# Seminar Topics: Information Extraction

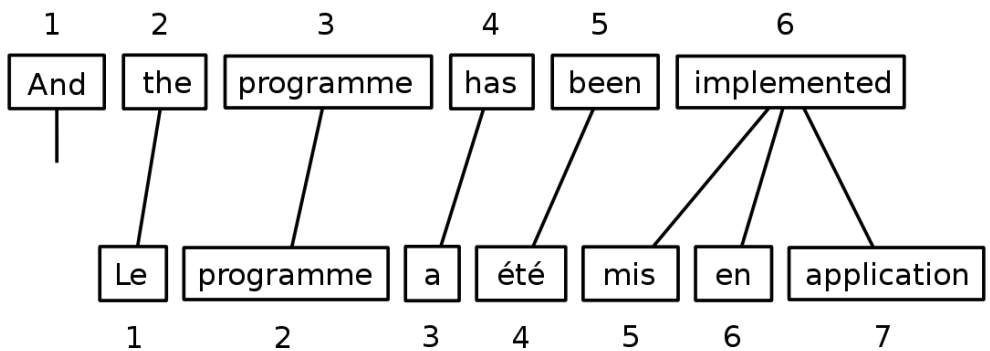
## English topics!

Viktor Hangya

hangyav@cis.lmu.de

# 1. Parallel sentence extraction

- There are 7000+ languages on the world
- Most IE approaches focus on a few of them only, e.g. English
- Parallel sentences are useful to transfer information from one language to another
  - Machine translation
  - Annotated data projection
  - Multilingual text representations
- Parallel sentences are expensive to create but we can mine them from the web automatically



# 1. Parallel sentence extraction

1. Introduction, feature based model:
   - Smith and Toutanova, 2010, **Extracting Parallel Sentences from Comparable Corpora using Document Level Alignment** *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*

2. Neural models:
   - Grégoire and Langlais, 2017, **A Deep Neural Network Approach To Parallel Sentence Extraction** *arXiv*
   - Artetxe and Schwenk, 2019, **Margin-based Parallel Corpus Mining with Multilingual Sentence Embeddings** *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*

3. Unsupervised model:
   - Keung et al., 2020, **Unsupervised Bitext Mining and Translation via Self-Trained Contextual Embeddings** *arXiv*

# 2. Target-level sentiment analysis

- Sentiment analysis: extract sentiment polarity of opinions:
  - Positive: I'm happy.
  - Negative: I'm sad.
  - Neutral: The sky is blue.
- Target-level: Opinions can be different given the target entity:
  - Android is better than iOS.
  - The food was great but the service was awful.

# 2. Target-level sentiment analysis

1. Introduction, feature based model:
   - Kiritchenko et al., 2014, **NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews** *Proceedings of the 8th International Workshop on Semantic Evaluation*

2. Neural models:
   - Wang et al., 2016, **Attention-based LSTM for Aspect-level Sentiment Classification** *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*
   - Liu and Zhang, 2017, **Attention Modeling for Targeted Sentiment** *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*

3. Target specific word meanings:
   - Li et al., 2018, **Transformation Networks for Target-Oriented Sentiment Classification** *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*

# 3. Relation Extraction and Classification in Scientific Documents

- Automatically identify relevant domain-specific semantic relations in a corpus of scientific publications
  - a new **method** is proposed for a **task**
  - a **phenomenon** is found in a certain **context**
  - **results** of different **experiments** are compared to each other
- Used for e.g.:
  - build knowledge-graphs
  - do a more detailed search
- The topic covers one shared-task:
  - SemEval-2018 Task 7

# 3. Relation Extraction and Classification in Scientific Documents

1. Overview:
   - Gábor et al., 2018, **SemEval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers** *Proceedings of the 12th International Workshop on Semantic Evaluation*
2. ClaiRE + UC3M-NII:
   - Hettinger et al., 2018, **ClaiRE at SemEval-2018 Task 7: Classification of Relations using Embeddings** *Proceedings of the 12th International Workshop on Semantic Evaluation*
   - Suárez-Paniagua et al., 2018, **UC3M-NII Team at SemEval-2018 Task 7: Semantic Relation Classification in Scientific Papers via Convolutional Neural Network** *Proceedings of the 12th International Workshop on Semantic Evaluation*
3. ETH-DS3Lab + Bf3R:
   - Rotsztejn et al., 2018, **ETH-DS3Lab at SemEval-2018 Task 7: Effectively Combining Recurrent and Convolutional Neural Networks for Relation Classification and Extraction** *Proceedings of the 12th International Workshop on Semantic Evaluation*
   - Neves et al., 2018, **Bf3R at SemEval-2018 Task 7: Evaluating Two Relation Extraction Tools for Finding Semantic Relations in Biomedical Abstracts** *Proceedings of the 12th International Workshop on Semantic Evaluation*

# 4. Novel and Emerging Entity Recognition

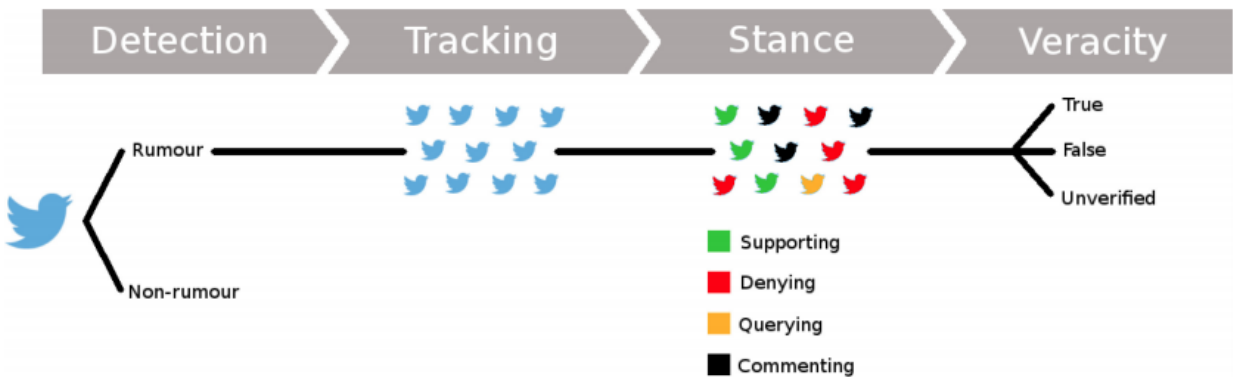
- NER systems perform well on data similar to what they were trained on and can detect frequent well behaving NEs
- New NEs emerge day-by-day which are often hard to detect for humans as well:
  - Tweet: *so.. kktny in 30 mins?!*
  - kktny: Kourtney and Kim Take New York (TV series)
- The topic covers one shared-task:
  - WNUT 2017

# 4. Novel and Emerging Entity Recognition

1. Overview:
   - Derczynski et al., 2017, **Results of the WNUT2017 Shared Task on Novel and Emerging Entity Recognition** *Proceedings of the 3rd Workshop on Noisy User-generated Text*
2. Arcada + FLYTXT:
   - Jansson and Liu, 2017, **Distributed Representation, LDA Topic Modelling and Deep Learning for Emerging Named Entity Recognition from Social Media** *Proceedings of the 3rd Workshop on Noisy User-generated Text*
   - Sikdar and Gambäck, 2017, **A Feature-based Ensemble Approach to Recognition of Emerging and Rare Named Entities** *Proceedings of the 3rd Workshop on Noisy User-generated Text*
3. Drexel-CCI + SJTU-Adapt:
   - Williams and Santia, 2017, **Context-Sensitive Recognition for Emerging and Rare Entities** *Proceedings of the 3rd Workshop on Noisy User-generated Text*
   - Lin et al., 2017, **Multi-channel BiLSTM-CRF Model for Emerging Named Entity Recognition in Social Media** *Proceedings of the 3rd Workshop on Noisy User-generated Text*

# 5. Rumor verification

- A large amount of information is share on social media real time
- But not all of them are true or verified
- Automatic rumor detection can help moderate social media platforms



# 5. Rumor verification

1. Introduction:
   - Zubiaga et al., 2018, **Detection and Resolution of Rumours in Social Media: A Survey** *ACM Computing Surveys*

2. Approaches for the individual subtasks:
   - Zubiaga et al., 2017, **Exploiting Context for Rumour Detection in Social Media** *International Conference on Social Informatics*
   - Enayet and El-Beltagy, 2017, **NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter** *Proceedings of the 11th International Workshop on Semantic Evaluations*

3. Joint approach:
   - Kochkina et al., 2018, **All-in-one: Multi-task Learning for Rumour Verification** *Proceedings of the 27th International Conference on Computational Linguistics*