What to expect

- Brief history of LLMs

- A bit of insight into the inner workings of LLMs

- Contemporary LLMs (and contemporary problems)

- Build intuition on…
  - Why they perform great on some tasks.
  - Why they suck on some tasks.

- Starting point for tool / model selection for a task

- Know how to run an LLM locally

A very condensed history of LLMs

Distributional hypothesis

"*a word is characterized by the company it keeps*"

- J. R. Firth (1950)

"*Words that occur in the same contexts tend to have similar meanings*"

- Z. S. Harris (1954)

Early precursors

Rule based

- Georgetown-IBM experiment 1952-54, translation russian to english

Very limited, 250 words vocabulary.

Early precursors

Parsing grammar

- Syntactic Structures by Noam Chomsky (1957)

Understanding grammar != understanding the message

Not everything that is grammatically correct is also a meaningful sentence.

“*Colorless green ideas sleep furiously.*”

- Noam Chomsky (1957)

Not everything that is grammatically correct is also a meaningful sentence.

“*Colorless green ideas sleep furiously.*”

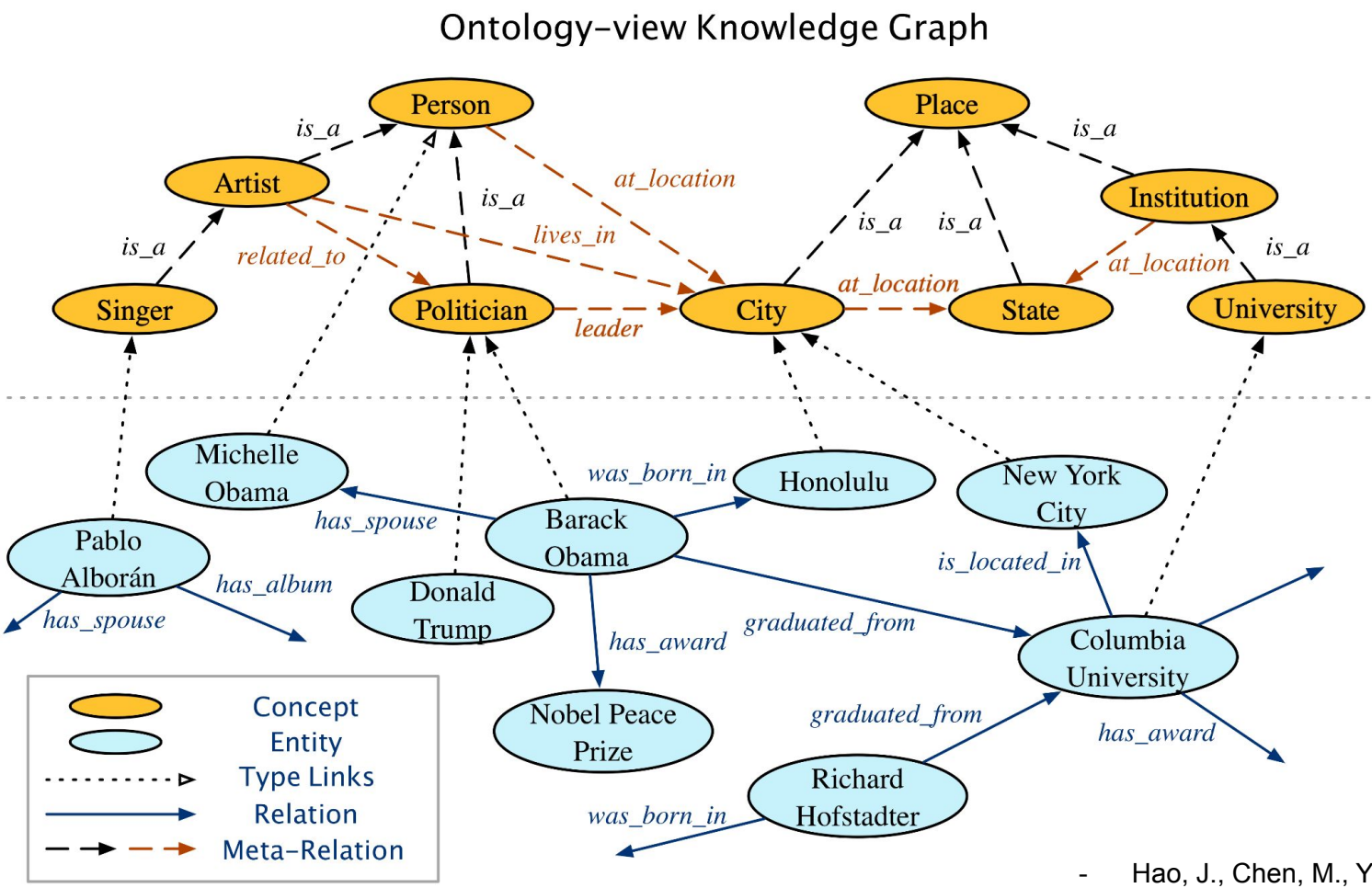
- Noam Chomsky (1957)

Sidetrack - Knowledge representation

Semantic networks to encapsulate knowledge

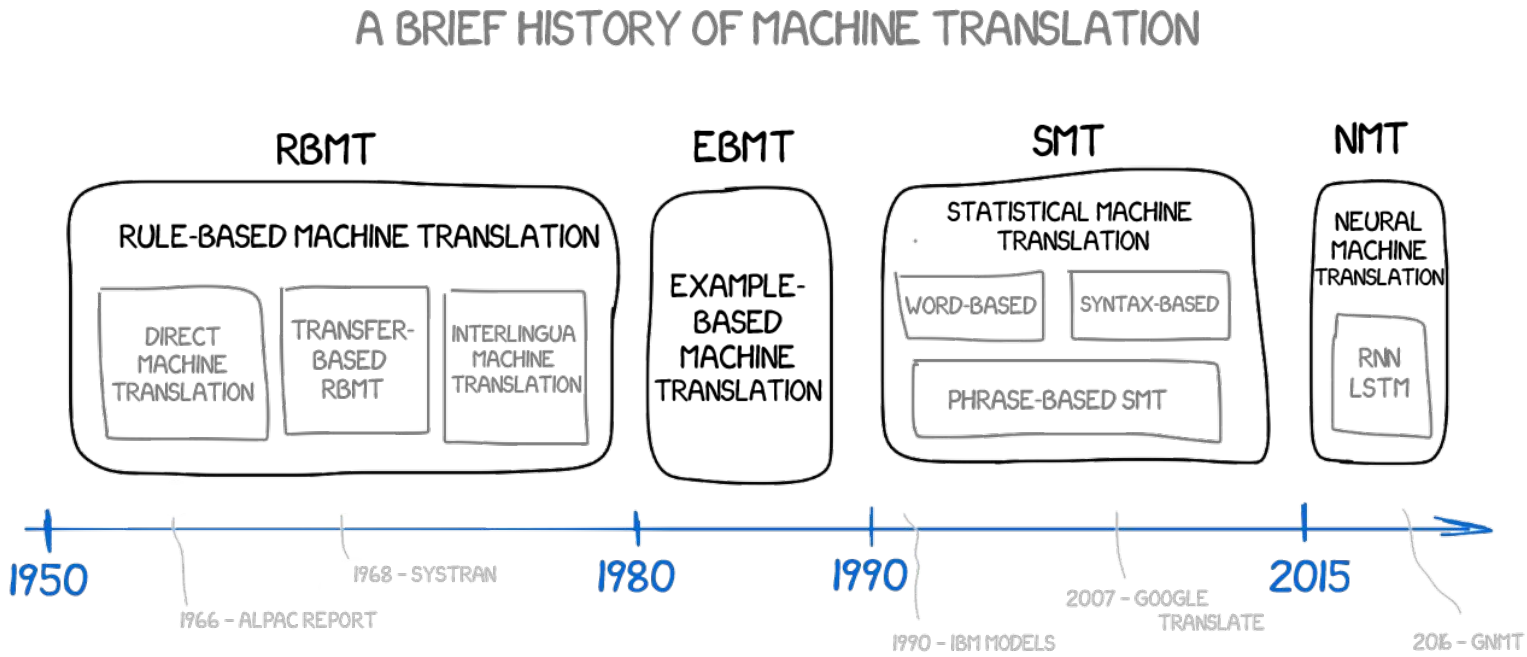
Ontologies

- Model knowledge using instances, classes, relations, and attributes



- Hao, J., Chen, M., Yu, W., Sun, Y. and Wang, W., 2019, July. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts.

Back to the history of LLMs…



https://vas3k.com/blog/machine_translation/

More recent History

2001 - Neural language models

2008 - Multi-task learning

2013 - Word embeddings

2013 - Neural networks for NLP

2014 - Sequence-to-sequence models

2015 - Attention

2015 - Memory-based networks

2018 - Pretrained language models

https://ruder.io/a-review-of-the-recent-history-of-nlp/