Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

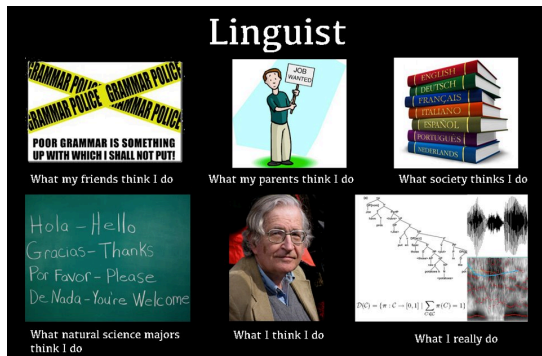# Grammar engineering in computational linguistics

Olga Zamaraeva

Universidade da Coruña
CITIC

April 26 2022

# Introduction

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- ▶ Olga Zamaraeva, PhD in Linguistics, University of Washington 2021
- ▶ I am a postdoc working in the field of computational linguistics
- ▶ I will talk about the field* in general
- ▶ ...and just a little about the specifics of my subfield



Also, don't check this out: https://xkcd.com/114/

# Computational linguistics

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

CompLing == NLP?    CompLing ∩ NLP = ∅?

---

▶ Linguistics: Study of human language 🌐

# Computational linguistics

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

CompLing == NLP?    CompLing $\cap$ NLP $= \emptyset$?

---

- Linguistics: Study of human language 🌐
- Computational linguistics:

# Computational linguistics

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

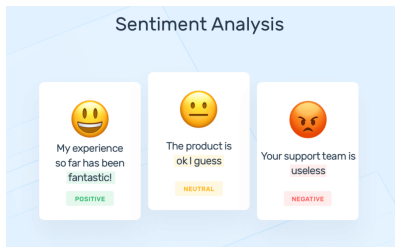$$\text{CompLing} == \text{NLP}? \quad \text{CompLing} \cap \text{NLP} = \emptyset?$$

- Linguistics: Study of human language 🌐
- Computational linguistics:
  - Applying computational methods to **linguistics**
    - a subfield of linguistics

# Computational linguistics

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

$$\text{CompLing} == \text{NLP}? \quad \text{CompLing} \cap \text{NLP} = \emptyset?$$

- Linguistics: Study of human language 🌐
- Computational linguistics:
  - Applying computational methods to **linguistics**
    - a subfield of linguistics
  - Natural language processing (NLP)
    - a field **of its own** (related to artificial intelligence)

# Computational linguistics

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

CompLing == NLP?    CompLing $\cap$ NLP $= \emptyset$?

---

- Linguistics: Study of human language 🌐
- Computational linguistics:
  - Applying computational methods to **linguistics**
    - a subfield of linguistics
  - Natural language processing (NLP)
    - a field **of its own** (related to artificial intelligence)
  - using "CompLing" for both currently feels like a historic artifact
    - ...though not an accident, since both fields have things in common

---

CompLing != NLP    CompLing $\cap$ NLP != $\emptyset$

# Natural language processing

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- Task-oriented, e.g.:
  - Given text/speech recording, make **automatic predictions** about **new similar data**:
  - Is a review good or bad?
  - Is this text a good description of this picture?
  - Is this sentence a good response to this question?
- A RQ: What can be learned about the world through the lens of language data? (Yatskar, p.c.)



https://www.analyticsvidhya.com/blog/2022/01/sentiment-analysis-with-lstm/

# Linguistics (computational or not)

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- What are some natural laws in how **children** acquire language?
- How is acquiring language by **adults** different?
- In what ways are languages of the world **fundamentally** similar?
- What are **people's attitudes** to various language behaviors?



https://languagelog.ldc.upenn.edu/nll/?p=29480

# Linguistics (computational) & NLP

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- The two fields are not the same but there is common ground
  - Recently, a lot of interest in NLP to linguistics, in particular language acquisition
  - Can neural networks help us understand language acquisition better?

# Linguistics (computational) & NLP

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- The two fields are not the same but there is common ground
  - Recently, a lot of interest in NLP to linguistics, in particular language acquisition
  - Can neural networks help us understand language acquisition better?
- Me: more of a linguist but collaborate with the NLP community

# Linguistics (computational) & NLP

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- The two fields are not the same but there is common ground
  - Recently, a lot of interest in NLP to linguistics, in particular language acquisition
  - Can neural networks help us understand language acquisition better?
- Me: more of a linguist but collaborate with the NLP community
  - My main interests are in the area of **formal** computational grammars

# Linguistics (computational) & NLP

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- The two fields are not the same but there is common ground
  - Recently, a lot of interest in NLP to linguistics, in particular language acquisition
  - Can neural networks help us understand language acquisition better?
- Me: more of a linguist but collaborate with the NLP community
  - My main interests are in the area of **formal** computational grammars
- Methodologies:

# Linguistics (computational) & NLP

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- The two fields are not the same but there is common ground
  - Recently, a lot of interest in NLP to linguistics, in particular language acquisition
  - Can neural networks help us understand language acquisition better?
- Me: more of a linguist but collaborate with the NLP community
  - My main interests are in the area of **formal** computational grammars
- Methodologies:
  - Linguistics: Formal and statistical, also social science methods

# Linguistics (computational) & NLP

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- The two fields are not the same but there is common ground
  - Recently, a lot of interest in NLP to linguistics, in particular language acquisition
  - Can neural networks help us understand language acquisition better?
- Me: more of a linguist but collaborate with the NLP community
  - My main interests are in the area of **formal** computational grammars
- Methodologies:
  - Linguistics: Formal and statistical, also social science methods
  - NLP: Largely statistical but this used to vary more

# Methodology: Formal or Statistical?

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

Rafael's School of Athens (*circa* 1510)

▶ Formal ("rule-based")
▶ Statistical (automatic learning)

# Methodology: Formal or Statistical?

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

Rafael's School of Athens (*circa* 1510)

▶ Formal ("rule-based")
▶ Statistical (automatic learning)
▶ Both approaches use patterns in data! And both can miss something.

# Methodology: Formal or Statistical?

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

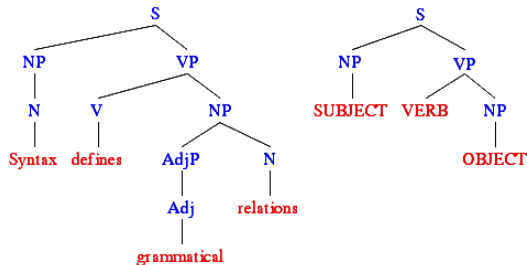Rafael's School of Athens (*circa* 1510)

- ▶ Formal ("rule-based")
- ▶ Statistical (automatic learning)
- ▶ Both approaches use patterns in data! And both can miss something.
- ▶ **Formal** is to **statistical** how **rationalist** is to **empiricist**

# Methodology: Formal or Statistical?

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

Rafael's School of Athens (*circa* 1510)

- ▶ Formal ("rule-based")
- ▶ Statistical (automatic learning)
- ▶ Both approaches use patterns in data! And both can miss something.
- ▶ **Formal** is to **statistical** how **rationalist** is to **empiricist**
- ▶ **Rationalist** is to **empiricist** how **idealist** is to **materialist**

# Methodology: Formal or Statistical?

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions
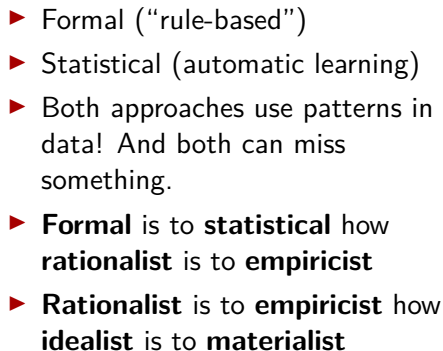
Rafael's School of Athens (*circa* 1510)

- Formal ("rule-based")
- Statistical (automatic learning)
- Both approaches use patterns in data! And both can miss something.
- **Formal** is to **statistical** how **rationalist** is to **empiricist**
- **Rationalist** is to **empiricist** how **idealist** is to **materialist**
- Ancient debate with no right answer

# Syntactic theory: A formal study of sentence structure

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions
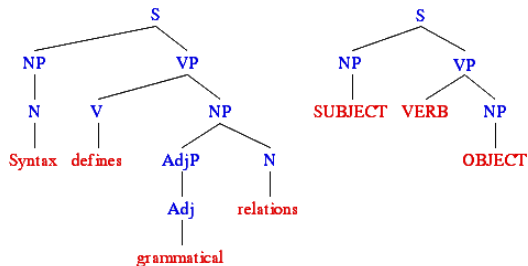
- ▶ Goal: Formally describe a system that predicts grammaticality

# Syntactic theory: A formal study of sentence structure

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- ▶ Goal: Formally describe a system that predicts grammaticality
    - ▶ Accept grammatical strings and reject ungrammatical ones
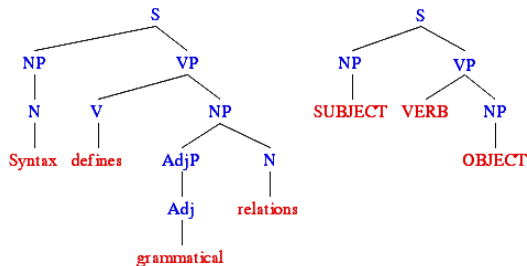    - ▶ e.g. *They read books.* is good but *Books they read.* is not, etc.

# Syntactic theory: A formal study of sentence structure
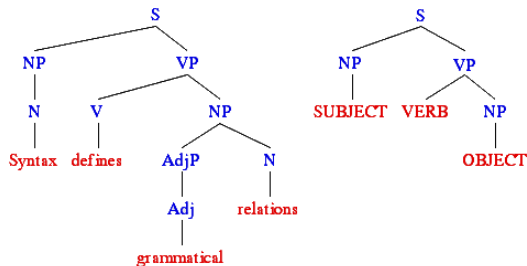
Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- ▶ Goal: Formally describe a system that predicts grammaticality
  - ▶ Accept grammatical strings and reject ungrammatical ones
  - ▶ e.g. *They read books.* is good but *Books they read.* is not, etc.
  - ▶ Attempt to **reason** about the contrasts

# Syntactic theory: A formal study of sentence structure
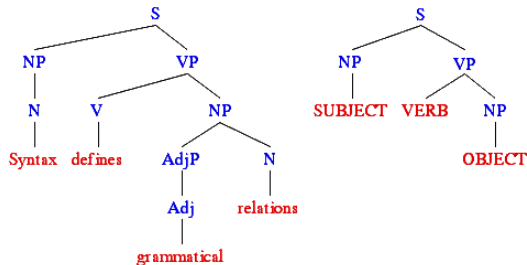
Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- ▶ Goal: Formally describe a system that predicts grammaticality
    - ▶ Accept grammatical strings and reject ungrammatical ones
    - ▶ e.g. *They read books.* is good but *Books they read.* is not, etc.
    - ▶ Attempt to **reason** about the contrasts
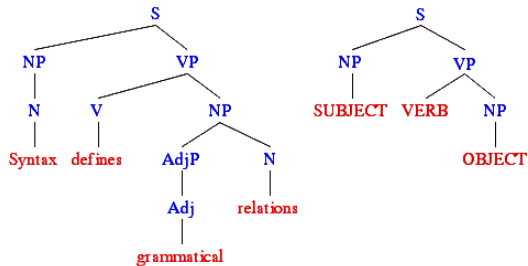    - ▶ Desiderata: plausibility, "simplicity", parsimony etc

# Syntactic theory: A formal study of sentence structure

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- ▶ Goal: Formally describe a system that predicts grammaticality
  - ▶ Accept grammatical strings and reject ungrammatical ones
  - ▶ e.g. *They read books.* is good but *Books they read.* is not, etc.
  - ▶ Attempt to **reason** about the contrasts
  - ▶ Desiderata: plausibility, "simplicity", parsimony etc
- ▶ Could I learn such a theory automatically from raw data?

https://koine-greek.com/2017/05/08/a-brief-history-of-syntactic-theory-early-chomsky/

# My field: Grammar Engineering

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- Come up with a syntactic theory
- Implement the theory on the computer
  - Now can rigorously test the theory against data

https://koine-greek.com/2017/05/08/a-brief-history-of-syntactic-theory-early-chomsky/

9

# A theory: Constraint-based syntax

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- ▶ Sentence structure is a feature structure (graph)
- ▶ A sentence is possible if all constraints unify
- ▶ Below: the Lexical Functional Grammar (LFG) formalism
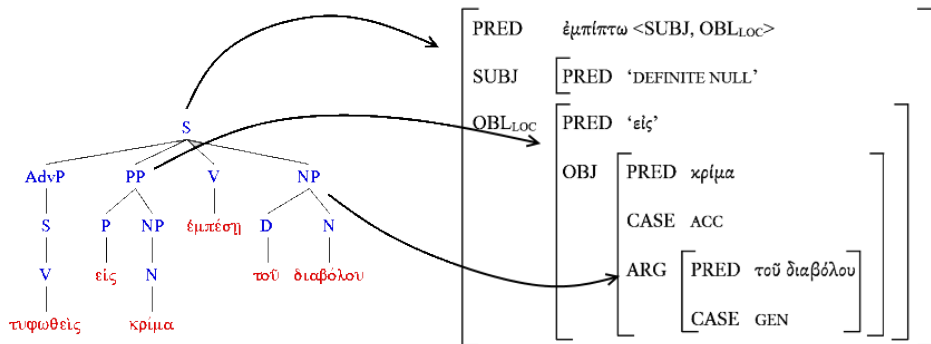


https://koine-greek.com/2017/05/22/a-brief-history-of-syntactic-theory-parallel-contraint-based-syntax/

# Summary: Grammar engineering for linguistic hypothesis testing

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- Syntax is complex; verifying structures by hand is intractable

# Summary: Grammar engineering for linguistic hypothesis testing

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- Syntax is complex; verifying structures by hand is intractable
- For rigorous testing, formal theories should be implemented on the computer

# Summary: Grammar engineering for linguistic hypothesis testing

Assembling Syntax

Olga Zamaraeva

Intro

Computational
linguistics

Grammar
Engineering

Automatic
grammar
suggestions

- Syntax is complex; verifying structures by hand is intractable
- For rigorous testing, formal theories should be implemented on the computer
- Can then **apply** implemented grammars:
  - Creating datasets for NLP training and evaluation
  - Tasks requiring high precision

# Summary: Grammar engineering for linguistic hypothesis testing
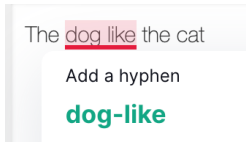
Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- Syntax is complex; verifying structures by hand is intractable
- For rigorous testing, formal theories should be implemented on the computer
- Can then **apply** implemented grammars:
  - Creating datasets for NLP training and evaluation
  - Tasks requiring high precision
- Demo: https://delph-in.github.io/delphin-viz/demo/

# Application: Grammar engineering for grammar coaching

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

The dog like the cat

Add a hyphen

**dog-like**

▶ Statistical systems are imprecise ☝

# Application: Grammar engineering for grammar coaching

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

The dog like the cat

Add a hyphen

**dog-like**

- ▶ Statistical systems are imprecise ☝
- ▶ Grammars: Incorporate 'error productions'

# Application: Grammar engineering for grammar coaching

Assembling Syntax
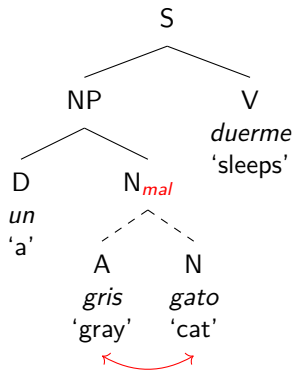
Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering
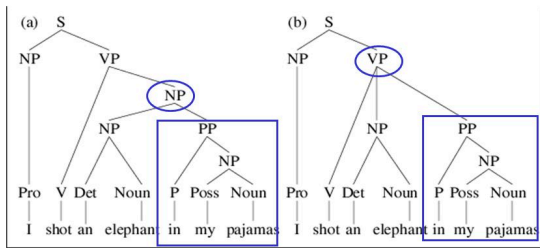
Automatic grammar suggestions

The dog like the cat

Add a hyphen

**dog-like**

- ▶ Statistical systems are imprecise ☝
- ▶ Grammars: Incorporate 'error productions'
- ▶ Map error productions to **useful** suggestions and feedback
    - ▶ The feedback needs to be rule-based and rigorous syntactic theory is better than ad-hoc

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

# Grammar engineering for grammar coaching

- My project for the next couple years here:
  - Similar systems exist for English and Chinese

# Natural language parsing

Assembling Syntax

Olga Zamaraeva

Intro
Computational linguistics
Grammar Engineering
Automatic grammar suggestions

- Parsing = mapping a string to its structure
- Constraint-based parsing is slow
- This year, I work with Prof. Gómez-Rodríguez on a technique to improve the speed
- This will later allow me to build a more realistic grammar coaching system



"*I shot an elephant in my pajamas*"

https://slideplayer.com/slide/4875662/

# That's it!

Assembling Syntax

Olga Zamaraeva

Intro

Computational linguistics

Grammar Engineering

Automatic grammar suggestions

- Thanks for listening!
- Come chat with me in the office at the ground floor!
- olga.zamaraeva@gmail.com
- Questions? If not, let's revisit the memes :)