European Commission

The State of Neural Machine Translation

Technology, Integration and Forecast

Christian Eisold, berns language consulting GmbH

European Commission

History of MT

RBMT

SYSTRAN
ECMT
PROMT
Google
APERTIUM
Promsit
LUCY

SMT

Moses
SYSTRAN
Google
Microsoft
MT@EC
KantanMT
Omniscien

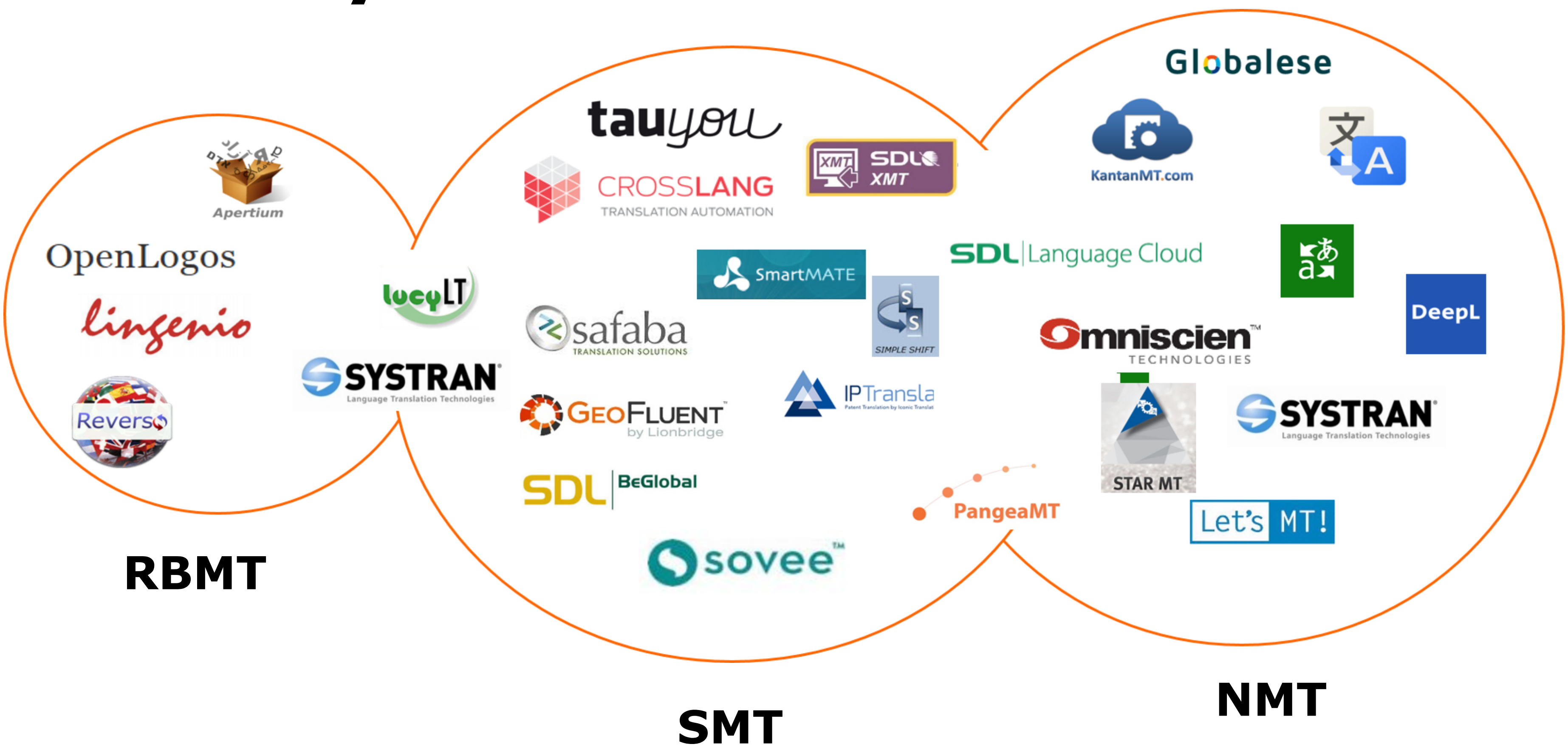
NMT

KantanMT
SYSTRAN
Omniscien
Globalese
Microsoft
Google
deepL
eTranslation

1968 — 1990 — 2000 — 2016 — today

European Commission

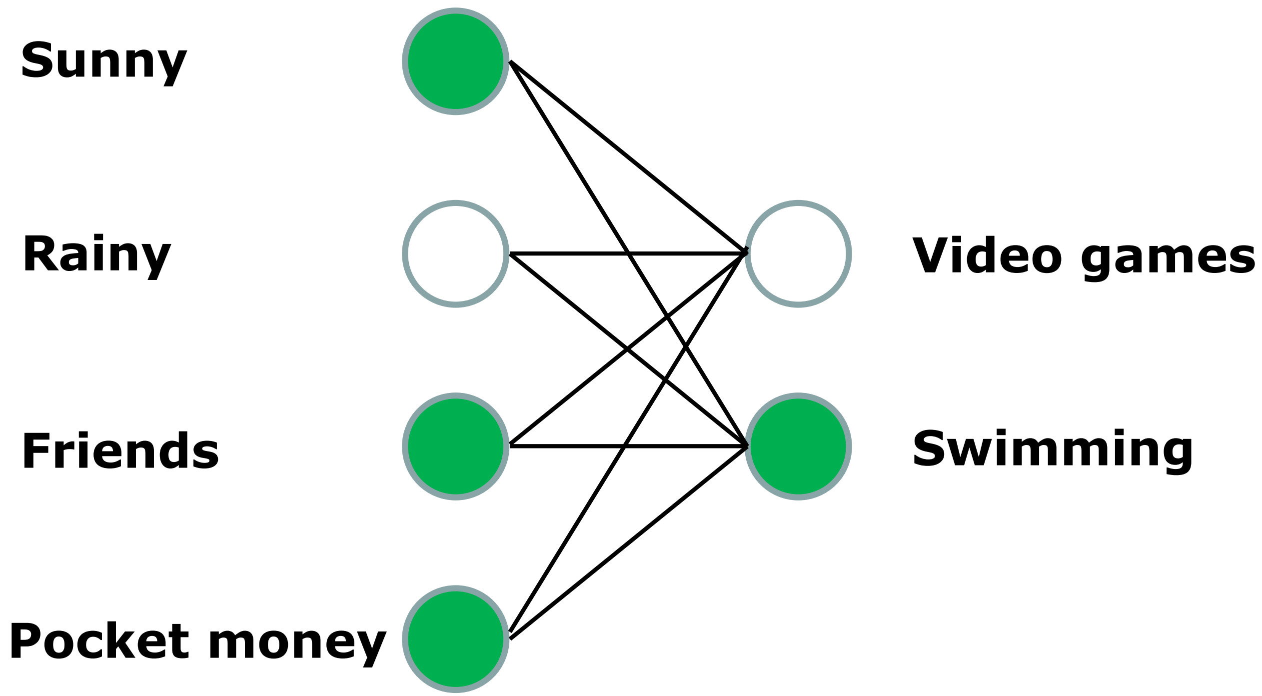
History of MT



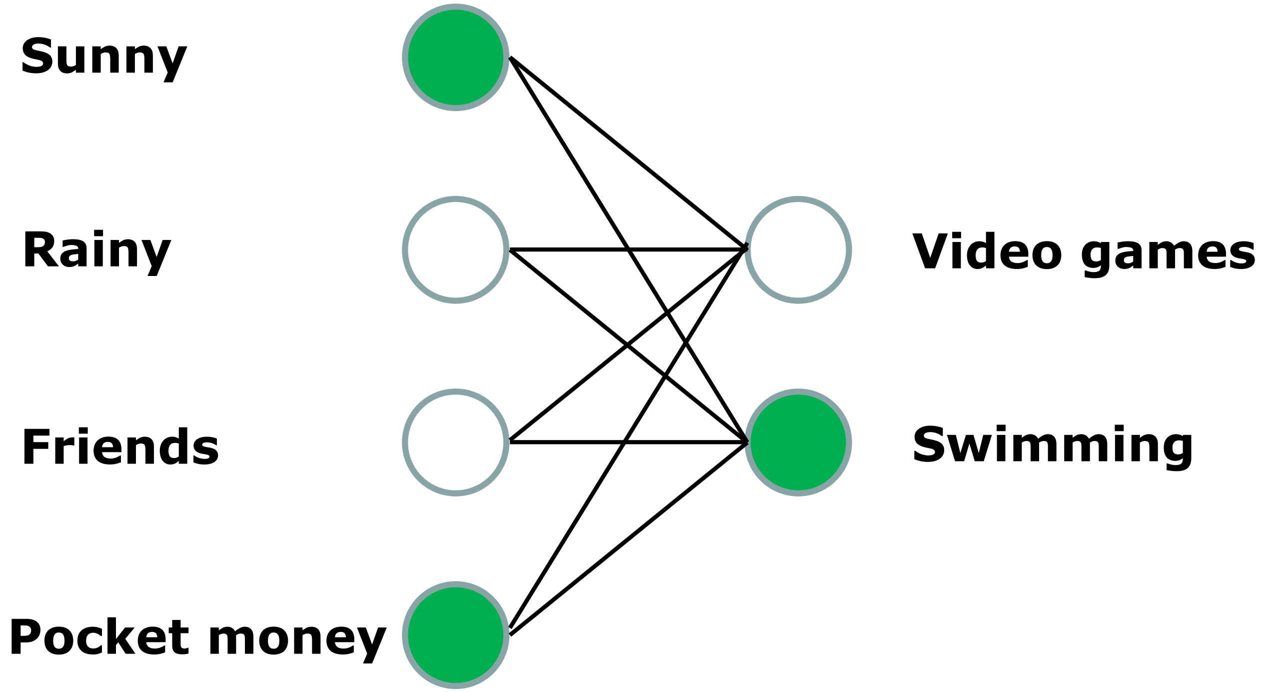
Neural networks

How they work & what they are capable of

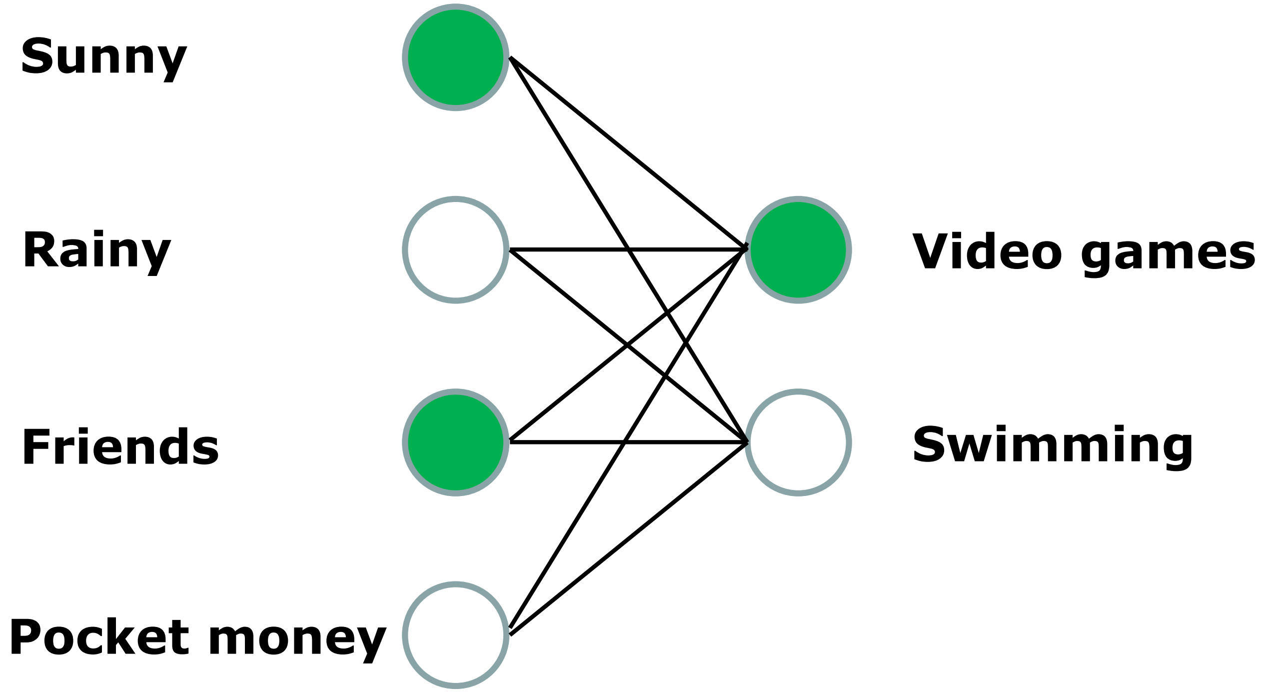
Neural Networks - Classification



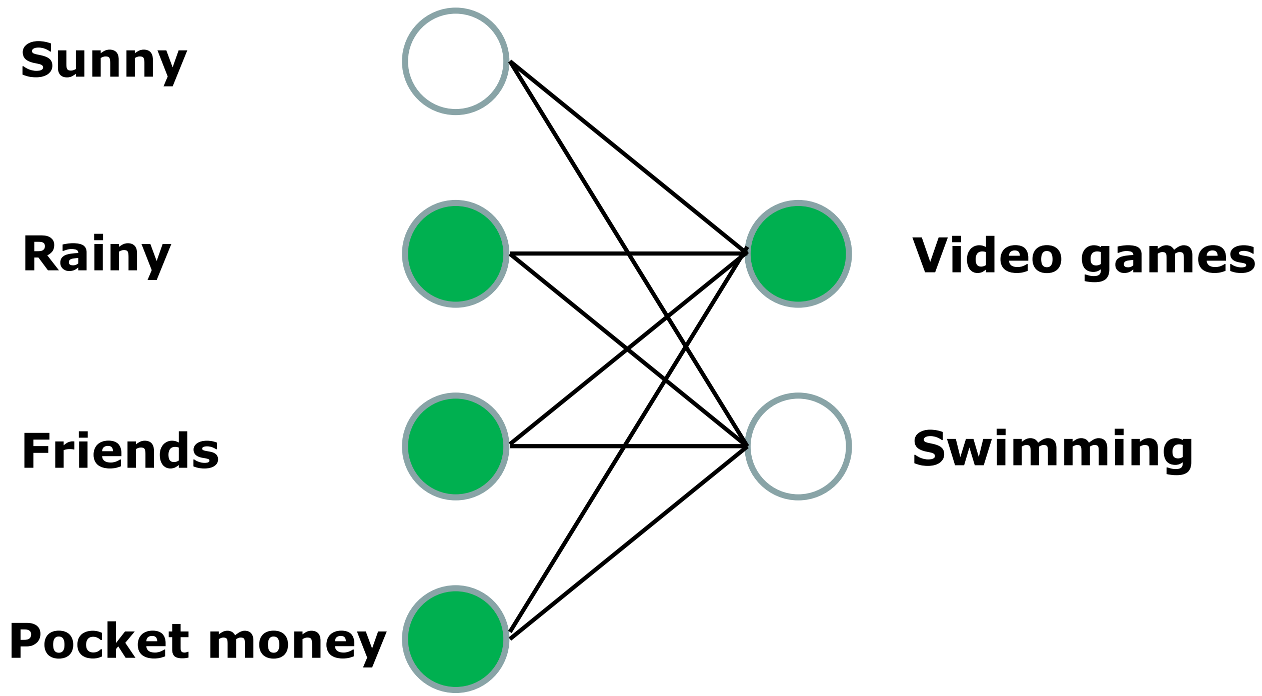
Neural Networks - Classification



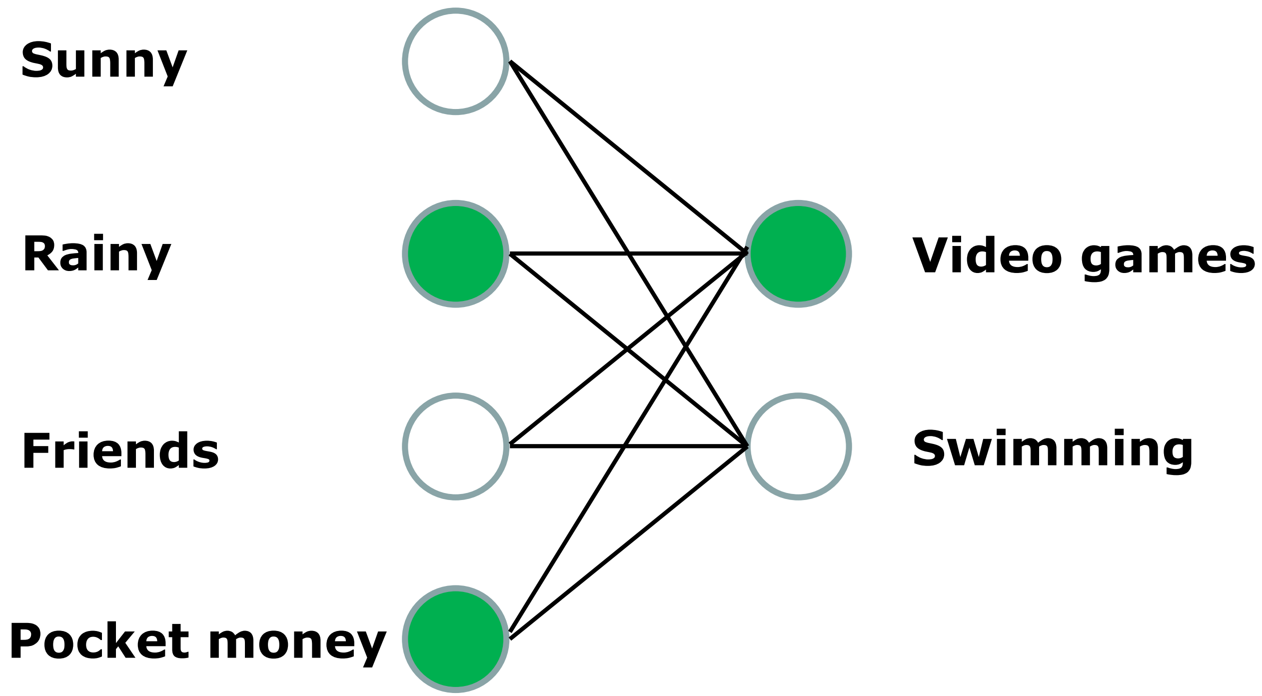
Neural Networks - Classification



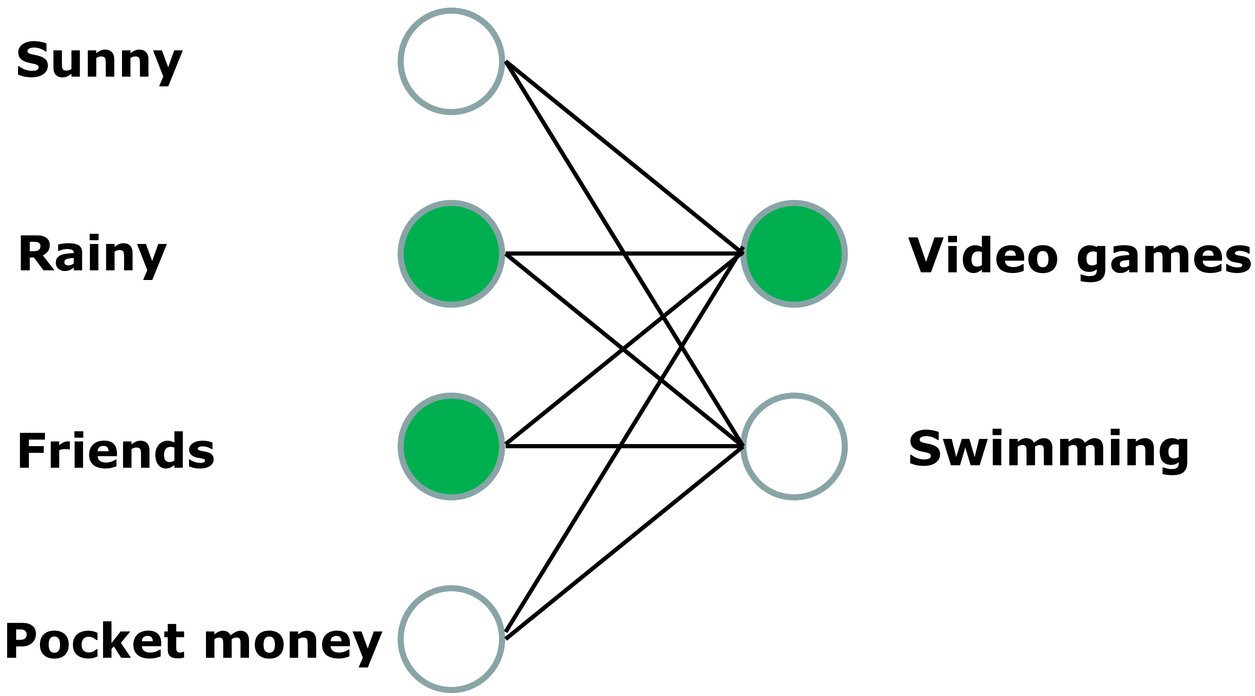
Neural Networks - Classification



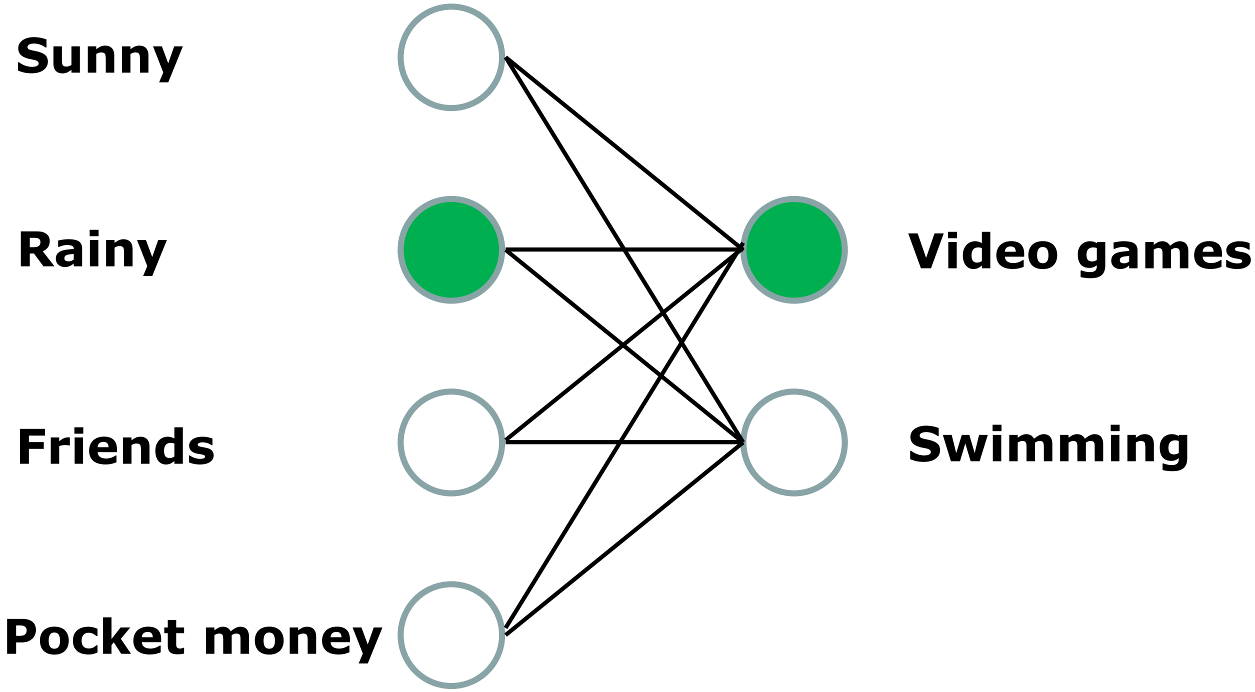
Neural Networks - Classification



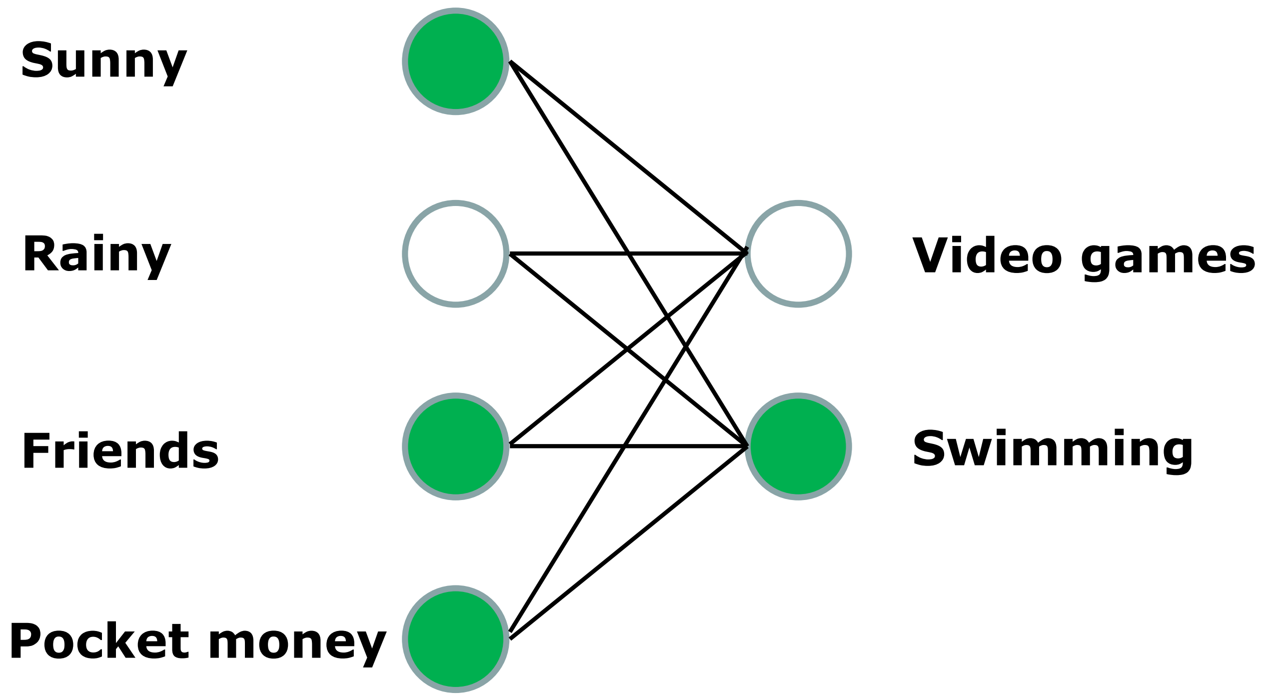
Neural Networks - Classification



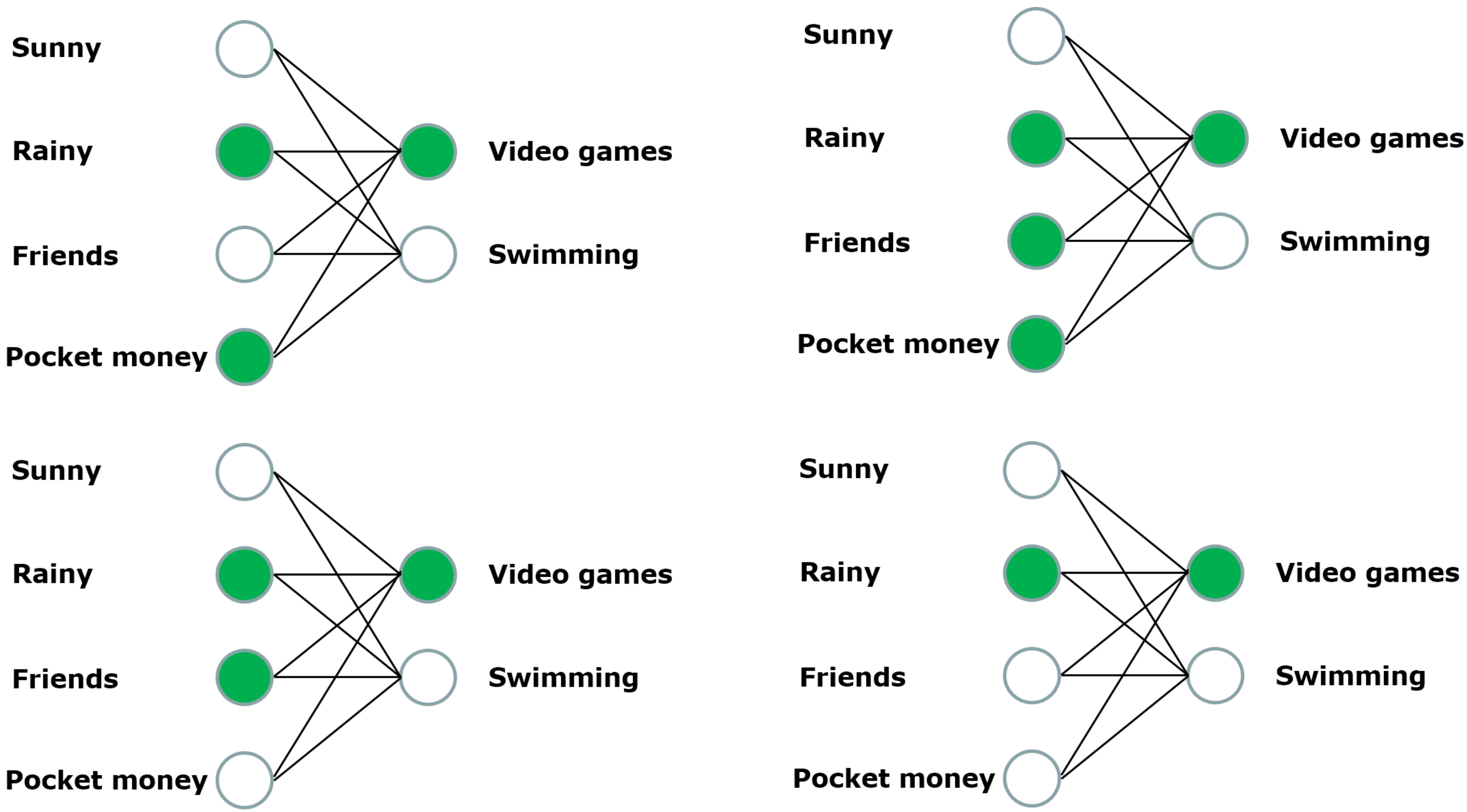
Neural Networks - Classification



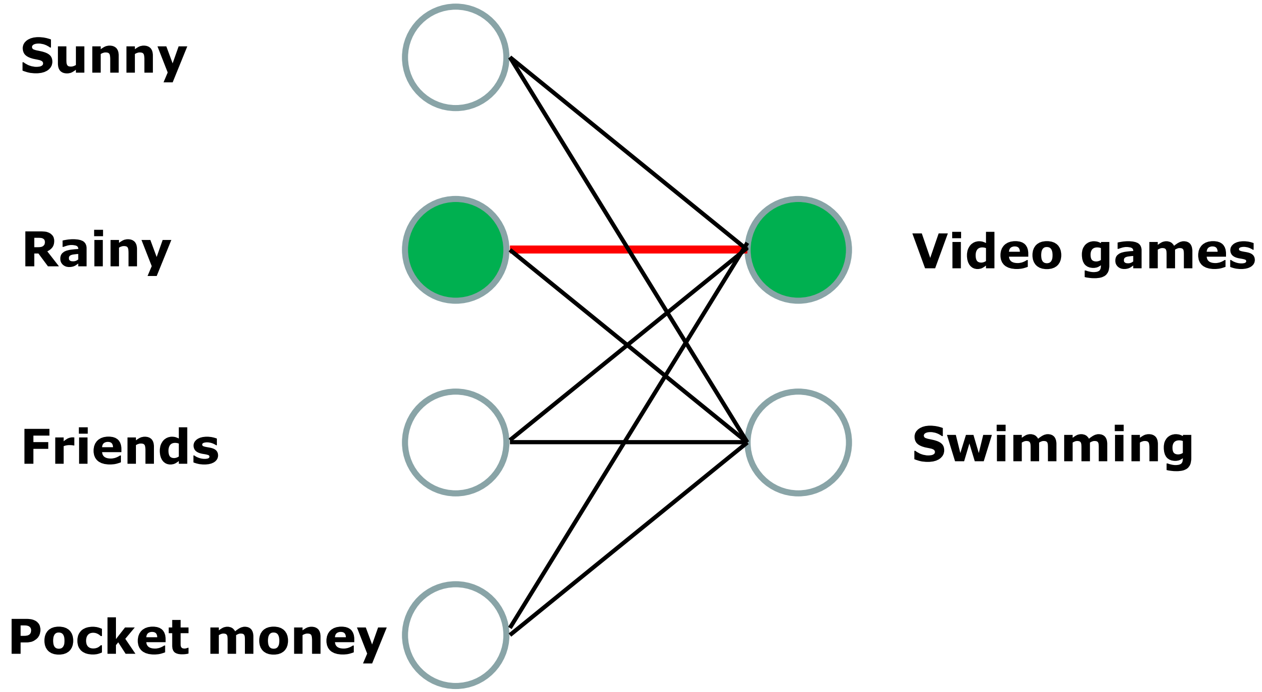
Neural Networks - Classification



Neural Networks - Classification



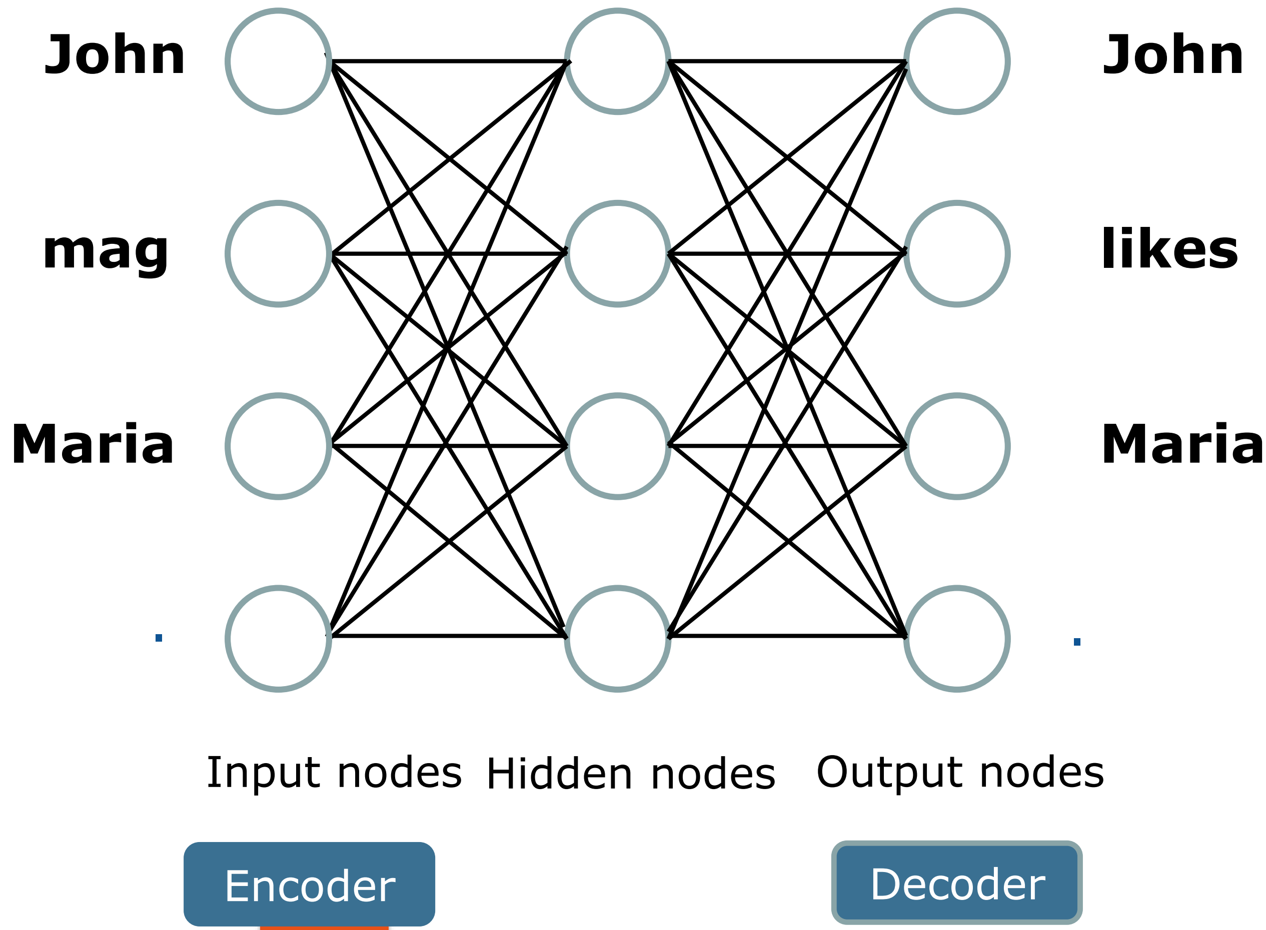
Neural Networks - Classification



Learning by weighting connections!

NMT – Neural Networks

A fully trained NMT Engine has found optimized weights for connections between source and target words in all training sentences!



NMT Terminology

Neuron

Nodes in a network corresponding to an activation function (fire vs. do not fire)

Layer

Single layer in the network, consisting of nodes (neurons) connected to other layers. The more layers, the deeper the network!

NMT Terminology

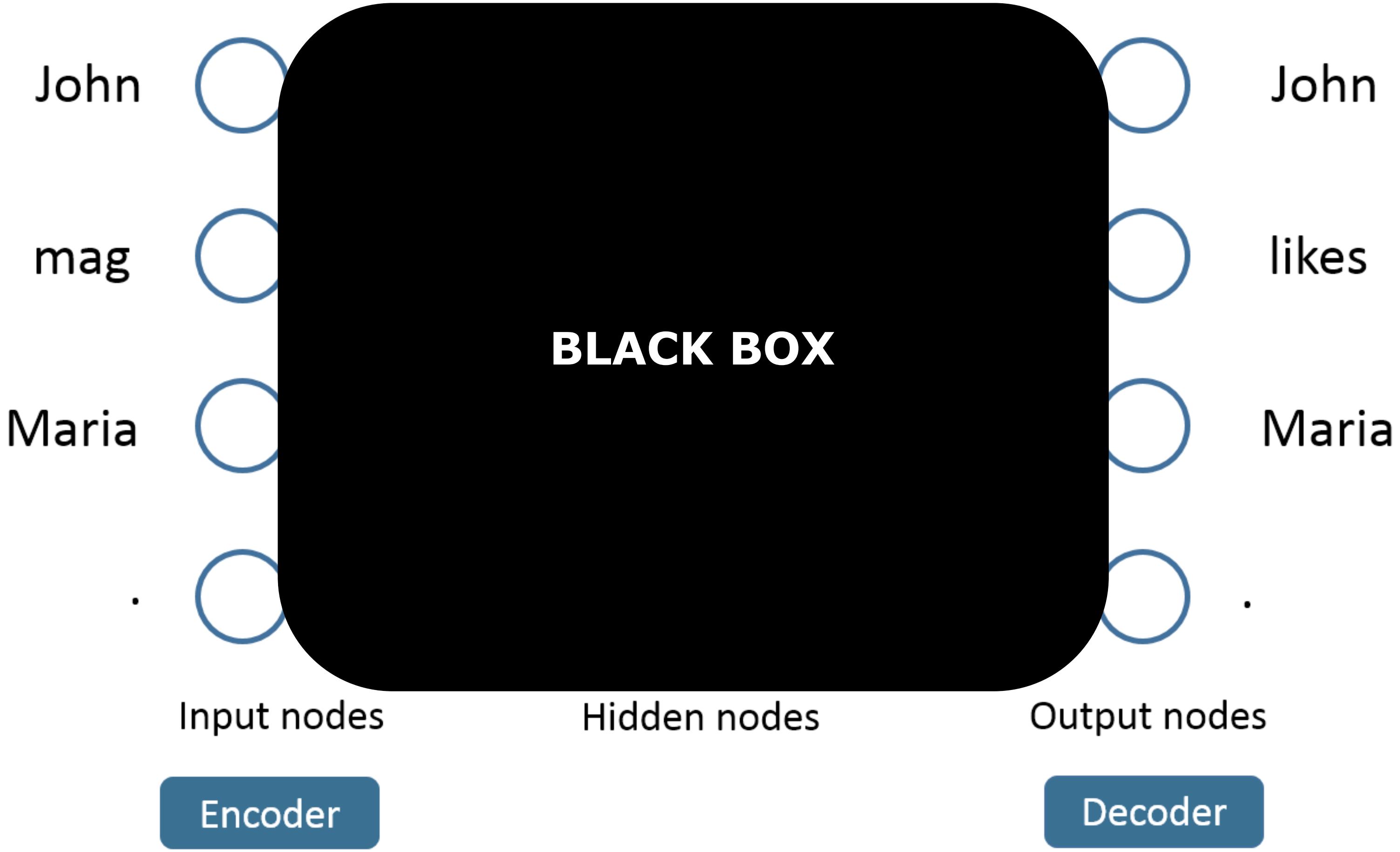
Encoder

Network that converts the initial sentence into a machine-readable representation (word vectors)

Decoder

Network that translates the internal representation of the sentence (from the encoder) into target language strings.

NMT – Neural Networks



Word embeddings

Words are vectors!

John

0.99	0.74	0.22	0.05	0.18	0.32	0.02	0.62	0.87	0.33	...

mag

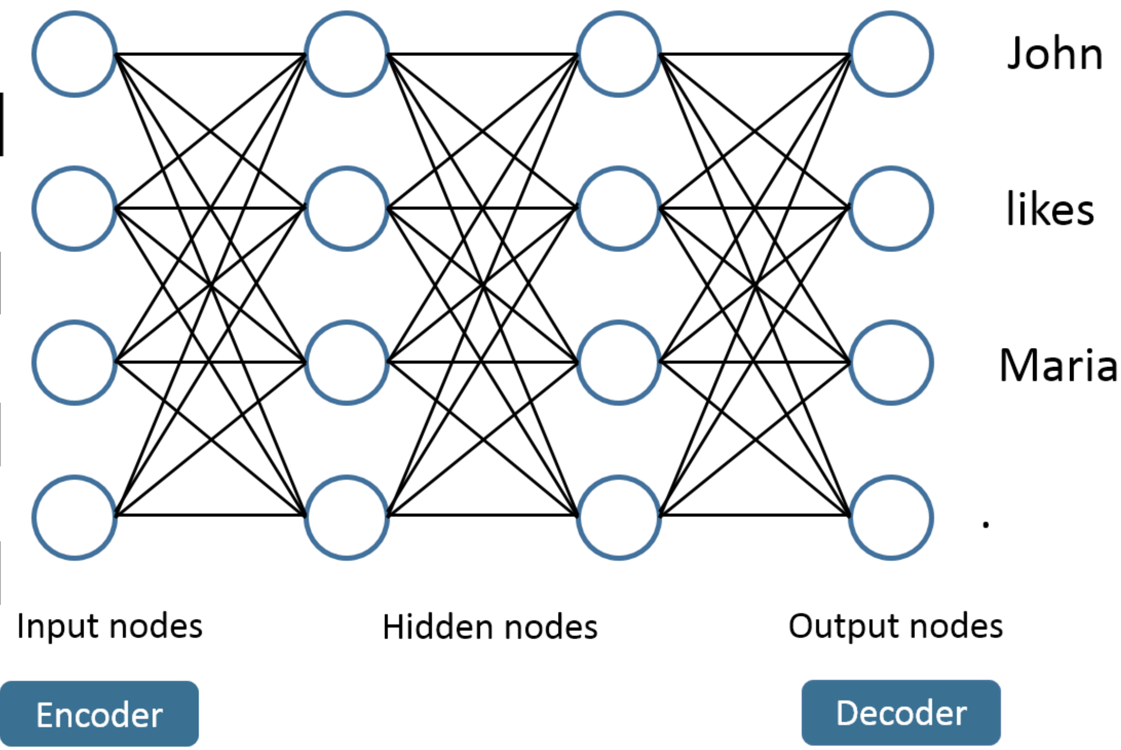
0.72	0.83	0.09	0.26	0.01	0.47	0.99	0.34	0.11	0.27	...

Maria

0.37	0.18	0.32	0.01	0.03	0.33	0.22	0.32	0.02	0.22	...

.

0.01	0.04	0.05	0.01	0.01	0.01	0.99	0.02	0.02	0.03	...



European Commission

Word vectors in NMT

Similar meanings of words are expressed by vector similarity!

John likes Maria.

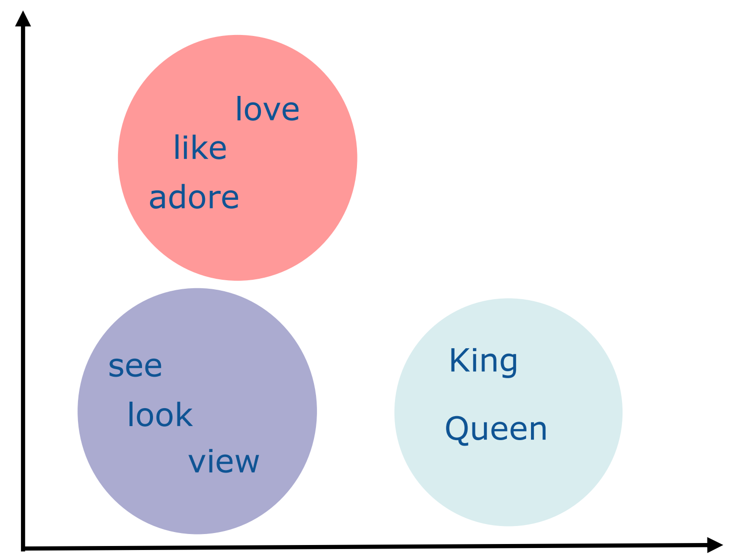
John loves Maria.

John adores Maria.

Men like to drink.

A king is a man.

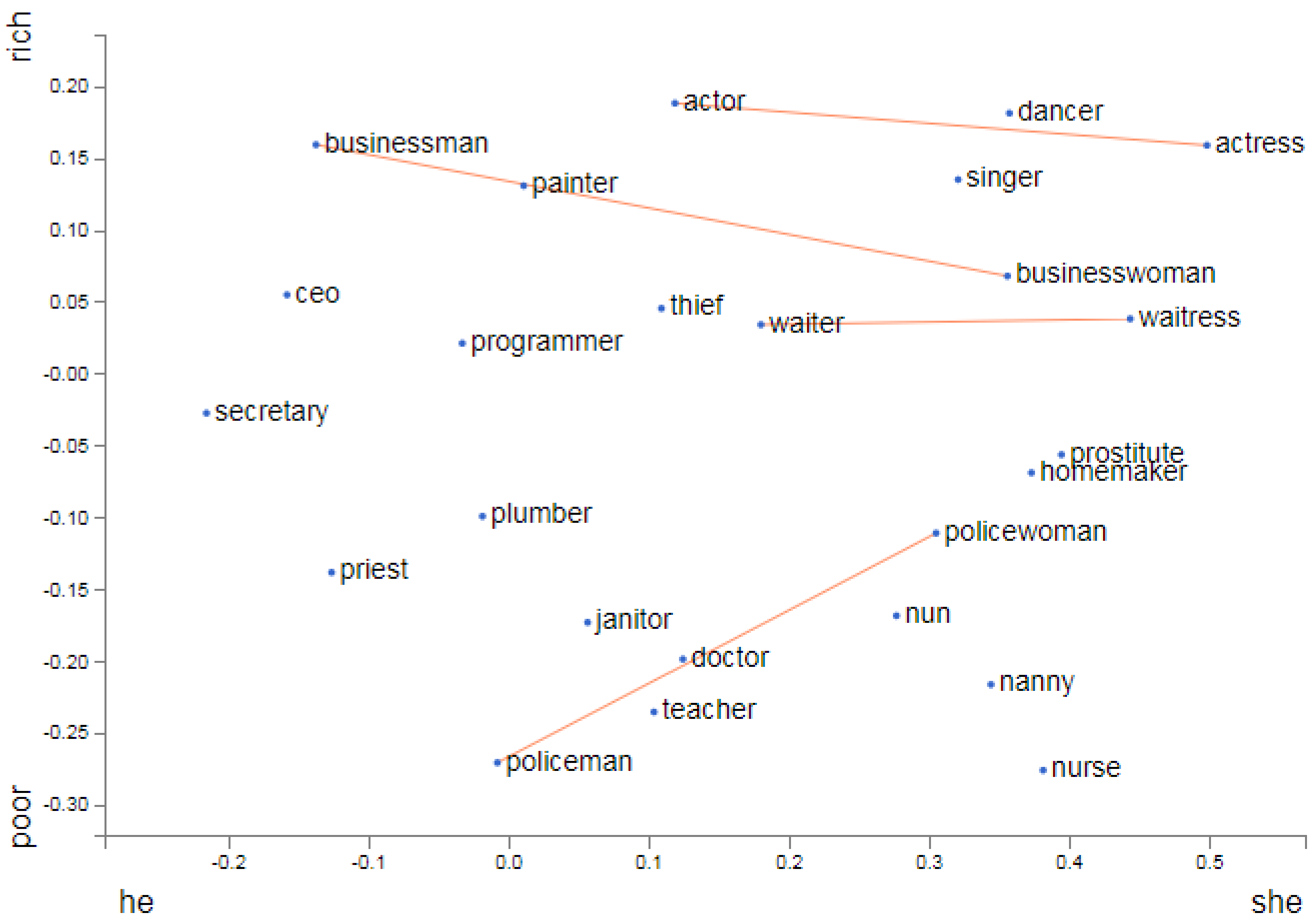
⇨ Kings like to drink



Vectors are learned <u>unsupervised</u> by word contexts!

Abstractions over concepts! High costs!

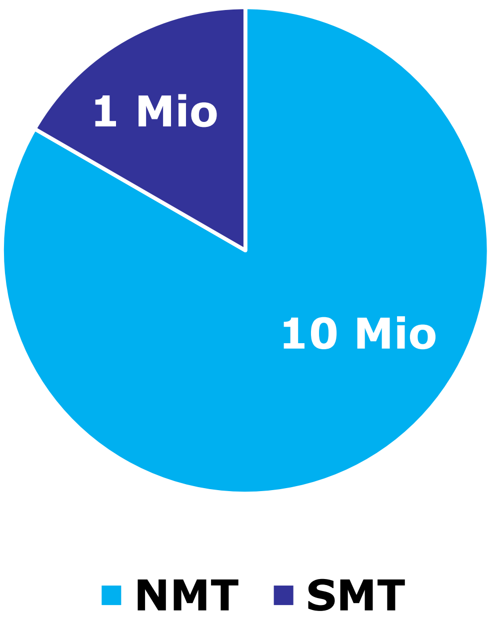
Word vectors in NMT



Training NMT

Resources, Technology, Method & Domain Adaptation

Training NMT Engines



- **Internal texts**
- **Texts from MT vendor**
- **Free texts**
- **Charged texts**
- **Back translation**
  Low resource languages

TBX ✓

HTML
TXT

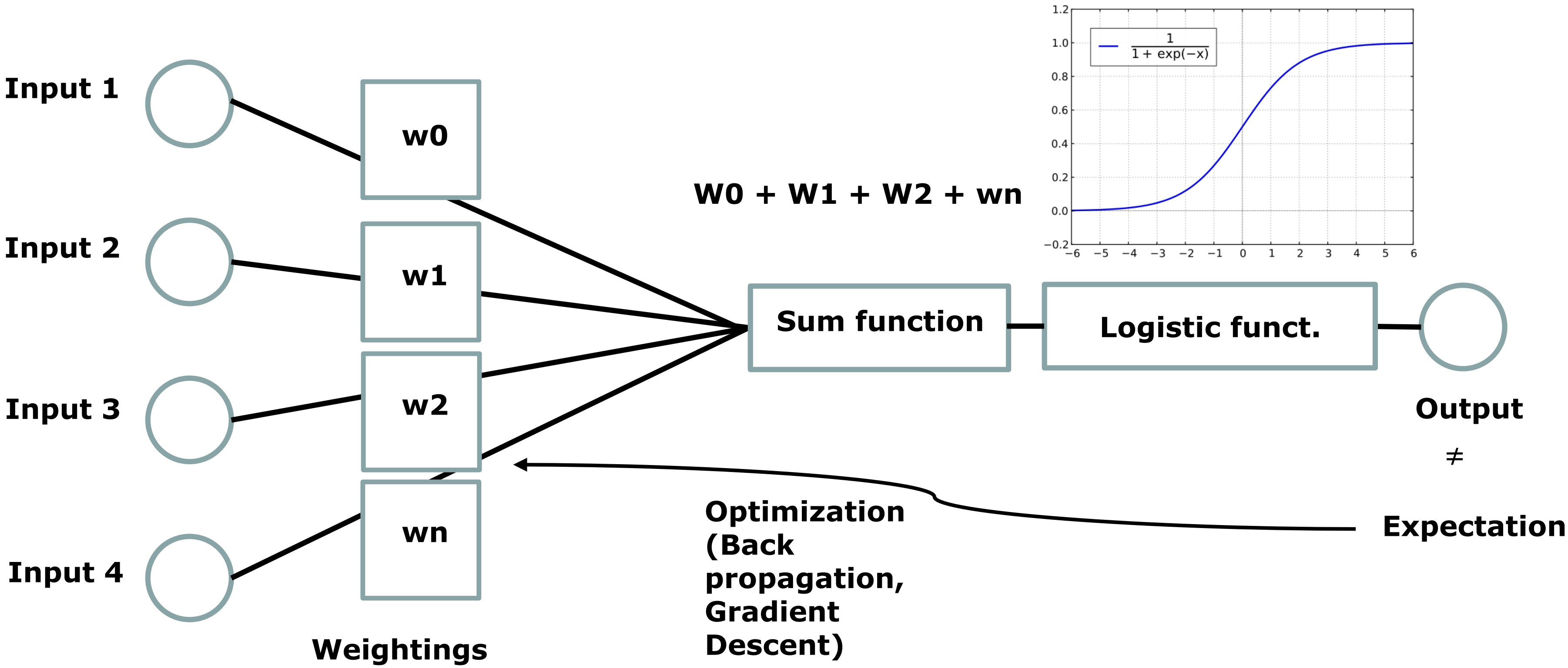
Alignment

NMT Technology

Frameworks
OpenNMT
Nematus
Marian NMT
SpacY
…

Parameters
14 Tokenizer parameters
12 Training parameters
12 Translation parameters
36 Model paramaters
26 Data parameters

Training NMT Networks



Neural Networks - Training

Training a neural network

1. **Initialize weightings (randomized/preset)**
2. **Activation of neurons (Forward propagation)**
3. **Measure errors**
4. **Pass back errors (Back propagation) and optimize weightings**
5. **Repeat 1. – 4. for the whole data set (1 Epoch)**

Domain Adaptation

Single-Domain-Engines

Law

Transportation

Marketing

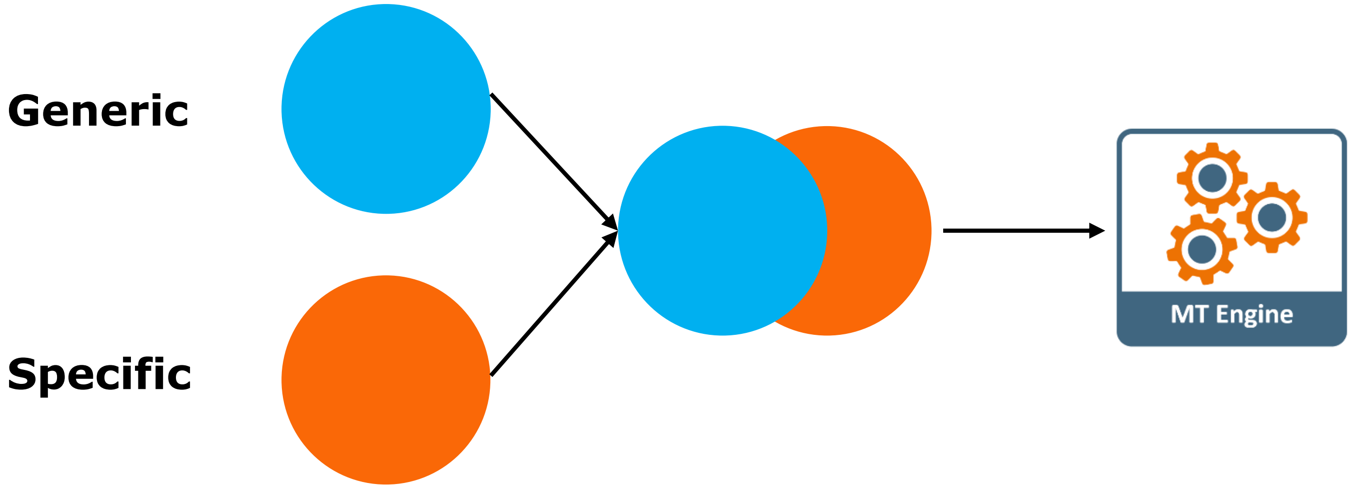
Domain Adaptation

Goal

Creation of an engine that correctly translates domain-specific texts and general language

Intuitive approach

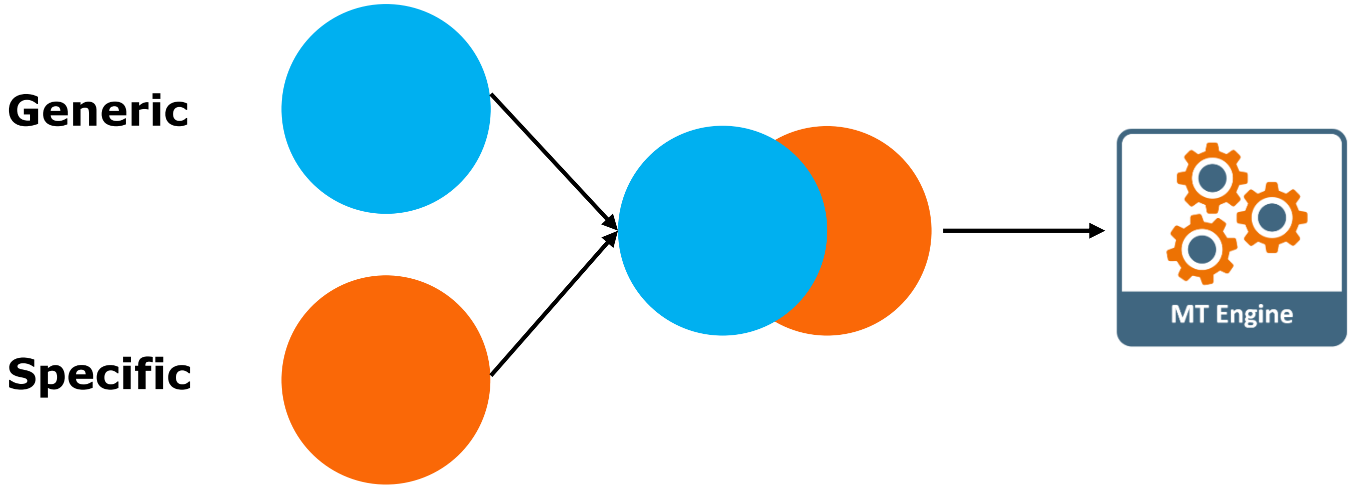
Training with combined generic and specific data



Domain Adaptation

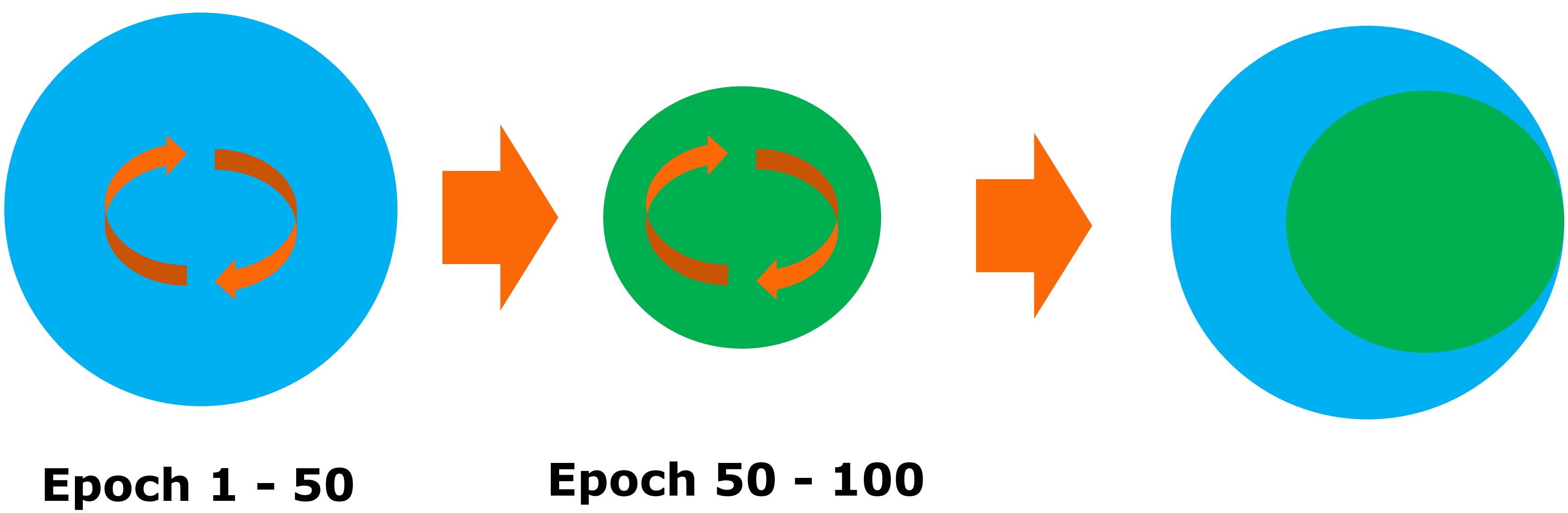
Problems

- Generic data dominates in-domain-data
- One engine per domain and LP (Language pairs x domains)



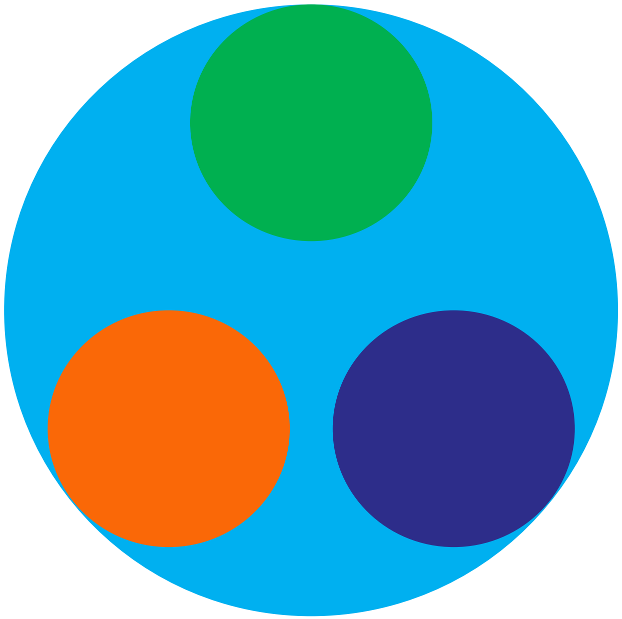
Domain Adaptation

Single-Domain-Engines with generic base



Domain Adaptation

Multi-Domain-Engine



- **Domains are treated like language pairs!**
- **Sentences are domain tagged**

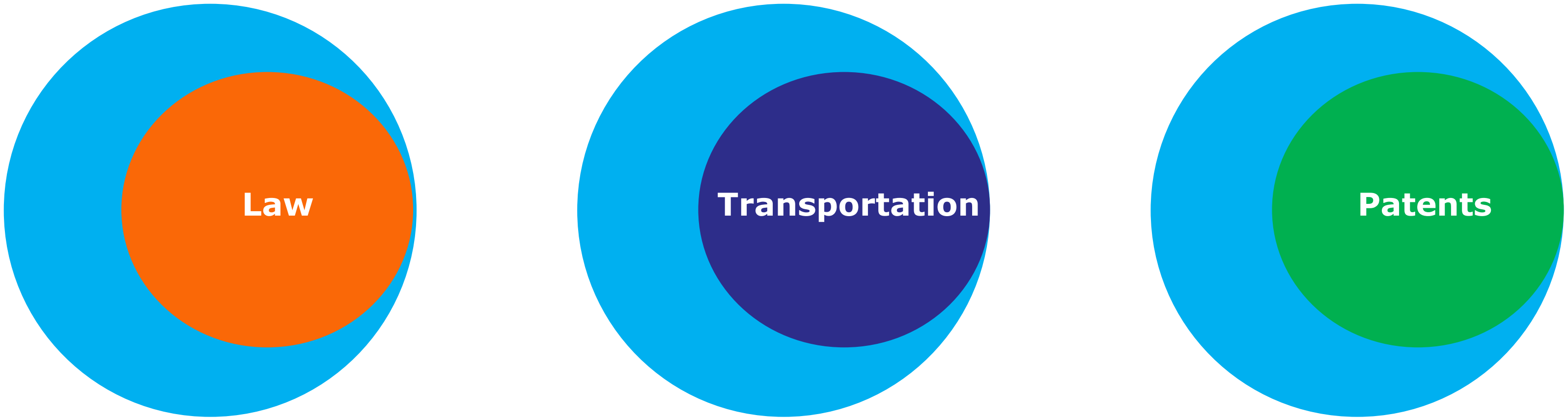
<LAE> This is a text on law

<PAT> This is a text on patents

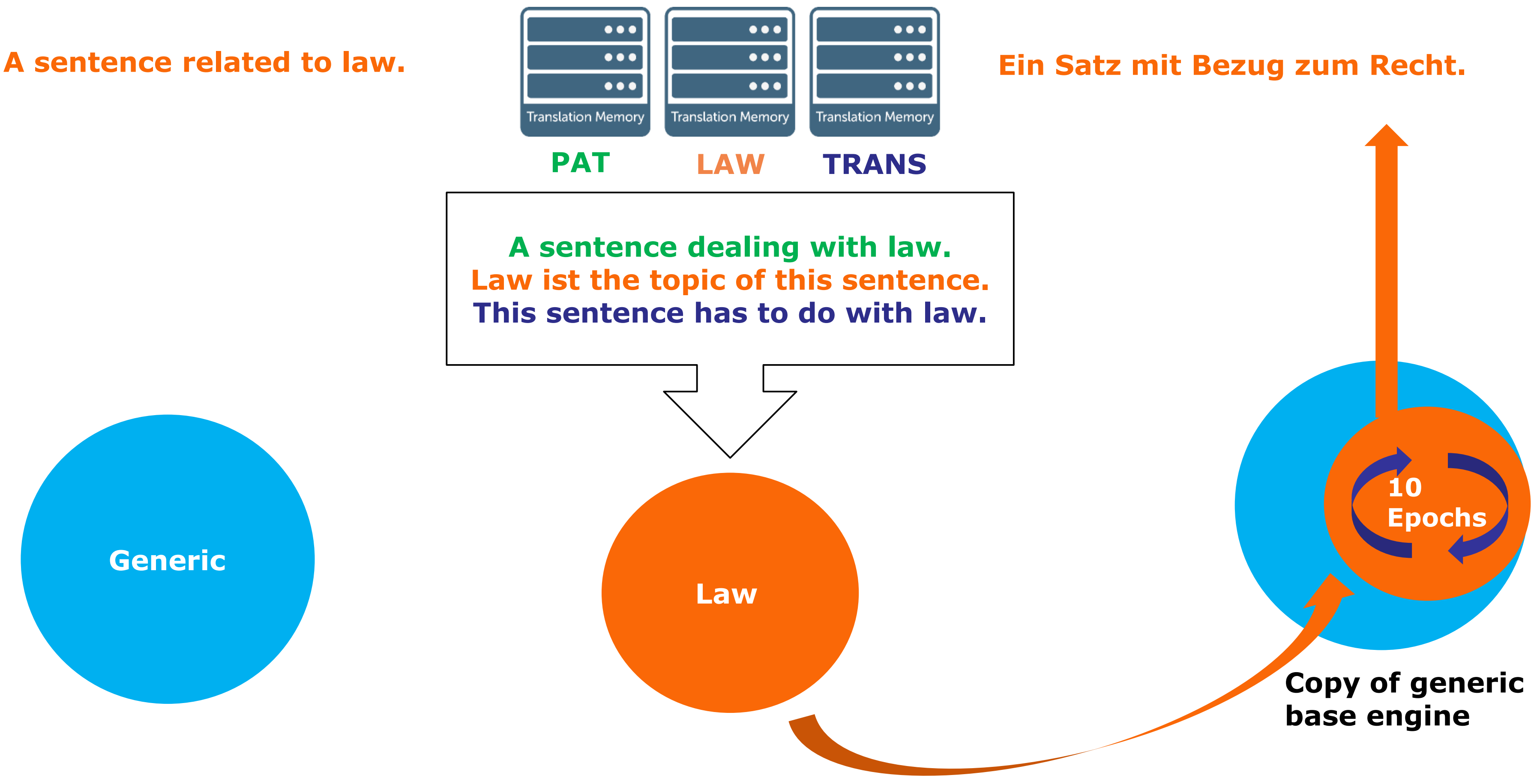
<TRANS> This is a text on transportation

Domain Adaptation

Single-Domain-Engines with generic base



Domain Adaptation (dynamic)



Benefits & Problems of NMT

SMT vs. NMT

SMT

Robustness

Idioms

Controlled language

Adequacy

Terminology

NMT

Morphology/ Inflection

Context awareness

Fluency

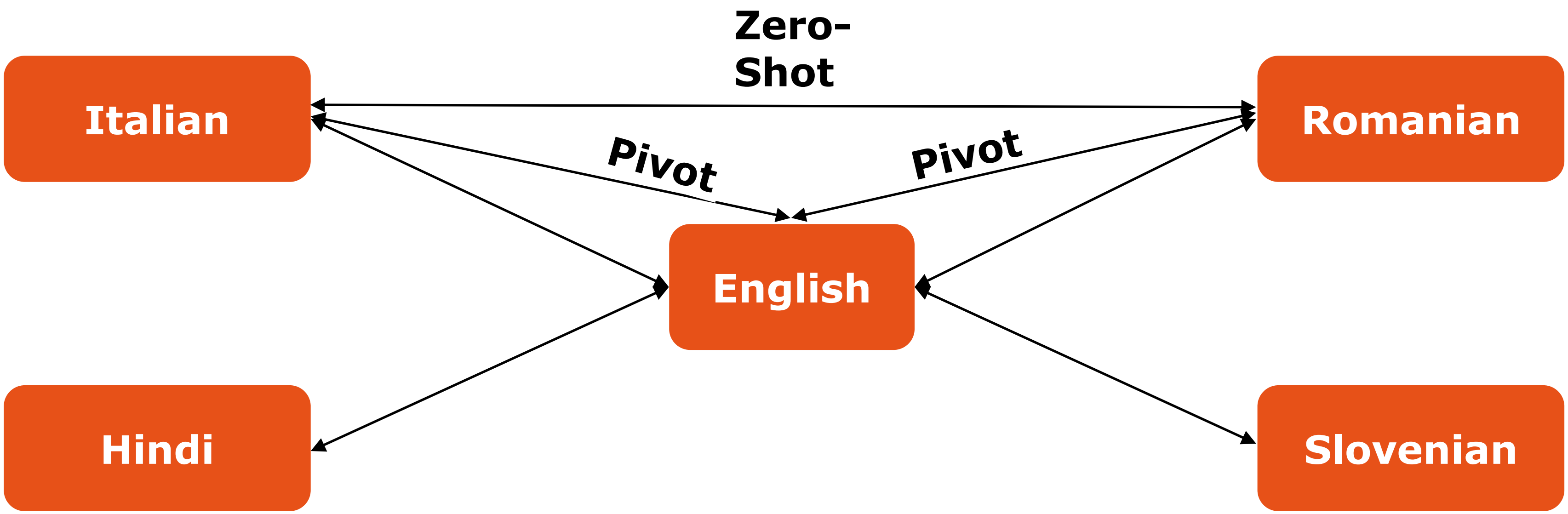
Word order

Bridging

Linguistic knownledge

Coherence

NMT Bridging



Problems with neural MT

Needs lots of data

Fluency can be deceptive

Terminology can not be integrated separately

Long sentences (> 20w)

Evaluation of NMT Output

BLEU has been established as the standard for SMT evaluations.

But: NMT is not like SMT!

The same meaning can be expressed in different ways:

Chancelor Merkel has said...
Merkel has said...
The chancellor of Germany said...

MT at the European Commission

iADAATPA, eTranslation

MT at the European Commission

2013

MT@EC

- 1 billion Segments from Euramis
- 24 languages
- 552 language pairs
- Domain specific

2017

eTranslation

Enables multilingualism for DSIs:

- API, Web UI
- Speed improvemens
- NMT

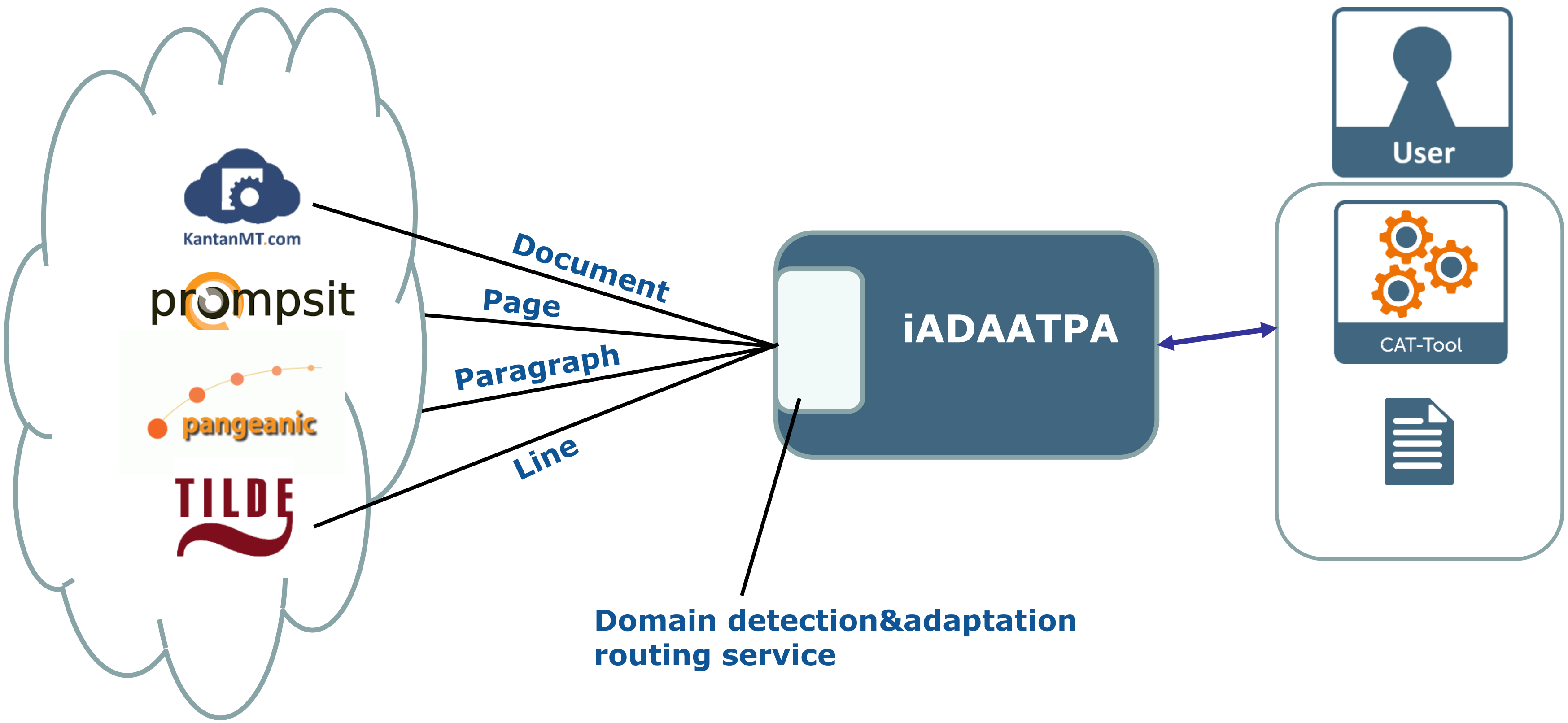
eTranslation TermBank

2018

iADAATPA

Intelligent, Automatic Domain Adapted Automated Translation for Public Administrations

MT at the European Commission



Single connection point to technologies and MT vendors