



What is Natural Language Processing, and what can it do?



Dr. Basil Ell

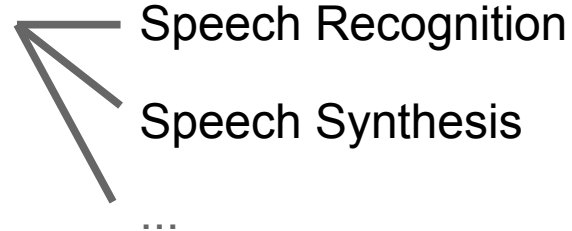
basile@ifi.uio.no

**Post-doctoral researcher at SIRIUS (University of Oslo)
and Bielefeld University, Germany**



Written Natural Language Processing

Spoken Natural Language Processing





Applications

What is it that you want to achieve using NLP?



Applications

(Aspect-Based) Sentiment Analysis / Opinion Mining

The serrated portion of the blade is sharp^{pos} but the straight edge is marginal at best^{neg}.



Applications

Question Answering (over a structured Knowledge Base)

“Give me all cosmonauts”, Database = DBpedia

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
PREFIX dbo: <http://dbpedia.org/ontology/>
```

```
PREFIX res: <http://dbpedia.org/resource/>
```

```
SELECT DISTINCT ?uri WHERE {  
    ?uri rdf:type dbo:Astronaut .  
    { ?uri dbo:nationality res:Russia . }  
    UNION  
    { ?uri dbo:nationality res:Soviet_Union . }  
}
```



Applications

Reading Comprehension / Question Answering

Question:

The pathology in the pancreas of patients with type 1 diabetes is:

Candidate Answers:

- a. Islet cell hyperplasia
- b. Islet cell necrosis
- c. Interstitial calcification
- d. Interstitial fibrosis
- e. Islet cell vacuolar degeneration



Applications

Natural Language Generation (from a structured Knowledge Base)

```
dbr:Curtain_(Novel) dbo:author dbr:Agatha_Christie .  
dbr:Curtain_(Novel) rdf:type dbo:Book .  
dbr:Curtain_(Novel) rdfs:label "Curtain (novel)"@en .  
dbr:Curtain_(Novel) dbp:releaseDate "September 1975"@en .  
dbr:Agatha_Christie rdf:type dbo:Writer .  
dbr:Agatha_Christie rdfs:label "Agatha Christie"@en .  
dbo:Book rdfs:label "book"@en .
```

→ “Curtain is a book by Agatha Christie published in September 1975.”



Applications

Argument Search

What are the top PRO / CON arguments regarding nuclear energy?

Found 124 arguments (77 pro; 47 con) in 13 documents (classified 335 sentences in 4.26 s)

PRO

english.donga.com / Feb 14, 2016

Nuclear energy is economically competitive in that LNG costs 187 won (0.17 US dollar) to produce one kilowatt of electricity and nuclear energy costs 39.2 won (0.04 dollar). 99.7%

PRO

ukessays.com / Feb 6, 2016

Professional essay writersOne of the main benefits of nuclear energy is that nuclear energy is one of the clean ways to produce energy. 99.7%

PRO

ukessays.com / Feb 6, 2016

CON

ukessays.com / Feb 6, 2016

The developing of nuclear power plant also not really feasible as around 15 to 20 years is needed to develop only single power plant. 99.37%

CON

ukessays.com / Feb 6, 2016

They also make the locations of nuclear power plant as a target area to disrupt the power supply as it can cause a large-scale devastation. 99.37%

CON

ukessays.com / Feb 6, 2016



Tasks

Typical tasks involved in realizing the applications



Tasks

Sentence Boundary Detection

My name is Jonas E. Smith. Please turn to p. 55.

Tokenization

Mr. O'Neill thinks that the boys ' stories about Chile
's capital are n't amusing.

Part-of-speech Tagging

Mr./NNP O'Neill/NNP thinks/VBZ that/IN the/DT ...

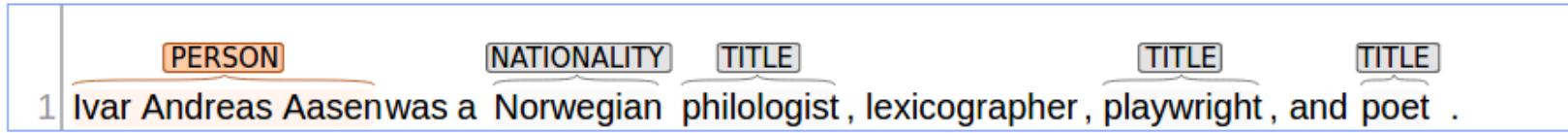
Lemmatization (lexeme to lemma)

Mr. O'Neill **think** that the **boy** ' **story** about ...



Tasks

Named Entity Recognition



Coreference Resolution

"I voted for Nader because he was most aligned with my values," she said.

A B B A A

Relation Extraction

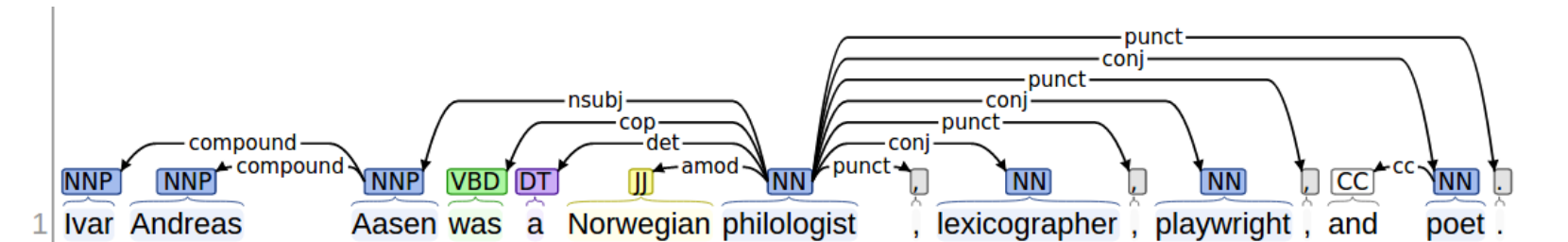
Barack Obama was born in Honolulu, Hawaii.

→ (Barack Obama, bornInCity, Honolulu)



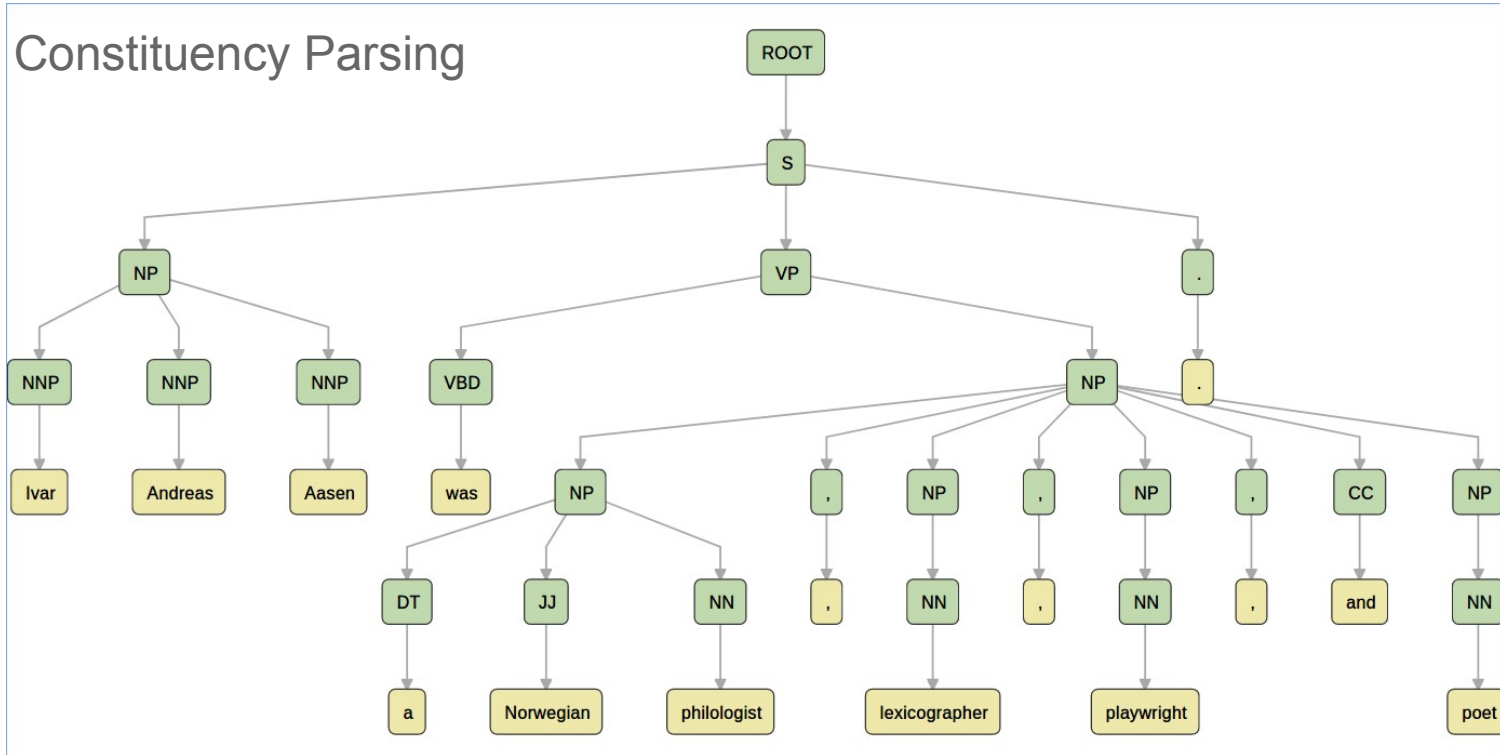
Tasks

Dependency Parsing





Tasks





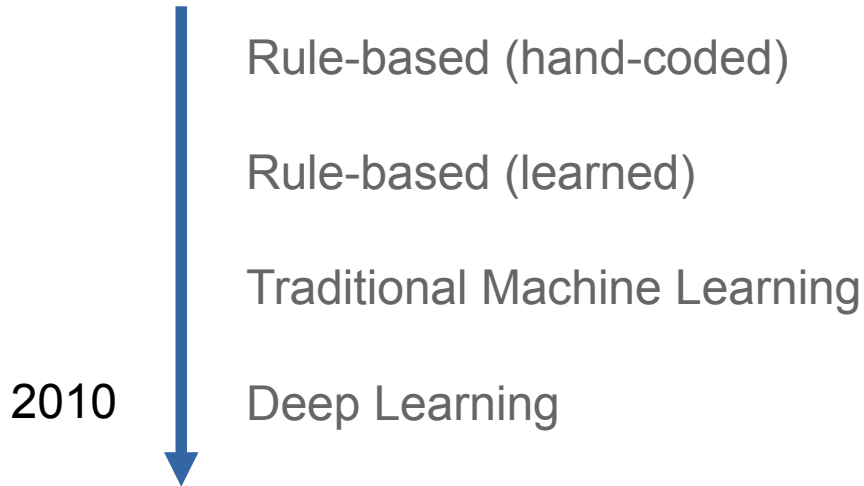
Written Natural Language Processing – Further Applications and Tasks

Argument Mining, Dialogue, Discourse Parsing, Entity Linking, Event Detection, Fake News Detection, Grammatical Error Correction, Grammar Induction, Information Extraction (for Knowledge Base Population), Machine Translation, Morphological Analysis, Natural Language Understanding, Ontology Learning, Paraphrasing, Question Generation, Reading Comprehension, Semantic Parsing, Semantic Role Labeling, Simplification, Stance Detection, Summarization, Terminology Extraction, Textual Similarity, Topic Modeling, Word Sense Disambiguation, ...

See also: <https://paperswithcode.com/area/natural-language-processing>



Written Natural Language Processing – History of Approaches





GPT-3 – Generative Pre-trained Transformer 3 (OpenAI)

Unsupervised

Language model: predicts the likelihood of a sentence existing in the world
175 billion parameters – the largest language model ever created

Q: What is human life expectancy in the United States?

A: Human life expectancy in the United States is 78 years.

Q: What are two reasons that a dog might be in a bad mood?

A: Two reasons that a dog might be in a bad mood are if it is hungry or if it is hot.

Q: Why don't animals have three legs?

A: Animals don't have three legs because they would fall over.

Q: How do you sporgle a morgle?

A: You sporgle a morgle by using a sporgle.

Q: How many rainbows does it take to jump from Hawaii to seventeen?

A: It takes two rainbows to jump from Hawaii to seventeen.

<https://lacker.io/ai/2020/07/06/giving-gpt-3-a-turing-test.html>

Brown, Tom B., et al. "Language models are few-shot learners." arXiv preprint arXiv:2005.14165 (2020).



Tweet



Chet Haase

@chethaase



A Machine Learning algorithm
walks into a bar.
The bartender asks, "What'll you
have?"
The algorithm says, "What's
everyone else having?"

6:54 PM · 01 Nov 17

3,749 Retweets **7,605** Likes



Open Questions

Do humans require as much training data to generate language?

Do language models have deeper understanding?

How to ensure that a language model does not fail in a critical situation?

How to combine knowledge representation and machine learning to enable domain-adapted ML and to require less training data?