Multimodal and multilingual MT

… and how this could be interesting for translators and translation studies …

MeMAD

Methods for Managing
Audiovisual Data

Things to talk about

Our projects related to machine translation

What we have done so far (WMT, IWSLT, ...)

Prototypes and ideas

Open questions and discussions

MeMAD
Methods for Managing
Audiovisual Data

Our Projects

Semantics & MT

FOTRAN
Found in Translation

1,000 languages

Audiovisual content & MT

Data collection & MT

fiskmö
finsk-svensk korpus & maskinöversättning

2 languages (fi/sv)

Machine Translation

MeMAD
Methods for Managing Audiovisual Data

6 languages (de/en/es/fi/fr/nl)
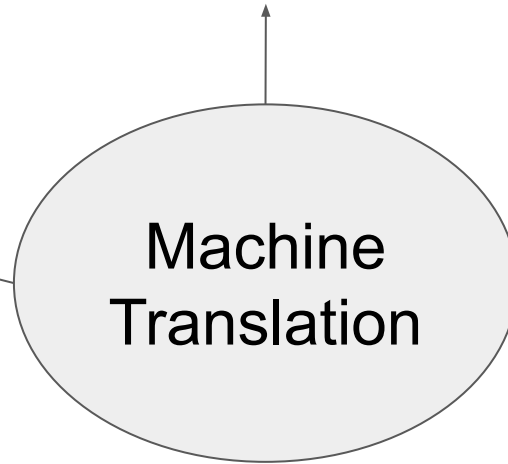
MeMAD
Methods for Managing Audiovisual Data

FoTran

meaning
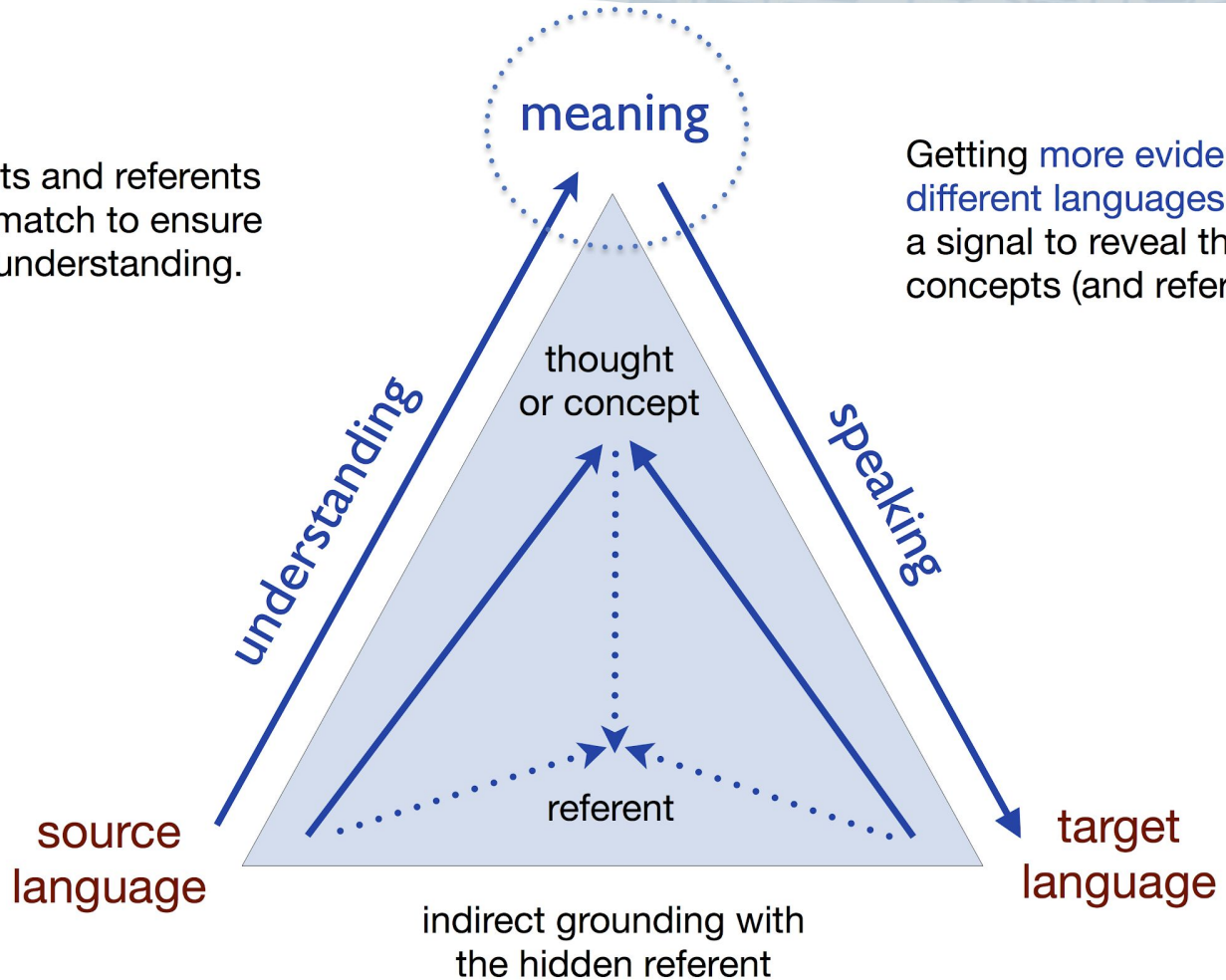
Concepts and referents should match to ensure mutual understanding.

Getting more evidence from different languages provides a signal to reveal the hidden concepts (and referents).

thought or concept

understanding

speaking

referent

source language

target language

indirect grounding with the hidden referent

FoTran

human translations

meaning

representations in semantic spaces

understanding

speaking

dense vector-based representation

encoder

decoder

neural MT

source language

target language

Learning Algorithm
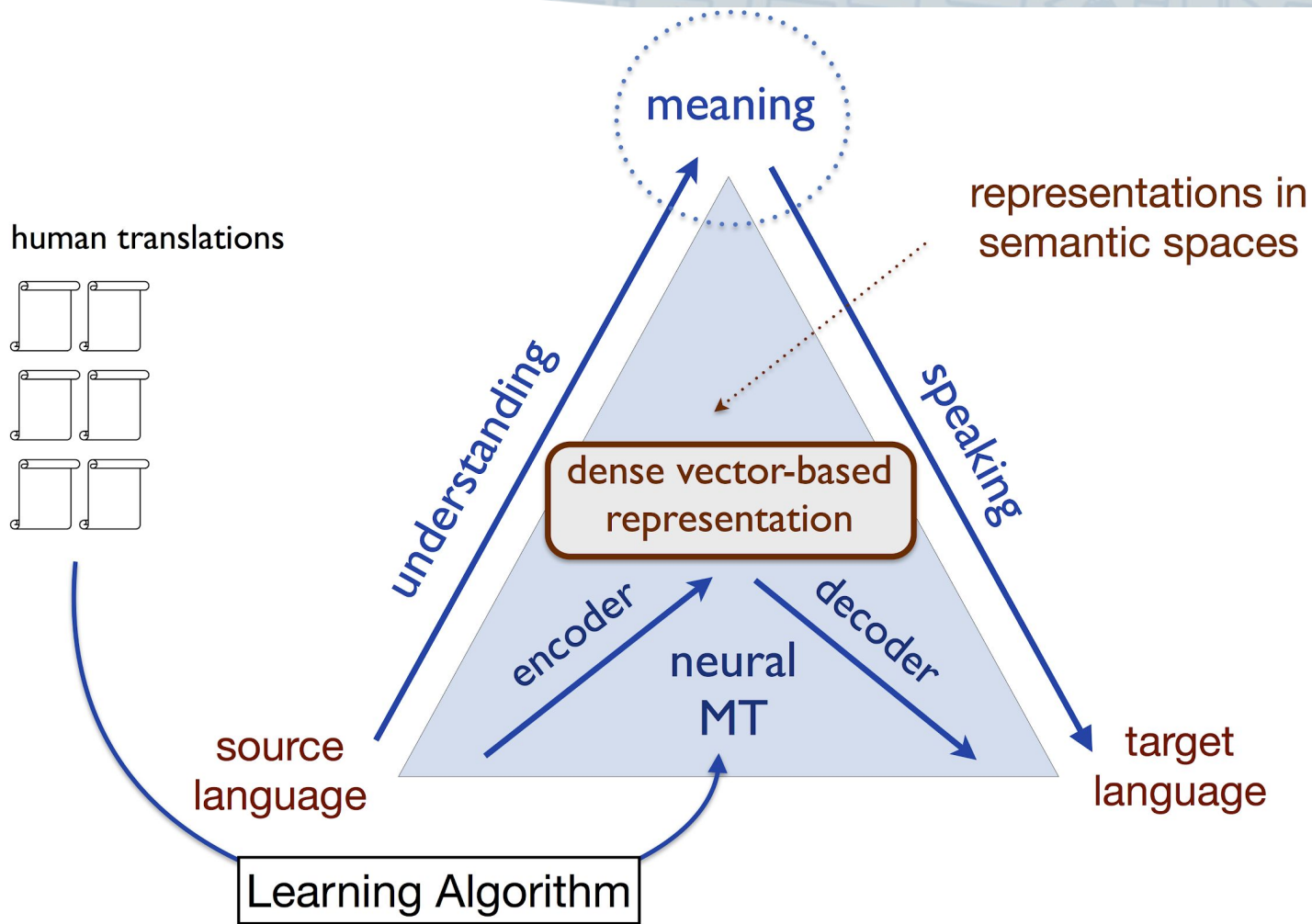
FoTran

human translations in
many languages
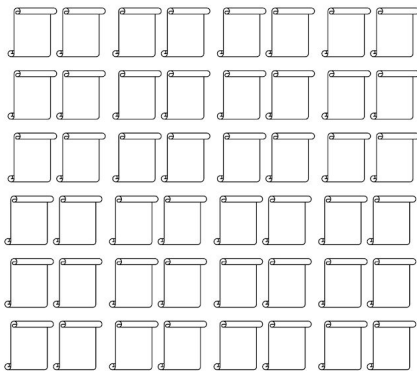
meaning

LIMR: language-independent
meaning representation?

NLI   paraphrase   linguistic
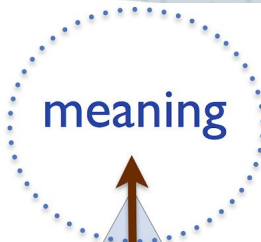   detection   properties

...

MT
model

source and   target languages

Learning Algorithm

F⊙TRAN
Found in Translation
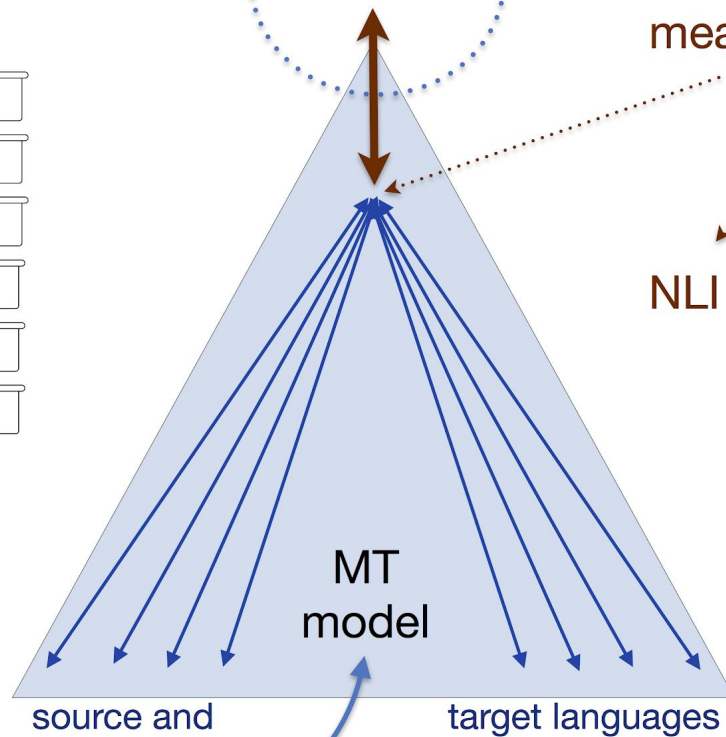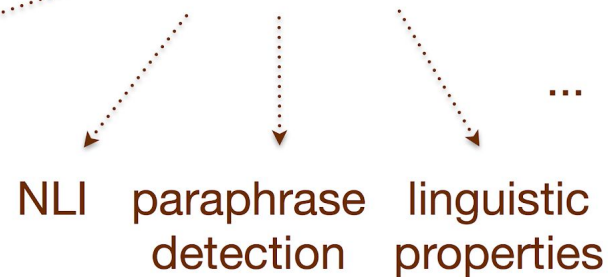
Building a massive parallel corpus for Finnish and Swedish

- Public data (research in linguistics and translation studies) and MT training data
- Collected from the public sector and private organisations

Creating a public translation engine of high quality

- General-purpose machine translation
- Customized translation engines

MeMAD
Methods for Managing
Audiovisual Data

| Finnish | Swedish / Norwegian / Danish | detect language |

| Finnish | Swedish | Norwegian | Danish |

translate

Support the project:

upload translation memory (.tmx, .xliff)

translation memory:　Choose File　no file selected

your email (optional):　［　　　　　　　　］

upload

upload translated documents (.xml, .html, .txt, .pdf, .doc)

original:　Choose File　no file selected

translation:　Choose File　no file selected

your email (optional):　［　　　　　　　　］

upload

Parallel Corpus Workbench

My corpora `+`

Search corpus: _____

newtest   `joerg20 ⬍`   `settings`   `refresh`

uploads `+`	monolingual	parallel `+`

fi.tar.gz
html
   2.html
sv.tar.gz
	en
	   2.xml
	fi
	sv
	   7.Days.in.Entebbe.xml
	   Action.Point.xml
	   Avengers.Infinity.War.xml
	   Deadpool.2.xml
	   El.aviso.xml
	   Flavors.of.Youth.xml
	   Ghostland.xml
	   Hereditary.xml
	   Hurricane.Bianca.2.From.Russia.with.Hate.xml
	   I.Feel.Pretty.xml
	   In.Darkness.xml
	   Jurassic.world. Fallen.Kingdom.xml
	   Lake.Placid.Legacy.xml
	   Life.of.the.Party.xml
	   Oceans.8.xml
	   Patient.Zero.xml
	   Perdida.xml
	   Saa.laenge.jeg.lever.xml
	   Siberia.xml
		fi-sv
		   7.Days.in.Entebbe.xml
		   Action.Point.xml
		   Avengers.Infinity.War.xml
		   Deadpool.2.xml
		   El.aviso.xml
		   Flavors.of.Youth.xml
		   Ghostland.xml
		   Hereditary.xml
		   Hurricane.Bianca.2.From.Russia.with.Hate.xml
		   I.Feel.Pretty.xml
		   In.Darkness.xml
		   Lake.Placid.Legacy.xml
		   Life.of.the.Party.xml
		   Oceans.8.xml
		   Patient.Zero.xml
		   Perdida.xml
		   Saa.laenge.jeg.lever.xml
		   Siberia.xml
		   Sierra.Burgess.Is.a.Loser.xml
		   Snake.Outta.Compton.xml
		   Status.Update.xml
		   Taxi.5.xml

My corpora   +

Search corpus:

Europarl   username   settings

opens interactive
sentence aligner

parallel	edit metadata re-align file1.xml edit download

en-fi
en-sv
en-zh
fi-sv
   subdir
      file1.xml
      file2.xml
      file3.xml
fi-zh
sv-zh
	s9.3 De sovjetiska trupperna är i f ärd med att l ämna Afghanistan . │ The Soviet troops are in the process of withdrawing from Afghanistan . s9.3
	
	s10.1 I flertalet av de regionala konflikterna har F örenta nationerna kunnat g öra viktiga insatser . │ In the majority of the regional conflicts the United Nations has been able to make important contributions . s10.1
	
	s10.2 Sverige medverkar bland annat i FN: s observat örsgrupper i Afghanistan och i Iran/ Irak . │ Sweden is participating in , among other things , the United Nations Military Observer Groups in Afghanistan and in Iran/ Iraq . s10.2
	
	s11.1 Det är med gl ädje vi mottagit beskedet att Nobels fredspris i år g år till FN: s fredsbevarande styrkor │ We were pleased to learn that the Nobel Peace Prize is being awarded this year to the United Nations peace- keeping forces . s11.1
	
	s11.2 Sverige har deltagit i s å gott som samtliga FN- insatser av detta slag . │ Sweden has participated in practically all the United Nations operations of this kind . s11.2
	
	s11.3 Det är ocks å hedrande f ör Sverige att generalsekreteraren , i samr åd med parterna , utsett v år FN- ambassad ör till sin personlige representant vid fredsf örhandlingarna mellan Iran och Irak . │ It is also an honour for Sweden that the Secretary- General , in consultation with the parties , has appointed our United Nations Ambassador to be his personal representative at the peace negotiations between Iran and Iraq . s11.3
	
	s12.1 Det f örnedrande apartheidsystemet i Sydafrika m åste avskaffas . │ The degrading apartheid system in South Africa must be abolished . s12.1

MeMAD
Methods for Managing
Audiovisual Data

Interactive Sentence Alignment

Sentence Alignment / subtest2018_fi-sv_The.Jurassic.Games

≥5 char ⇕   ≤10 sentences ⇕   cognates

change corpus   link ⇕   reset   save   align

| 61 | Hienot kulissit . | Hela befolkningssnittet finns med . | 61 |
62	Tämän mennessä paras .	En lysande uppsättning .	62
63	Seuraa kannibaalia .	Den bästa hittills .	63
		Följ kannibalen .	64
64	- Pysy poissa , idiootti !	- Håll dig undan , din skit !	65
65	- Älä satuta minua !	- Gör mig inte illa !	66
66	Ole kiltti !	Snälla !	67

MeMAD
Methods for Managing
Audiovisual Data

In MeMAD: Helsinki Leads WP4

- Development of **multilingual** MT with **multimodal** input.
- Implementation and training of **neural machine translation** models covering at least six EU languages, both minor and major.
- Development of **discourse-oriented machine translation** optimised for the dynamics of the narrative in the audiovisual data streams.
- Providing support for **cross-lingual content retrieval** based on automatic content analysis.

(from WMT - multimodal MT task)
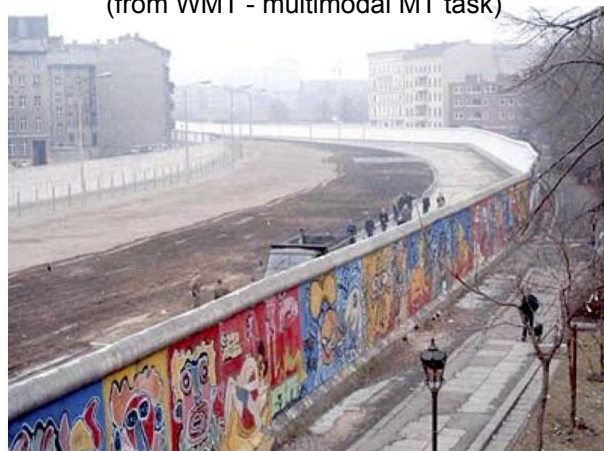
En:	A *wall* divided the city.	
De 1:	Eine *Wand* teilte die Stadt.	✕
De 2:	Eine *Mauer* teilte die Stadt.	✓

MeMAD
Methods for Managing
Audiovisual Data

What we have done so far

Image caption translation

- WMT18 multimodal translation task
  (EN text + image → CS/DE/FR text)   → skipped Czech

Speech-to-text translation

- IWSLT18 speech translation task
  (EN audio → DE text)

MeMAD
Methods for Managing
Audiovisual Data

The MeMAD Submission to the WMT18 Multimodal Translation Task

- Collaboration between
  - Aalto University
  - University of Helsinki
  - EURECOM
- Paper accepted to Conference for Machine Translation
- Preprint available https://arxiv.org/abs/1808.10802v2
- Open source, code available on github Waino/OpenNMT-py (branch develop_mmod)

MeMAD
Methods for Managing
Audiovisual Data

WMT: Multi-modal translation task

A bird flies over the water → Model → Ein Vogel fliegt über das Wasser

MeMAD
Methods for Managing
Audiovisual Data

WMT: The submitted system

- We adapt the Transformer (Vaswani et al., 2017) architecture
- Global image features extracted from Detectron, a pre-trained object detection and localization neural network.
- Multi-lingual training: a single model trained to translate into both languages simultaneously, then finetuned for each language separately.
- Ensemble of 3 independently trained models.

MeMAD
Methods for Managing
Audiovisual Data

WMT: Data sets

Data set	images	en	de	fr	sentences
Multi30k	✓	✓	✓	✓	29k
MS-COCO	✓	✓	+	+	616k
OpenSubtitles		✓	✓	✓	23M/42M
				1M, 3M, and 6M subsets used.	

Table 1: Summary of data set sizes. ✓means attribute is present in original data. + means data set augmented in this work.

MeMAD
Methods for Managing
Audiovisual Data

EN → DE	BLEU ↑	Meteor ↑	TER ↓	LTA ↑
•MeMAD_1_FLICKR_DE_MeMAD-OpenNMT-mmod_U (P)	38.5	56.6	44.5	47.49
CUNI_1_FLICKR_DE_NeuralMonkeyTextual_U	32.5	52.3	50.8	46.37
CUNI_1_FLICKR_DE_NeuralMonkeyImagination_U (P)	32.2	51.7	51.7	47.21
UMONS_1_FLICKR_DE_DeepGru_C (P)	31.1	51.6	53.4	48.04
LIUMCVC_1_FLICKR_DE_NMTEnsemble_C (P)	31.1	51.5	52.6	46.65
LIUMCVC_1_FLICKR_DE_MNMTEnsemble_C (P)	31.4	51.4	52.1	45.81
OSU-BD_1_FLICKR_DE_RLNMT_C (P)	32.3	50.9	49.9	45.25
OSU-BD_1_FLICKR_DE_RLMIX_C	32.0	50.7	49.6	46.09
SHEF_1_DE_LT_C	30.4	50.7	53.0	48.04
SHEF_1_DE_MLT_C (P)	30.4	50.7	53.0	48.32
SHEF1_1_DE_ENMT_C	30.8	50.7	52.4	44.41
SHEF1_1_DE_MFS_C (P)	30.3	50.7	53.1	48.32
LIUMCVC_1_FLICKR_DE_MNMTSingle_C	28.8	49.9	55.6	45.25
LIUMCVC_1_FLICKR_DE_NMTSingle_C	29.5	49.9	54.3	47.77
Baseline	27.6	47.4	55.2	45.25
AFRL-OHIO-STATE_1_FLICKR_DE_4COMBO_U (P)	24.3	45.4	58.6	46.09
AFRL-OHIO-STATE_1_FLICKR_DE_2IMPROVE_U	10.0	25.4	79.0	25.42

EN → FR	BLEU ↑	Meteor ↑	TER ↓	LTA ↑
●MeMAD_1_FLICKR_FR_MeMAD-OpenNMT-mmod_U (P)	44.1	64.3	36.9	73.08
CUNI_1_FLICKR_FR_NeuralMonkeyTextual_U	40.6	61.0	40.7	68.44
CUNI_1_FLICKR_FR_NeuralMonkeyImagination_U (P)	40.4	60.7	40.7	69.29
UMONS_1_FLICKR_FR_DeepGru_C (P)	39.2	60.0	41.8	68.82
LIUMCVC_1_FLICKR_FR_MNMTEnsemble_C (P)	39.5	59.9	41.7	68.53
LIUMCVC_1_FLICKR_FR_NMTEnsemble_C (P)	39.1	59.8	41.9	68.44
SHEF_1_FR_LT_C	38.8	59.8	41.5	69.57
SHEF_1_FR_MLT_C (P)	38.9	59.8	41.5	69.86
SHEF1_1_FR_ENMT_C	38.9	59.8	41.2	67.87
SHEF1_1_FR_MFS_C (P)	38.8	59.7	41.6	67.58
OSU-BD_1_FLICKR_FR_RLNMT_C (P)	39.0	59.5	41.2	68.91
OSU-BD_1_FLICKR_FR_RLMIX_C	38.6	59.3	41.5	67.68
LIUMCVC_1_FLICKR_FR_MNMTSingle_C	37.9	58.5	43.4	67.77
LIUMCVC_1_FLICKR_FR_NMTSingle_C	37.6	58.4	43.2	67.11
Baseline	36.3	56.9	54.3	66.26

MeMAD
Methods for Managing
Audiovisual Data

WMT: Human evaluation (direct assessment)

#	Ave %	Ave z	English→German System
1	91.7	0.69	gold_DE_1
2	87.2	0.479	MeMAD_MeMAD-OpenNMT-mmod_U
3	73.5	-0.046	SHEF_1_DE_MLT_C
	73.8	-0.066	CUNI_NeuralMonkeyImagination_U
	72.6	-0.078	SHEF1_1_DE_MFS_C
	71.6	-0.08	LIUMCVC_MNMTEnsemble_C
	72.1	-0.11	UMONS_DeepGru_C
	72.5	-0.112	LIUMCVC_NMTEnsemble_C
	71.1	-0.179	OSU-BD_RLNMT_C
	68.6	-0.206	AFRL-OHIO-STATE_4COMBO_U
	67.4	-0.272	baseline_DE

#	Ave %	Ave z	English→French System
1	90.3	0.487	gold_FR_1
2	86.8	0.349	MeMAD_MeMAD-OpenNMT-mmod_U
3	78.5	0.047	CUNI_NeuralMonkeyImagination_U
	77.3	-0.005	UMONS_DeepGru_C
	74.9	-0.05	LIUMCVC_NMTEnsemble_C
	74.9	-0.075	SHEF1_1_FR_MFS_C
	74.5	-0.088	SHEF_1_FR_MLT_C
	73.0	-0.11	LIUMCVC_MNMTEnsemble_C
	74.4	-0.12	OSU-BD_RLNMT_C
	66.0	-0.376	baseline_FR

MeMAD
Methods for Managing
Audiovisual Data

WMT: Results

- We had the **best** system for En→De and En→Fr, by a wide margin
- However, the effect of the visual features is small.
- The largest gains come from the quality of the underlying text-only NMT system.
- Use of additional in-domain (COCO) and selected out-of-domain (OpenSubtitles) data is effective.

MeMAD
Methods for Managing Audiovisual Data

WMT: Visual "disambiguation"

Figure 2: Image 117 was translated correctly as feminine "eine besitzerin steht still und ihr brauner hund rennt auf sie zu ." when not using the image features, but as masculine "ein besitzer ..." when using them. The English text contains the word "her". The person in the image has short hair and is wearing pants.

MeMAD
Methods for Managing
Audiovisual Data

MeMAD at IWSLT 2018

Speech translation task

- Input : English audio (TED talks)
- Output: German text

Tracks

- Pipeline approach: English audio → ASR→ MT → German text
- End-to-end approach: English audio → German text

MeMAD
Methods for Managing
Audiovisual Data

IWSLT: Data sets

Task-specific:

- TED Speech Translation: EN audio + transcripts, DE translations
- TED-LIUM: EN audio + transcripts
- TED-WIT3: EN transcripts + translations

Other:

- WMT data, OpenSubtitles2018 (all text only)

MeMAD
Methods for Managing
Audiovisual Data

IWSLT: Pipeline model

- Baseline: NN following the Listen, Spell, and Attend (LAS) architecture

- Acoustic model: Hybrid TDNN-HMM
- Language model: 4-gram
- Based on the Kaldi toolkit

Model	Dev WER	Test WER
TDNN + large 4-gram	8.24	8.83
LAS	15.83	16.16

MeMAD
Methods for Managing
Audiovisual Data

IWSLT: Pipeline model

ASR output translation model

- NMT with attentional RNNs (amun) or transformer
- Experiments with OpenNMT-py and Marian-NMT
- Sub-selection of ASR hypotheses

Training data	Model	BLEU
TED-ASR-TOP-1	AMUN	16.65
TED-ASR-TOP-10	AMUN	16.28
TED-ASR-TOP-50	AMUN	15.88
TED-ASR-TOP-1	TRANSFORMER	18.25
TED-ASR-TOP-10	TRANSFORMER	17.90
TED-ASR-TOP-50	TRANSFORMER	18.14

MeMAD
Methods for Managing
Audiovisual Data

Optimise NMT for ASR Output Translation

Problem: Standard MT is not trained for noisy ASR output

- Run ASR on all TED talks → ASR-to-German TED-trans
- Simulate ASR for text-only data: train NMT for English → ASR-English

Training data	Model	BLEU
TED-ASR-TOP-10	AMUN	61.87
TED-ASR-TOP-10	TRANSFORMER	61.91
TED-ASR-TOP-50	AMUN	61.82

- Translate all English OpenSubtitles2018 to ASR-like English

MeMAD
Methods for Managing
Audiovisual Data

What does this do?

Another example:
- Original: I'm a child of **1984**,
- ASR-like: i am a child of **nineteen eighty four**

Original	Because in the summer of 2006, the E.U. Commission tabled a directive.
ASR-REF	because in the summer of two thousand and six the e u commission tabled directive
ASR-OUT	because in the summer of two thousand and six you commission tabled a directive

Original	Stasi was the secret police in East Germany.
ASR-REF	what is the secret police in east germany
ASR-OUT	stars he was the secret police in east germany

MeMAD
Methods for Managing
Audiovisual Data

Translating the Development Test Set

The effect of "ASR-like translation" on MT performance

	BLEU	
Training data	Untuned	Tuned
TED-ASR-Top-10+Subs	20.44	20.58
TED-ASR-Top-10+Subs-ASR	19.79	20.80

MeMAD
Methods for Managing
Audiovisual Data

IWSLT: Final Submission

Training data	BLEU
TED-ASR-Top-10	14.34
TED-ASR-Top-10+Subs	16.45
TED-ASR-Top-10+Subs-ASR	15.80

System	BLEU	TER	BEER	characTER	BLEU(CI)	TER(CI)	#Words
			Baseline condition				
TIIC	28.09	55.74	54.73	84.72	29.44	53.73	39611
USTC-NEL	26.47	58.03	52.69	92.24	27.86	55.98	38372
ALIBABA	22.36	63.03	51.77	69.26	24.23	60.22	39751
Apptec	21.45	64.12	51.56	63.47	22.72	61.69	41210
KIT	19.44	67.94	50.61	58.16	20.78	65.52	42128
AFRL	17.24	69.10	49.23	64.27	18.37	66.78	41155
MeMAD	15.8	74.51	47.01	82.56	17.13	72.00	41848

MeMAD
Methods for Managing
Audiovisual Data

IWSLT: End-to-End Model

Work in progress...

MeMAD

Methods for Managing
Audiovisual Data

IWSLT: End-to-End Model

Work in progress… and we're not the only one.

System	BLEU	TER	BEER	characTER	BLEU(CI)	TER(CI)	#Words
Baseline condition							
TIIC	28.09	55.74	54.73	84.72	29.44	53.73	39611
USTC-NEL	26.47	58.03	52.69	92.24	27.86	55.98	38372
ALIBABA	22.36	63.03	51.77	69.26	24.23	60.22	39751
Apptec	21.45	64.12	51.56	63.47	22.72	61.69	41210
KIT	19.44	67.94	50.61	58.16	20.78	65.52	42128
AFRL	17.24	69.10	49.23	64.27	18.37	66.78	41155
MeMAD	15.8	74.51	47.01	82.56	17.13	72.00	41848
End-to-End condition							
USTC-NEL	19.4	68.20	48.77	87.30	20.77	65.73	41372
FBK	10.24	78.20	40.68	129.47	11.16	76.38	36627
KIT	8.4	88.54	41.48	80.38	9.22	86.55	44155
JHU	5.45	89.59	35.46	99.89	6.09	88.20	40932

MeMAD
Methods for Managing
Audiovisual Data

Discussions ...

MeMAD

Methods for Managing
Audiovisual Data

memad.eu
info@memad.eu

@memadproject

MeMAD Project