# Catching the Falling Knife of NMT

Ondřej Bojar
bojar@ufal.mff.cuni.cz
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University, Prague

November 1, 2017

# Catching the Falling Knife

When markets fall they fall by gravity.
There is no level one can calculate as bottom.
One needs to wait till the markets just fall and bottom out.
http://www.niveza.in/stock-news/learn-investing/dont-catch-the-falling-knife

# Recent WMT History

2013

English-Czech

| # | score | range | system |
|---|---|---|---|
| 1 | 0.580 | 1-2 | CU-BOJAR |
| | 0.578 | 1-2 | CU-DEPFIX |
| 3 | 0.562 | 3 | ONLINE-B |
| 4 | 0.525 | 4 | UEDIN |
| 5 | 0.505 | 5-7 | CU-ZEMAN |
| | 0.502 | 5-7 | MES |
| | 0.499 | 5-8 | ONLINE-A |
| | 0.484 | 7-9 | CU-PHRASEFIX |
| | 0.476 | 8-9 | CU-TECTOMT |
| 10 | 0.457 | 10-11 | COMMERCIAL-1 |
| | 0.450 | 10-11 | COMMERCIAL-2 |
| 12 | 0.389 | 12 | SHEF-WPROA |

# Recent WMT History

## 2013

E

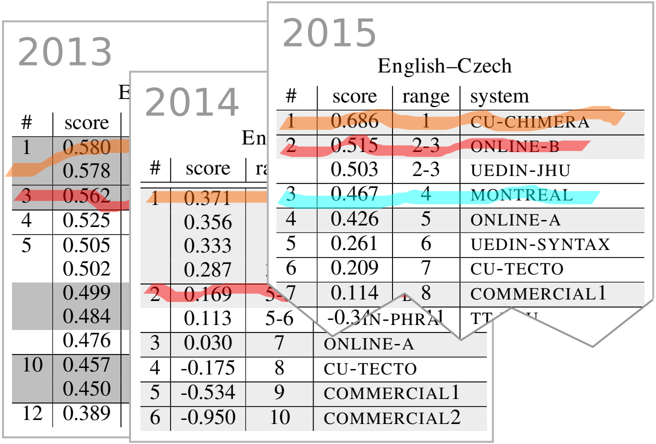
| # | score |
|---|---|
| 1 | 0.580 |
| | 0.578 |
| 3 | 0.562 |
| 4 | 0.525 |
| 5 | 0.505 |
| | 0.502 |
| | 0.499 |
| | 0.484 |
| | 0.476 |
| 10 | 0.457 |
| | 0.450 |
| 12 | 0.389 |

## 2014

English–Czech

| # | score | range | system |
|---|---|---|---|
| 1 | 0.371 | 1-3 | CU-DEPFIX |
| | 0.356 | 1-3 | UEDIN-UNCNSTR |
| | 0.333 | 1-4 | CU-BOJAR |
| | 0.287 | 3-4 | CU-FUNKY |
| 2 | 0.169 | 5-6 | ONLINE-B |
| | 0.113 | 5-6 | UEDIN-PHRASE |
| 3 | 0.030 | 7 | ONLINE-A |
| 4 | -0.175 | 8 | CU-TECTO |
| 5 | -0.534 | 9 | COMMERCIAL1 |
| 6 | -0.950 | 10 | COMMERCIAL2 |

# Recent WMT History



2013

| # | score |
|---|---|
| 1 | 0.580 |
| | 0.578 |
| 3 | 0.562 |
| 4 | 0.525 |
| 5 | 0.505 |
| | 0.502 |
| | 0.499 |
| | 0.484 |
| | 0.476 |
| 10 | 0.457 |
| | 0.450 |
| 12 | 0.389 |

2014

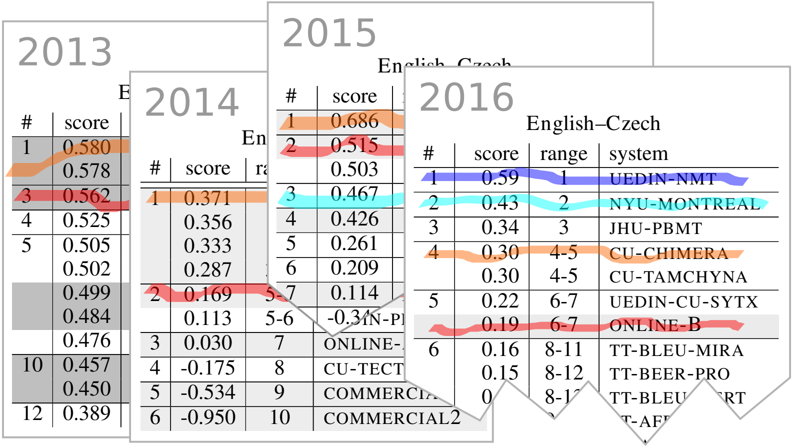
| # | score | ra | |
|---|---|---|---|
| 1 | 0.371 | | |
| | 0.356 | | |
| | 0.333 | | |
| | 0.287 | | |
| 2 | 0.169 | 5- | |
| | 0.113 | 5-6 | |
| 3 | 0.030 | 7 | ONLINE-A |
| 4 | -0.175 | 8 | CU-TECTO |
| 5 | -0.534 | 9 | COMMERCIAL1 |
| 6 | -0.950 | 10 | COMMERCIAL2 |

2015

English–Czech

| # | score | range | system |
|---|---|---|---|
| 1 | 0.686 | 1 | CU-CHIMERA |
| 2 | 0.515 | 2-3 | ONLINE-B |
| | 0.503 | 2-3 | UEDIN-JHU |
| 3 | 0.467 | 4 | MONTREAL |
| 4 | 0.426 | 5 | ONLINE-A |
| 5 | 0.261 | 6 | UEDIN-SYNTAX |
| 6 | 0.209 | 7 | CU-TECTO |
| 7 | 0.114 | 8 | COMMERCIAL1 |

# Recent WMT History



# CUNI Collective Efforts for WMT17

- Neural Monkey (Helcl and Libovický, 2017).
- NMT Training Task (Bojar et al., 2017).

- BPE, Learning rate and other meta-parameters.
- Batch sizing (smaller/larger/variable).
- Additional training objective:
  - Targetting GIZA++ alignments.
  - Scoring the *set* of produced words, disregarding position.
- Minibatch bucketing.
- Curriculum learning. (Kocmi and Bojar, 2017)
- Pre-trained embeddings.
- Domain adaptation: Subsample for Testset / Each Doc.
- Neural sys combination: Concatenative/Multi-encoder.

# … If Gains, then Mediocre …

- Neural Monkey (Helcl and Libovický, 2017).
- NMT Training Task (Bojar et al., 2017).

- ~~BPE, Learning rate and other meta-parameters.~~
- ~~Batch sizing (smaller/larger/variable).~~
- Additional training objective:
  - ~~Targetting GIZA++ alignments.~~
  - ~~Scoring the *set* of produced words, disregarding position.~~
- ~~Minibatch bucketing.~~
- ~~Curriculum learning.~~ (Kocmi and Bojar, 2017)
- ~~Pre-trained embeddings.~~
- Domain adaptation: Subsample for Testset / ~~Each Doc.~~
- ~~Neural sys combination: Concatenative/Multi-encoder.~~

# … But it Later Worked for Others!

- Neural Monkey (Helcl and Libovický, 2017).
- NMT Training Task (Bojar et al., 2017).

- ~~BPE, Learning rate and other meta-parameters.~~
- ~~Batch sizing (smaller/larger/variable).~~
- Additional training objective:
  - Targetting GIZA++ alignments.
  - Scoring the *set* of produced words, disregarding position.
- ~~Minibatch bucketing.~~
- ~~Curriculum learning.~~ (Kocmi and Bojar, 2017)
- ~~Pre-trained embeddings.~~
- Domain adaptation: Subsample for Testset / Each Doc.
- Neural sys combination: Concatenative/Multi-encoder.

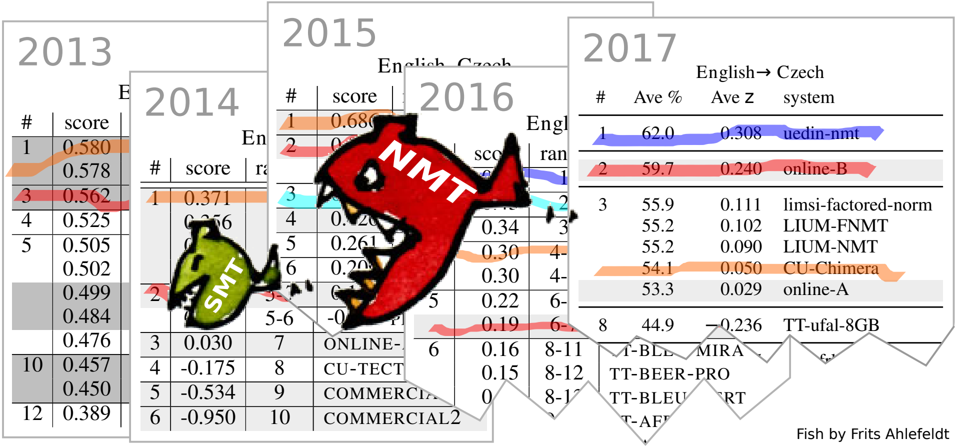
# Our WMT17 System

... so we sticked to phrase-based MT backbone:

- Moses system with several phrase tables:
  - Standard corpus-based one (synthetic mononews only!).
  - Output of TectoMT for the test set.
  - **Output of Nematus 2016 and Neural Monkey 2017.**
- Followed by Depfix (Rosa et al., 2012).
  - Fixing agreement.
  - Recovering lost negation.

All details in Sudarikov et al. (2017).

# … And the Result:




Fish by Frits Ahlefeldt

# We Were Hoping to Be the Second!

| | Manual | | Automatic Scores | | | | |
|---|---|---|---|---|---|---|---|
| # | Ave % | Ave z | BLEU | TER | CharacTER | BEER | System |
| 1 | **62.0** | **0.308** | **22.8** | **0.667** | **0.588** | **0.540** | uedin-nmt |
| 2 | **59.7** | **0.240** | 20.1 | 0.703 | 0.612 | 0.519 | online-B |
| 3 | 55.9 | 0.111 | 20.2 | **0.696** | 0.607 | **0.524** | limsi-factored |
| | 55.2 | 0.102 | 20.0 | 0.699 | - | - | LIUM-FNMT |
| | 55.2 | 0.090 | 20.2 | 0.701 | **0.605** | 0.522 | LIUM-NMT |
| | 54.1 | 0.050 | **20.5** | **0.696** | 0.624 | 0.523 | **CU-Chimera** |
| | 53.3 | 0.029 | 16.6 | 0.743 | 0.637 | 0.503 | online-A |
| 8 | 41.9 | -0.327 | 16.2 | 0.757 | 0.697 | 0.485 | PJATK |

Automatic scores by `http://matrix.statmt.org/`.

# Outline

1. How good UEDIN's WMT17 outputs are in fact.
   - Remaining errors.
   - News vs. doc-level phenomena.
2. An empirical comparison of toolkits.
3. What NMT offers to computational linguistics.

# Is UEDIN NMT That Much Better?

| SRC | 28-Year-Old Chef Found Dead at San Francisco Mall |
|---|---|
| | 28letý šéfkuchař Found Dead v San Francisco Mall |
| | Osmadvacetiletý šéfkuchař nalezen mrtev v obchodě v San Francisku |

# Is UEDIN NMT That Much Better?

| | |
|---|---|
| SRC | 28-Year-Old Chef Found Dead at San Francisco Mall |
| MT | 28letý šéfkuchař Found Dead v San Francisco Mall |
| REF | Osmadvacetiletý šéfkuchař nalezen mrtev v obchodě v San Francisku |

# Is UEDIN NMT That Much Better?

| | |
|---|---|
| SRC | 28-Year-Old Chef Found Dead at San Francisco Mall |
| MT | 28letý šéfkuchař Found Dead v San Francisco Mall |
| REF | Osmadvacetiletý šéfkuchař nalezen mrtev v obchodě v San Francisku |

| | |
|---|---|
| SRC | A 28-year-old chef who had recently moved to San Francisco was found dead in the stairwell of a local mall this week. |
| | Osmadvacetiletý kuchař, který se nedávno přestěhoval do San Francisca, byl tento týden nalezen mrtvý na schodišti místního obchodního centra. |
| | Osmadvacetiletý šéfkuchař, který se nedávno přistěhoval do San Franciska, byl tento týden $\emptyset$ schodech místního obchodu. |

# Is UEDIN NMT That Much Better?

| | |
|---|---|
| SRC | 28-Year-Old Chef Found Dead at San Francisco Mall |
| MT | 28letý šéfkuchař Found Dead v San Francisco Mall |
| REF | Osmadvacetiletý šéfkuchař nalezen mrtev v obchodě v San Francisku |

| | |
|---|---|
| SRC | A 28-year-old chef who had recently moved to San Francisco was found dead in the stairwell of a local mall this week. |
| MT | Osmadvacetiletý kuchař, který se nedávno přestěhoval do San Francisca, byl tento týden nalezen mrtvý na schodišti místního obchodního centra. |
| REF | Osmadvacetiletý šéfkuchař, který se nedávno přistěhoval do San Franciska, byl tento týden ∅ schodech místního obchodu. |

# Is UEDIN NMT That Much Better? (2/4)

| | |
|---|---|
| SRC | A spokesperson for Sons & Daughters **said** they were "shocked and devastated" by his death. |
| | Mluvčí společnosti Sons & Daughters **uvedla**, že jsou jeho smrtí "šokováni a zdrceni". |
| | Mluvčí restaurace Sons & Daughters **řekl**, že jsou jeho smrtí „šokováni a zničeni“. |

# Is UEDIN NMT That Much Better? (2/4)

| | |
|---|---|
| SRC | A spokesperson for Sons & Daughters **said** they were ”shocked and devastated” by his death. |
| MT | Mluvčí společnosti Sons & Daughters **uvedla**, že jsou jeho smrtí ”šokováni a zdrceni”. |
| REF | Mluvčí restaurace Sons & Daughters **řekl**, že jsou jeho smrtí „šokováni a zničeni“. |

# Is UEDIN NMT That Much Better? (2/4)

| | |
|---|---|
| SRC | A spokesperson for Sons & Daughters **said** they were "shocked and devastated" by his death. |
| MT | Mluvčí společnosti Sons & Daughters **uvedla**, že jsou jeho smrtí "šokováni a zdrceni". |
| REF | Mluvčí restaurace Sons & Daughters **řekl**, že jsou jeho smrtí „šokováni a zničeni“. |

| | |
|---|---|
| SRC | "He found an apartment, he was dating a girl," Louis Galici**a** told KGO. |
| | „Našel si byt, chodil s dívkou,“ řekl Louis Galici**a** pro KGO. |
| | "Našel si byt, chodil s holkou," řekl Louis Galici**e** KGO. |

# Is UEDIN NMT That Much Better? (2/4)

| | |
|---|---|
| SRC | A spokesperson for Sons & Daughters **said** they were "shocked and devastated" by his death. |
| MT | Mluvčí společnosti Sons & Daughters **uvedla**, že jsou jeho smrtí "šokováni a zdrceni". |
| REF | Mluvčí restaurace Sons & Daughters **řekl**, že jsou jeho smrtí „šokováni a zničeni“. |

| | |
|---|---|
| SRC | "He found an apartment, he was dating a girl," Louis Galici**a** told KGO. |
| REF | „Našel si byt, chodil s dívkou,“ řekl Louis Galici**a** pro KGO. |
| MT | "Našel si byt, chodil s holkou," řekl Louis Galici**e** KGO. |

# Is UEDIN NMT That Much Better? (3/4)

| | |
|---|---|
| SRC | The police arrested two men, who on Tuesday attacked a thirty-five-year-old man with a knife and a machete. |
| | Policie **obvinila** dva útočníky, kteří v úterý v centru Olomouce napadli nožem a mačetou pětatřicetiletého muže. |
| | Policie **zatkla** dva muže, kteří v úterý napadli pětatřicetiletého muže nožem a mačetou. |

# Is UEDIN NMT That Much Better? (3/4)

| | |
|---|---|
| SRC | The police arrested two men, who on Tuesday attacked a thirty-five-year-old man with a knife and a machete. |
| REF | Policie **obvinila** dva útočníky, kteří v úterý v centru Olomouce napadli nožem a mačetou pětatřicetiletého muže. |
| MT | Policie **zatkla** dva muže, kteří v úterý napadli pětatřicetiletého muže nožem a mačetou. |

# Is UEDIN NMT That Much Better? (3/4)

| | |
|---|---|
| SRC | The police arrested two men, who on Tuesday attacked a thirty-five-year-old man with a knife and a machete. |
| REF | Policie **obvinila** dva útočníky, kteří v úterý v centru Olomouce napadli nožem a mačetou pětatřicetiletého muže. |
| MT | Policie **zatkla** dva muže, kteří v úterý napadli pětatřicetiletého muže nožem a mačetou. |

| | |
|---|---|
| SRC | There were creative differences on the set and a disagreement. |
| | Došlo ke vzniku kreativních rozdílů na scéně a k neshodám. |
| | Na place byly tvůrčí rozdíly a neshody. |

# Is UEDIN NMT That Much Better? (3/4)

| | |
|---|---|
| SRC | The police arrested two men, who on Tuesday attacked a thirty-five-year-old man with a knife and a machete. |
| REF | Policie **obvinila** dva útočníky, kteří v úterý v centru Olomouce napadli nožem a mačetou pětatřicetiletého muže. |
| MT | Policie **zatkla** dva muže, kteří v úterý napadli pětatřicetiletého muže nožem a mačetou. |

| | |
|---|---|
| SRC | There were creative differences on the set and a disagreement. |
| REF | Došlo ke vzniku kreativních rozdílů na scéně a k neshodám. |
| MT | Na place byly tvůrčí rozdíly a neshody. |

# Is UEDIN NMT That Much Better? (4/4)

| | |
|---|---|
| SRC | **Economy Secretary** Keith Brown visited the site today and was among the first to walk from the land on to the bridge. |
| | **Ekonomický tajemník** Keith Brown stavbu dnes navštívil a byl mezi prvními, kteří přišli z pevniny na most. |
| | **Ministr hospodářství** Keith Brown dnes místo navštívil a byl mezi prvními, kteří vyšli ze země na most. |

# Is UEDIN NMT That Much Better? (4/4)

| | |
|---|---|
| SRC | **Economy Secretary** Keith Brown visited the site today and was among the first to walk from the land on to the bridge. |
| REF | **Ekonomický tajemník** Keith Brown stavbu dnes navštívil a byl mezi prvními, kteří přišli z pevniny na most. |
| MT | **Ministr hospodářství** Keith Brown dnes místo navštívil a byl mezi prvními, kteří vyšli ze země na most. |

# Luckily ;-), Catastropic Errors Happen

Also WMT17 UEDIN outputs (but not easy to spot):

| | |
|---|---|
| SRC | ... said Frank initially stayed in **hostels**... |
| MT | ... řekl, že Frank původně zůstal v **Budějovicích**... |
| ↳*Gloss* | ... said that Frank initically stayed in **Budweis**... |
| SRC | Most of the **Clintons'** income... |
| MT | Většinu příjmů **Kliniky**... |
| ↳*Gloss* | Most of the income of the **Clinic**... |
| SRC | The 63-year-old has now been made a special repres |
| MT | 63letý **mladík** se nyní stal zvláštním zástupcem... |
| ↳*Gloss* | The 63-year-old **youngster** has now become a speci |

# Catastropic Errors Happen (2/2)

| | |
|---|---|
| SRC | Criminal Minds star Thomas Gibson sacked after hitting producer |
| REF | Thomas Gibson, hvězda seriálu Myšlenky zločince, byl propuštěn po té, co uhodil režiséra |
| MT | **Kriminalisté Minsku** hvězdu Thomase Gibsona **vyhostili** po **zásahu** producenta |
| ↳*Gloss* | **Minsk criminal investigators** have **expelled** the star Thomas Gibson after **striking** the producer |

| | |
|---|---|
| SRC | ...add to that its long-standing grudge... |
| REF | ...přidejte k tomu svou dlouholetou nenávist... |
| MT | ...přidejte k tomu svou dlouholetou **záštitu**... |
| ↳*Gloss* | ...add to that its long-standing **auspices**...<br>(grudge = zášť → záštita = auspices) |

# UEDIN at WMT17

- Our small annotation of up to 185 sentences.
- Blind mix: reference or MT.

Real MT was assumed to be:

| | OB | DM | DV |
|---|---|---|---|
| MT | 142 (76.8 %) | 86 (77.5 %) | 72 (87.8 %) |
| didn't know | 34 (18.4 %) | 9 (8.1 %) | 6 (7.3 %) |
| human | 9 (4.9 %) | 16 (14.4 %) | 4 (4.9 %) |
| Total | 185 (100.0 %) | 111 (100.0 %) | 82 (100.0 %) |

$\Rightarrow$ 10–20% of outputs indistinguishable from humans.

# UEDIN at WMT17

- Our small annotation of up to 185 sentences.
- Blind mix: reference or MT.

Real MT was assumed to be:

| | OB | DM | DV |
|---|---|---|---|
| MT | 142 (76.8 %) | 86 (77.5 %) | 72 (87.8 %) |
| didn't know | 34 (18.4 %) | 9 (8.1 %) | 6 (7.3 %) |
| human | 9 (4.9 %) | 16 (14.4 %) | 4 (4.9 %) |
| Total | 185 (100.0 %) | 111 (100.0 %) | 82 (100.0 %) |

⇒ 10–20% of outputs indistinguishable from humans.

| | OB | DM | DV |
|---|---|---|---|
| almost flawless | 17 (9.19 %) | 2 (1.80 %) | 0 (0 %) |
| flawless | 82 (44.32 %) | 37 (33.33 %) | 27 (32.93 %) |

⇒ 30–50% of outputs flawless or almost flawless.

# Errors Flagged

| # | % | Error Type |
|---|---|---|
| 35 | 26.1 | lexical error |
| | | ↳ 13 carriages, 2 altercations, 2 decks, 2 plantings, ... |
| 22 | 16.4 | notNice |
| 10 | 7.5 | **namedEntity** |
| 12 | 9.0 | **world knowledge needed or helpful** |
| 7 | 5.2 | terminology |
| | | ↳ winter wheat, winter barley, emergency kill cord, ... |
| 7 | 5.2 | grammar |
| 6 | 4.5 | extra |
| 5 | 3.7 | minor |
| 5 | 3.7 | anaphora |
| 3 | 2.2 | valency |
| 2 | 1.5 | global sentence structure |
| 2 | 1.5 | units (CZK $\neq$ GBP) |
| 2 | 1.5 | BPE |
| 2 | 1.5 | SRL |
| 14 | 10.4 | Other, 1 occurrence each |
| 134 | 100.0 | Total flags |

# Negation in Nematus en→cs

Manual analysis of HimL (medical) and news by Rudolf Rosa:

| | Annotated | Negated | | Meaning Correct | | Error | |
|---|---|---|---|---|---|---|---|
| | | No | Yes | Yes | No | in Negation | Elsewhere |
| Czech sents | 298 | 237 | 61 | 55 | 6 | 2 | 4 |
| % of Annotated | | 79.5% | 20.5% | 18.5% | 2.0% | **0.7%** | 1.3% |
| % of Negated | | | 100% | 90% | 10% | 3.3% | 6.6% |

- Errors in negation very rare.
- In many cases, hard negation phenomena correct.
- No single error in Czech double negation.
- Lexicalized negation also handled perfectly:
  - was slurring = mluvila nesrozumitelně, recently = nedávno, unfortunately = bohužel, homeless = bezdomovce, failing to coordinate = nekoordinovala, rather than = nikoliv.
- In total, only 2 clear errors:
  - 1 missing negation,
  - 1 incorrect negation scope (due to subject-object marking error).

# Doc-Level Effects in News?

Does MT have to consider cross-sentence phenomena?
Manual annotation of 40–92 “paragraphs” from WMT11:

- 4 consecutive sentences per “paragraph”.
- 4 manual versions of each sentence:
  - Original Czech / translation from English
  - Translation from German (3 different translations).
- Some paragraphs “clean”, some “mixed” (each sentence coming from a different source).
- Blind annotation to identify clean vs mixed.
  - 71–78% of “mixed” paragraphs marked as clean.
  - 17–21% of “clean” paragraphs marked as mixed.

$\Rightarrow$ in up to 80% sents, source probably captures everything.
$\Rightarrow$ in up to 20%, humans seem to produce incoherent text.
$\Rightarrow$ News domain exhibits too few cross-sentence links.

# Call for WMT18 Test Suites

Burlot and Yvon (2017): test suite with automatic checks.

1. Create contrastive source sentence pairs.
2. Have everyone translate them.
3. (Automatically) check if the desired phenomenon is handled as expected.

# Call for WMT18 Test Suites

Burlot and Yvon (2017): test suite with automatic checks.

1. Create contrastive source sentence pairs.
2. Have everyone translate them.
3. (Automatically) check if the desired phenomenon is handled as expected.

My goal for WMT18:

- Extend the standard 3k news test sentences with *your* contributions:
  - Contrastive source sentence pairs.
  - Automatic checks of outputs.
- Participants will translate everything.
- *You* will then evaluate your portion of the test set.

$\Rightarrow$ Collectively, we will focus on many specific things.

# Take-Home Message #1

In large-data settings:

- NMT has sufficiently resolved:
  - morphology,
  - negation.
- Remaining errors concern primarily:
  - world knowledge,
  - terminology,
  - rare words (incl. named entities),
  - anaphora.

Dedicated test suites needed:

- otherwise we'd be evaluating generally solid outputs.

Contact me to extend WMT18 test set with your data.

# Open-Source Tools

| | Toolkit / Language |
|---|---|
| DL4MT | Theano / Python |
| ↳ **Nematus** | Theano / Python |
| **Marian (incl. AmuNMT)** | C++ |
| seq2seq_attn | Torch / Lua |
| ↳ **OpenNMT** | Torch / Lua+Python |
| Lamtram | DyNet / C++ |
| **Neural Monkey** | Tensorflow / Python |
| Google seq2seq | Tensorflow / Python |
| **Google tensor2tensor (Transformer)** | Tensorflow / Python |

- Hard to choose one, all have their goods and bads.
- There is always a big cost of getting it running.

A more complete list by Jon Dehdari: `https://github.com/jonsafari/nmt-list`

# Nematus vs. Neural Monkey

| System | Greedy BLEU | Beam BLEU | Training Time |
|---|---|---|---|
| Neural Monkey | 21.13 | 22.74 | 4d07h |
| Neural Monkey + ReLu | 21.97 | 23.14 | 4d17h |
| Neural Monkey + ReLu + Softmax Fix | 21.73 | 22.99 | 4d18h |
| Nematus closest setup | 22.77 | 24.32 | 10d |

- Nematus better in BLEU but two times slower.
- Neural Monkey got 1.5 BLEU improvement from:

    `initializer=tf.random_uniform_initializer(-0.5, 0.5)`

    vs. `initializer=tf.random_uniform_initializer(-0.5, 0-5)`
- ~2 BLEU point from TF change from 0.11 to 1.0
    - One reason is ReLU becoming the default activation function.

Common: Czech→English, CzEng 1.6 limited to 30M sent pairs. Fixed BPE 30k. (BLEU evaluated on the BPE). Encoder: emb 512; max length 50; RNN size 1000; GRU with no dropout. Decoder: emb 512; max length 50; RNN size 1000; conditional GRU with no dropout. Optimization: Adam with learning rate $10^{-4}$ optimized on the cross entropy. Batch size was 60. Beam search: maximum steps 50; length normalization 0.6; beam size 20. NM run on GeForce 1080Ti, Nematus run on GeForce 1080.

# A Little Messy Empirical Comparison

- Czech→English, CzEng 1.6 limited to 30M sent. pairs.
- Fixed BPE 30k. (BLEU evaluated on the BPE).

| System | BLEU | Steps | Training Time |
|---|---|---|---|
| Deep Nematus 2017 | 28.96 | 30M (1 ep) | **25d8h** |
| Transformer 201k*2k | 27.84 | 201k (∼1 ep) | **1d06h** |
| Shallow Nematus (best known) | 25.93 | 30M (1 ep) | 10d |
| Marian 1 GPU | 24.65 | 29M (∼1 ep) | **1d11h** |
| Neural Monkey (not best setup) | 23.14 | 30M (1 ep) | 5d17h |
| OpenNMT default setting, 4 GPU | 21.72 | 30M (1 ep) | 15h |
| Transformer 8*GPU 843k*1k + avg | 32.68 | ∼4.2 ep | 7d00h |
| Transformer 1820k*2k + avg | 31.85 | ∼9.0 ep | 11d04h |
| Transformer 1383k*2k + avg | 31.72 | ∼6.9 ep | 8d12h |
| Transformer 699k*2k + avg | 30.86 | ∼3.4 ep | 4d08h |
| Transformer 699k*2k | 30.43 | ∼3.4 ep | 4d08h |
| Marian 1 GPU | 26.66 | 124M (∼4.1 ep) | 4d16h |
| OpenNMT default, 4 GPU | 26.05 | ∼17 ep | |

1 ep = 30 M sent pairs, 1 step = 1 sent. pair *or* 2048 BPE tokens

# Take-Home Message #2

- Terrible amount of man/GPU time easily wasted in chasing toolkits and baselines.

⇒ Pick your favourite one and stick with it.
⇒ Try to respect the choice of others when reviewing.

My personal picks: Marian, tensor2tensor, Neural Monkey

# Take-Home Message #2

- Terrible amount of man/GPU time easily wasted in chasing toolkits and baselines.

⇒ Pick your favourite one and stick with it.
⇒ Try to respect the choice of others when reviewing.

My personal picks: Marian, tensor2tensor, Neural Monkey

*Now* I understand the frustration of trying to catch up with the WMT benchmark.

# Take-Home Message #2

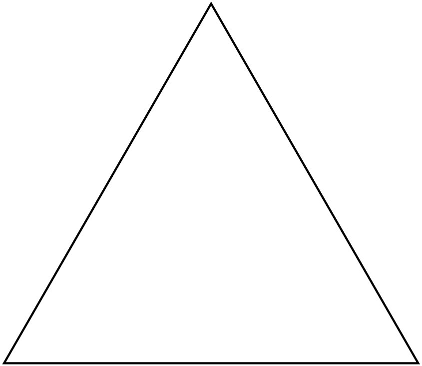
- Terrible amount of man/GPU time easily wasted in chasing toolkits and baselines.

⇒ Pick your favourite one and stick with it.
⇒ Try to respect the choice of others when reviewing.

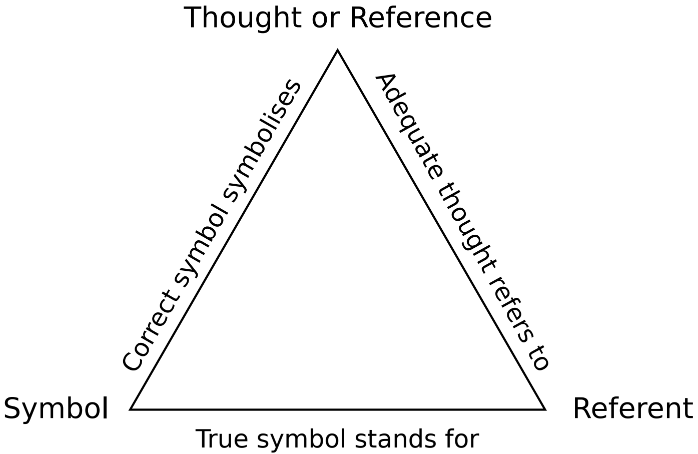
My personal picks: Marian, tensor2tensor, Neural Monkey

*Now* I understand the frustration of trying to catch up with the WMT benchmark.

Anyone interested in a really constrained translation task?

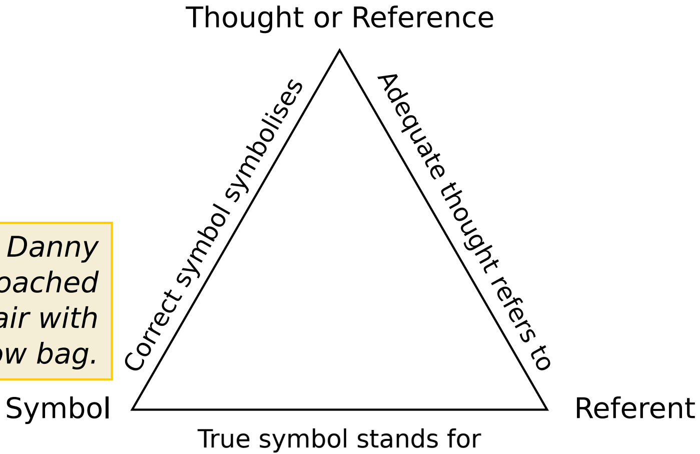


# Semiotic Triangle by Ogden and Richards

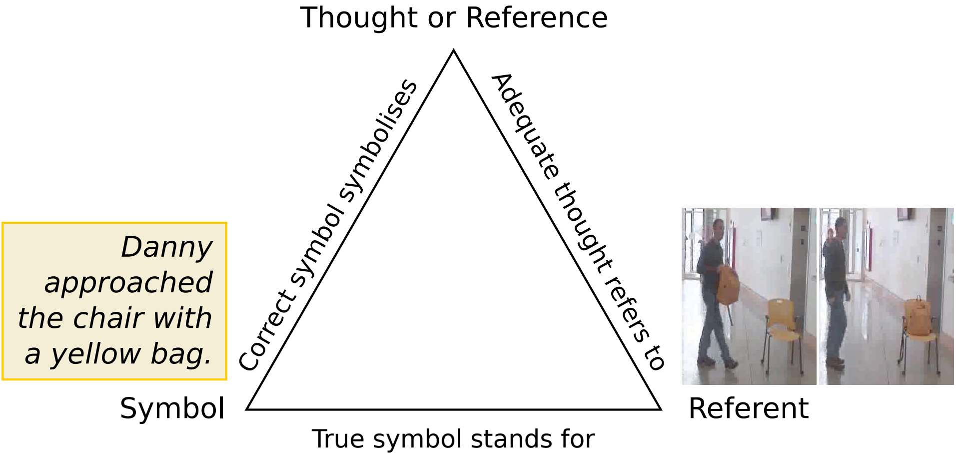


# Semiotic Triangle by Ogden and Richards

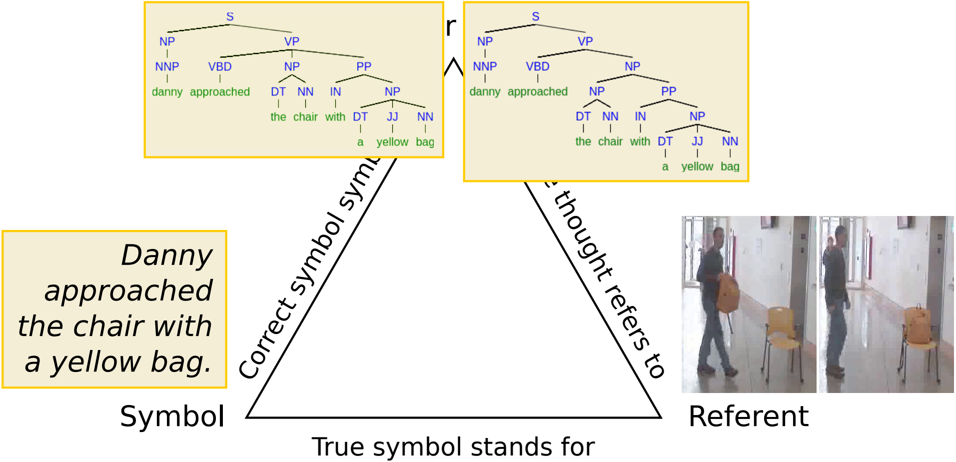
*Danny approached the chair with a yellow bag.*



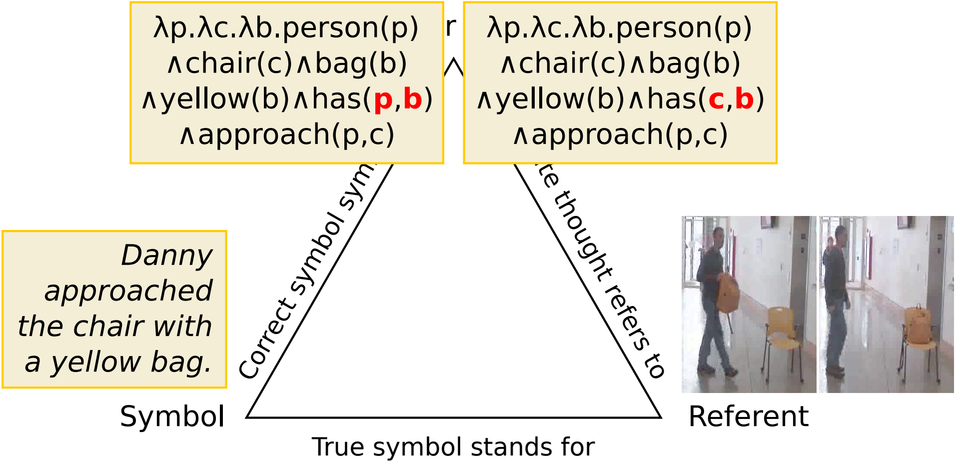
# Semiotic Triangle by Ogden and Richards



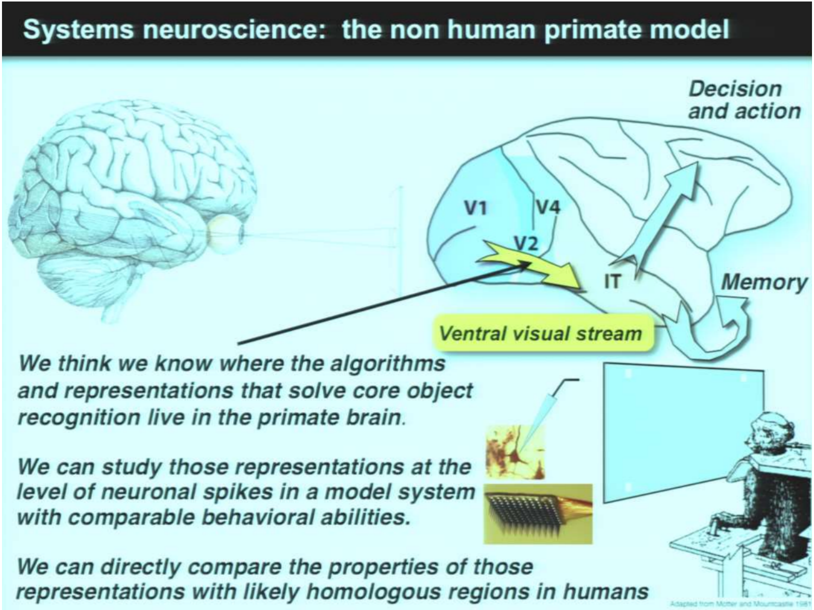
# Semiotic Triangle by Ogden and Richards



# Semiotic Triangle by Ogden and Richards



# DiCarlo NIPS 2013 Tutorial on Vision



# From Vision to Language

DiCarlo (2013): Human object recognition explained by:

- Recording apes' neuronal activity
  and attaching a single-layer NN to interpret it
- Measuring human performance

... on the same object recognition tasks.

- and relating them.

# From Vision to Language

DiCarlo (2013): Human object recognition explained by:

- Recording apes' neuronal activity
  and attaching a single-layer NN to interpret it
- Measuring human performance

... on the same object recognition tasks.

- and relating them.

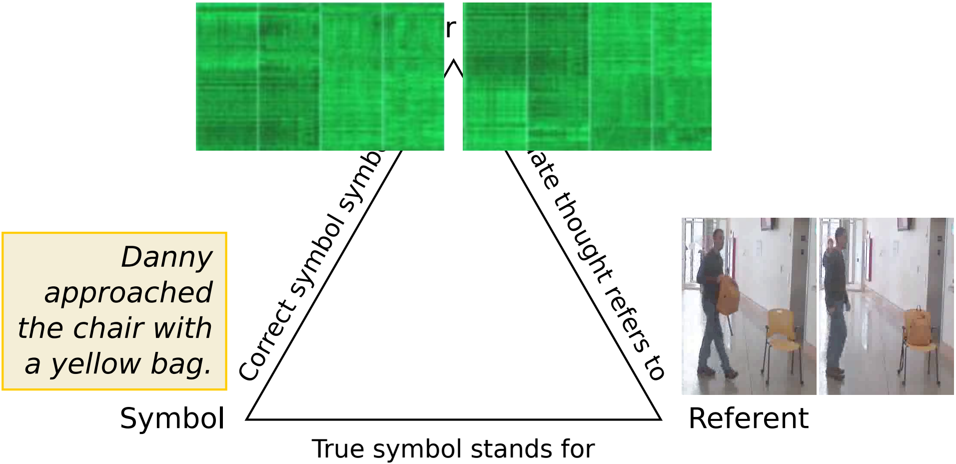
Proposal: Instead of catching the falling knife:

- Record NMT/NN behaviour (all parameters accessible)
- and human behaviour, possibly recording:
  - Objective: reading studies, eye-tracking, ...
  - Subjective: introspection.

... on the same language processing tasks.

- and relate them.

# Semiotic Triangle by Ogden and Richards



# Some Techniques of NN Inspection

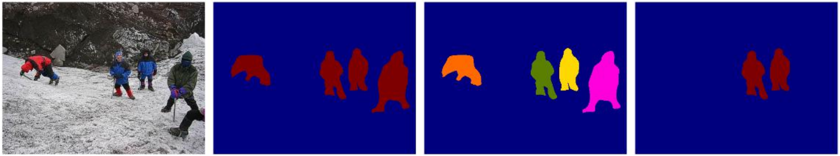
- MicroNNs, e.g. Shi et al. (2016) learning length.
- Lobotomy (Li et al., 2016).
- Observing activations, attentions...
- Exploring representation space.
  - t-SNE and PCA for sentence pairs
  - Translation by search = similarity in meaning reflected in space
  - Attaching an NN to see if it can infer:
    - POS or morphology from NMT
    - Subject-Verb agreement (Linzen et al. TACL/EACL 2017)
- Linguistic exploration:
  - Various test suites (Burlot 2017, Burchhardt MQM, Lingeval97).
  - Stanford Natural Language Inference (SNLI) `https://nlp.stanford.edu/projects/snli/`
  - Paraphrases (Dreyer and Marcu, 2012; Bojar et al., 2013).
- Comparing representations (Nili et al., 2014).

# Aspects of Meaning

- **Meaning is a coarsening:**
  - Pictures: Semantic segmentation ("reverse raytracing")
  - Programs: The output they give (caveat: undecidable).
  - CL: Reference to real world? Speaker's intention?
- Meaning can be shifted, modified.
- Meanings can be compared.
- Meaning is generally compositional.
  - (along the linguistic structure).
- Pragmatics: Named entities, numbers, anaphora...
- Expressions are ambiguous.
- Meanings are vague.
- **Are meanings stateful?**
- **Are meanings continuous?**

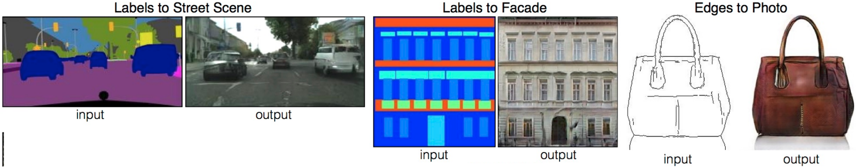
# Meaning as a Coarsening

## Semantic Segmentation of Pictures



(a) input image (b) object class segmentation of class ***people*** (c) object instance segmentation of class ***people*** (d) segmentation from expression *"people in blue coat"*

## ... and generating back with pix2pix:




Illustrations from `http://www.cs.toronto.edu/~tingwuwang/semantic_segmentation.pdf` and `https://github.com/phillipi/pix2pix`.

# Meaning Statefulness

Stateful Meaning Representation:

- “The state of mind after having read this and produced this output so far.”
- Corresponds to models with attention.
- Btw needed to interpret humour (Gluscevskij, 2017).

Stateless Meaning Representation:

- Points correspond to meanings.
  - As in models without attention.

# Continuous Spaces in NMT

- … are plentiful:
  - Word, sentence, document embeddings…
- … (probably) crucially depend on the task:
  - NMT vs. QA vs. summarization vs. sentiment …
- … (probably) depend on the architecture.
- … (obviously) depend on the point in the architecture.

Desired properties of continuous sentence representations:
(by Schwenk and Douze (2017)):

- semantic closeness,
- multilingual closeness, incl. across many languages,
- preservation of content (task-specific).

… plus the properties I listed three slides back.

# Is Sentence Meaning Continuous?

## 10k–100k sentence paraphrases in English and Czech

(Dreyer and Marcu, 2012; Bojar et al., 2013)

*Premiere of Iraq Nuri al-Maliki was given an excuse by President Bush, who expressed his confidence in him, and he stated that the circumstances are complicated.*

# Is Sentence Meaning Continuous?

## 10k–100k sentence paraphrases in English and Czech

(Dreyer and Marcu, 2012; Bojar et al., 2013)

*Premiere of Iraq Nuri al-Maliki was given an excuse by President Bush, who expressed his confidence in him, and he stated that the circumstances are complicated.*

*President Bush said that he trusts in Nouri Maliki, head of government of Iraq, and he stated that he finds an excuse for him "because the situation is tricky".*

*Head of cabinet of Iraq Nuri al-Maliki was given an excuse by President Bush, who expressed his trust in him, and he indicated that the circumstances are difficult.*

*Iraq's head of cabinet Nuri al-Maliki was given a reason by President Bush, who expressed his trust in him, and he indicated that the case is tricky.*

*President Bush said that he has faith in Iraqi head of cabinet Nouri al-Maliki, and he stated that he finds an excuse for him "for the case is complicated".*

# Is Sentence Meaning Continuous?

## 10k–100k sentence paraphrases in English and Czech

(Dreyer and Marcu, 2012; Bojar et al., 2013)

*Premiere of Iraq Nuri al-Maliki was given an excuse by President Bush, who expressed his confidence in him, and he stated that the circumstances are complicated.*
*President Bush said that he trusts in Nouri Maliki, head of government of Iraq, and he stated that he finds an excuse for him "because the situation is tricky".*
*Head of cabinet of Iraq Nuri al-Maliki was given an excuse by President Bush, who expressed his trust in him, and he indicated that the circumstances are difficult.*
*Iraq's head of cabinet Nuri al-Maliki was given a reason by President Bush, who expressed his trust in him, and he indicated that the case is tricky.*
*President Bush said that he has faith in Iraqi head of cabinet Nouri al-Maliki, and he stated that he finds an excuse for him "for the case is complicated".*

Q: Are all these paraphrases close in sent embedding spaces?

Q: How entagled are manifolds of *different* sents?

… pushing Holger Schwenk to work on this with me.

# Examining Continuous Space of Sents.

Stages of Space Mapping:

1. Propose directions of exploration.
2. Generate seed pairs of sentences for each of the directions.
3. Collect specimens along the proposed directions:
   - interpolation, a “sentence in between”,
   - extrapolation, “a sentence further in the hinted direction”.
   - Allow people to say “impossible”.
4. Validate the relations.
5. Create the partially ordered set.
6. Search for a manifold covering the ordered set.

Work in progress with Chris Callison-Burch.

# Directions of Exploration (1/2)

- Politeness.
- Tense.
- Verity: How much the speaker believes the message.
- Modality: Willingness/Ability of the speaker to do it.
- “Counting” / Generic Numerals, Scalar adjectives.
  - I saw a handful of people there. / a big crowd / a massive crowd.
  - freezing / cold / chilly
- “Negation”, but not only reversing the main predicate.
- Complexity / simplicity, Length.

# Directions of Exploration (2/2)

- Specificity / Generality, Vagueness.
  - Geese fly / Geese migrate / Geese migrate south / The Canadian geese flew over the pond at friendly Farms in their southward migration.
  - Hammer the hook into the wall. / Put the hook on the wall. / Do the thingy in there.
- Contextual boundness.
  - Give it to him. / Give the parcel to the man at the counter. / Give your parcel to the operator at the post office.
- High/low style/English/class.
  - Hey y'all it's a nice day ain't it?
  - Greetings! Lovely weather we are having.

Thanks to Sarka Zikanova for some of the ideas.
Looking forward for any other ideas you can suggest.

# First Results of Getting Pairs

| | |
|---|---|
| Can you please give me a minute? | Could you leave me alone? |
| Close the door. | Close the damn door man |
| Can you help me find something? | I need you to help me get some |
| May I talk to Mary? | Is Mary here? |
| I'm sorry-I don't believe we have met. | Who the hell are you? |
| Can you move so I can see the screen? | You aren't made of glass, you k |
| Will you kindly exit? | I do not want you here! |
| Would you please get the mail? | Get the mail! |
| Can I help you? | What do you want? |
| Can you please help me with this? | Get over here and help me! |
| Can you make me breakfast? | Why are you not making me br |
| I tried to call were you busy? | You never answer your phone. |

# First Results of Midpointing (1/3)

| Can you help me find something? |
|---|
| Find this for me. |
| Help me find something. |
| Please help me find something. |
| Will you help me? |
| Would you help me look? |
| Your assistance in finding something is required. |
| I need you to help me get something. |

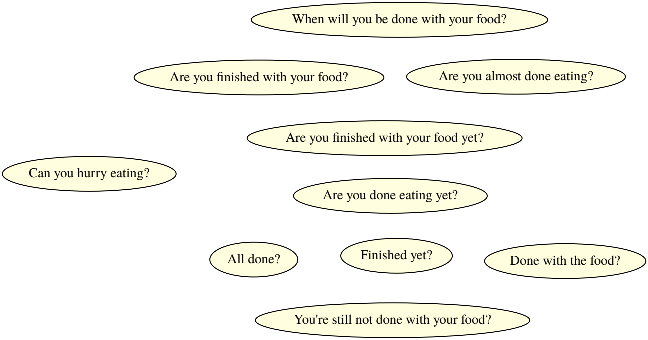
# First Results of Midpointing (2/3)

| Can you please give me a minute? |
| --- |
| Come back later |
| Give me a minute. |
| Hey give me a minute. |
| I'd like a minute alone. |
| I need a minute to myself. |
| I need more time. |
| One minute. |
| One moment. |
| Please wait. |
| Could you leave me alone? |

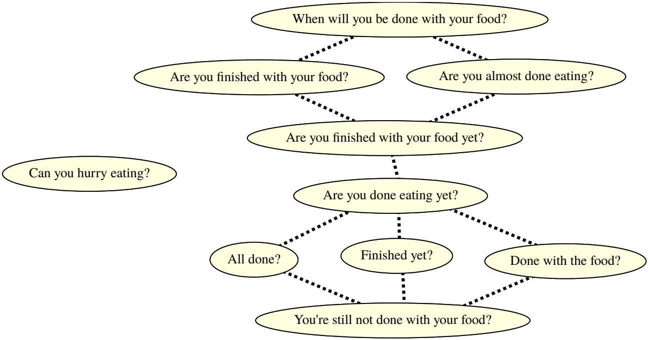
# First Results of Midpointing (3/3)

| Can you move so I can see the screen? |
|---|
| Blocking the view, friend. |
| Can you move a bit? |
| Can you please move? |
| Could you move a little bit, you're blocking the screen. |
| Hey can you move. |
| I can't see, can you move a little? |
| Move your blocking the screen |
| Please move. |
| You aren't made of glass, you know. |

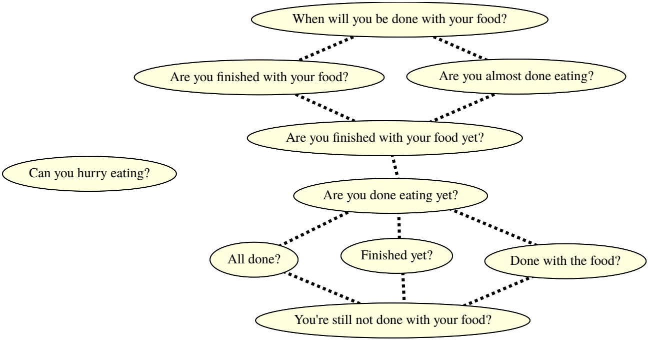
# Collect All Variations

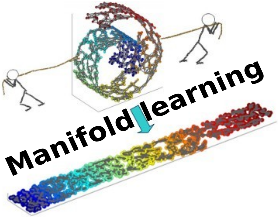


# Ask Crowd to Partially Sort Them

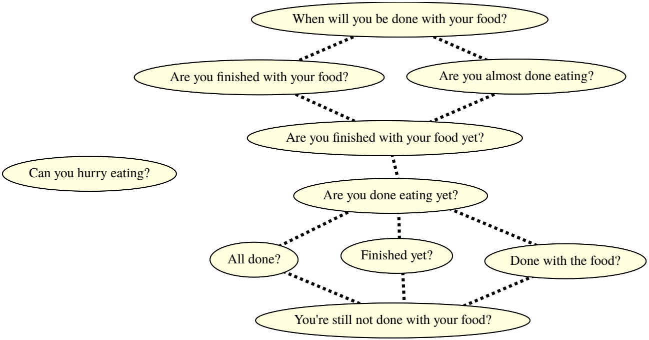


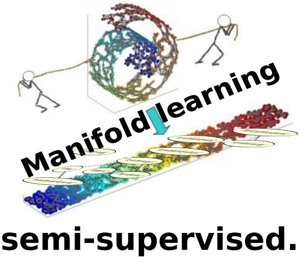
# Find Methods for Manifold Learning



# Match Posets with Learned Manifolds



# Take-Home Message #3


Fish by Frits Ahlefeldt

# Summary

- Neural MT reaches and can surpass humans.
  - Catastrophic errors still possible.
- As a side-effect, continuous representations are learned.
- Insight in vision thanks to relating
  ape, computer and human vision.

# Summary

- Neural MT reaches and can surpass humans.
  - Catastrophic errors still possible.
- As a side-effect, continuous representations are learned.
- Insight in vision thanks to relating
  ape, computer and human vision.

- Computational linguistics has plenty of data.
- Other data can be relatively easily obtained.

$\Rightarrow$ Let's train NMT/NLU systems and dissect them.

# Summary

- Neural MT reaches and can surpass humans.
  - Catastrophic errors still possible.
- As a side-effect, continuous representations are learned.
- Insight in vision thanks to relating
  ape, computer and human vision.

- Computational linguistics has plenty of data.
- Other data can be relatively easily obtained.

⇒ Let's train NMT/NLU systems and dissect them.

FinMT FiNMT MT. Fin? I'm not afraid.

# References

Ondřej Bojar, Matouš Macháček, Aleš Tamchyna, and Daniel Zeman. 2013. Scratching the Surface of Possible Translations. In *Proc. of TSD 2013*, Lecture Notes in Artificial Intelligence, Berlin / Heidelberg. Západočeská univerzita v Plzni, Springer Verlag.

Ondřej Bojar, Jindřich Helcl, Tom Kocmi, Jindřich Libovický, and Tomáš Musil. 2017. Results of the WMT17 Neural MT Training Task. In *Proceedings of the Second Conference on Machine Translation*, Copenhagen, Denmark, September. Association for Computational Linguistics.

Franck Burlot and François Yvon. 2017. Evaluating the morphological competence of machine translation systems. In *Proceedings of the Second Conference on Machine Translation, Volume 1: Research Papers*, pages 43–55, Copenhagen, Denmark, September. Association for Computational Linguistics.

J. J. DiCarlo. 2013. Mechanisms underlying visual object recognition: Humans vs.neurons vs. machines. NIPS Tutorial.

Markus Dreyer and Daniel Marcu. 2012. HyTER: Meaning-Equivalent Semantics for Translation Evaluation. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 162–171, Montréal, Canada, June. Association for Computational Linguistics.

Dmitrij Gluscevskij. 2017. Methodological issues and prospects of semiotics of humour. *Sign Systems Studies*, 45(1/2):137–151.

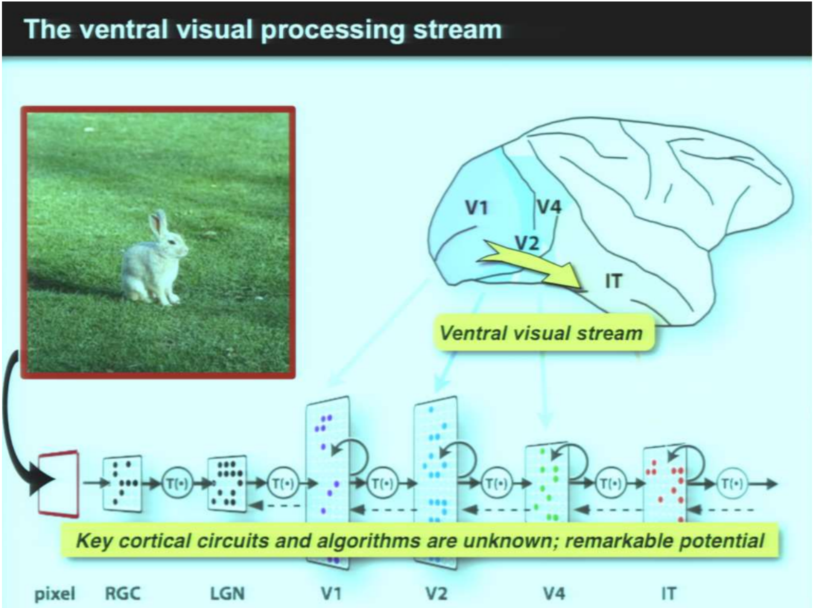
Jindřich Helcl and Jindřich Libovický. 2017. Neural Monkey: An open-source tool for sequence learning. *The Prague Bulletin of Mathematical Linguistics*, 107:5–17.

Tom Kocmi and Ondřej Bojar. 2017. Curriculum Learning and Minibatch Bucketing in Neural Machine Translation. In *Proceedings of Recent Advances in NLP (RANLP 2017)*.

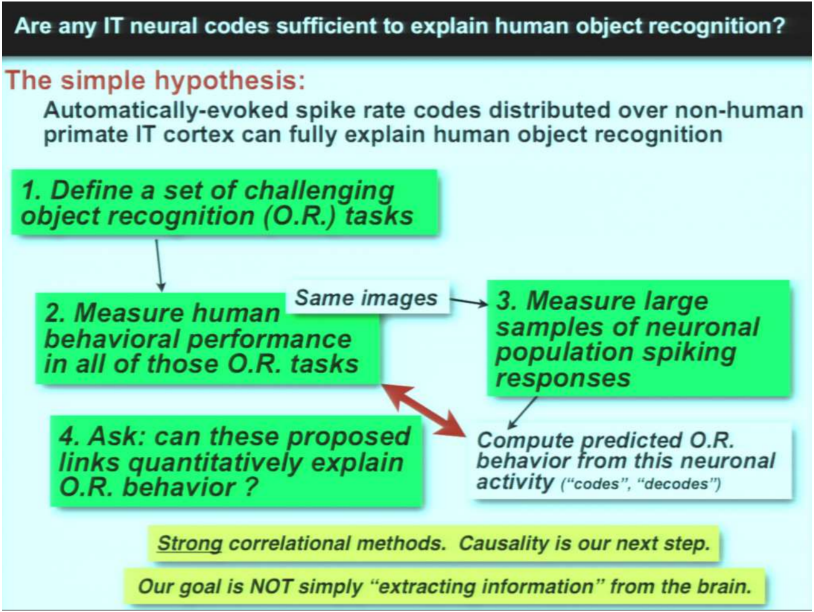
Jiwei Li, Will Monroe, and Dan Jurafsky. 2016. Understanding neural networks through representation erasure. *CoRR*, abs/1612.08220.

Hamed Nili, Cai Wingfield, Alexander Walther, Li Su, William Marslen-Wilson, and Nikolaus

# DiCarlo NIPS 2013 Tutorial on Vision



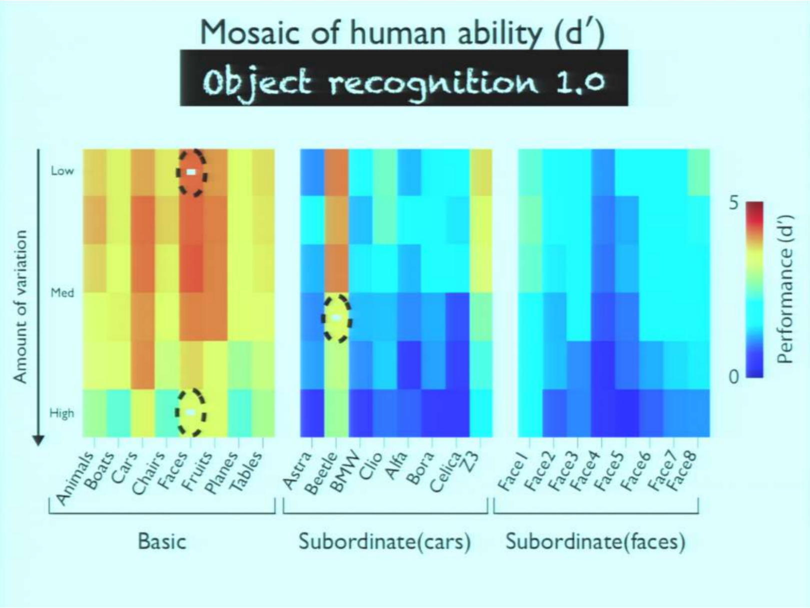
# DiCarlo NIPS 2013 Tutorial on Vision



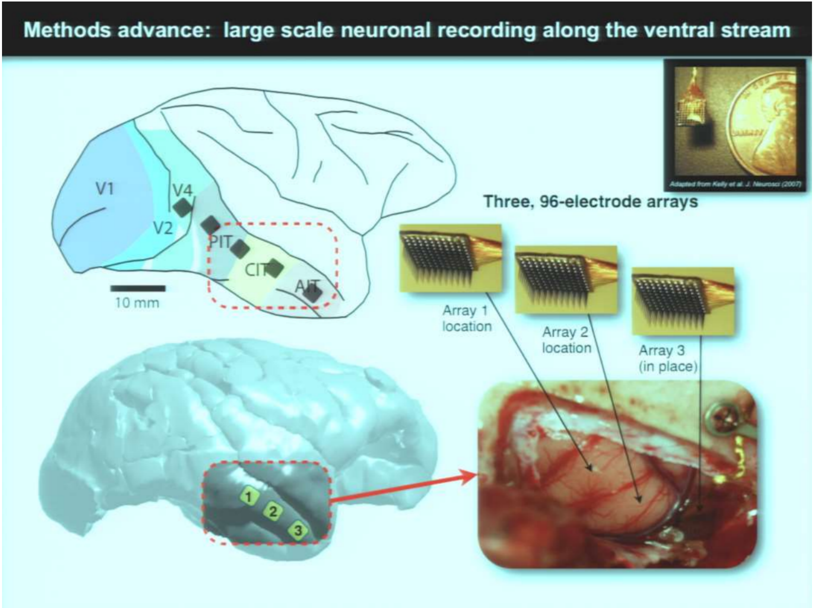
# DiCarlo NIPS 2013 Tutorial on Vision

- 64 objects, can generate as many images as we like
- full parametric control
- “natural” statistics
- uncorrelated, new background every image
- not fully “natural” by design -- challenging for computer vision, doable by humans

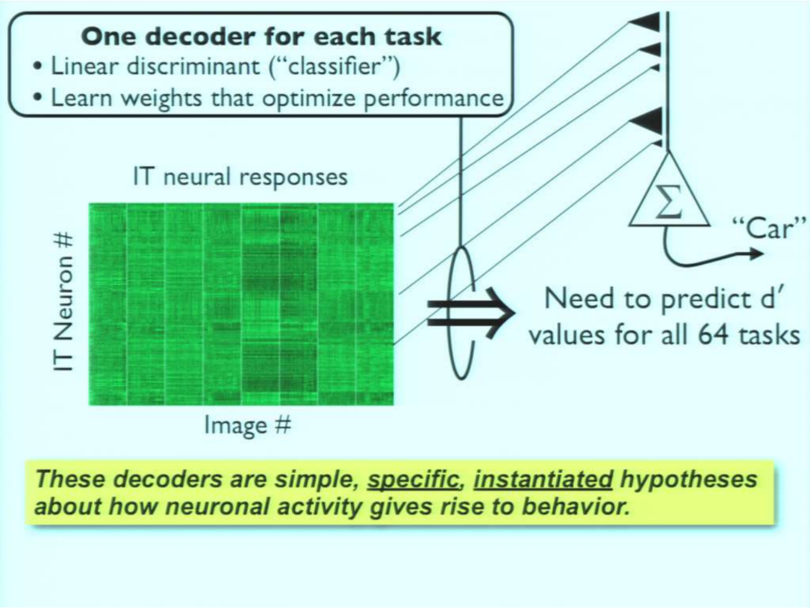
# DiCarlo NIPS 2013 Tutorial on Vision



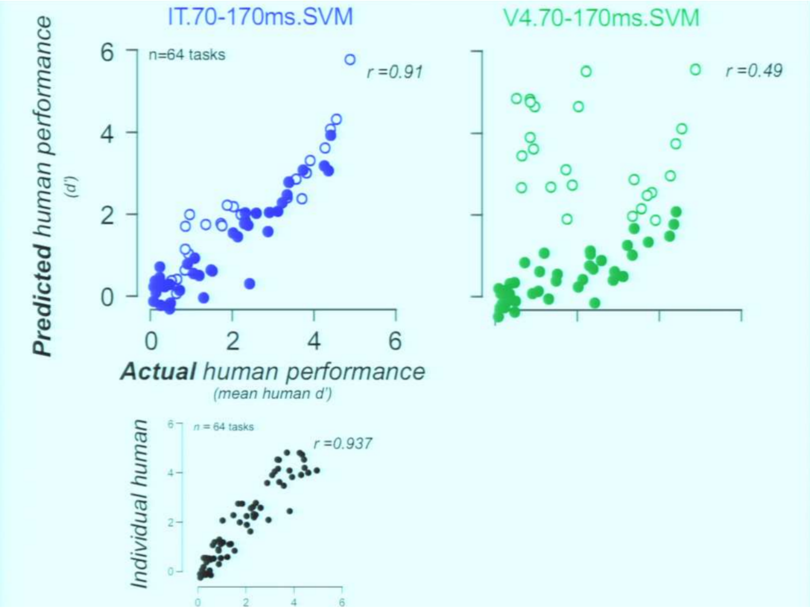
# DiCarlo NIPS 2013 Tutorial on Vision



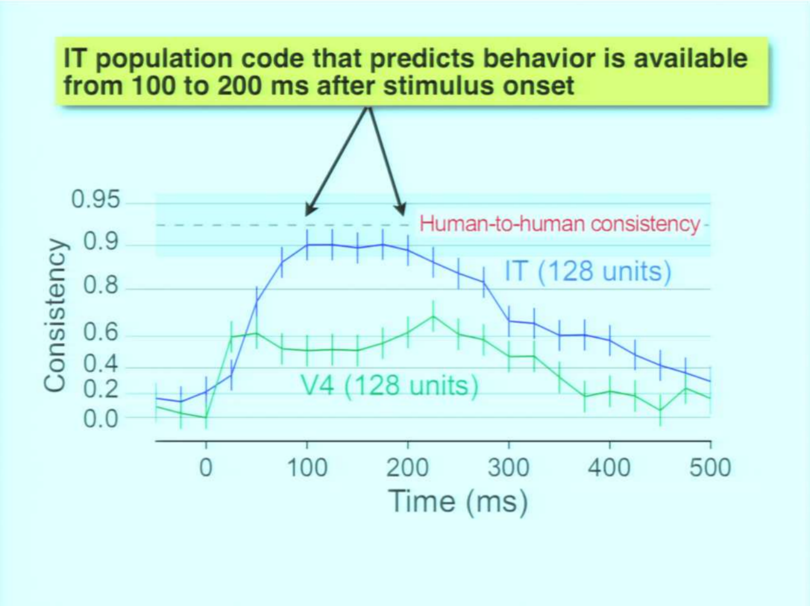
# DiCarlo NIPS 2013 Tutorial on Vision



# DiCarlo NIPS 2013 Tutorial on Vision



# DiCarlo NIPS 2013 Tutorial on Vision



# DiCarlo NIPS 2013 Tutorial on Vision