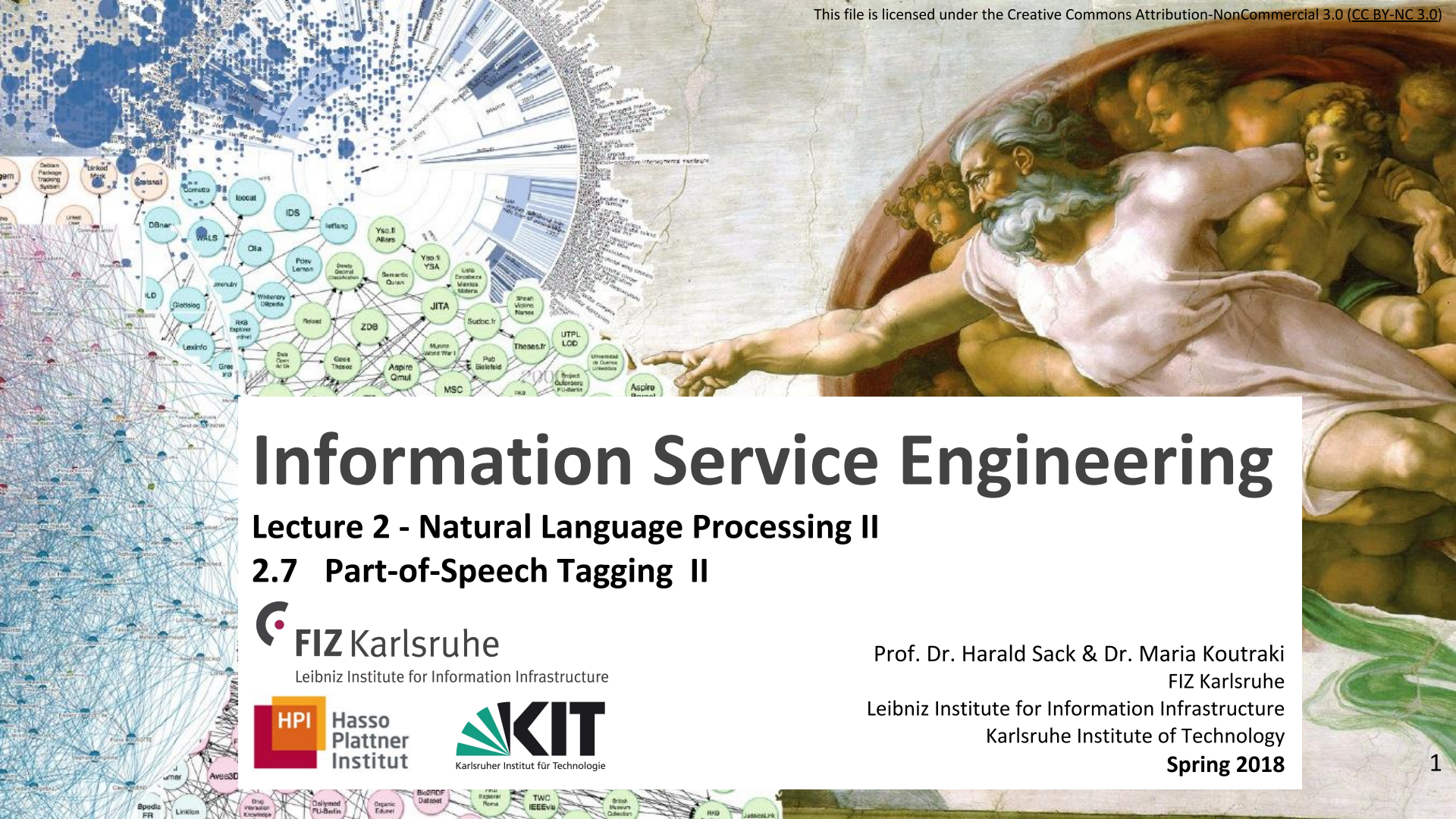




Information Service Engineering

Lecture 2 - Natural Language Processing II

2.7 Part-of-Speech Tagging II



FIZ Karlsruhe

Leibniz Institute for Information Infrastructure



Karlsruher Institut für Technologie

Prof. Dr. Harald Sack & Dr. Maria Koutraki
FIZ Karlsruhe

Leibniz Institute for Information Infrastructure
Karlsruhe Institute of Technology
Spring 2018

Information Service Engineering

Lecture 2: Natural Language Processing II

2.1 Finite State Automata

2.2 Finite State Transducers

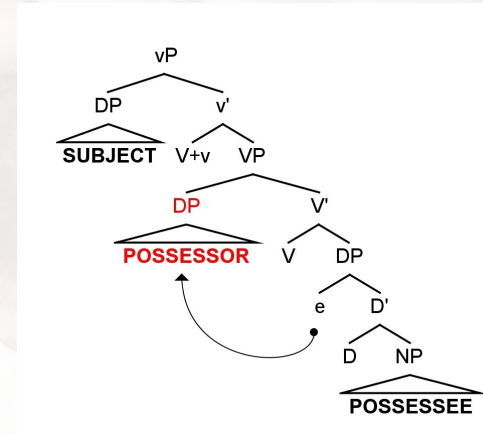
2.3 Language Model and N-Grams

2.4 Language Model and N-Grams II

2.5 Language Model and N-Grams III

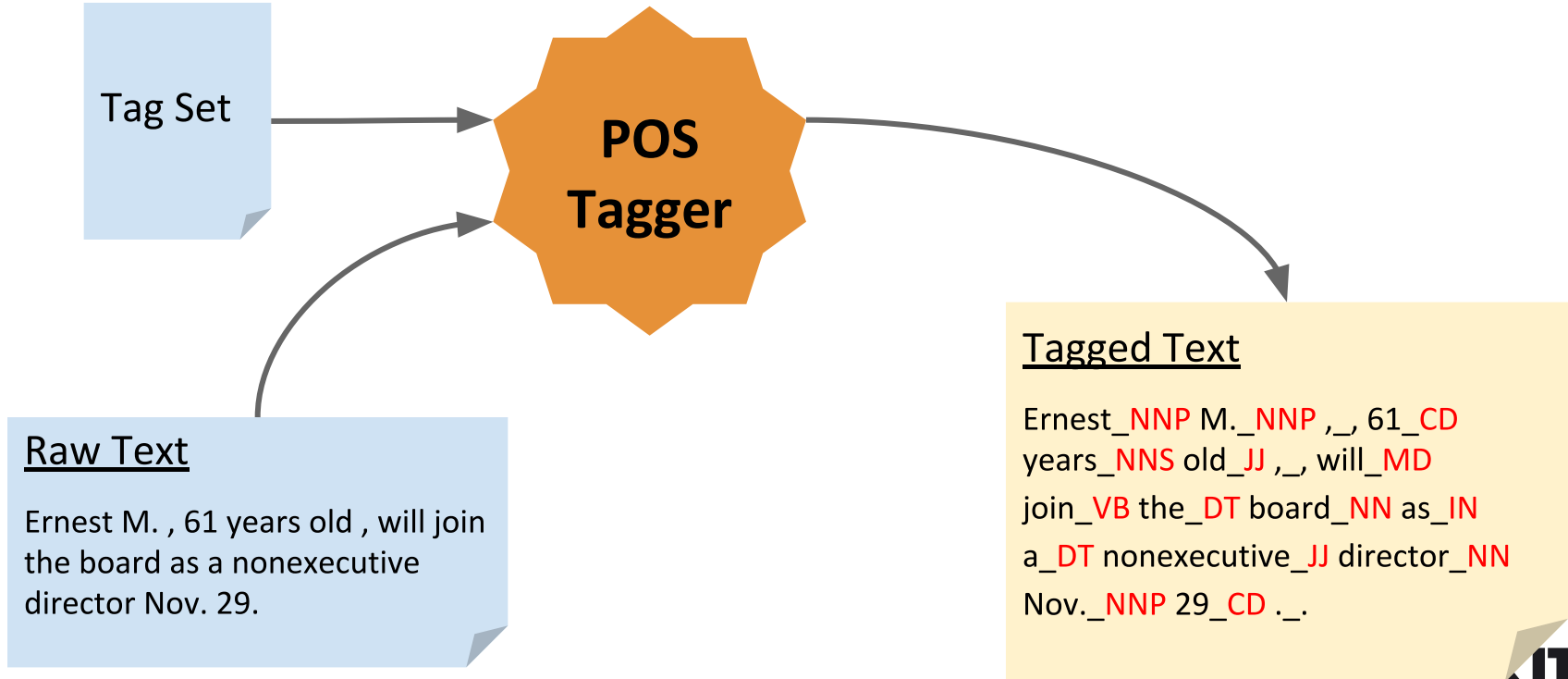
2.6 Part-of-Speech Tagging

2.7 Part-of-Speech Tagging II



Part-of-Speech Tagging

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II



POS Tagging - The Baseline

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

1. Tagging **unambiguous words** with the **correct label**
 2. Tagging **ambiguous words** with their **most frequent label**
 3. Tagging **unknown words** as a **noun**
-
- This method (*Baseline*) performs around **90% accuracy**
 - **State-of-the-art POS tagger** achieve about **97% accuracy**
 - **Humans** (*Ceiling*) perform around **97% accuracy**

More sophisticated POS Tagging

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- *time flies like an arrow*
 - time (NN/VBZ)
 - flies (VBZ/NNS)
 - like (VBZ/IN/RBZ/CC/JJ)
 - an (DT)
 - arrow (NN/VBZ)
- Since there is ambiguity, we cannot simply look up the correct POS in a dictionary.

More sophisticated POS Tagging

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- *time flies like an arrow*
 - time (NN/VBZ)
 - flies (VBZ/NNS)
 - like (VBZ/IN/RBZ/CC/JJ)
 - an (DT)
 - arrow (NN/VBZ)
- time_[NN] flies_[VBZ] like_[IN] an_[DT] arrow_[NN]

Algorithms for POS Tagging

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- **Rule-based POS Tagging**
 - 2-step approach:
 1. Dictionary to assign a list of potential tags
 2. Hand-written rules to restrict to a POS tag
- **Stochastic (Probabilistic Tagging)**
 - Hidden Markov Models
- **Transformation Based Tagging**
 - Combination of Rule-based and Stochastic Tagging
 - Rules are learned from data

Stochastic POS Tagging

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- What is the **most likely sequence of tags** for the **given sequence of words w** ?

Stochastic POS Tagging

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- What is the **most likely sequence of tags** for the **given sequence of words w**?
- Let's try:

෧/DT ෧෪/NN ෦ 10/VBZ ෧෧ ෦ 5/VBG ෧/DT
෧෧ 0/NN ෪/.

෧/DT ෧෪ 4/NN ෦ 10/VBZ 9 1 5 5 0 5/VBG ෪/.

෧/DT ෧෪ 5/NN ෦ 10/VBZ 10 0 5/෪ 0 5/VBG ෪/.

෧/DT ෧෧ 7 7 5/JJ ෧ 0 9/NN

Stochastic POS Tagging

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- What is the **most likely sequence of tags** for the **given sequence of words w** ?
- Let's try:

ශ්‍රී/DT ලූ/NN ෧෦/VBZ නිශ්ශ්‍රී/NN ෧෦/VBG ශ්‍රී/DT
 නිශ්ශ්‍රී/NN /./

VBZ 9 1 5 5 0 5 VBG 1

෧/DT ෧6 5/NN 0 10/VBZ 10 0 5 ෧0 5 ෧/VBG ෧/.

ෆ/DT ෆෆෆෆෆෆෆ/JJ ෆෆෆෆෆෆෆ/NN

- Now, what is the most likely tag sequence for

၆ **မိုး**(၇)(၇)**၆** **ကွေး**(၀) **ဒုဗ္ဗ**(၁၀) (၁၀)(၀)(၅)-**လူ**(၀)(၅)-

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

Stochastic POS Tagging

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- How do you predict tags?
- Two types of information are useful
 - Relations between **words and tags**
 - as e.g., **dog**/NN, **fox**/NN, **a**/DT, **the**/DT
 - Relations between **tags and tags**
 - as e.g., **DT NN**, **DT JJ NN**...

Stochastic POS Tagging

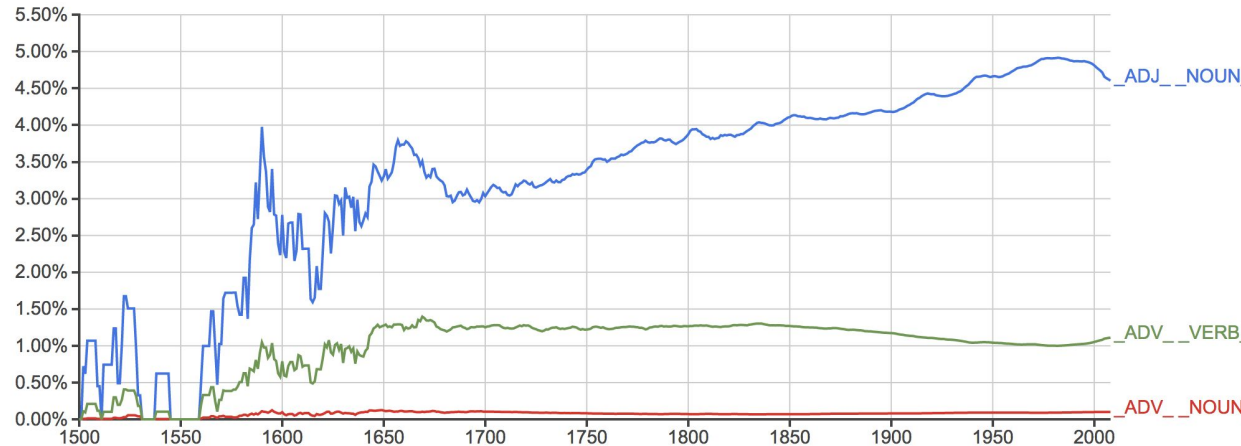
Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- Model POS tagging as **sequence tagging problem**
- Some POS-Tag sequences are more likely than others

Google Books Ngram Viewer

Graph these comma-separated phrases: ☐ case-insensitive
between 1500 and 2008 from the corpus English with smoothing of 3 [Search lots of books](#)

[Google N-gram Viewer Example](#)



(click on line/label for focus)

logy

How to resolve Tag Ambiguity?

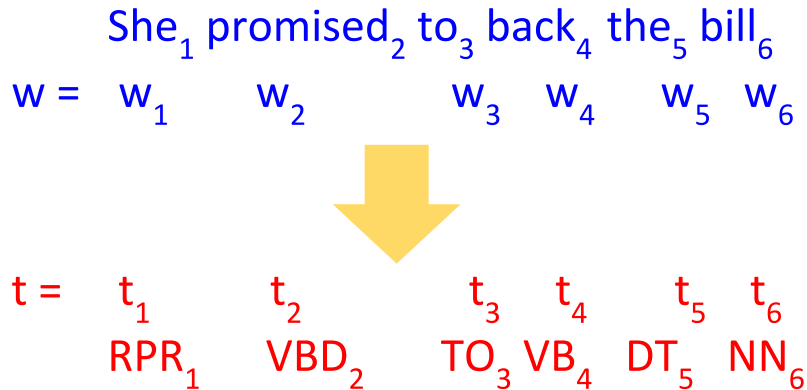
Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- Making a decision based on:
 - **Current Observation:**
 - Word (W_0): „35-years-old“
 - Prefix, Suffix: „computation“ „comp“, „ation“
 - Lowercased word: „New“ „new“
 - Word shape: „35-years-old“ „d-a-a“
 - **Surrounding observations**
 - Words (W_{+1} , W_{-1})
 - **Previous decisions**
 - POS tags (T_{-1} , T_{-2})

Hidden Markov Models (HMM)

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- Finding the best **sequence of tags** $(t_1, \dots, t_n)$ that corresponds to the **sequence of observations** $(w_1, \dots, w_n)$



Hidden Markov Models (HMM)

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- Finding the best **sequence of tags** $(t_1, \dots, t_n)$ that corresponds to the **sequence of observations** $(w_1, \dots, w_n)$
- Probabilistic View
 - Considering all possible sequences of tags
 - **Choosing** the tag sequence from this universe of sequences, which is **most probable given the observation sequence**

$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} P(t_1^n | w_1^n)$$

Hidden Markov Models (HMM)

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- Using the **Bayes Rule**:

$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} P(t_1^n | w_1^n)$$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$P(t_1^n | w_1^n) = \frac{P(w_1^n | t_1^n) \cdot P(t_1^n)}{P(w_1^n)}$$

$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} P(w_1^n | t_1^n) \cdot P(t_1^n)$$

likelihood
of word string

prior probability
of tag sequence

Hidden Markov Models (HMM)

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II

- Using the **Markov Assumption**:

$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} P(w_1^n | t_1^n) \cdot P(t_1^n)$$

Emission Probability

$$P(w_1^n | t_1^n) \simeq \prod_{i=1}^n P(w_i | t_i)$$

only depends on its POS tag, **independent of other words**

Transition Probability

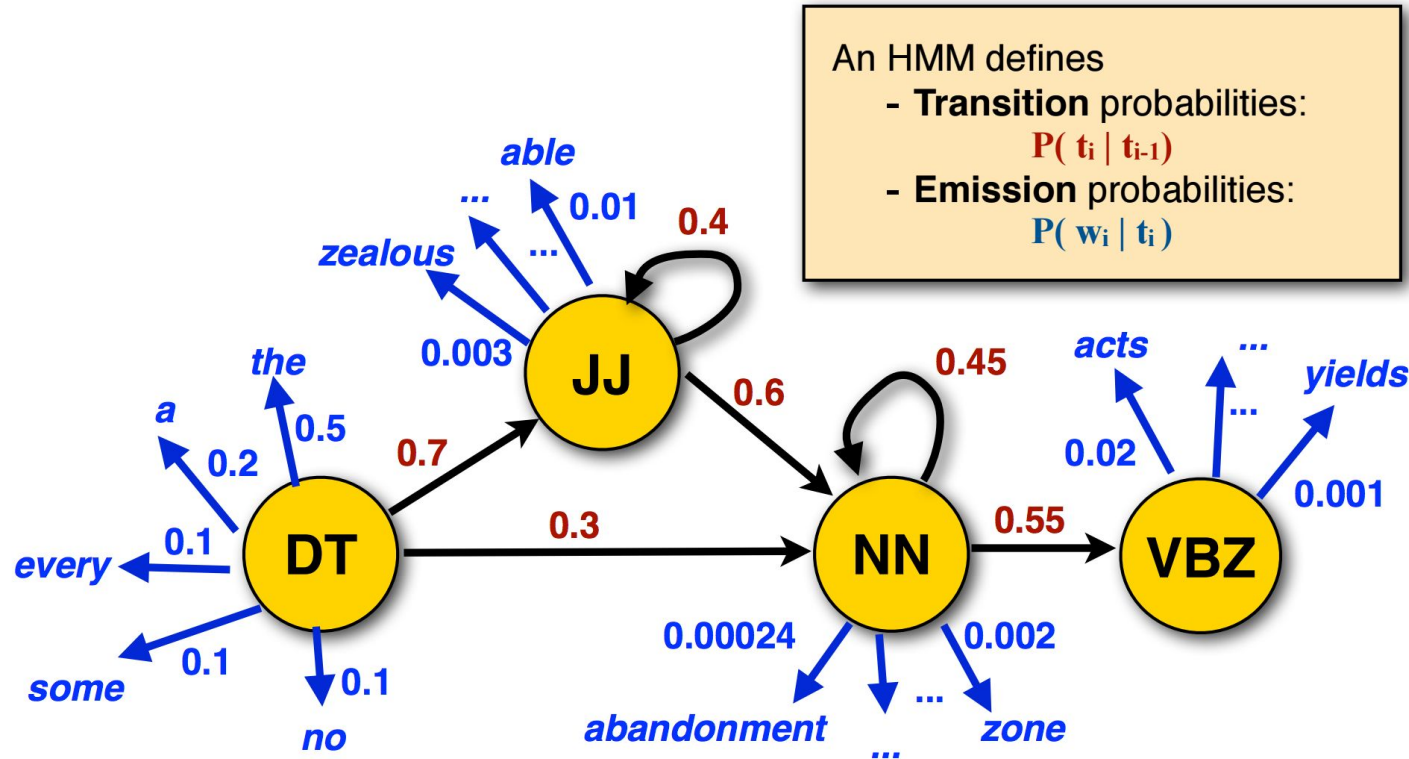
$$P(t_1^n) \simeq \prod_{i=1}^n P(t_i | t_{i-1})$$

only depends on previous POS tag, i.e. **bigram**

$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} \prod_{i=1}^n P(w_i | t_i) \cdot P(t_i | t_{i-1})$$

Hidden Markov Models (HMM)

Information Service Engineering - Lecture 2 Natural Language Processing II - 2.7 Part-of-Speech Tagging II



Information Service Engineering

Lecture 2: Natural Language Processing II

2.1 Finite State Automata

2.2 Finite State Transducers

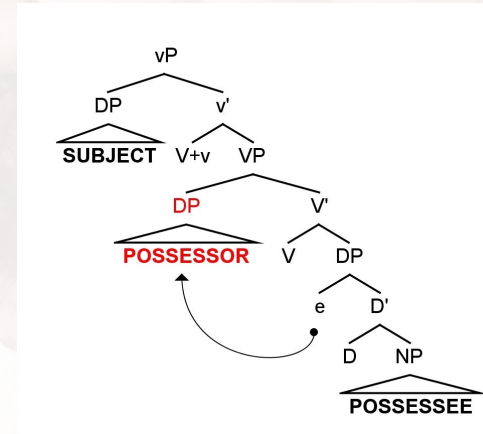
2.3 Language Model and N-Grams

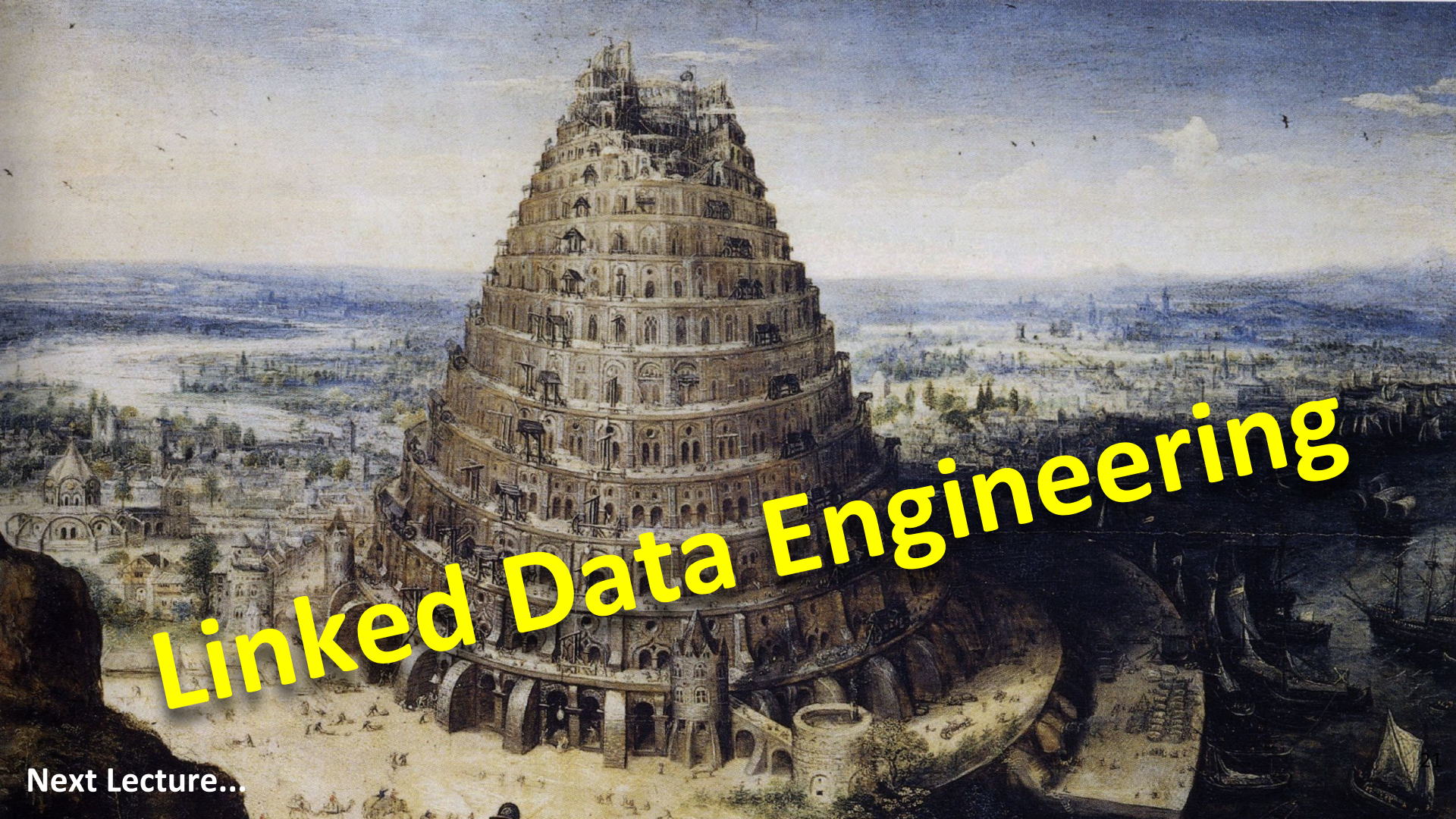
2.4 Language Model and N-Grams II

2.5 Language Model and N-Grams III

2.6 Part-of-Speech Tagging

2.7 Part-of-Speech Tagging II





Linked Data Engineering

Next Lecture...

Picture References:

- P.21: Lucas van Valckenborch, The Tower of Babel (1594)
[https://commons.wikimedia.org/wiki/File:The Alchemist by Pieter Brueghel the Younger.jpg](https://commons.wikimedia.org/wiki/File:The_Alchemist_by_Pieter_Brueghel_the_Younger.jpg)