Natural Language Processing
CSE 40657/60657
Fall 2023
Three Goals

- Machine translation
- Natural language interfaces
- Information extraction
Three Goals

- Machine translation
- Natural language interfaces
- Information extraction
Three Goals

- Machine translation
- Natural language interfaces
- Information extraction
Three Goals

machine translation

natural language interfaces

information extraction
Machine Translation

The Georgetown-IBM machine translation experiment, 1954
Q: Please write me a sonnet on the subject of the Forth Bridge.
A: Count me out on this one. I never could write poetry.
Q: Add 34957 to 70764
A: (Pause about 30 seconds and then give as answer) 105621.
Q: Do you play chess?
A: Yes.
Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?
A: (After a pause of 15 seconds) R-R8 mate.


Amidst the rolling waves and salty air,
A marvel stands with strength that ne'er abates,
The Forth Bridge, a symbol grand and rare,
Connecting lands through iron-laden gates.

Its mighty spans in red embrace the sky,
A testament to human skill and might,
A feat of engineering that draws the eye,
A bridge that spans both day and starry night.

For decades now, it's stood with grace untold,
Its towers rise, unyielding 'gainst the tide,
A link 'twixt shores where tales of old are told,
A legacy of progress, far and wide.

So let us marvel at this bridge so grand,
A union of the land and sea and man.
Natural Language Interfaces

Q: Please write me a sonnet on the subject of the Forth Bridge.
A: Count me out on this one. I never could write poetry.
Q: Add 34957 to 70764.
A: (Pause about 30 seconds and then give as answer) 105621.
Q: Do you play chess?
A: Yes.
Q: I have K at my K1, and no other pieces. You have only K at K8 and R at R1. It is your move. What do you play?
A: (After a pause of 15 seconds) R-R8 mate.


Do you play chess?
I don't play chess myself, but I'm familiar with the rules and strategies of the game. If you have any questions about chess or need assistance with understanding certain aspects of the game, feel free to ask!

I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

In the situation you've described, I would play Rb1#. This move would place my rook at R1 and deliver checkmate to your king at K1, ending the game.
Information Extraction

Please explain the concept of presence according to St. Thomas Aquinas.

St. Thomas Aquinas' concept of presence extends beyond the doctrine of the Eucharist and also encompasses the broader philosophical understanding of presence and existence. In his philosophy, presence is closely linked to the metaphysical concepts of substance and accidents.

1. **Substance and Accidents**: Aquinas distinguishes between the substance and accidents of a thing. The substance is the underlying, unchanging reality or essence of an object, while the accidents are the mutable, external characteristics that we perceive through our senses, such as color, taste, size, and shape.

2. **Real and Objective Presence**: Aquinas' philosophy emphasizes the objective reality of things. He believed that things exist in themselves, independently of our perceptions or thoughts about them. In this context,
However…

Please write a sentence that starts and ends with the same word.

"Time" is both a constant companion and an enigmatic force, shaping the very existence of our universe.

Does "0110101" have an even or odd number of "1"s?

The binary number "0110101" has an odd number of "1"s. There are a total of 4 "1"s in the sequence.

Please insert "freakin" into the word "fantastic."

Sure, the modified word would be "fanfreakintastic."

Why can't it be "fantasfreakintic"?

You're absolutely right, "fantasfreakintic" would also be a valid way to insert "freakin" into the word "fantastic." My previous response was just one possible option, but your suggestion works perfectly fine as well!

Please insert "freakin" into the word "computational."

Certainly! The modified word would be "comfreakintational."
However...

What is wrong with the sentence "Who was John wondering whether could eat peanuts?"

In response to the question, "Who would like ice cream," would you say, "I would!" or "I'd!?"

The sentence "Who was John wondering whether could eat peanuts?" is a bit awkward and unclear. The main issue is with the placement of "whether" and the lack of a clear subject for the action. To make the sentence clearer, you could rephrase it as:

"John was wondering whether someone could eat peanuts."

Both "I would!" and "I'd!" are correct responses to the question "Who would like ice cream?" They both convey the idea that you would like ice cream. "I'd" is a contraction of "I would," so either response is appropriate and grammatically accurate. The choice between the two depends on your personal preference and the level of informality you want to convey.
Two Big Ideas

<table>
<thead>
<tr>
<th></th>
<th>Linguistics</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Two Big Ideas

Linguistics  Machine Learning

1950  1970  1990  2010

ALPAC report, 1966
A noble spirit embiggens the smallest man.
A noble spirit embiggens the smallest man

**Syntax** -> **Semantics**

**Words** -> **Lexical Semantics**

**Morphemes** -> **Morpho-semantics**

**Text** -> **Phonemes**

**Noisy Text** -> **Speech**
Predicate: `enlarge`

**Roleset id:** `enlarge.01`, *make bigger*, **Source:** , **vncls:** , **framnet:**

`enlarge.01`: ENLARGE-V NOTES: Preposition 'by' takes precedence over EXT tag. (from enlarge.01-v) ENLARGED-J NOTES: Based on Thyme and Sharp data. Comparison to enlarge.01. Enlarge.01 is a member of VN class other_cos-45.4. Framed by Katie. (from enlarged.01-j)

**Roles:**
- **Arg0-PAG**: *causer of biggenment* (vnrole: 45.4-agent)
- **Arg1-PPT**: *thing getting bigger* (vnrole: 45.4-patient)
- **Arg2-EXT**: EXT
- **Arg3-DIR**: *start point*
- **Arg4-GOL**: *end point*
## Two Big Ideas

<table>
<thead>
<tr>
<th>1950</th>
<th>1970</th>
<th>1990</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistics</td>
<td></td>
<td>Machine Learning</td>
<td></td>
</tr>
</tbody>
</table>
Two Big Ideas

DARPA Speech and Natural Language Workshop, 1989
The hurricane moved from the west to the east. The system reached hurricane intensity on November 15 when it was south of Jamaica, and then proceeded to pass over the southern part of Hispaniola and Puerto Rico. On November 17, Leni strengthened rapidly in the northeastern Caribbean and reached a maximum wind speed of 250 km/h south of St. Croix, U.S. Virgin Islands. The hurricane weakened during its passage through the Leeward Islands and dissipated in the open waters of the Atlantic Ocean on November 23. Before passing through the northern small islands, Leni generated large waves that claimed two lives in northern Colombia. Heavy rains caused significant damage to crops in southeastern Puerto Rico. Hurricane Leni made landfall with its highest intensity over St. Croix Island, where there were widespread flooding and erosion, but the damage was only of "moderate" degree. The U.S. and Puerto Rico suffered losses of 330 million dollars. The highest rainfall was recorded in the French territory of Martinique with 867 mm. The island lost 3 people and over 200 properties were destroyed. A further property loss was experienced in Antigua and Barbuda, where 1 person was killed, and floods and erosion occurred. The heavy rains also contaminated the local water supplies. As far south as Grenada, the storm caused significant damage, with large waves cutting off联系 between several towns and the capital. Due to the massive losses caused by Hurricane Leni, its name was retired from the Atlantic hurricane name list, and it will not be used again in future North Atlantic tropical cyclone naming.

The 10th Panchen Lama, also known as the 10th Panchen Lama, was the abbot of the Gelugpa, or Yellow Hat, order of藏传佛教, with the second highest position in the teachings of the Gelugpa order, after the Dalai Lama. In June 1949, after the approval of the acting President of the Republic of China, Chiang Kai-shek, and the establishment ceremony on August 10, he formally succeeded. In September of the same year, after the People's Liberation Army captured Xining, the 10th Panchen Lama actively contacted the government, and expressed his support for the central government. In 1951, both the 10th Panchen Lama and the 14th Dalai Lama signed the "Agreement on the Peaceful Liberation of Tibet." In 1959, after the riots in Tibet, the 10th Panchen Lama decided to remain in Tibet, accepting the appointment as the president of the Tibet Autonomous Region. Due to his disapproval of the issues arising from the democratic reforms implemented in Tibet by the Chinese government. The 10th Panchen Lama wrote the "Seven Thousand Words Book" in 1962, criticizing the Chinese government's policies in Tibet. In 1964, he was accused of "counter-revolutionary, anti-socialist, and conspiring to闹革命" and had all of his positions stripped. In 1966, the "Panchen Group" was persecuted and imprisoned. It was not until 1978 that he was released with the permission of Deng Xiaoping, the then leader of China. After release, he recovered his positions, and implemented a series of reconstruction and reform programs in Tibet, which were supported by many Tibetans.

Münster is a city in the northern part of the German state of North Rhine-Westphalia. It is the cultural center of the region and the capital of the Münster district. It served as the site of the Münster Rebellion, which occurred during the German Reformation, and it is also known as the place where the 1648 Peace of Westphalia was signed. Today, the city is known as the "Bicycle Capital of Germany."
In the case where the appositive is an IP or CP, on the surface the IP-NP or CP-NP is a constituent,
and therefore it is a word-level category whereas both head nouns in a noun-noun compound are phrase-level
categories. The former case, however, is different from a relative clause in that the latter but not the former has a
predication when it is not a verb phrase, and to the arguments (if there is any) of the VP a dash tag is attached
to indicate the predication relationship between the head noun and the argument.

To sum up, predication is represented by attaching dash tags to the subject, to the purpose of this guideline, by predication we only mean the relationship between
the subject and the predicate. Predication is represented non-configurationally by attaching the functional tag
-SBJ to the subject at the IP level, attaching functional tags indicating thematic roles to the arguments that are the complements
of the verb (-OBJ, -EXT, etc.), when the predicate is a VP. Since a VP is always the focus of the sentence.

Another scenario where the VP is not the predicate is when the verb is the
questionable and negatable, with the help of the copula in this case, the predicate is the QP. Adjuncts are attached at the VP level
possible non-verbal predicates in Chinese are NPs, QPs and PPs. They occur when
the verb is a non-verb.

1.1.1.1 The subject
NP predicates:

When the predicate is a VP, the VP is generally formed by the verb and its
adjuncts. When the VP is not the predicate:

(i) Either the verb is the copula (is, does, can, etc.) and therefore do not assign
a dash tag -PRD. However, each of the arguments do get a dash tag. When the
verb phrase present it is assumed that a null verb is present and it projects layer of VP
external to the predicate:

1.1.1.1.2 Verbal predicates:

Another scenario where the VP is not the predicate is when the verb is the

1.1.1.2 The predicate

Superficially this looks similar to noun-noun compound in which the first noun

1.1.1.2.1 Adjectival predicates:

Another scenario where the VP is not the predicate is when the verb is the

Another scenario where the VP is not the predicate is when the verb is the

Another scenario where the VP is not the predicate is when the verb is the

Another scenario where the VP is not the predicate is when the verb is the
11月17日，莱尼在加勒比海东北部快速强化，在美属维尔京群岛的圣克洛伊岛以南约34公里海域达到每小时250公里的最大风速。飓风在途经背风群岛期间逐渐减弱，最终于11月23日在大西洋开放海域上空消散。行经小安的列斯群岛前，莱尼产生的大浪夺走了哥伦比亚北部两人的生命。狂风暴雨导致波多黎各东南部的农作物严重受损。暴雨还导致圣马丁岛法国领地的警署，有3人遇难，超过200套物业被毁。附近的安提瓜和巴布达也有1人丧生，该国西南部也被严重破坏，大浪导致多个城镇和首府受到影响，其名称“莱尼”因此从大西洋风暴命名名单上退役，以后将不再在北大西洋热带气旋命名时使用。第十世班禅额尔德尼是藏传佛教格鲁派第10世班禅额尔德尼活佛，在西藏宗教地位仅次于达赖喇嘛。1949年6月3日在时任国民政府代总统李宗仁批准后，于8月10日坐床典礼后正式继任。同年9月，解放军占领西宁后，十世班禅主动与中共联系，表示拥护中央政府。1951年，十世班禅、十四世达赖喇嘛与中央政府签订《关于和平解放西藏办法的协议》。1959年，爆发藏区骚乱后，十世班禅主动与中共联系，表示拥护中央政府。1962年，十世班禅因对中共在藏区实施民主改革中产生的问题的不满，向中央政府提出严正交涉。1964年，十世班禅被“最大的反动农奴主之一”并被剥夺了一切职务。1966年，十世班禅因“班禅集团”受到迫害入狱。直到1978年，在得到时任中华人民共和国最高领导人邓小平的允许后，十世班禅得以出狱。出狱后，十世班禅恢复职务，并在西藏实施了涉及藏族文化、宗教、语言、习俗的重建与改革，得到了许多藏族人的爱戴。明斯特是德国北莱茵-威斯特法伦州北部的城市，是威斯特法伦地区的文化中心。亦是明斯特行政区的首府。明斯特在德国宗教改革期间是明斯特反叛的发源地，此地也是1648年标志着三十年战争结束的《威斯特伐利亚和约》签署地。今天，该城以德国的自行车之都闻名。
Natural Language Processing

Computers process massive amounts of information every day in the form of human language. Although they do not understand it, they can learn how to do things like answer questions about it, or translate it into other languages. This course is a systematic introduction to the ideas that form the foundation of current language technologies and research into future language technologies.

The official prerequisite is CSE 20312. Students should be experienced with writing substantial programs in Python. The course also makes use of finite automata, context-free grammars, basic linear algebra, multivariable differential calculus, and probability theory. Ideally, students should have taken CSE 30151, Math 10560, and ACMS 30440, but please contact the instructor if you have questions about the necessary background.

Staff

The best way to contact the teaching staff is on Campuswire.