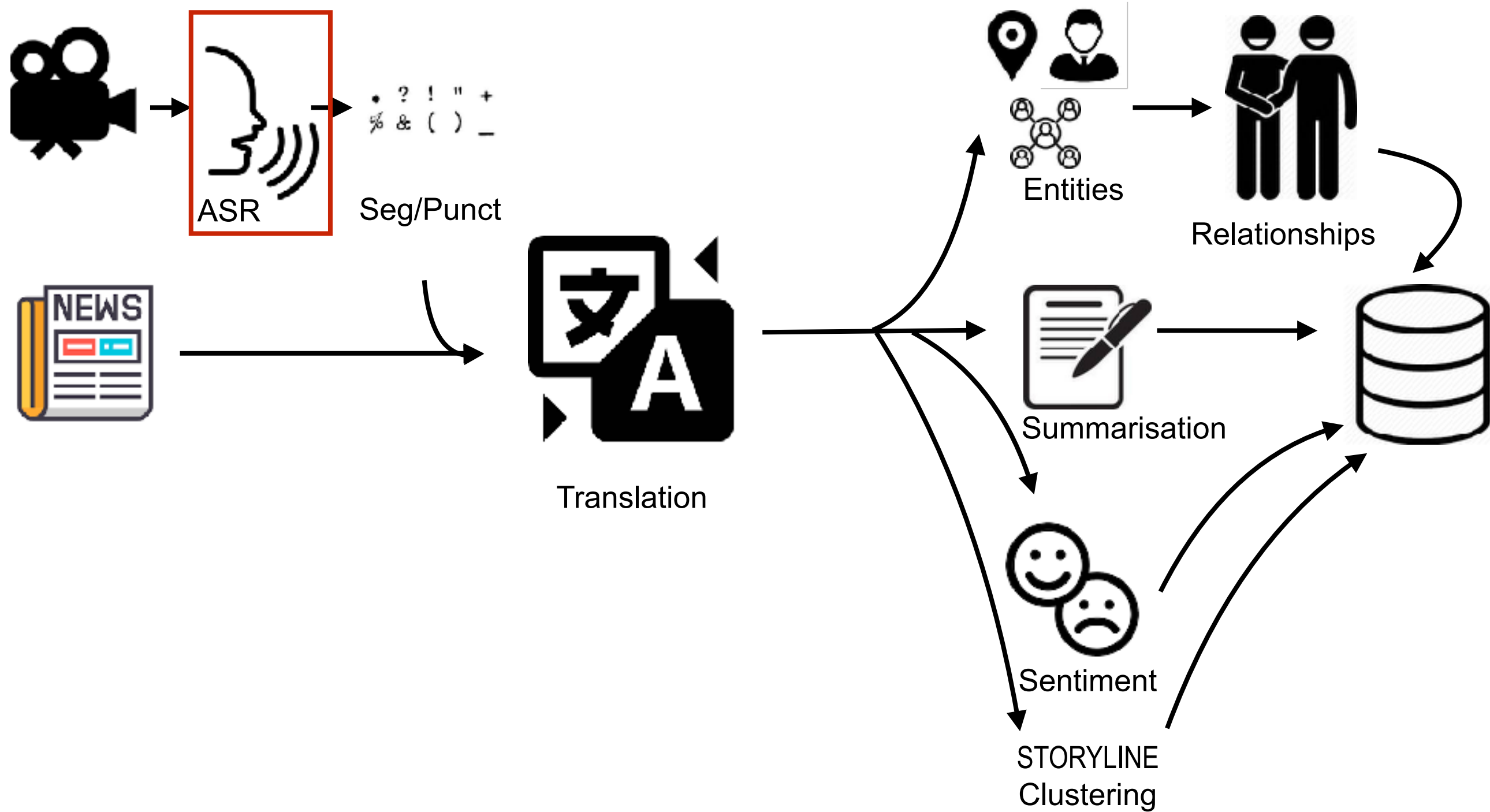
- Last week tested 24 hours of processing of 62 broadcast channels
- One AWS p3.8xlarge (244.0 GiB RAM, 32 vCPUs, 4 Tesla V100) to serve the data in HLS format
- Three AWS m5.24xlarge (384.0 GiB RAM, 96 vCPUs) to process it

SUMMA Channels



Country	Name	Language
Qatar	Al-Jazeera English	English
Iran	Al-Alam TV	Arabic
United Arab Emirates	Al-Arabiya TV	Arabic
Algeria	Algerian TV	Arabic
Algeria	Algerian TV Channel 3 (A3C)	Arabic
Iraq	Iraqia News TV	Arabic
Iraq	Al-Sharqiyah News TV	Arabic
Germany	German ARD TV	German
Korea (South)	Arirang TV	English
United Kingdom	BBC1	English
United Kingdom	BBC News Channel	English
United Kingdom	BBC Parliament TV channel	English
United Kingdom	BBC World Service	English
United Kingdom	BBC World TV	English
Belarus	Belarus 24 TV	Russian
Bahrain	Bahrain TV	Arabic
Egypt	Al-Misriyah TV	Arabic
China	China Global Television Network (CGTN, formerly CCTV)	English
Russia	Channel One TV	Russian
Russia	Channel One TV Worldwide (for Europe)	Russian
Algeria	Ennahar TV	Arabic
France	France 24 English Television	English
Lebanon	Future News TV	Arabic
Saudi Arabia	Al-Ikhabriyah TV	Arabic
Iran	Islamic Republic of Iran News Network	Persian
Iran	Vision of the Islamic Republic of Iran Network 1	Persian
Iran	Vision of the Islamic Republic of Iran Network 2	Persian
Kuwait	Kuwait TV 1	Arabic
Lebanon	LBC Sat TV	Arabic
Libya	Libya 218 TV	Arabic
Libya	Libya's Channel TV	Arabic
Libya	Libya al-Hadath TV	Arabic
Libya	Al-Rasmiyah (Official) TV	Arabic
Libya	Libya al-Mostakbal TV	Arabic
Morocco	2M Monde	Arabic
Lebanon	Al-Manar TV	Arabic
Mauritania	Mauritanian TV	Arabic
Lebanon	Al-Mayadeen TV	Arabic
Morocco	Al Aoula TV	Arabic
Libya	Al-Naba TV	Arabic
Egypt	Nile News TV	Arabic
Russia	NTV Mir	Russian
Oman	Oman Television	Arabic
Ukraine	One Plus One TV	Ukrainian
Palestinian Territory, Occupied	Palestinian Satellite Channel TV	Arabic
Iran	Press TV	English
Qatar	Al-Jazeera TV	Arabic
Qatar	Qatar TV	Arabic
Russia	RTR Planeta TV	Russian
Russia	RT	English
Russia	Rusiya al-Yawm	Arabic
Yemen	Yemen TV	Arabic
Yemen	Republic of Yemen TV	Arabic
United Arab Emirates	Sky News Arabia	Arabic
Syria	Al-Ikhabriyah al-Suriyah TV	Arabic
Syria	Syria TV	Arabic
Tunisia	National Tunisian TV	Arabic
Libya	Libya (Al-Ahrar) TV	Arabic
Ukraine	ICTV television	Ukrainian
Ukraine	5 Kanal TV	Ukrainian
Russia	Rossiia 24 news channel	Russian
Russia	Zvezda TV	Russian

SUMMA Pipeline



Unlocking Audio Visual Content

Media Item | Trending Stor

Not Secure demo.summa-project.eu/#/trending/media-items/0169ad96-e583-4460-b5d8-29742b346b67

Media Item

DW Arab Live Street Chunk.

Added 4 days ago (2018-10-26 12:16 UTC)
 Changed 4 days ago (2018-10-26 12:16 UTC)
 Source: [http://www.dw.com/en/egyptian-activist-asks-for-justice-for-her-son/50288544](#)

Topics:
 history (5.8%) terrorism (4.2%) united states (2.7%) conflict (2.2%) arts (2%) israel (1.8%)
 opinion (1.7%) syria (1.3%) culture (1.3%) society (1%)

Summary:
 A young tribe needs support, especially from the younger generations, because the struggle for the ancestral home and, will decline the future of them, as well as the ascending pace of events in the digital world, your program helps you to maintain a clear vision and show, for you, the latest developments are what deserve knowledge, and what causes laughter, as well as the facilities of industry, and the consumer, moments, my return, celebrates the first International Day, its new use, completely electric power, and the death of a death, is it an accident?

Transcript:
 Original Live (ar) English Original (en)

كناخ قوله لوى الى الدم وبخاصة من الأوجال الثانية لأن الصراع من أول وهن الأجداد سوف يتورس مسائل هم أيضا ...
 رفعة الأحداث المتسارعة في العالم الرقمي يرتكز بناءهم على الثقافة على روية وانسجام ويرض لهم تحت
 التطورات هو ما ينعين المعرفة وما يثر الحسنة كما ورائق السعادة والسعادة لطول عهدي نعتك. نعتك العلي الأثرى جازها
 البنية موت شريفة والكامل وريحية وانا من هي سيرة كهرتية نعم في جزايا جيدة ليد الانتشار هم مسيرهم حسن خيرة بلغة
 في اعتدال العافية مع أربعة منجوعين وفخية الأثرى لوجين الأثرى، بوجه خيرة إننا لوجية أجداد الحد من الفلاح القوي في
 العلم كود أجواء الحرب القوية وما هي كعالت الانتحاب الأثرى، من حد تصاريخ القوية لوجوا بعد كين أريه بعد كين
 وأربعين فلقا على ذلك اليوم على لانية وهي كل مكان لعتار [...] يمكن تمثيلها أيضا أيضا ثلثة خيرة تقصن أيضا
 أيضا نعت العافية مع عتيدة من نعتي من تحرب سوي الأثمن [...] في كثرين القين وعبدة على حررت السبية كفتنا
 إلى هذا المكان حيث يوجد لنس أن [...] حمة نعت من توفير على ثلثة [...] الحية في العالم
 الرقمي وحد برادون هذا صور أثار والأيقون القوية القوية الكبرية لثلاثة لثلاثة على الأيقون القوية بارة ما صجوجت الهدية
 المعنوية القوية نعتي بعتوا بالاحياء على معلومات من الانترنت من مواقع لوقاب الجراء القوية القوية حرض سمرحة في خدمة
 المصمم والنشط [...] سبب القبان له شكك ميسون ملك ميسون لعدة ثلثة نعتي تقرها [...] شكك لعتنا
 فريد من نوعة بصمات الأصابع تكلك لأمل إعتدة نعتل لبرامج الحرب تكلم مجموعة الهندسة المعمارية القوية تكلم سنج بلقتان
 وذلك كشارب كمن كمن بوقت وقوع الهجوم نعتي، المجرحة لدم حروقة الهجوم هو كمال بومان بومان وثقة نعتك صور
 الإقرار الأمطاحة القوية القوية القوية القوية [...] حرو شهود القبان يتم لوجوا في صور 000 الأثرى
 وبصاف إليها صير

Previous Chunk Media Item **Next Chunk Media Item**

Media Item	Topics	Time Added
DW Arab Live Street Ch...	egypt, opinion, conflict, pro...	- 5m 2018-10-26 02:11 UTC
DW Arab Live Street Ch...	history, terrorism, united ...	2018-10-26 02:16 UTC
DW Arab Live Street Ch...	internet, social media, ch...	+ 5m 2018-10-26 02:21 UTC
DW Arab Live Street Ch...	culture, arts, history, media...	+ 10m 2018-10-26 02:26 UTC
DW Arab Live Street Ch...	society, music, united state...	+ 10m 2018-10-26 02:31 UTC

Logged in as **Alexandra**
 Feedback

ASR across many languages

Language	Train Set	Train Size	WER In-domain	WER SUMMA
English	Multi-Genre Broadcast Challenge	1600 hours	26.1	
Arabic	Arabic Multi-Dialect Broadcast Challenge	1200 hours	14.7	
German	BNC radio news	160 hours	9.9	34.6
Spanish	Global Phone	17 hours	13.2	
Russian	Euronews	60 hours	14.4	

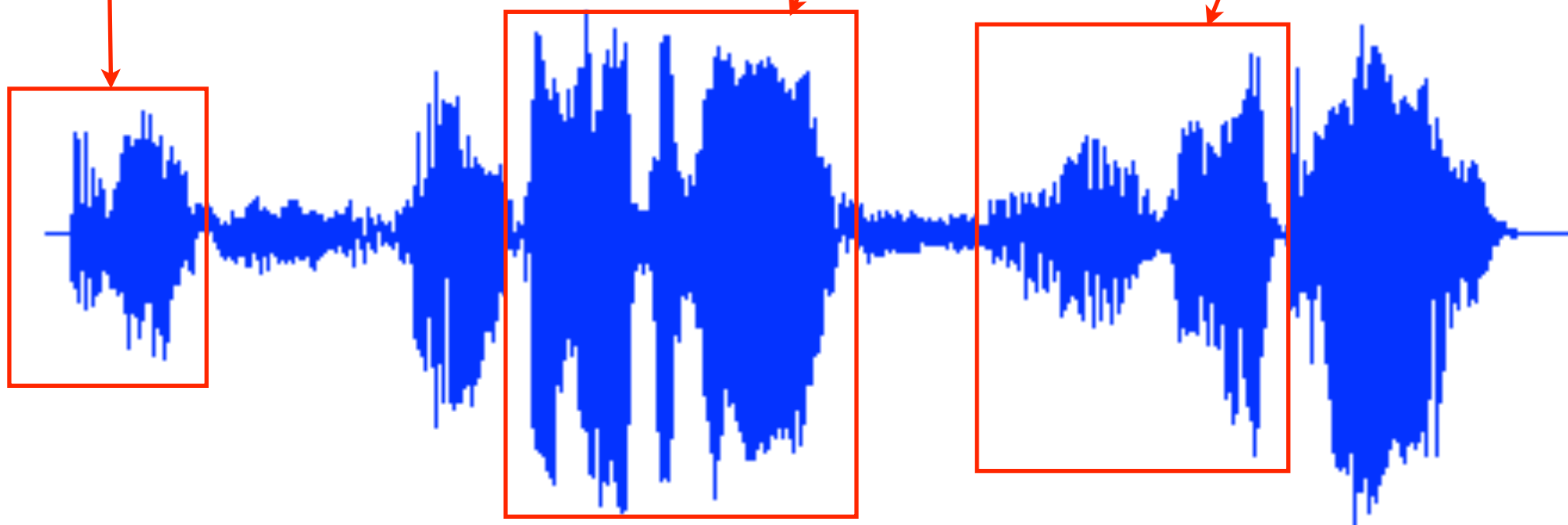
Added: Latvian, Ukranian, Farsi, and Portuguese

Training ASR Using Subtitles

Light supervision

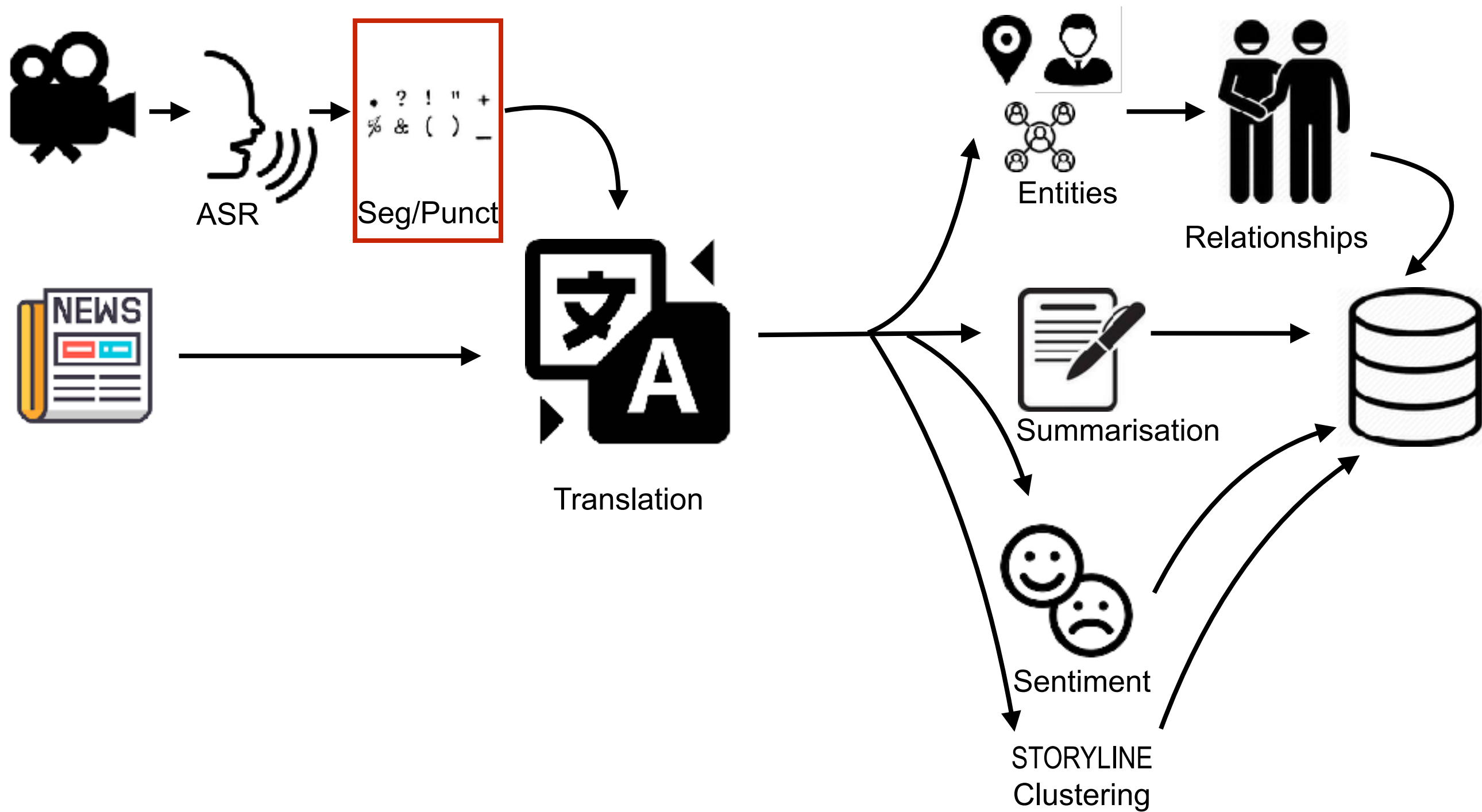
The heaviest rain is going to fall in and around Manchester

THE HEAVIEST TRAIN IS GOING TO FALL IN A MY MANCHESTER
AND THE SURROUNDING AREA

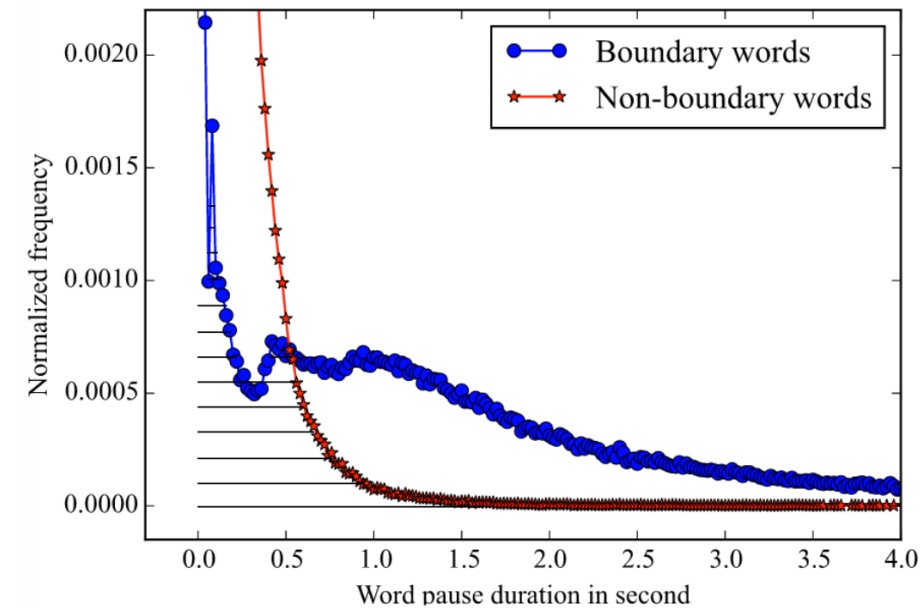
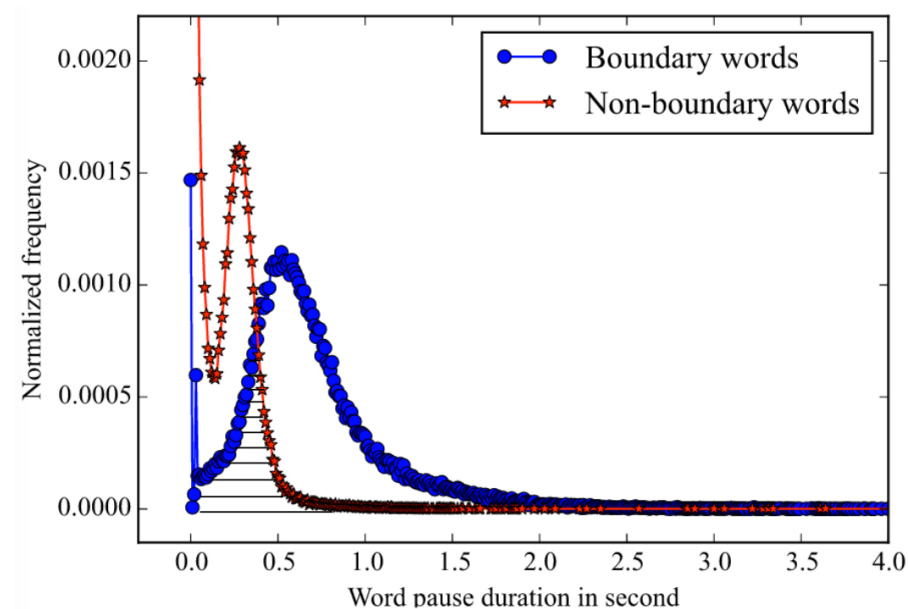


Matching Error Rate used as cut-off

SUMMA Pipeline



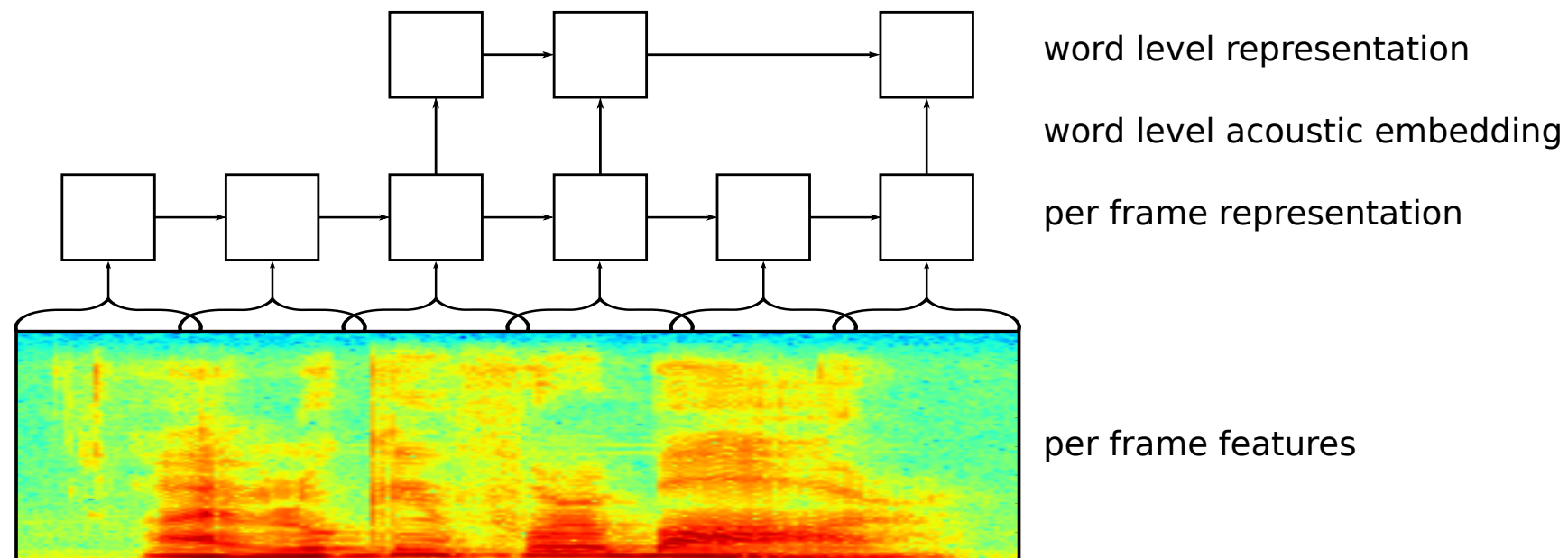
Dealing with ASR Output: Segmentation



Pause duration distribution on German BCN corpus (left) and English MGB corpus (right)

Yang Wang, Alexandre Nanchen, Alexandros Lazaridis, David Imseng, and Philip N.Garner.
Comparative study on sentence boundary prediction for german and english broadcast news.
Idiap-RR Idiap-RR-18-2017, Idiap, 7 2017

Dealing with ASR Output: Punctuation



Hierarchical encoder

In practice:

- use pause segmentation model to segment on long pauses
- use punctuation model trained on large amounts of text to add: , . ? !

Ondrej Klejch, Peter Bell, and Steve Renals. Sequence-to-sequence models for punctuated transcription combining lexical and acoustic features. ICASP, April 2017

SUMMA NLP Pipeline

Media Item

((pw)) Jazzer TV.

منه 21:33 نوبتي" إلى التخلي عن أي تشكيلات عسكرية أو أمنية | المجلس الانتقالي الجنوبي اليمني ال

Previous Chunk Media Item

Topics

news (7.6%) | oil (4.3%) | arab world (2.9%) | middle east (2.1%) | saudi arabia (2.1%) | international relations (2.0%) | africa (2.0%) | politics (2.0%) | economy (1.9%) | environment (1.5%)

Summary

Transcript

Original (he) | English | Original (ar)

إن الروس وسريين افترقا على بيع كميات إضافية في السوق بعد، وهذا لا يعني أنهم يبيعون الغاز الروسي إكنا ببيع النفط، وقد قلده ويوافق من وزير الطاقة السعودي، فوه إن الرئيس زعمه إقليمية الفتح إلى حارة جالين وسيمتد إلى برميل في كوروا قلبا طلب المستهلكين الفخشي إن سوق النفط العالمية التي إنداك هذا حاليا من جهة قبل بيوتك يوافق كوروا إن روسيا والسعودية... إن الرئيس زعمه إقليمية الفتح إلى حارة جالين وسيمتد إلى برميل في كوروا قلبا طلب المستهلكين الفخشي إن سوق النفط العالمية التي إنداك هذا حاليا من جهة قبل بيوتك يوافق كوروا إن روسيا والسعودية... إن الرئيس زعمه إقليمية الفتح إلى حارة جالين وسيمتد إلى برميل في كوروا قلبا طلب المستهلكين الفخشي إن سوق النفط العالمية التي إنداك هذا حاليا من جهة قبل بيوتك يوافق كوروا إن روسيا والسعودية...

SUMMA NLP Pipeline

- Home
- Live
- On Demand
- Recordings
- Library
- Logout

Media Item Added 22 minutes ago (2019-01-24 07:51 UTC)

((pw)) Jazzer TV. Source: http://www.24-7news.com/News/World/2019/01/24/07:51/170

Topic

energy (5.6%) | oil (4.3%) | oil world (3.5%) | middle east (3.1%) | saudi arabia (3.1%) | international relations (2.8%) | africa (2.4%) | politics (2.2%) | diplomacy (1.5%) | environment (1.2%)

Summary:

Transcript

Original (en) English Original (ar)

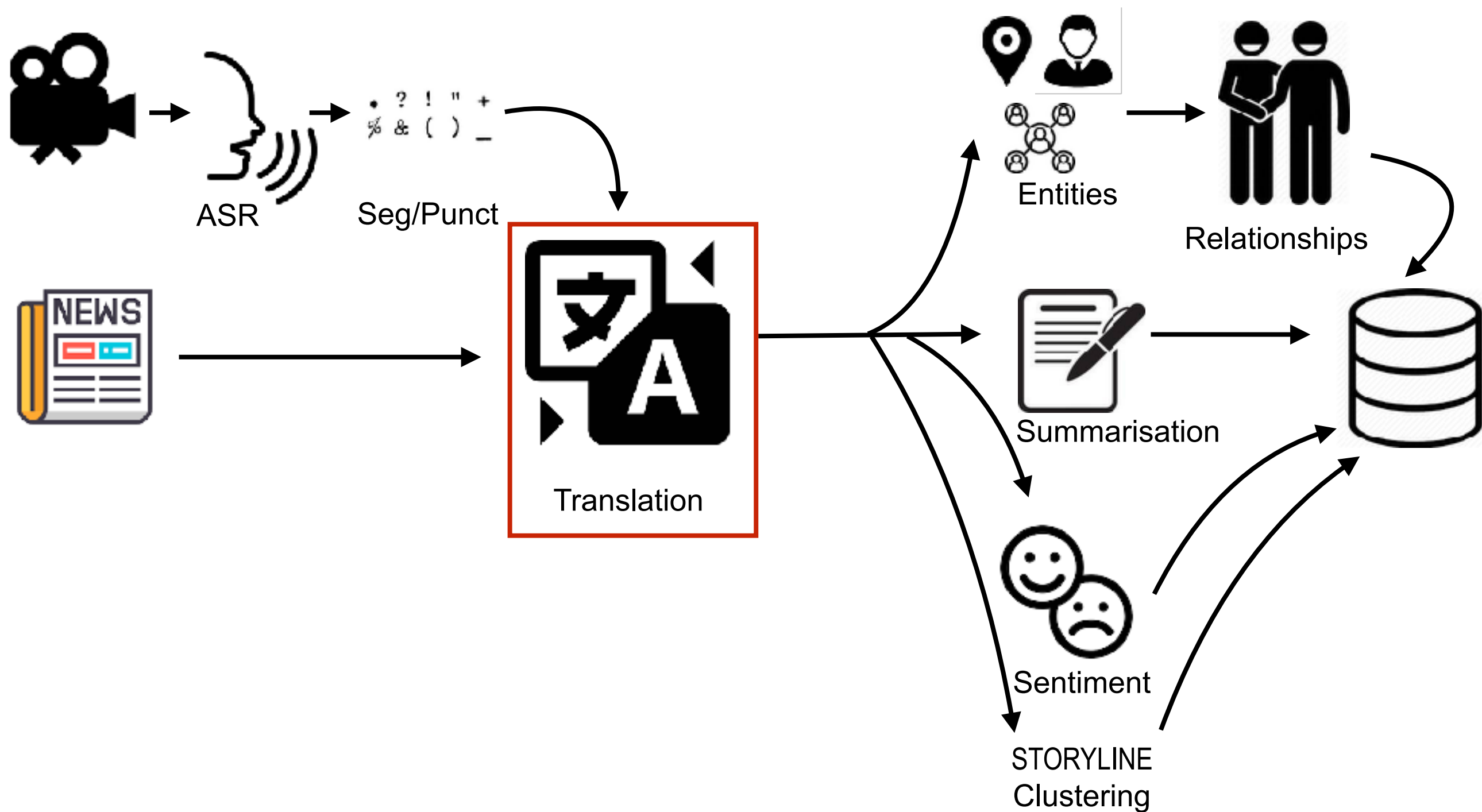
The Russian and Saudi agreed to pump additional quantities on the market quietly and not seem to be listening to the President's order, to pump more, and it caused the **Minister of energy** the **Saudi Arabia** said that **Riyadh** increased its oil production to 10,000 barrels in October to meet the growing consumer demand and that the international oil market is currently getting a good supply from the work of market. For his part, Russian President **Vladimir Putin** announced that his country may increase its oil production by a range of 200 to 300,000 barrels a day to make up for the possible shortage in the past and during a conference in **Moscow**, he was attended by senior officials in the international energy sector, including Saudi Minister of energy **Khalid Al-Falih**, President **Putin's** campaign against his American counterpart, without paying attention to the rise in oil prices. Geopolitical factors influence the price rise, this is truly effective, better not interfering, in dealings with the market and not resorting to the use of political tools in order to obtain competitive advantages. It is also preferable to move away from oil to other products to stabilize prices. The oil, as it was in the Soviet era, because this will not lead to positive results, the **Emir of Qatar**, **Sheikh Tamim bin Hamad Al-Thani**, is currently holding talks with the President of **Paraguay**, **Maria Ines**, at the **Recognition Palace**, dealing with bilateral relations and regional and international issues. The two sides expected to sign agreements aiming at strengthening relations between the two countries. **Sheikh Tamim bin Hamad Al-Thani** arrived in **Paraguay**, the third station on his tour in **Latin America**, coming from **Doha**. He held talks with the president, **Maria Ines**, in the **Kaakeri**, focused on bilateral relations and the development of cooperation between the two countries in several fields joining us from the capital of **Paraguay**, correspondents of **Al-Jazeera**, and he heard more details about him, the visit of the **Emir of Qatar to Paraguay** the talks are there, and a means at present, the expanded meeting is still holding negotiations between the **Emir** of the state of **Qatar**, **Sheikh Tamim bin Hamad Al-Thani**, the President of the **Paraguay Republic**, these talks, as you mentioned, deal with bilateral relations as well as issues regional and international after a few minutes from now, bilateral agreements will be signed between the governments of the two countries, and we have observed that this approach, which is the signature of bilateral agreements, is repeated in the statements of the **Emir of Qatar** accept the visit in **Ecuador**, as for the second one, currently, in **Panamá**, the individual, of course, all of these agreements are aimed at deepening, maintaining cooperation between the countries of **Qatar** and these countries in **Latin America**, and specifically cooperation in the economic and commercial aspects of tourism, health and culture, also, there was a remarkable thing, it is the handing over of the **Emir of Qatar** to the keys to the capitals of these countries less than an hour before the arrival of the **Emir of Qatar** to the airport is, in the capital in the capitals of the individual or the **Talibani**, the mayor handed over the **Emir of Qatar** the keys to the capitals before he heads here, where we are now at the **Republican Palace** to meet with the president of **Paraguay**. Of course, it is known to **Paraguay**, that its president who received the decision before, is, before, in the middle of last August, had decided before days of the embassy's embassy from **Jerusalem** to **Tel Aviv**, after it was only three and a half months after it was transferred to **Jerusalem**, the President decided to re-transfer it to **Tel Aviv** according to the explanation of the **Paraguay** with a view to intensifying peace process efforts for the **Middle East** peace process, thank you very much, and my standing, we were with us from the **Republican Palace** in the capital of **Paraguay**, the UN mission was continued with you, and in it, after the disaster of the Indonesian **Sulawesi** earthquake and tsunami, the remote areas are waiting for aid.

Alexandra Birch

MT and the Media

28

SUMMA Pipeline



Challenges for Translation in the Media

- Efficient, fast translation of large volumes
- High quality translation
- Story level translation
- Robust to different dialects
- Evaluation of for gisting



Challenges for Translation in the Media

- **Efficient, fast translation of large volumes**
- High quality translation
- Story level translation
- Robust to different dialects
- Evaluation of for gisting



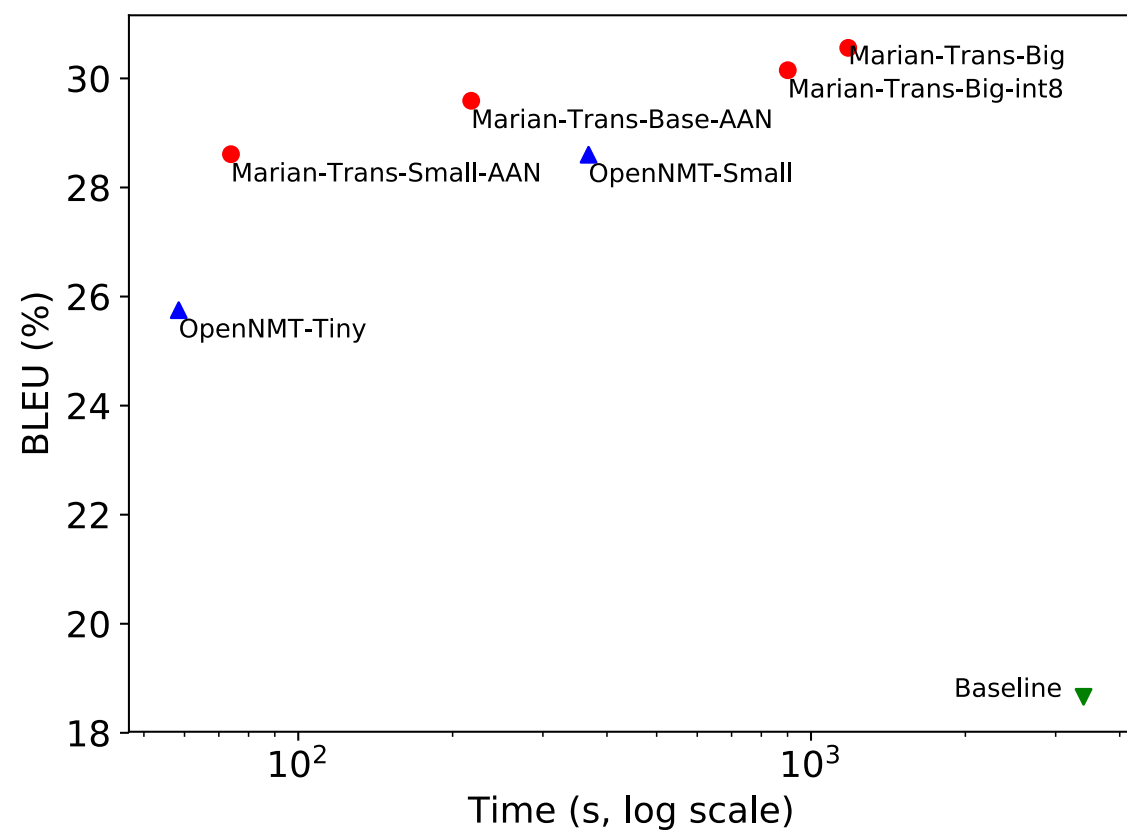
Marian

- Portable C++11 code with minimal dependencies (only CUDA and Boost)
- Custom auto-diff engine with dynamic graphs (similar to DyNet)
- Optimized towards NMT
- Multi-device, multi-node training, and decoding (GPU and CPU)
- Websites:
 - ✦ <http://marian-nmt.github.io> and
 - ✦ <https://github.com/marian-nmt/marian>

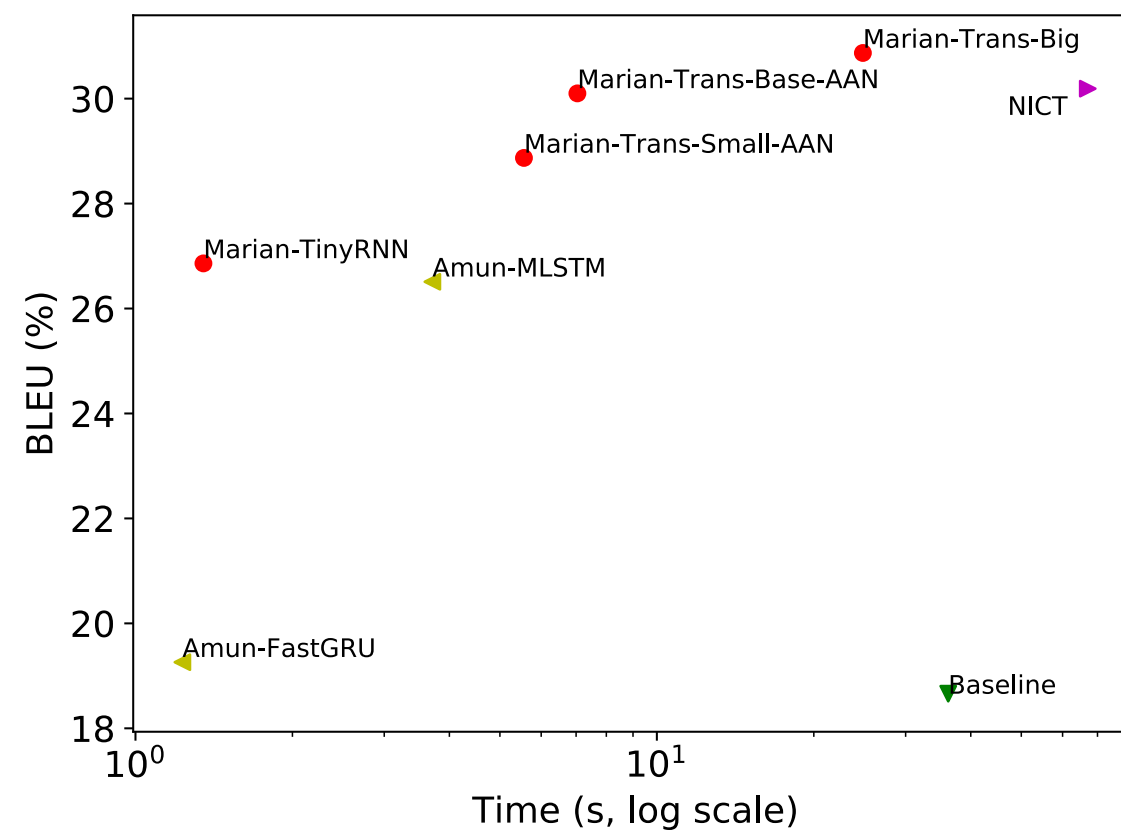
Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F. T. Martins, and Alexandra Birch

Marian: Fast Neural Machine Translation in C++. ACL 2018

WNMT 2017: Shared task on efficiency



(a) CPU Time vs. Accuracy



(b) GPU Time vs. Accuracy

Alexandra Birch, Andrew Finch, Graham Neubig, and Yusuke Oda (2018). Findings of the Second Workshop on Neural Machine Translation and Generation. ACL

Mentions of NMT toolkits: WMT proceedings



- Marian 169
- OpenNMT 89
- Tensor2tensor (+T2T) 59
- Sockeye 47
- Nematus 42
- Fairseq 8

Source: Marcin Junczys-Dowmunt

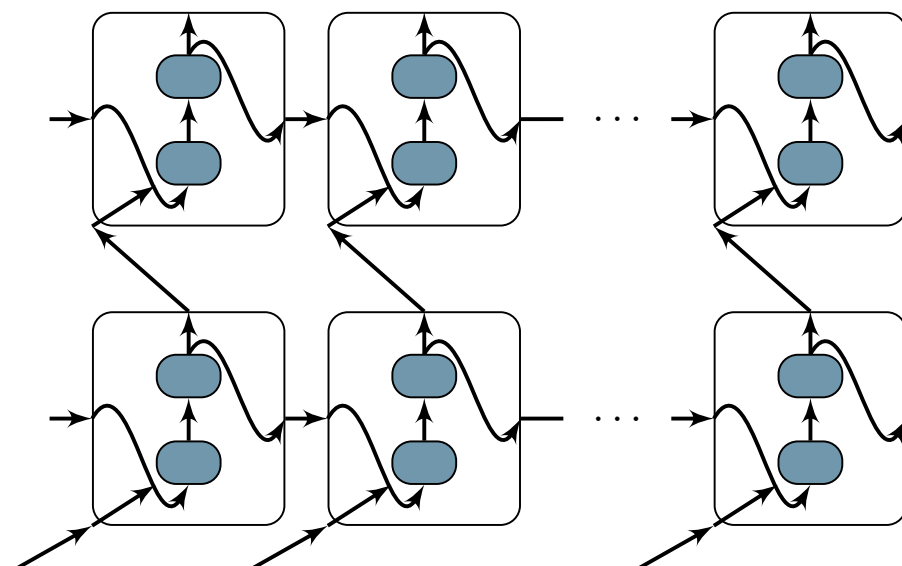
Challenges for Translation in the Media

- Efficient, fast translation of large volumes
- **High quality translation**
- Story level translation
- Robust to different dialects
- Evaluation of for gisting

Deeper Models

2017 WMT News Task

Constrained news task, including ties:



Bi-deep

	From English	To English
Turkish	Edinburgh	Edinburgh
Czech	Edinburgh	Edinburgh
Chinese	Edinburgh	Edinburgh
German	Munich	Edinburgh
Russian	Edinburgh	National Research Council
Latvian	Tilde	Edinburgh
Finnish	Helsinki	UPC

(Edinburgh did not participate in Finnish)

Antonio Valerio Miceli Barone, Jindrich Helcl, Rico Sennrich, Barry Haddow and Alexandra Birch. "Deep Architectures for Neural Machine Translation". In Proceedings of WMT, 2017

Rico Sennrich, Alexandra Birch, Anna Currey, Ulrich Germann, Barry Haddow, Kenneth Heafield, Antonio Valerio Miceli Barone, Philip Williams (2017). The University of Edinburgh's Neural MT Systems for WMT17

Challenges for Translation in the Media

- Efficient, fast translation of large volumes
- High quality translation
- **Story level translation**
- Robust to different dialects
- Evaluation of for gisting



Co-reference test set

Source:

context: Oh, I hate **flies**. Look, there's another one!

current sent.: Don't worry, I'll kill **it** for you.

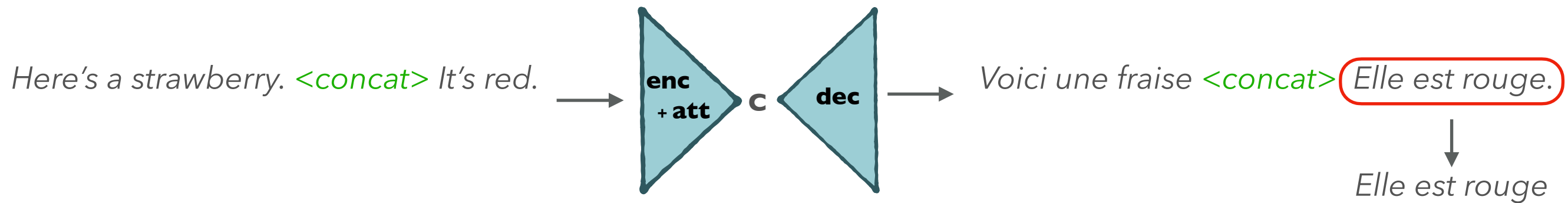
Target:

- 1 context: Ô je déteste les **mouches**. Regarde, il y en a une autre !
correct: T'inquiète, je **la** tuerai pour toi.
incorrect: T'inquiète, je **le** tuerai pour toi.
- 2 context: Ô je déteste les **mouchecons**. Regarde, il y en a un autre !
correct: T'inquiète, je **le** tuerai pour toi.
incorrect: T'inquiète, je **la** tuerai pour toi.

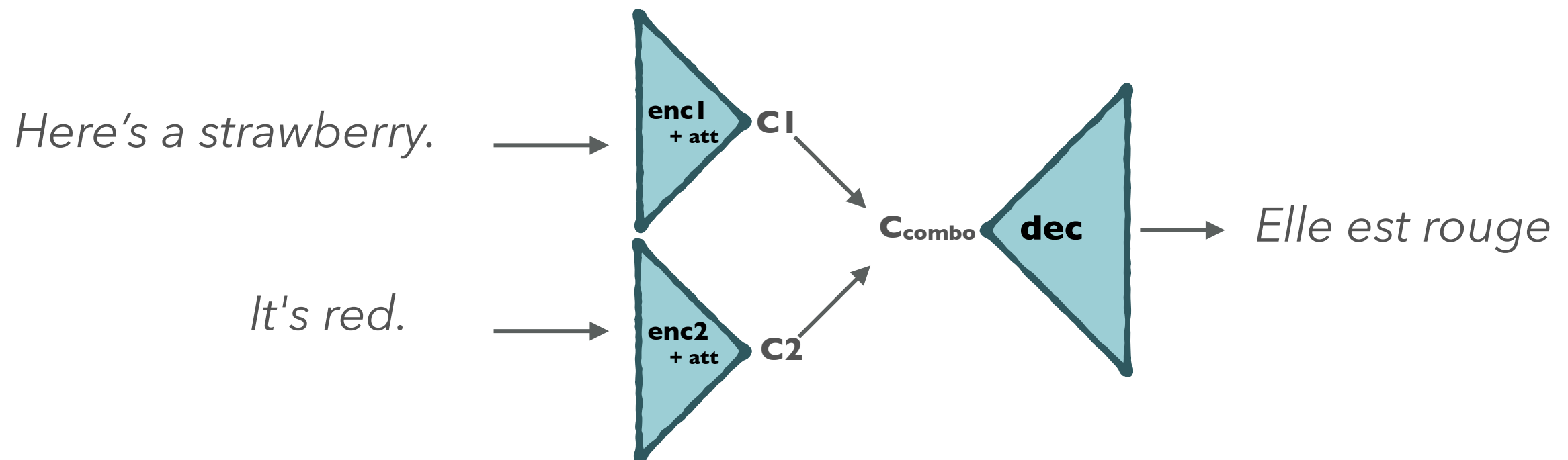
Rachel Bawden, Rico Sennrich, Alexandra Birch, and Barry Haddow (2018). Evaluating Discourse Phenomena in Neural Machine Translation. NAACL

Modelling Context

Concatenated Model:



Multi-source Model:





Context in Machine Translation

	BLEU	Co-reference	Ambiguity
Baseline	19.52	50.0	50.0
Multis. (prev. source)	20.22	50.0	53.0
Multis. (prev. target)	17.89	47.0	50.5
Concat. (src + tgt)	20.09	63.5	52.0
Multis. and Concat. (tgt)	20.85	72.5	57.0

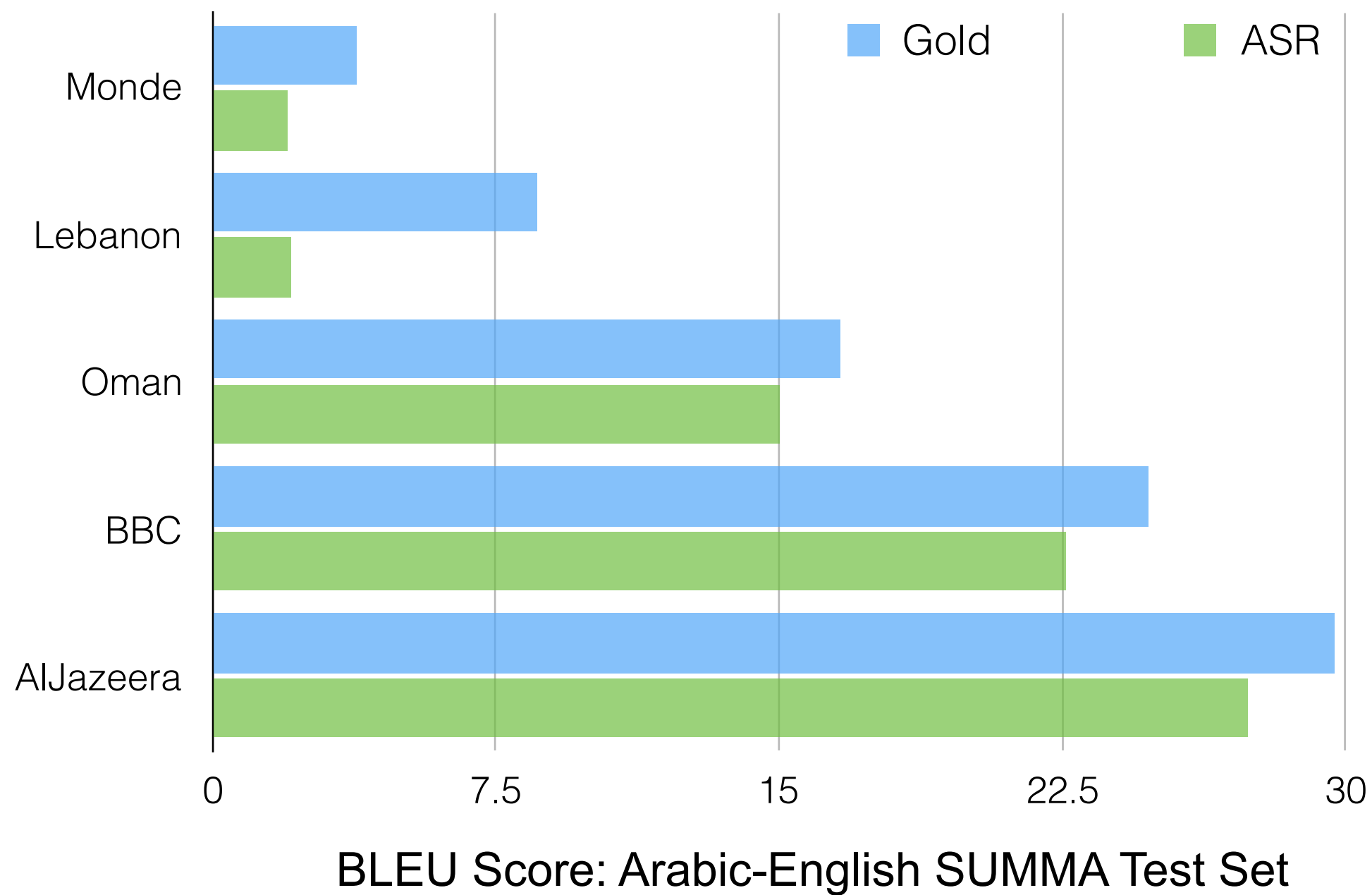
Rachel Bawden, Rico Sennrich, Alexandra Birch, and Barry Haddow (2018). Evaluating Discourse Phenomena in Neural Machine Translation. NAACL



Challenges for Translation in the Media

- Efficient, fast translation of large volumes
- High quality translation
- Story level translation
- **Robust to different dialects**
- Evaluation of for gisting

Dialectal Translation Arabic



Challenges for Translation in the Media

- Efficient, fast translation of large volumes
- High quality translation
- Story level translation
- Robust to different dialects
- **Evaluation of for gisting**



Evaluating Machine Translation for Gisting

Instructions: Fill each one of the gaps in the "problem sentence" at the bottom with the most fitting **single word**, using only information from the *hint text* (if there is one).

Hint text: (you might need to scroll to find some highlighted text)

The Federal Republic of Germany after 1945 experienced a huge economic boom, which was the economic basis for a stable democracy.

In the German Democratic Republic the socialist one-party dictatorship of the SED and the socialist planned economy have been introduced at the same time.

Until 1989, the GDR had therefore great economic problems.

The consequences had a major impact on life in the GDR.

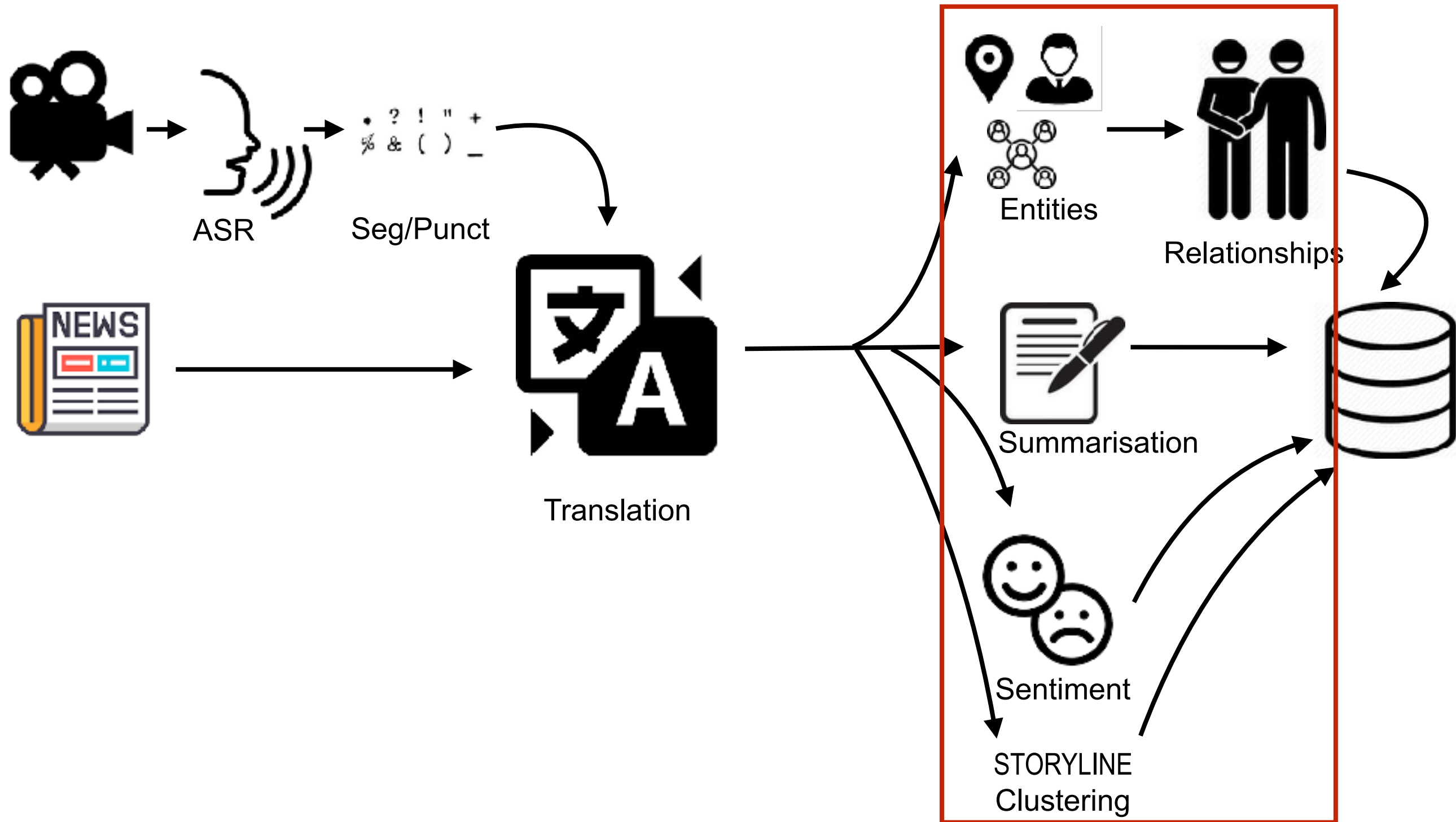
Problem sentence: At the same time in the German Democratic Republic , the socialist one-party dictatorship of the SED and

state-planned were introduced .

✓ Submit

Mikel Forcada, Carolina Scarton, Lucia Specia, Barry Haddow, and Alexandra Birch.
Exploring Gap Filling as a Cheaper Alternative to Reading Comprehension
Questionnaires when Evaluating Machine Translation for Gisting, WMT 2018

SUMMA Pipeline



Multilingual Clustering



Miranda, S., Znotiņš, A., Cohen, S. B., & Barzdins, G. (2018). Multilingual Clustering of Streaming News. EMNLP 2018

Multilingual Named Entity detection

Legal experts clash on whether Cohen plea threatens Trump

Aug 22, 2018 6:25:48 PM

Source: [FOX News](http://feeds.foxnews.com/~r/foxnews/most-popular/~3/6xbx-0meJSA/legal-experts-clash-on-whether-cohen-plea-threatens-trump.html) (<http://feeds.foxnews.com/~r/foxnews/most-popular/~3/6xbx-0meJSA/legal-experts-clash-on-whether-cohen-plea-threatens-trump.html>)

[Go to SUMMA](#)

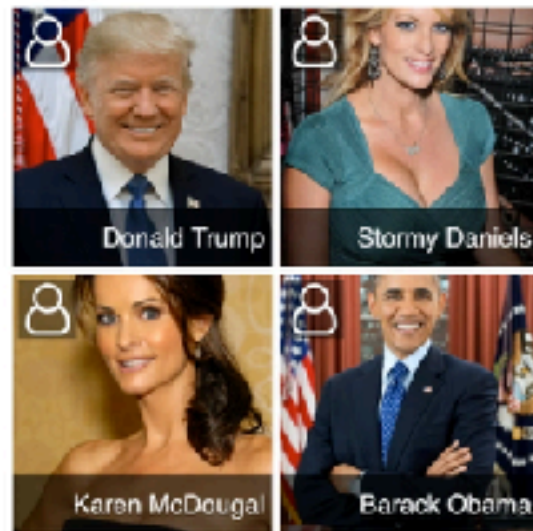
Summary:

- Michael Cohen's plea deal chucked a live political grenade into the debate over President Trump's legal exposure – but that debate is far from settled, as experts clash over whether his implication of the president in campaign finance violations will amount to anything.
- The president's former longtime personal attorney and self-described "fixer" entered a guilty plea with federal prosecutors on Tuesday, admitting to violating campaign finance laws by arranging hush money payments to adult film star Stormy Daniels and former Playboy model Karen McDougal "at the direction" of then-candidate Trump.

























Topics:

[donald trump](#)
[us](#)
[united states](#)
[election](#)
[media](#)
[republicans](#)
[elections](#)
[campaign](#)
[racism](#)
[republican](#)

Entities:



English

Michael Cohen 's plea deal chucked a live political grenade into the debate over President **Trump** 's legal exposure – but that debate is far from settled, as experts clash over whether his implication of the president in campaign finance violations will amount to anything. The president's former longtime personal attorney and self-described "fixer" entered a guilty plea with federal prosecutors on Tuesday, admitting to violating campaign finance laws by arranging hush money payments to adult film star **Stormy Daniels**  and former Playboy model **Karen McDougal**  "at the direction" of then-candidate **Trump** . **Trump** , though, claimed the move to pay off the two women was not a crime – while suggesting such allegations can be settled by fine. "**Michael Cohen**  plead guilty to two counts of campaign finance violations that are not a crime. **Obama**  had a big campaign finance violation and it was easily settled!" **Trump**  tweeted Wednesday morning. **Trump**  was referring to former President **Barack Obama** 's 2008 campaign, which was fined \$375,000 by the **Federal Election Commission**  for a series of missing notices for more than 1,800 contributions. They totaled \$1.8 million. But **Lanny Davis** , **Cohen** 's attorney, argued there is little room for interpretation here. "There is no question that he's committed a federal crime," **Davis**  told **Fox News**  and **America's Newsroom**  on Wednesday. He also argued that it's never been settled whether a sitting president can be indicted, despite suggestions to the contrary from **Trump**  allies. **Davis**  added that his client, under oath on Tuesday, admitted to making the "donations to keep quiet two women" at **Trump** 's direction. At issue is a \$130,000 hush money payment made by **Cohen**  to **Daniels**  in the weeks leading up to the presidential election in exchange for her silence about an alleged affair with **Trump** , and the **McDougal**  case, where she was paid \$150,000 by the

Afonso Mendes, David Nogueira, Samuel Broscheit, Filipe Aleixo, Pedro Balage, Rui Martins, Sebastiao Miranda, and Mariana S. C. Almeida
SUMMA at TAC Knowledge Base Population Task 2017



Ethics of SUMMA

- Legal Compliance GDPR (EU General Data Protection Regulation) May 2018
- Ethics Board:
 - ◆ Professor Mireille Hildebrandt, Vrije Universiteit Brussels
 - ◆ Dr Els Kindt, KU Leuven Centre for IT & IP Law (CiTiP)
 - ◆ Dr Julia Powles, Cornell Tech / University of Cambridge
 - ◆ Professor Lorna Woods, Essex University
- Protection of Personal Data and Privacy by Design: **Social Media**
 - ◆ Public vs private figures
 - ◆ Minimize risk: no videos or images
 - ◆ Aggregation of posts
 - ◆ No transfer of tweets between partners, only links

Broader Ethical Issues

- Open Source Project:
 - ◆ Dual Use: Military application
 - ◆ Research tool: Politics, Social Sciences
- Job security and automation:
 - ◆ More resilient to existing economic pressures
 - ◆ Augmentation not automation
- Media Bubble:
 - ◆ Allows us to access global stories and viewpoints

GoURMET



UNIVERSITY OF AMSTERDAM



Universitat d'Alacant
Universidad de Alicante

1 January 2019 – 31 December 2021



BBC World Service

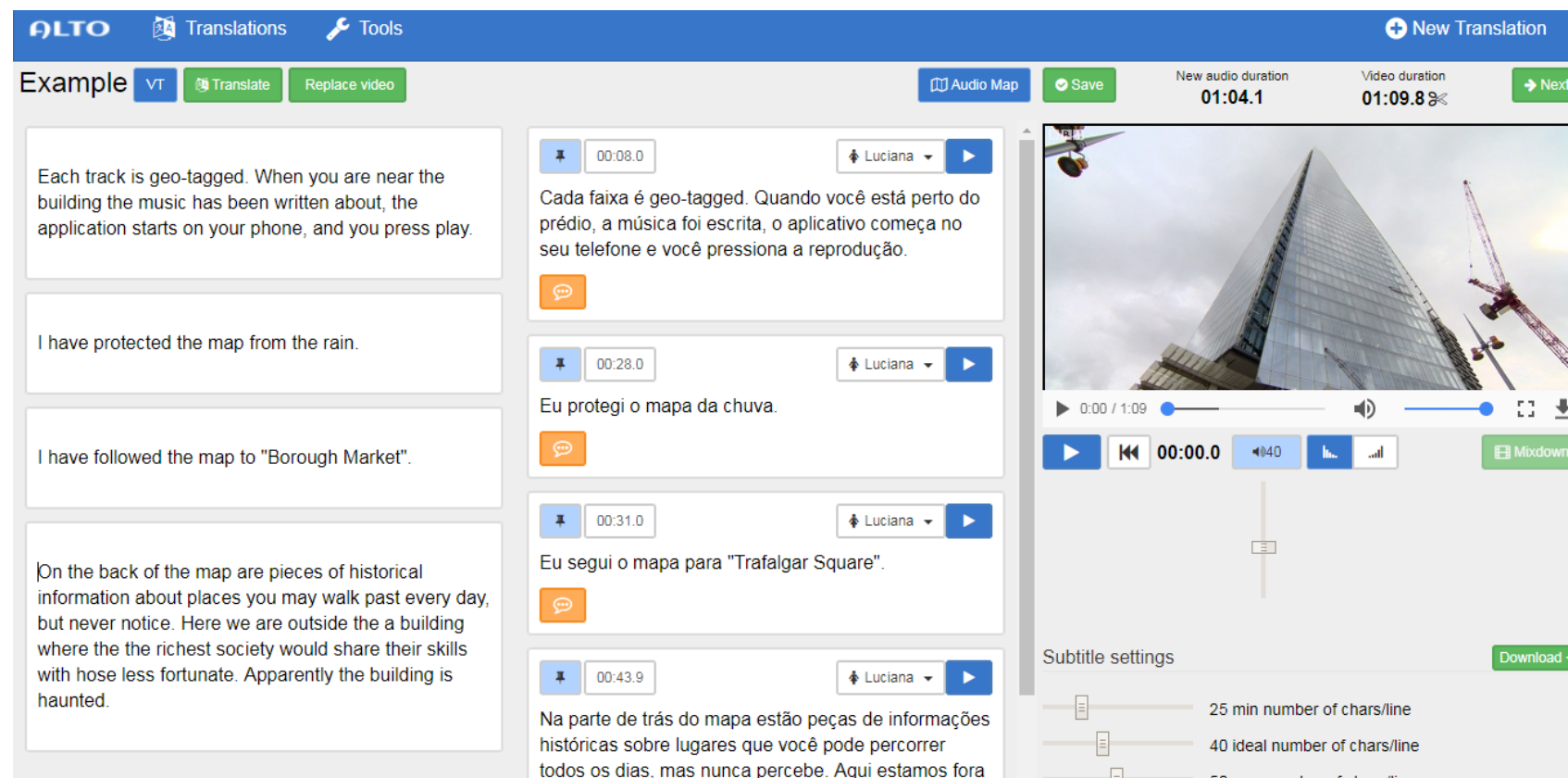
2017: Launched 12 new languages - Afaan Oromo, Amharic, Gujarati, Igbo, Korean, Marathi, Pidgin, Punjabi, Serbian, Telugu, Tigrinya, and Yoruba.



Goal: Reach 500 Million people by 2022

GoURMET: BBC World Service

Efficient content creation



The screenshot displays the ALTO web interface for content creation. The interface is divided into several sections:

- Header:** Includes the ALTO logo, navigation links for 'Translations' and 'Tools', and a 'New Translation' button.
- Example Section:** Contains a 'VT' button, 'Translate' and 'Replace video' buttons, and an 'Audio Map' button. It also shows 'New audio duration' (01:04.1) and 'Video duration' (01:09.8) with a 'Next' button.
- Main Content Area:** A list of text segments on the left, each with a corresponding audio track and subtitle on the right. The segments are:
 - Each track is geo-tagged. When you are near the building the music has been written about, the application starts on your phone, and you press play.
 - I have protected the map from the rain.
 - I have followed the map to "Borough Market".
 - On the back of the map are pieces of historical information about places you may walk past every day, but never notice. Here we are outside the a building where the the richest society would share their skills with hose less fortunate. Apparently the building is haunted.
- Video Player:** Shows a video of a tall building under construction. It includes a play button, a progress bar (0:00 / 1:09), a volume control, and a 'Mixdown' button.
- Subtitle Settings:** A section at the bottom right with a 'Download' button and three sliders for subtitle settings: '25 min number of chars/line', '40 ideal number of chars/line', and '50 max number of chars/line'.

Research problem: low-resource machine translation

Summary

- **How can NLP help the Media?**
 - ✦ Unlocking more content from different media types/languages
 - ✦ Making sense of deluge of data
 - ✦ Reach a wider audience
- **How can we deal with multilingual content?**
 - ✦ Translation: Faster, deeper, with context
 - ✦ Multilingual NLP
- **Context of EU Projects:**
 - ✦ **SUMMA**: Automated Media Monitoring Platform
 - ✦ **GoURMET**: Low-resource Machine Translation



Thank you