

Beyond Word Embeddings: Semantic Representation Explained

Ari Bornstein
Senior CDA AI + Start Up



About Me

- Website: <http://thepythiccoder.com>
- Twitter: @pythiccoder
- Github: [aribornstein](#)





START-UP NATION



Open Source





Experiences Are Only As Good
As The Brains Behind them



Neural NLP Milestones



Sentiment Analysis

Extract information from your text

Use the demo below to experiment with the Text Analytics API. Pick one of our examples or provide your own. Identify the language, sentiment, key phrases, and entities (Preview) of your text by clicking "Analyze".

See it in action

I had a wonderful trip to Seattle and enjoyed seeing the Space Needle!

Analyze

Example - English - Positive

Example - English - Negative

Example - Spanish - Positive

Example - Spanish - Negative

Analyzed text

JSON

LANGUAGES: English (confidence: 100 %)

KEY PHRASES: Seattle, wonderful trip, Space Needle

SENTIMENT: 98 %

LINKED ENTITIES (PREVIEW): I had a wonderful trip to [Seattle](#) and enjoyed seeing [the Space Needle](#)!

Machine Reading Comprehension BiDAF

Machine Comprehension

Machine Comprehension (MC) answers natural language questions by selecting an answer span within an evidence text. The AllenNLP toolkit provides the following MC visualization, which can be used for any MC model in AllenNLP. This page demonstrates a reimplementation of [BiDAF \(Seo et al, 2017\)](#), or Bi-Directional Attention Flow, a widely used MC baseline that achieved state-of-the-art accuracies on [the SQuAD dataset](#) (Wikipedia sentences) in early 2017.

Enter text or

Passage

In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague, where he was to study. He arrived too late to enroll at Charles-Ferdinand University; he had never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures in philosophy at the university as an auditor but he did not receive grades for the courses.

Question

What city did Teslas move to in 1880?

RUN >

Answer

Prague

Passage Context

In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for **Prague**, where he was to study. He arrived too late to enroll at Charles-Ferdinand University; he had never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures in philosophy at the university as an auditor but he did not receive grades for the courses.

Model internals (beta)

Natural Language Inference

Textual Entailment

Textual Entailment (TE) takes a pair of sentences and predicts whether the facts in the first necessarily imply the facts in the second one. The AllenNLP toolkit provides the following TE visualization, which can be run for any TE model you develop. This page demonstrates a reimplementation of [the decomposable attention model \(Parikh et al, 2017\)](#), which was state of the art for [the SNLI benchmark](#) (short sentences about visual scenes) in 2016. Rather than pre-trained Glove vectors, this model uses [ELMo embeddings](#), which are completely character based and improve performance by 2%

Enter text or

Premise

An interplanetary spacecraft is in orbit around a gas giant's icy moon.

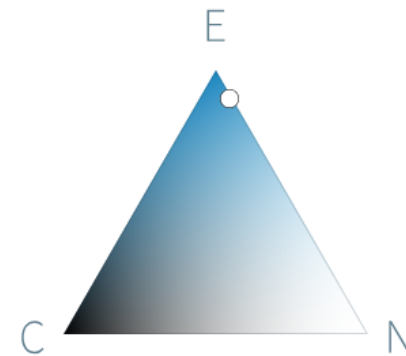
Hypothesis

The spacecraft has the ability to travel between planets.

RUN >

Summary

It is **likely** that the premise **entails** the hypothesis.



Judgement

Probability

Entailment

89.3%

Contradiction

0.8%

Neutral

9.8%

Model internals (beta)



Neural Machine Translation NMT

Source	Reference
Une fusillade a eu lieu à l'aéroport international de Los Angeles.	There was a shooting in Los Angeles International Airport.
Cette controverse croissante autour de l'agence a provoqué beaucoup de spéculations selon lesquelles l'incident de ce soir était le résultat d'une cyber-opération ciblée.	Such growing controversy surrounding the agency prompted early speculation that tonight's incident was the result of a targeted cyber operation.

What are Word Vectors?



“Words or phrases from the vocabulary are mapped to vectors of real numbers.”

Embedding Examples (10,0000 Feet)

Traditional Embeddings

- Bag of Words
- TF-IDF
- Distributional Embeddings
PMI

Neural

- Word2Vec CBOW & SkipGram
- Glove Vectors
- Fast Text

Modeling With Embeddings

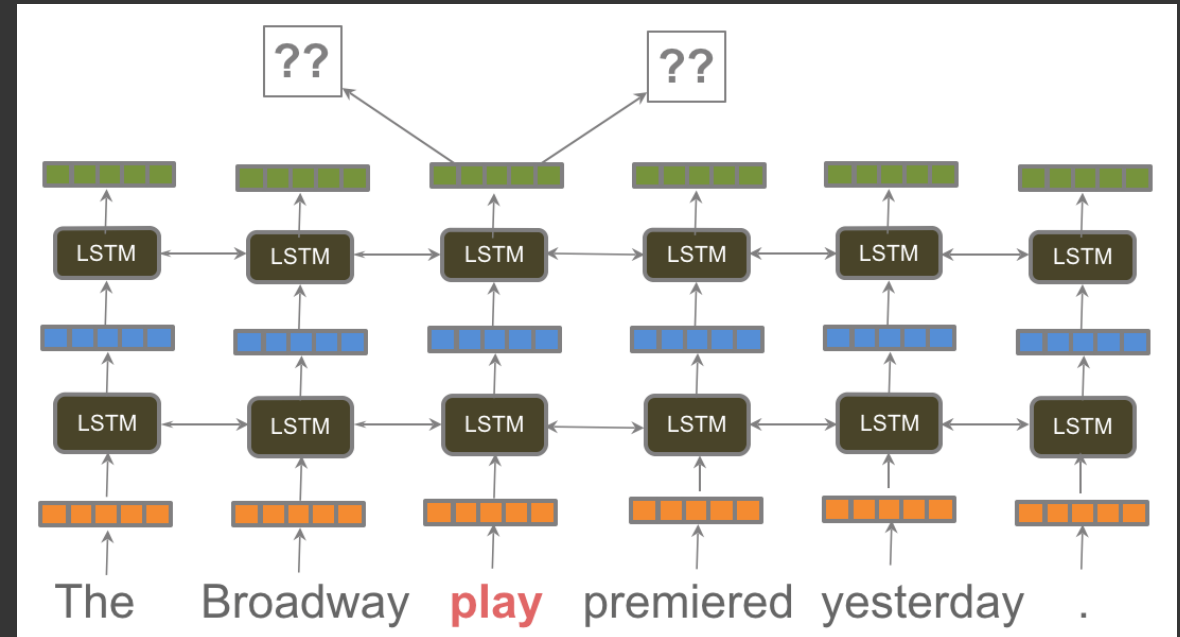
Traditional

- HMMs
- Logistic Regression
- Boosted Trees/ Forests
- Bayesian Methods

Neural

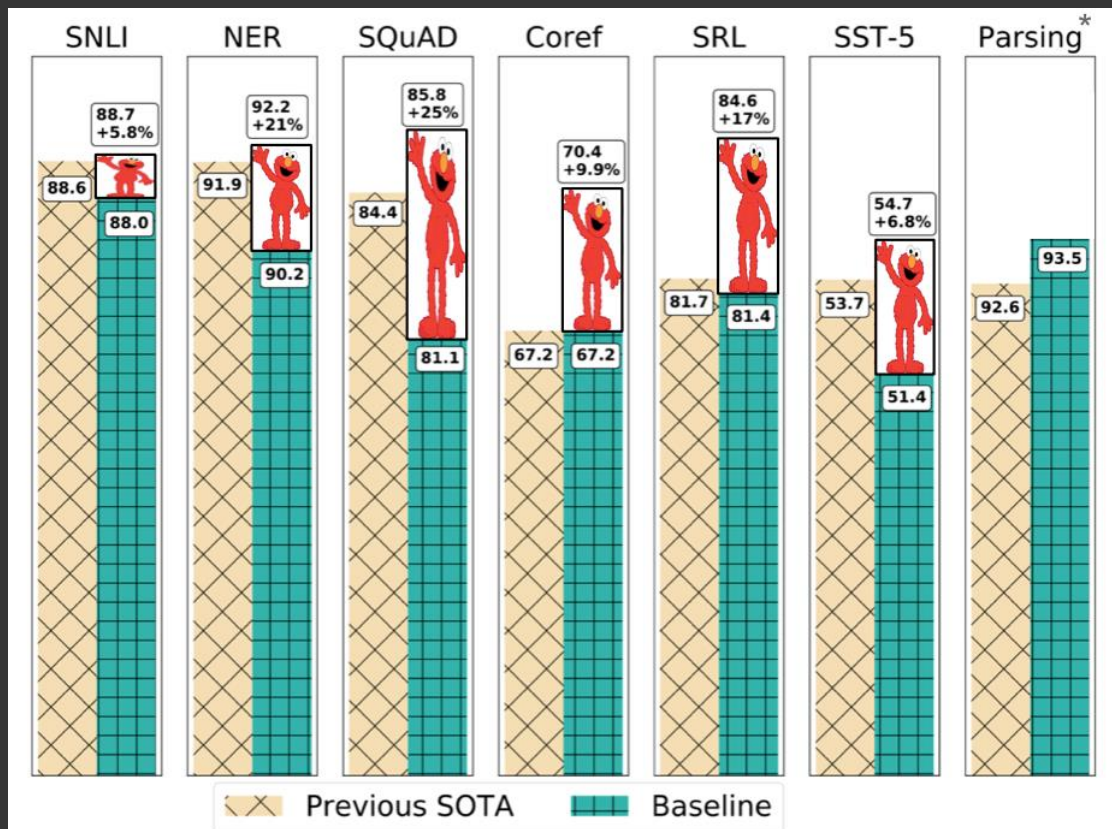
- Feed Forward
- 1D CNNs
- RNNs (GRU/LSTMs)
- Attention Mechanisms

Example ELMO

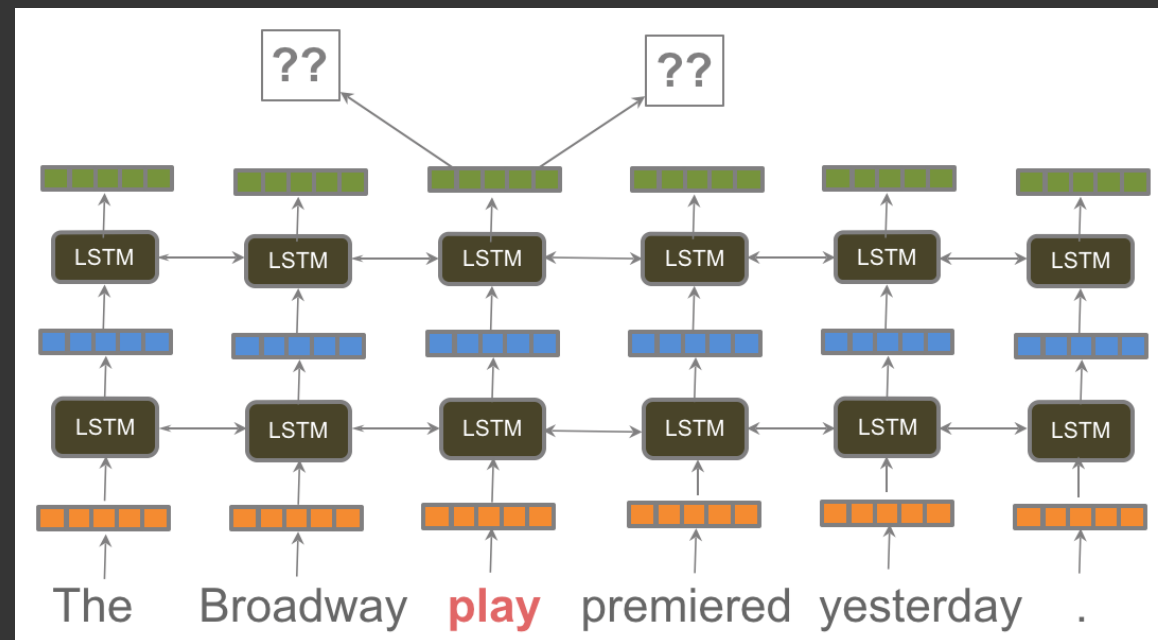


<http://runder.io/nlp-imagenet/>

Example ELMO



*Kitaev and Klein, ACL 2018 (see also Joshi et al., ACL 2018)



<http://runder.io/nlp-imagenet/>

Pitfalls of Neural NLP



Superficial Correlations in Models

```
text_to_sentiment("My name is Emily")
2.2286179364745311

text_to_sentiment("My name is Heather")
1.3976291151079159

text_to_sentiment("My name is Yvette")
0.98463802132985556

text_to_sentiment("My name is Shaniqua")
-0.47048131775890656
```

Rob Speer
[how-to-make-a-racist-ai-without-really-trying.ipynb](#)

question	answer
How many...?	2
Is/Are... ?	Yes
What sport...?	Tennis
What animal...?	Dog

```
text_to_sentiment("Let's go get Italian food")
2.0429166109408983

text_to_sentiment("Let's go get Chinese food")
1.4094033658140972

text_to_sentiment("Let's go get Mexican food")
0.38801985560121732
```

Rob Speer
[how-to-make-a-racist-ai-without-really-trying.ipynb](#)

From Agrawal et al.

Models Fail Under Adversarial Evaluation

Machine Comprehension

Machine Comprehension (MC) answers natural language questions by selecting an answer span within an evidence text. The AllenNLP toolkit provides the following MC visualization, which can be used for any MC model in AllenNLP. This page demonstrates a reimplementation of [BiDAF \(Seo et al, 2017\)](#), or Bi-Directional Attention Flow, a widely used MC baseline that achieved state-of-the-art accuracies on [the SQuAD dataset](#) (Wikipedia sentences) in early 2017.

Enter text or

Passage

In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague, where he was to study. He arrived too late to enroll at Charles-Ferdinand University; he had never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures in philosophy at the university as an auditor but he did not receive grades for the courses. Tadakatsu moved to the city of Chicago in 1881.

Question

What city did Teslas move to in 1880?

RUN >

Answer

Chicago

Passage Context

In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague, where he was to study. He arrived too late to enroll at Charles-Ferdinand University; he had never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures in philosophy at the university as an auditor but he did not receive grades for the courses. Tadakatsu moved to the city of **Chicago** in 1881.

Model internals (beta)

Not Robust to Changes in Semantic Structure

There is no pleasure in watching a child suffer.

Prediction: negative

83.1% accuracy



Syntactic paraphrase:

In watching the child suffer, there is no pleasure.

Prediction: positive



41.8% dev instances broken
(correct prediction becomes incorrect)

lyyer and collaborators broke the tree-structured
bidirectional LSTM sentiment classification model.

Struggle with Lexical Inference

Textual Entailment

Textual Entailment (TE) takes a pair of sentences and predicts whether the facts in the first necessarily imply the facts in the second one. The AllenNLP toolkit provides the following TE visualization, which can be run for any TE model you develop. This page demonstrates a reimplementation of [the decomposable attention model \(Parikh et al, 2017\)](#), which was state of the art for [the SNLI benchmark](#) (short sentences about visual scenes) in 2016. Rather than pre-trained Glove vectors, this model uses [ELMo embeddings](#), which are completely character based and improve performance by 2%

Enter text or

Premise

Jack and Jill climbed Everest

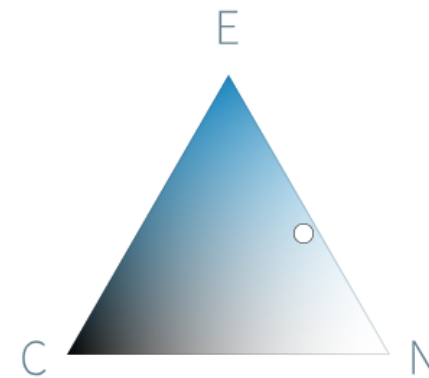
Hypothesis

Two hikers were on a mountain

RUN >

Summary

It is **somewhat likely** that there is **no correlation** between the premise and hypothesis.



Judgement

Probability

Entailment

43.5%

Contradiction

5.2%

Neutral

51.3%

Model internals (beta)



Innate Prior Debate



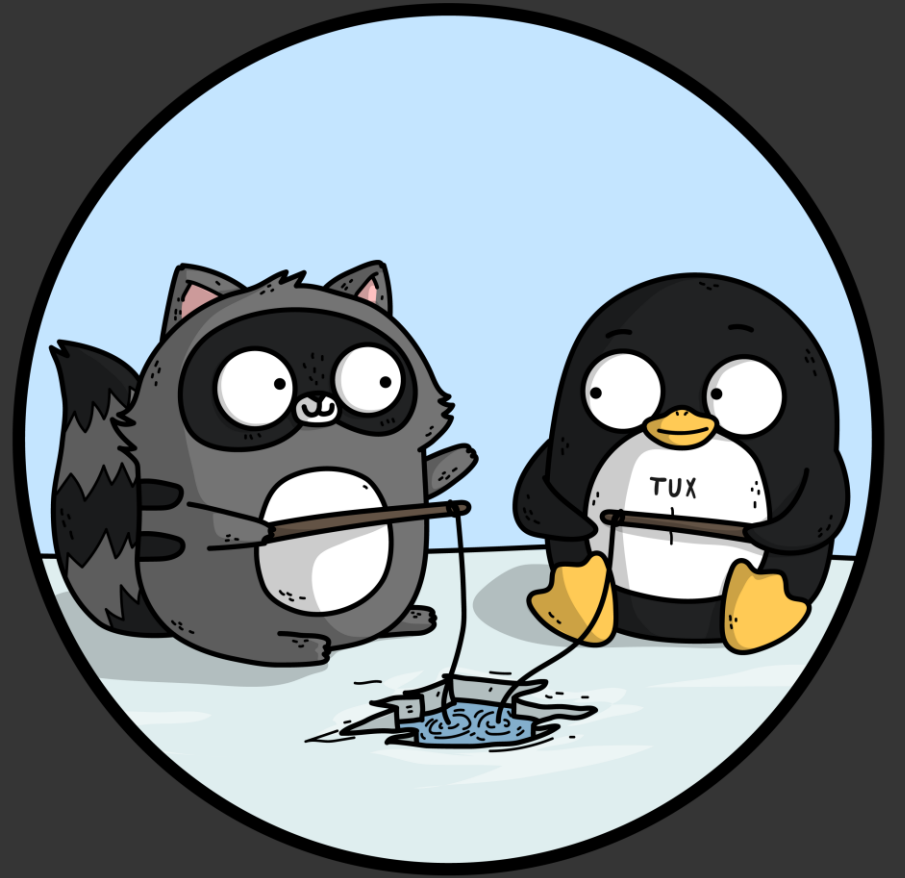
"Is structure a necessary Good
or Evil?"

Innate Prior Debate



"Is structure a necessary Good
or Evil?"

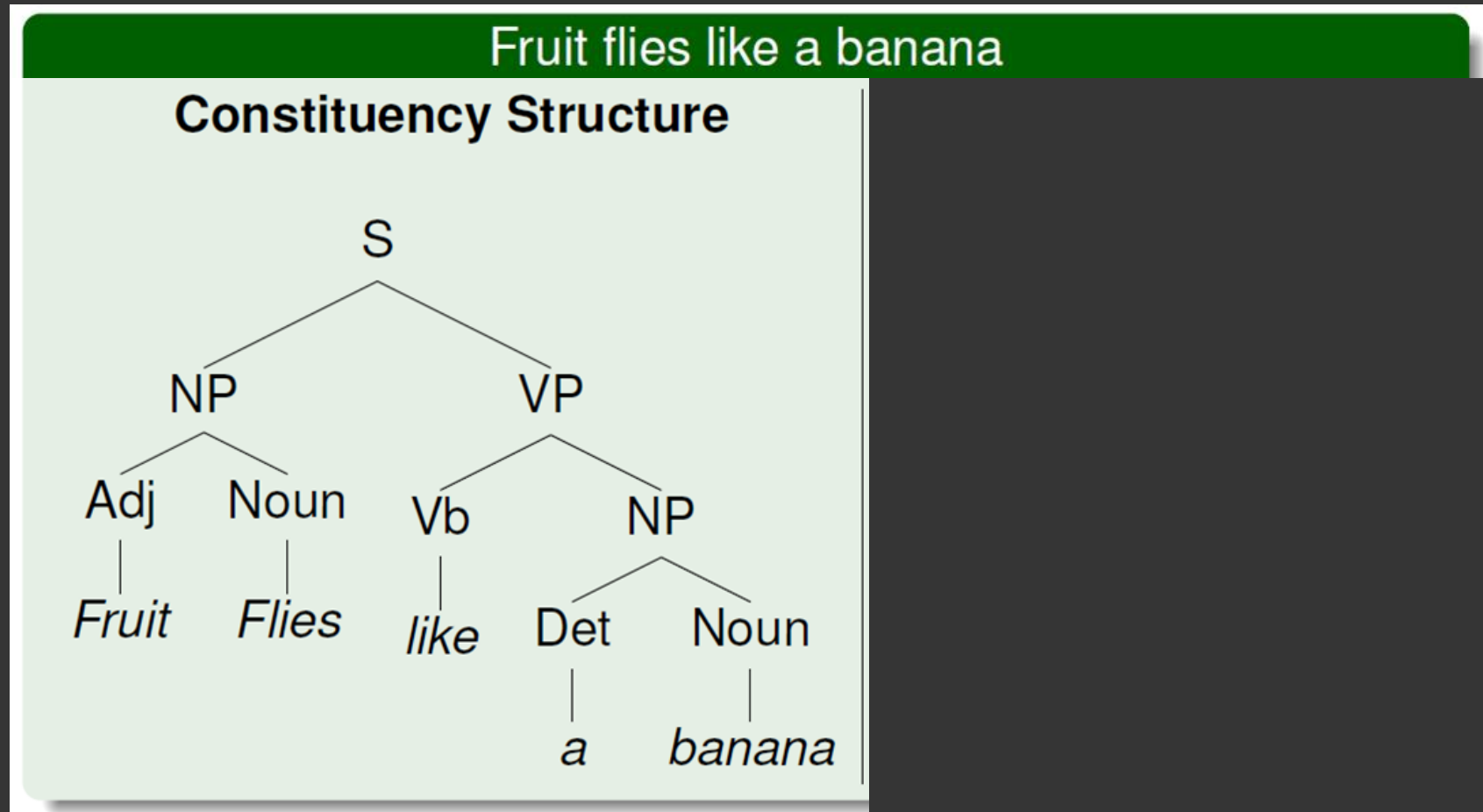
Introducing Semantic Structure



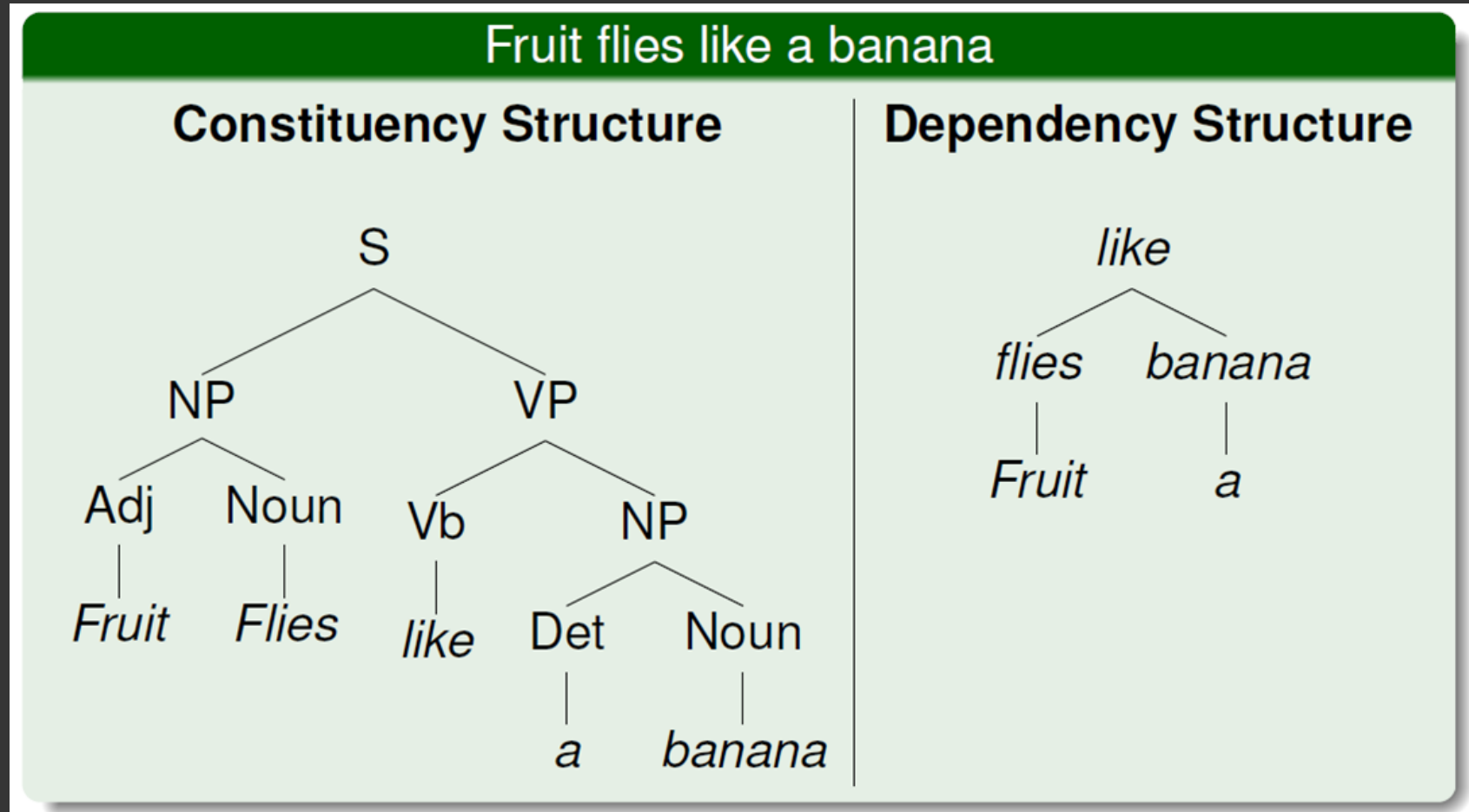
Dependency Parsing

Fruit flies like a banana

Dependency Parsing

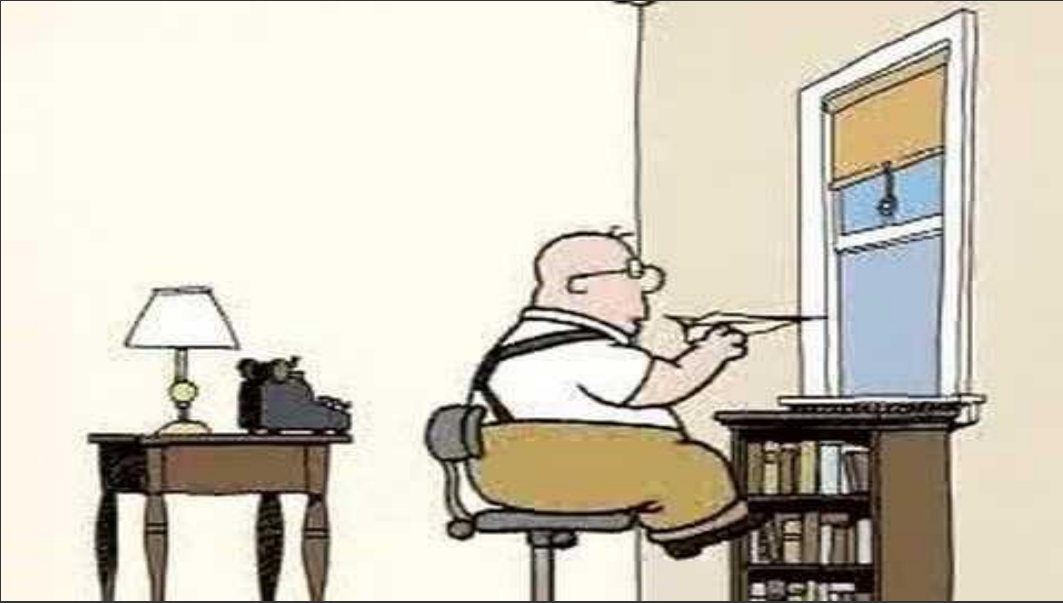


Dependency Parsing



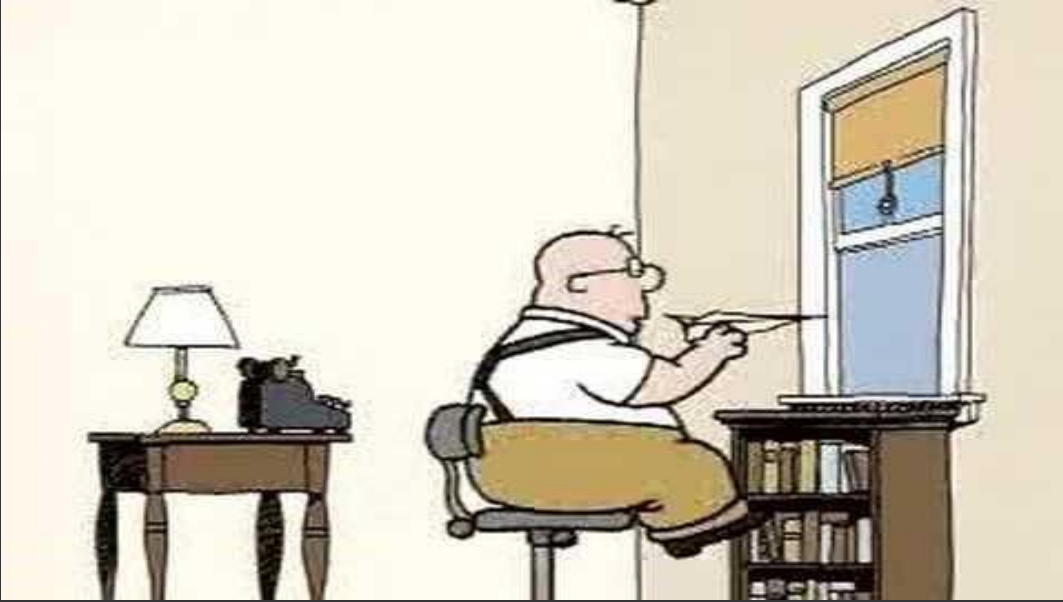
SRL

- Semantic Role Labeling aims to recover the predicate-argument structure of a sentence



The tale of Mr. Morton gives a great intro to Subject predicate structure. Whatever the predicate says he does.

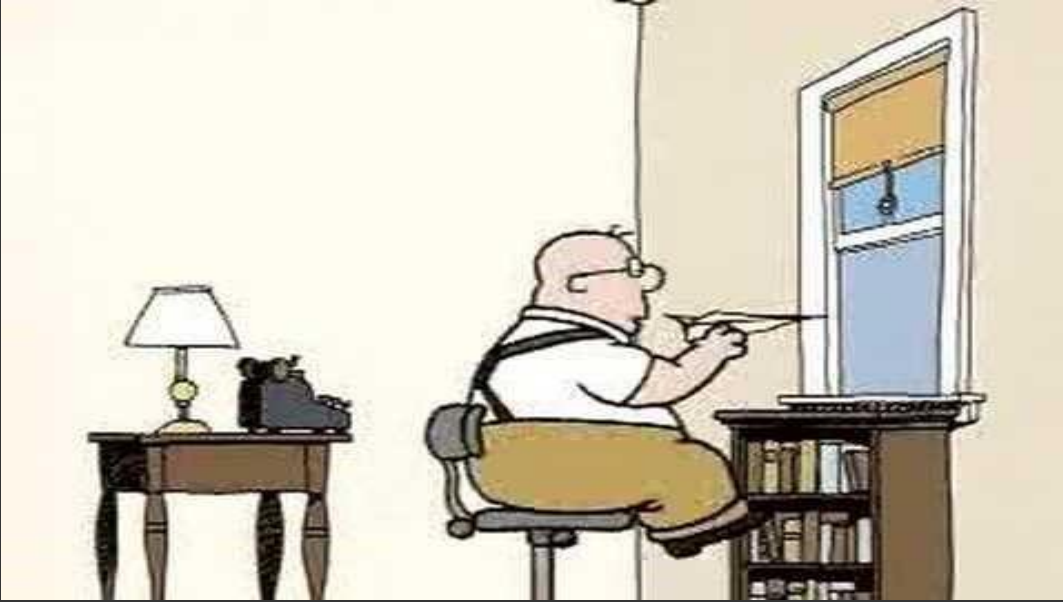
SRL



The tale of Mr. Morton gives a great intro to Subject predicate structure. Whatever the predicate says he does.

- Semantic Role Labeling aims to recover the predicate-argument structure of a sentence
- I.e who did what to whom, when, why, where and how.

SRL



The tale of Mr. Morton gives a great intro to Subject predicate structure. Whatever the predicate says he does.

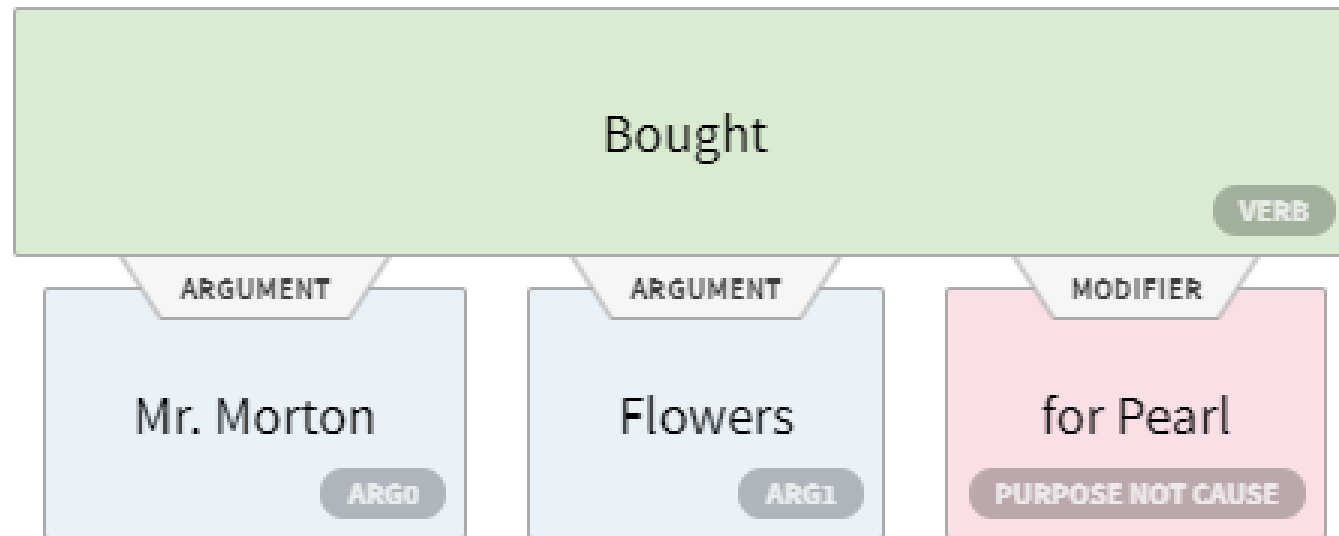
- Semantic Role Labeling aims to recover the **predicate-argument** structure of a sentence
- I.e who did what to whom, when, why, where and how.
- Think of a predicate as a function and the semantic roles as class typed arguments.

SRL

- Semantic Role Labeling aims to identify the predicate-structure of a



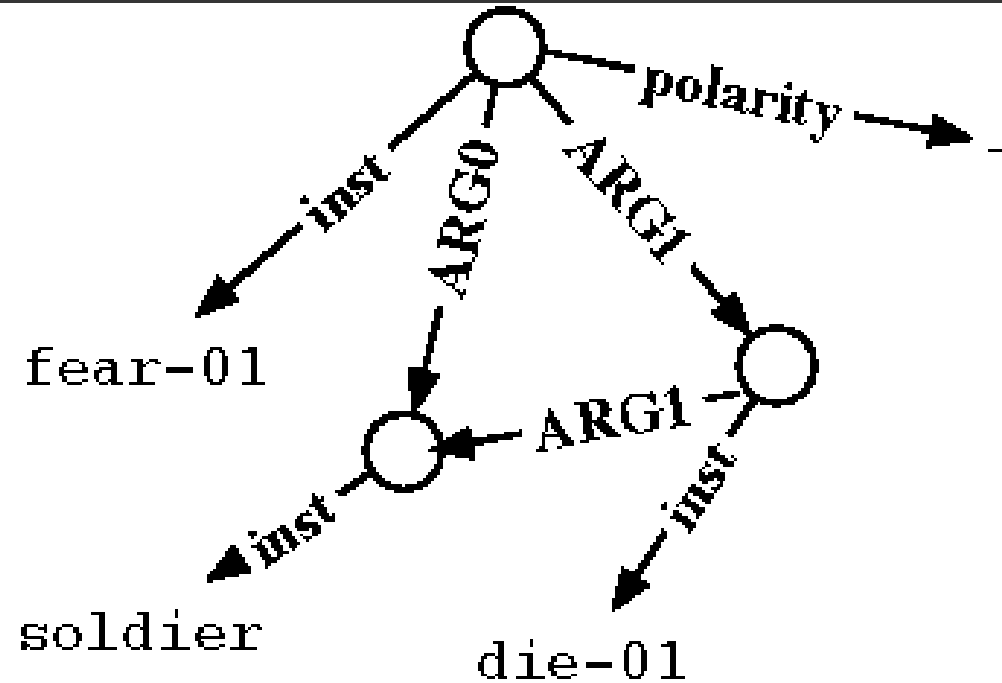
The tale of Mr. Morton's philosophy
Subject predicate structure. What ever the
predicate says he does.



roles as class typed
arguments.

AMR

- AMR Abstracts Semantic Structure from the text



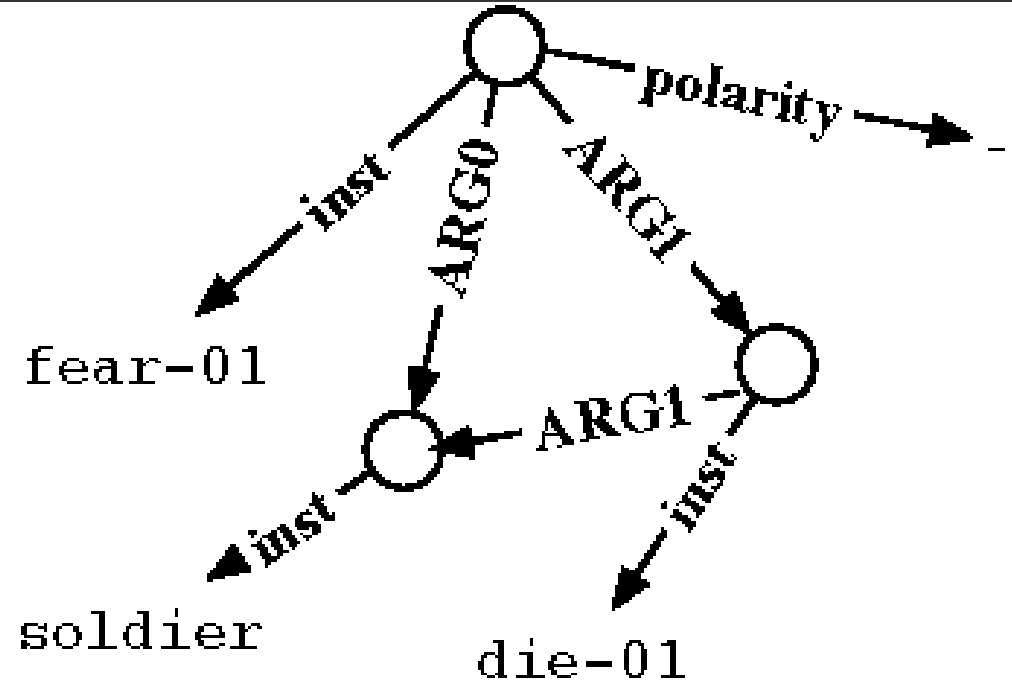
The soldier was not afraid of dying.

The soldier was not afraid to die.

The soldier did not fear death

AMR

- AMR Abstracts Semantic Structure from the text
- JAMR Parser



The soldier was not afraid of dying.
The soldier was not afraid to die.
The soldier did not fear death

SDP

- Preserves Sentence Structure

SDP

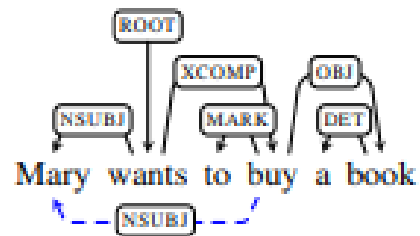
- Preserves Sentence Structure
- Deep Parse full predicate argument structure

SDP

- Preserves Sentence Structure
- Deep Parse full predicate argument structure
- Semantic Generalization

SDP

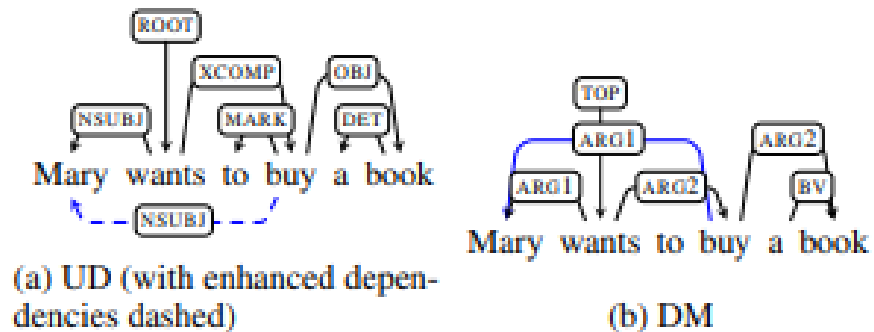
- Preserves Sentence Structure
- Deep Parse full predicate argument structure
- Semantic Generalization



(a) UD (with enhanced dependencies dashed)

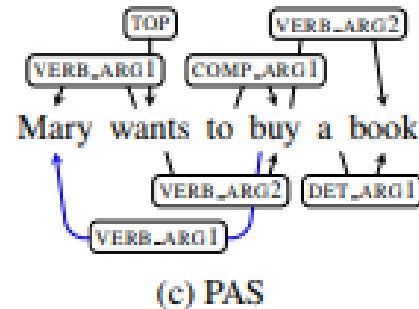
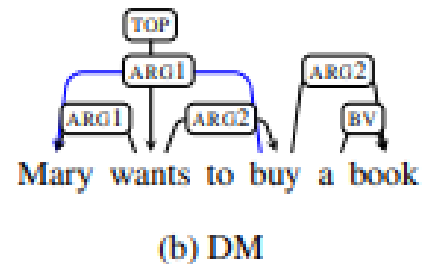
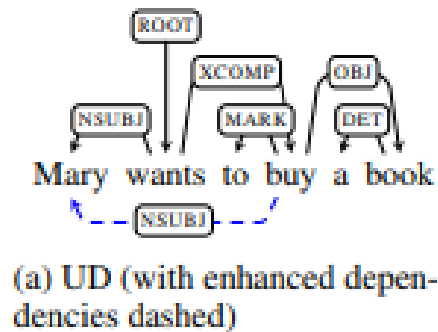
SDP

- Preserves Sentence Structure
- Deep Parse full predicate argument structure
- Semantic Generalization



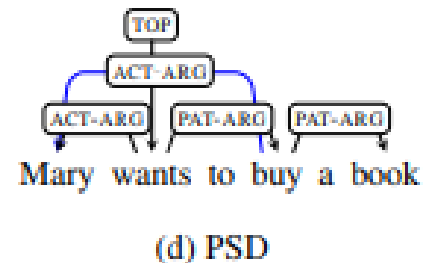
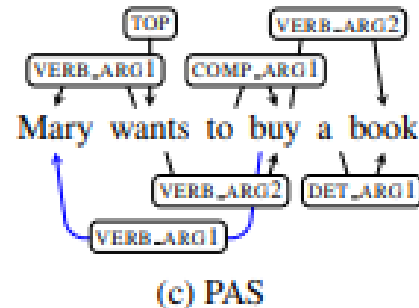
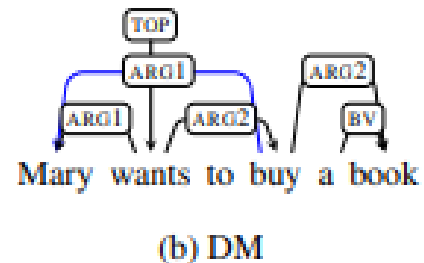
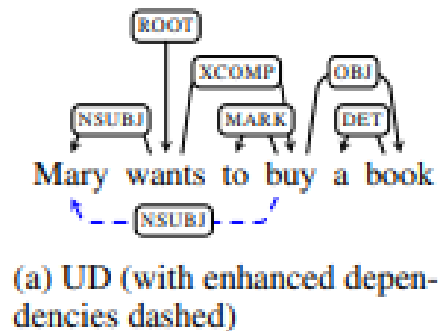
SDP

- Preserves Sentence Structure
- Deep Parse full predicate argument structure
- Semantic Generalization



SDP

- Preserves Sentence Structure
- Deep Parse full predicate argument structure
- Semantic Generalization

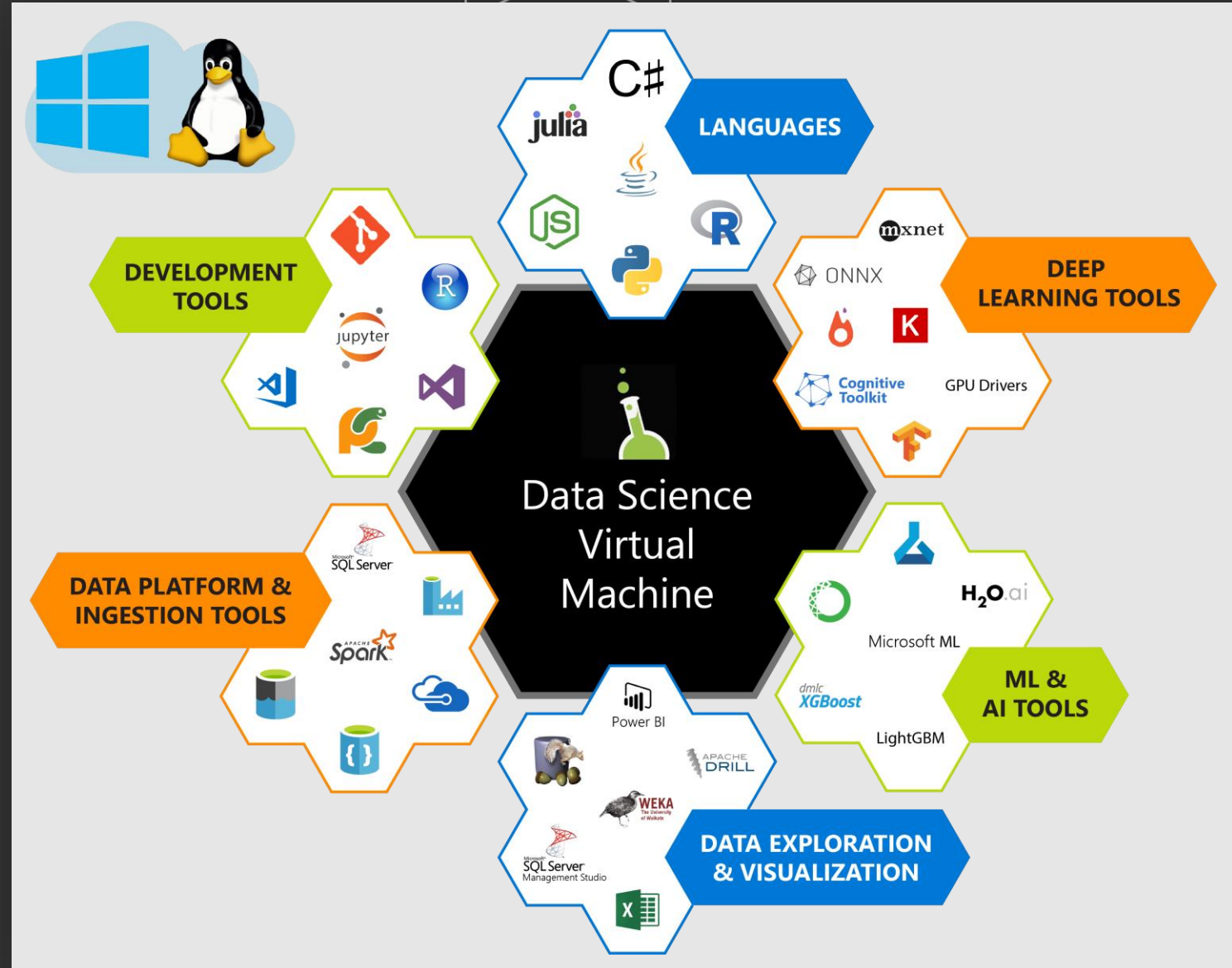


SDP Cloud Demo



Data Science Virtual Machines (DSVM)

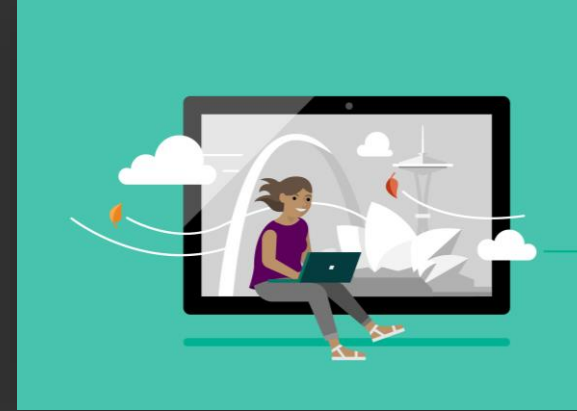
Pre-Configured environments
in the cloud for
Data Science & AI Modeling,
Development & Deployment.



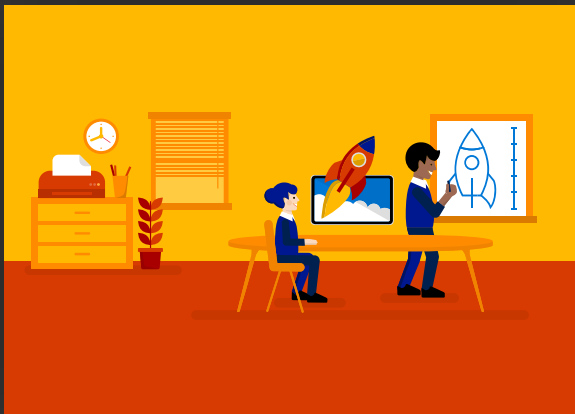
Why Data Science VMs?



Elastic analytics
desktop in the
cloud



Get started
quickly on Azure
Machine Learning



Pre-setup
environment at
work and school



Priced the same
as the base
Azure VM

Azure Machine Learning Overview

Prepare

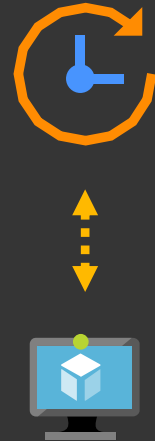


Prepare
Data

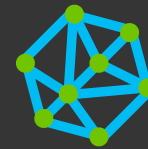
Experiment



Build model
(your favorite IDE)



Train & Test
Model



Register &
Manage Model

Deploy



Build
Image



Deploy Service &
Monitor Model

PyNeurbo Demo



Questions ???



