TECHNISCHE UNIVERSITÄT DRESDEN

SEMANTIC COMPUTING

Lecture 1: Welcome and Introduction

Dagmar Gromann

International Center For Computational Logic

TU Dresden, 19 October 2018

Welcome to Semantic Computing!

Organization

- **Lecture:** Friday, 9.20 – 10.50 am, ABP/E005
- **Tutorial:** Friday, 11.10 – 12.40 am, APB-E065
- **Lecture free periods:**
  Christmas: Saturday, 22 December 2018 to Sunday, 6 January 2019
- **Contact details:**
  E-mail: dagmar_gromann@tu-dresden.de
  Office: ABP 2034
  Office hour: individual – please contact me per e-mail
- **Exam:**
  Individual oral exams by end of February

Organization - Materials and Platforms

- **Web page with Materials:** `https://iccl.inf.tu-dresden.de/web/Semantic_Computing_(SS2018)`
- **Tutorial materials:** Google Colab Jupyter Notebooks at `https://github.com/dgromann/SemComp_WS2018` (also linked from the main web page)
- **Test yourself and discuss:** E-learning at OPAL `https://bildungsportal.sachsen.de/opal/dmz/`

Lectures

The **lectures** provide a theoretical introduction to central concepts of semantic computing

- definition of semantic computing and its interaction with other fields
- overview of core problems addressed
- understanding of most important tasks
- broad overview of employed core technologies

Tutorials

The **tutorials** focus on hands-on exercises of lecture contents, which means

- practical implementations of theoretical lecture contents
- autonomous work on provided tasks in small groups under supervision
- presentations of results (model solution will also be provided)
- implementation and presentation in-class

For this it is necessary to

- take notes during lecture
- ask questions during lecture if something is not entirely clear

Tutorials - Motivation

Why those hands-on exercises?

- active work on lecture materials is more effective
- develop a deeper understanding of theoretical concepts
- learn autonomous work
- practice team-working skills
- strengthen and broaden software development skills
- learn how to present and explain your work

Overview

Overview

Semantics

Philosophical and linguistic study of meaning

Semantic Computing

Semantic computing is a highly interdisciplinary field that addresses the core problem of understanding the meanings of computational contents (audio, video, text, software, etc.) and their representation in a machine-readable format. Within this lecture, we are only interested in natural language and mostly textual data.

Main tasks of Semantic Computing

Analysis

identify and represent meaning

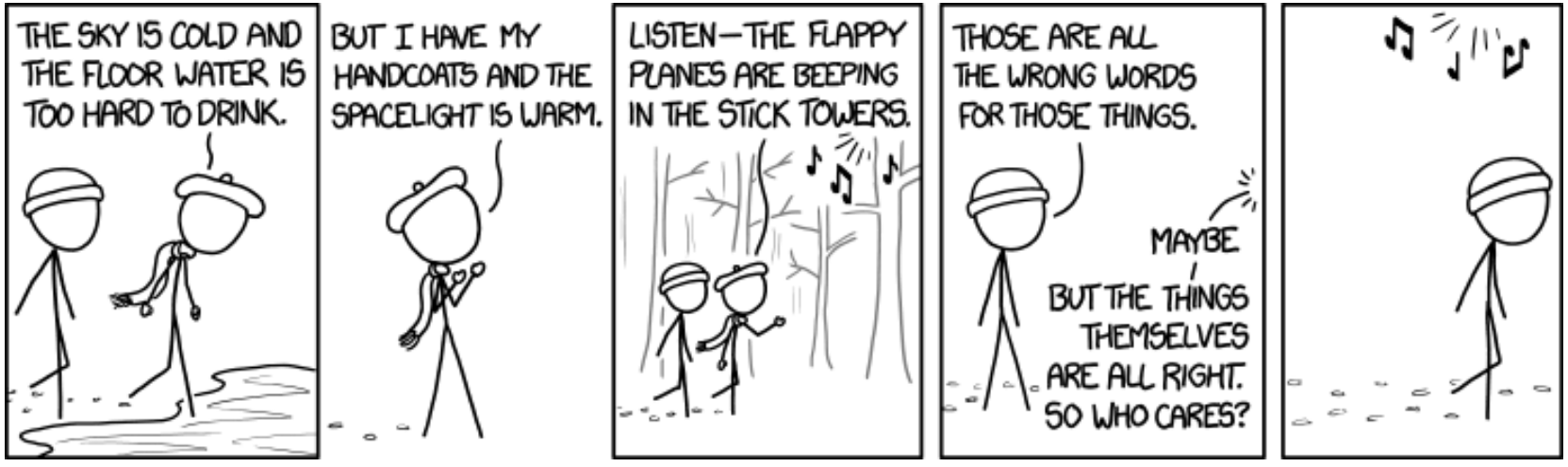
Integration

integrate semantics from various, heterogeneous sources

Application

Utilize identified, represented, and potentially integrated semantics to solve specific problems

What is meaning?



Source: https://xkcd.com/1322/

What is meaning?

This question has kept philosophers busy across centuries, so the following examples of theories of meaning are quite high-level.

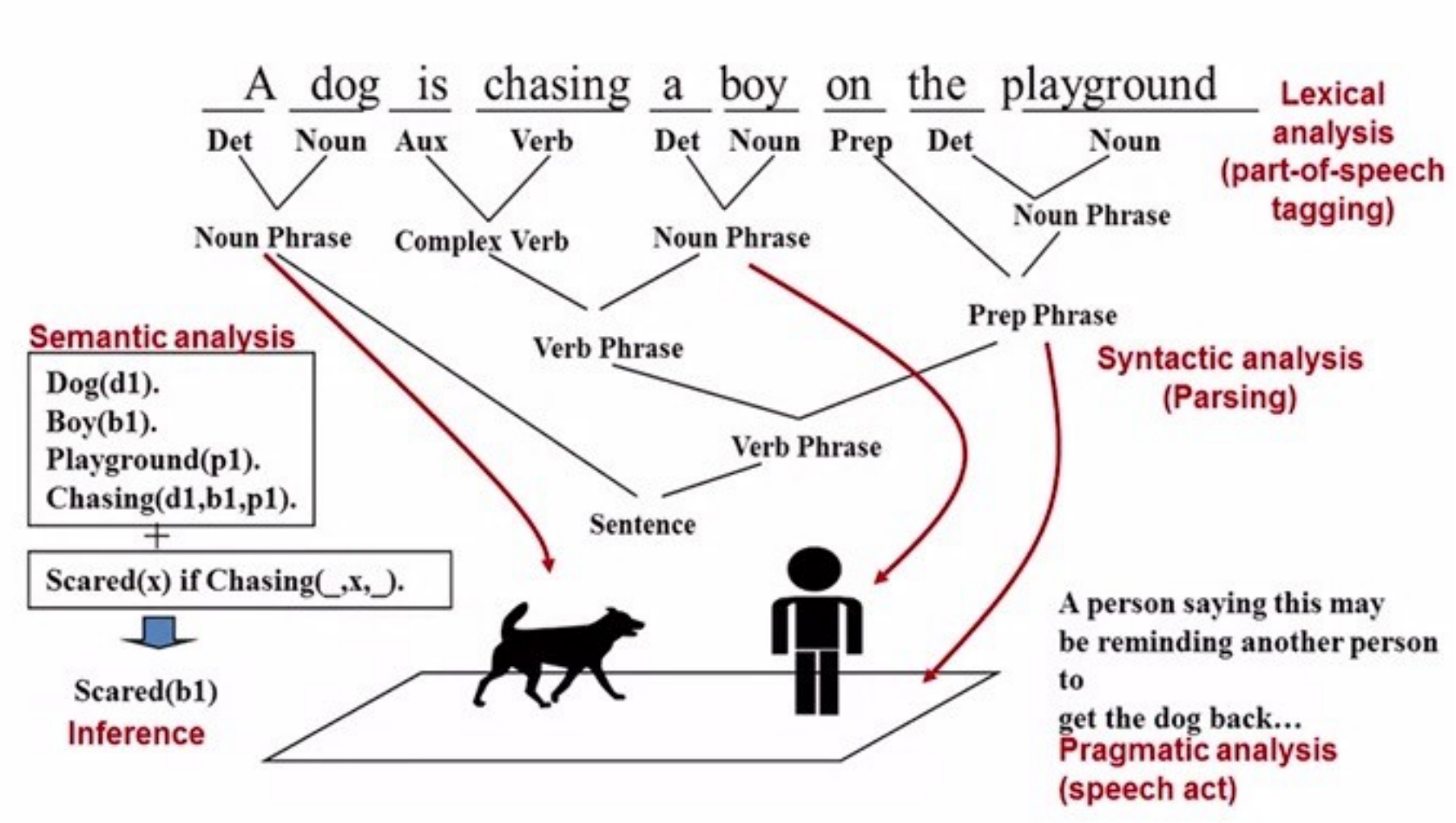
- **Referential Theory**: words obtain their meaning in reference to real-world objects, e.g. Saul Kripke or John Searle
- **Logical positivism**: only statements that can be empirically verified can be cognitively meaningful; reduce all knowledge to logical foundations, e.g. Rudolf Carnap
- **Meaning as Truth Condition**: a sentence can be reduced to being true or false, e.g. John Davidson
- **Linguistic Philosophy**: we should study language in use and in everyday situation to understand its meaning, e.g. Ludwig WIttgenstein or John L. Austin
- **many more**

Linguistic Meaning

- Lexical semantics studies the meaning of words...
  - by references to concepts and real-world objects (semiotics, knowledge representation, etc.)
  - by “the company they keep”, i.e., their context (distributed and distributional semantics, etc.)
- Grammatical semantics studies the meaning of...
  - syntactic categories, e.g. yellow hat vs. glowing yellow vs. yellowing leaves
  - morphemes, e.g. ending -ed of walked
- Formal semantics
  - formalizes meaning of natural language in a logical system (e.g. propositional or predicate calculi, etc.)
- many more

Source: Cruse, A. (2011). Meaning in language: An introduction to semantics and pragmatics. Oxford University Press.

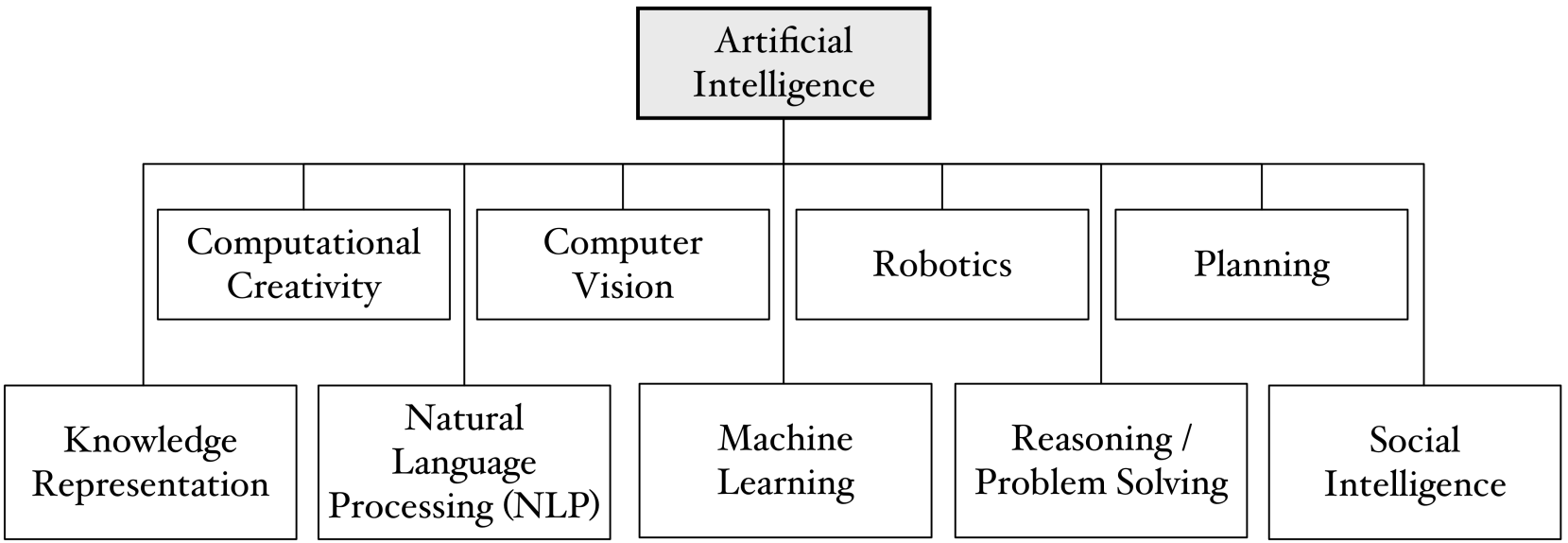
An example


A dog is chasing a boy on the playground
Det Noun Aux Verb Det Noun Prep Det Noun
Lexical analysis (part-of-speech tagging)
Noun Phrase
Complex Verb
Noun Phrase
Noun Phrase
Prep Phrase
Verb Phrase
Verb Phrase
Sentence
Syntactic analysis (Parsing)
Semantic analysis
Dog(d1).
Boy(b1).
Playground(p1).
Chasing(d1,b1,p1).
+
Scared(x) if Chasing(_,x,_).
Scared(b1)
Inference
A person saying this may be reminding another person to get the dog back…
Pragmatic analysis (speech act)

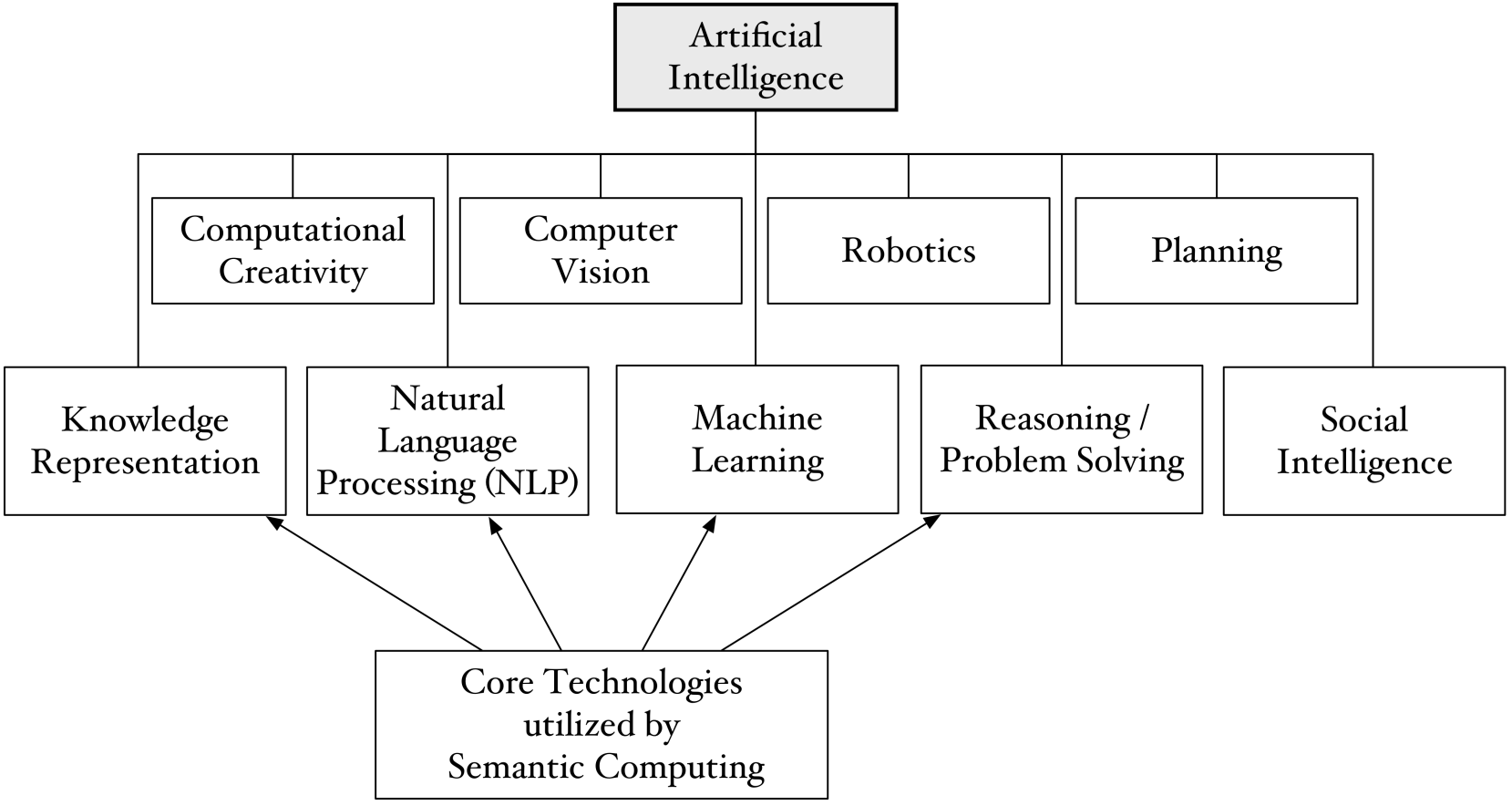
Core problems

- **Symbol grounding problem:** How symbols (words) obtain and derive their meaning and what meaning really is.
- **Natural Language Understanding:** Parsing human language to enable machine language comprehension (AI-complete)
- **Machine Learning:** Ability of machines to learn from data without being explicitly programmed
- **Further problems:** How to identify and extract information (information extraction), how to represent identified meaning, how to identify user intentions, and many more

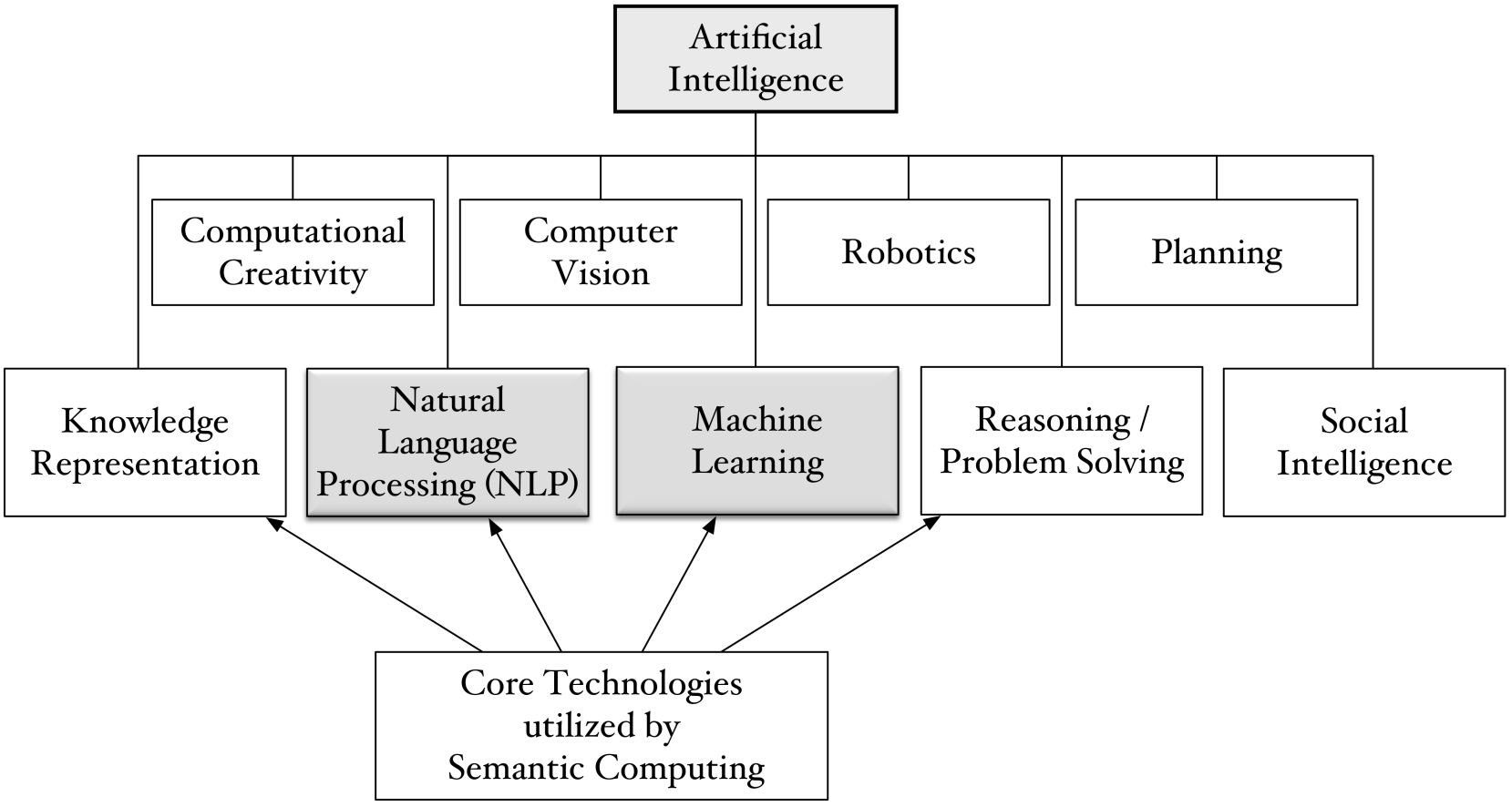
Positioning



Positioning



Main Focus of this Lecture



Showcases

Some showcases of semantic computing

- Virtual assistants, e.g. Google Duplex
- IBM's Watson playing Jeopardy
- AlphaGo winning Go competition

Overview of Lecture

Part 1: Linguistics and Natural Language Processing (NLP)

- Introduction to Linguistics
- Natural Language Processing

Part 2: Machine Learning

- Distributional and distributed semantics
- Paradigms: supervised, unsupervised, semi-supervised, reinforcement learning
- Applications: classification, regression, clustering
- Neural Networks: Feed-forward, RNN, LSTM

Part 3: Meaning and Knowledge Representation

- Ontology learning from text
- Knowledge graphs (link to lecture)

What I hope to teach?

- A basic understanding of effective methods used in semantic computing to analyze, understand, represent natural language
- The bigger picture of natural language processing and linguistic methods including machine learning approaches
- A basic understanding of core contributions of semantic computing to other fields, such as ontology learning (from text)

Introduction to Linguistics

Linguistics

Linguistics

is the scientific study of language and its structure, including the study of grammar, syntax, phonetics, and meaning.

What is a natural language?

A discrete/categorical system of symbols that combine to convey meaning. It has evolved naturally and historically in humans through use and repetition without planning. A natural language is different from a constructed, formal language, such as a programming language (C, Python, Java, etc.), arithmetic language, or a language used to study logic.

Branches of Linguistics

Basic distinction between theoretical and applied linguistics.

Theoretical Linguistics

- Generative linguistics: grammar as a mental system that generates sentences
- Cognitive linguistics: connection of language and cognition
- Historical linguistics: change of language over time
- Etymology: word histories and origins
- Sociolinguistics: effect of society on language
- ...

Branches of Linguistics

Basic distinction between theoretical and applied linguistics.

Applied Linguistics

- Computational linguistics: applied computing to language analysis, translation, processing
- Psycholinguistics: psychological and neurobiological factors that enable language acquisition, comprehension, use
- Neurolinguistics: connection of neurology and language
- Forensic linguistics: linguistic studies in law and justice
- ...

Branches of Linguistics

Branches by linguistic structure studied	
Phonetics	study of physical aspects in speech production and comprehension
Phonology	study of systematic organization of sounds in a language
Morphology	study of the structure of parts of words (e.g. suffix, prefix, stem)
Syntax	how words combine to form grammatical structures
Semantics	how longer sequences (text, corpora) are composed in a language
Pragmatics	how the context of an utterance influences its meaning
Discourse analysis	the meaning of words, phrases, and their combinations in sentences
Linguistic typology	study of similarities and differences between languages

Branches of Linguistics

Branches by linguistic structure studied	
Phonetics	study of physical aspects in speech production and comprehension
Phonology	study of systematic organization of sounds in a language
Morphology	study of the structure of parts of words (e.g. suffix, prefix, stem)
Syntax	how words combine to form grammatical structures
Semantics	how longer sequences (text, corpora) are composed in a language
Pragmatics	how the context of an utterance influences its meaning
Discourse analysis	the meaning of words, phrases, and their combinations in sentences
Linguistic typology	study of similarities and differences between languages

Syntax

- Traditional grammarians:
  - Study rules and regularities in languages by analyzing parts of speech (verb, noun, adjective, etc.)
  - Formation of sentences is like putting beads on a string
- Chomskian turn (Noam Chomsky[1])
  - Linguistic competence: ability of a speaker to determine whether a sentence is correct and true or false (idealized capacity)
  - Linguistic performance: how we use language given restrictions, such as memory limitations or distractions

Which words are possible in English?

symbed, coured, amiable, fistry, maudlin, dehappy

[1] Chomsky, N. (1965). Aspects of the Theory of Syntax (Vol. 11). MIT press

Syntax

- Traditional grammarians:
  - Study rules and regularities in languages by analyzing parts of speech (verb, noun, adjective, etc.)
  - Formation of sentences is like putting beads on a string
- Chomskian turn (Noam Chomsky[1])
  - Linguistic competence: ability of a speaker to determine whether a sentence is correct and true or false (idealized capacity)
  - Linguistic performance: how we use language given restrictions, such as memory limitations or distractions

Which words are possible in English?

symbed, coured, **amiable**, fistry, **maudlin**, dehappy

[1] Chomsky, N. (1965). Aspects of the Theory of Syntax (Vol. 11). MIT press

Competence vs. Performance

The Far Side by Gary Larson

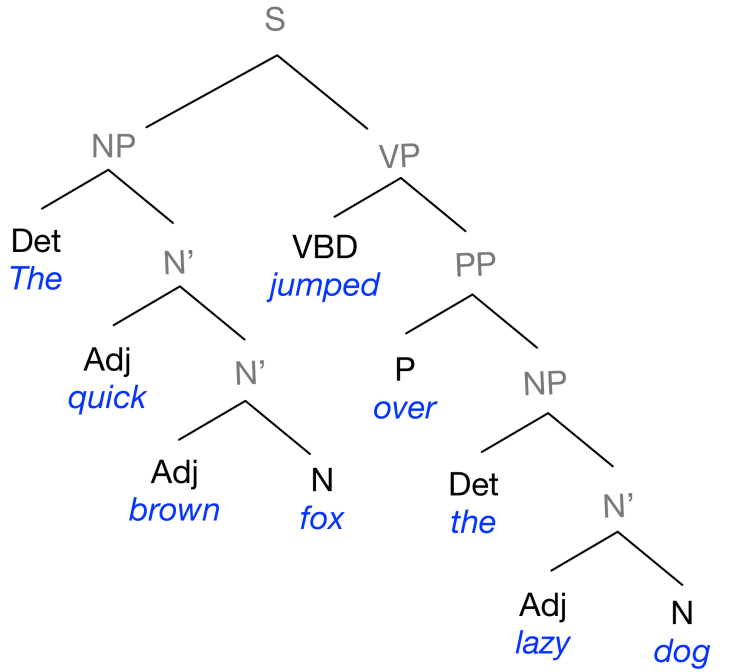
Syntactic structures

Sentences are composed by subparts called constituents, each of which belongs to a particular syntactic category. Trees can be used to represent constituency, syntactic categories, and word order.

Syntactic Categories

Noun Phrase (NP)
Verb Phrase (VB)
Prepositional Phrase (PP)
Adverbial Phrases (AdvP)
Determiner (D)
Auxiliary (A): have, may
Conjunction (Conj): and
Noun (N), Verb (V), Preposition (P), Adjective (Adj), Adverb (Adv)

Phrase Structure Tree



Types of Grammars

Phrase-structure grammar is only one way of representing syntactic relations. There are many other types of grammars studied by linguists, such as

- **Generative Grammar:** attempts to specify rules that allow for the production of all and only grammatical sentences in a language
- **Dependency grammar:** based on dependency relations as opposed to the constituency relations (more later)
- **Construction Grammar:** group of grammars that extend into semantics and pragmatics to achieve a pairing of form and function; language evolution (Link to example)
- ...

Semantics

The meaning of a sentence depends on the meanings of its part and on how they are combined (Principle of Compositionality, Gottlob Frege). Semantics is concerned with the literal meaning of words and sentences.

How to determine the meaning of a sentence?

1. The capital of Australia is Sydney/Canberra. (truth value)
2. The dog is on the rug. (truth values can change over time)
3. The person in the first row on the left plays golf. (we can understand a sentence even if we don't know its truth value)
4. Colorless green ideas sleep furiously. (we can determine nonsensical sentences)
5. She is a widow. (lexical entailment: the truth value of one expression follows from another, e.g. from this example we can entail "She must have been married.")

Word meaning

The meaning of a word can also be determined by its components, which are:

- Phoneme: smallest unit (sound) of language: /d/, /u:/
- Morpheme: smallest meaningful unit of language that cannot be further broken down: de-, car, -tion
- Lexeme: basic unit of meaning that persists across inflections (all headwords of a dictionary), e.g. “bring”, “bringing”, “brought” are forms of the same lexeme

Example

Phonetic: lead, progress, record
Morphological: un – fortunate – ly
Lexical: lead(1) = type of metal; lead(2) = guide in front

Selected Types of word meaning

- **Encyclopaedic meaning**: words provide access to a large inventory of structured knowledge (world knowledge)
- **Denotational meaning**: reference of a word to object/concept or its “dictionary definition” (signifier <-> signified)
- **Connotative meaning**: word meaning is understood by its cultural or emotional association (positive, negative, neutral conntation; e.g. “She’s a dragon” in Chinese and English)
- **Conceptual meaning**: word meaning is associated with the mental concepts it gives access to (e.g. prototype theory)
- **Distributional meaning**: “You shall know a word by the company it keeps” (J.R. Firth 1957: 11)[1]

[1] John Rupert Firth (1957). "A synopsis of linguistic theory 1930-1955." In Special Volume of the Philological Society. Oxford: Oxford University Press.

Ambiguity

Ambiguity: more than one interpretation is plausible

- **Lexical ambiguity** (polysemy): one word of a specific syntactic category can have several meanings, which is in this context called a lexical sense (e.g. “bank”: 1) financial institution, 2) ground bordering a lake or river, etc.)
- **Homonymy**: different words that are spelled and pronounced the same way (e.g. a “book” vs. to “book”)
- **Structural ambiguity**: one sentence can be interpreted in several ways (e.g. “The Pope’s baby steps on gays.”)

Which type of ambiguity?

“Each of us saw her duck.”

Ambiguity

Ambiguity: more than one interpretation is plausible

- **Lexical ambiguity** (polysemy): one word can have several meanings, which is in this context called a lexical sense (e.g. “bank” can be the financial institution or the ground bordering a lake or river)
- **Homonymy**: different words that are spelled and pronounced the same way (e.g. a “book” vs. to “book” a flight)
- **Structural ambiguity**: one sentence can be interpreted in several ways (e.g. “The Pope’s baby steps on gays.”)

Which type of ambiguity?

“Each of us saw her **duck.**”
(Homonymy: a) action of ducking, b) a bird))

Computational Example: Lexical Ambiguity

Traditional approach: build a taxonomy, such as WordNet, that provides a denotational definition of each word and represents their interrelations in a hierarchical structure[1].

```
from nltk.corpus import wordnet as wn
wn.synsets('bank')
```

Synset('bank.n.01'),
Synset('bank.n.02'),
Synset('bank.n.03'), ...

```
wn.synset('bank.n.02').definition()
```

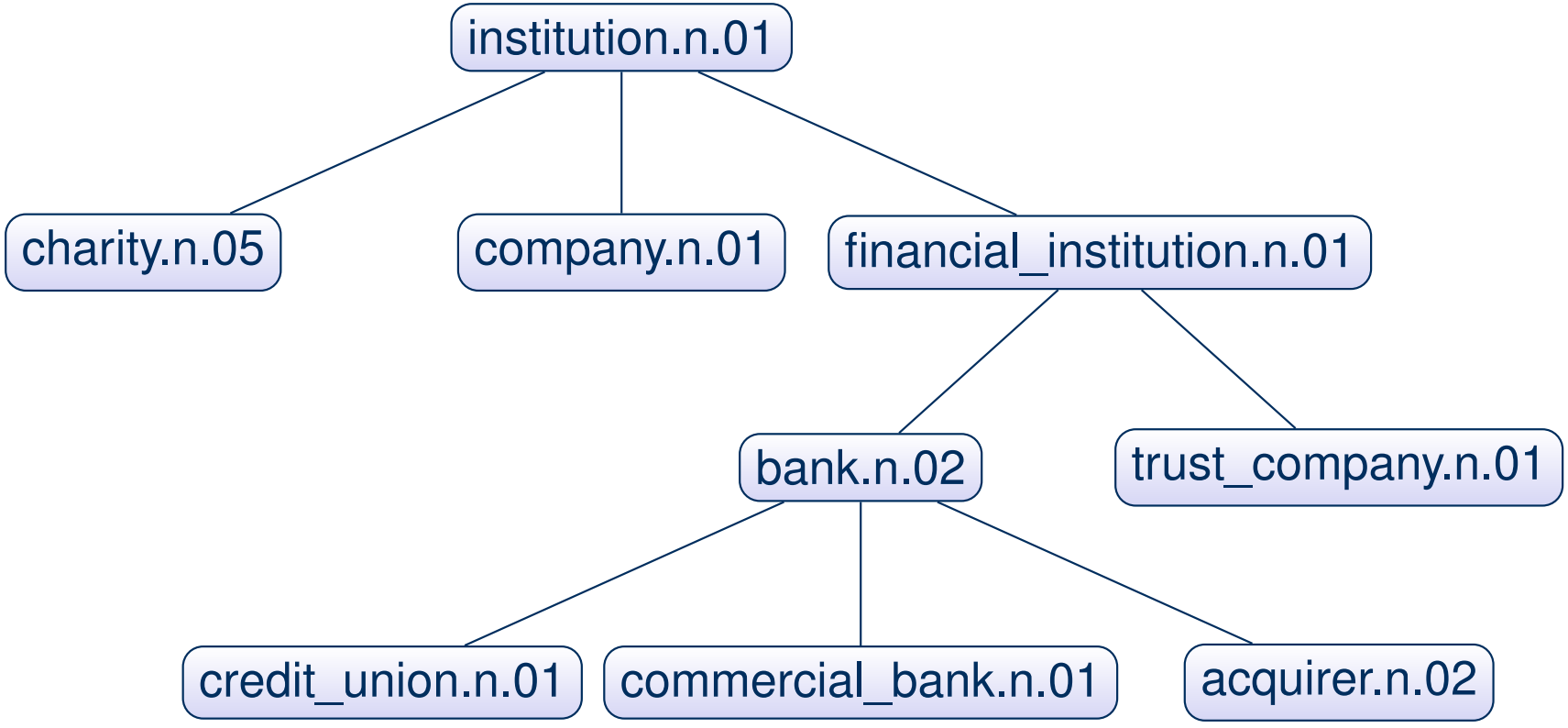
"a financial institution that ..."

```
bank = wn.synset('bank.n.02')
hypernyms = lambda s : s.hypernyms()
list(bank.closure(hypernyms))
```

Synset('financial_institution.n.01'),
Synset('organization.n.01'),
Synset('social_group.n.01'),
Synset('abstraction.n.06')
Synset('entity.n.01'),
...

[1] synset - a set of one or more synonyms, see http://wordnetweb.princeton.edu/perl/webwn

WordNet Hierarchy



Lexical-Semantic Relations

- Synonymy: (almost) identicial meaning (e.g. mud, sludge)
- Hyponymy: meaning inclusion (type-of, is-a, e.g. seagul is-a bird)
- Hypernymy: inverse hyponymy (hypernym: bird; hyponym: seagul)
- Meronymy: part-whole relations (e.g. car, tyre)
- Antonymy: opposite in meaning (e.g. big, small)
- ...

Tasks: Word Sense Disambiguation (WSD) with WordNet

Given a word and its context we want to automatically determine which of the WordNet senses is the context-appropriate.

"I need to go to the **bank** to deposit money."
Synset('bank.n.01') sloping land (especially beside a body of water)
Synset('depository_financial_institution.n.01') a financial institution that accepts deposits and channels the money into lending activities
Synset('bank.n.03') a long ridge or pile
....
Synset('bank.v.01') tip laterally
Synset('bank.v.05') be in the banking business
...

A Simple Lesk Algorithm for WSD

```
from nltk.corpus import wordnet as wn

def lesk(sentence, ambiguous_word):
        max_overlaps = 0
        lesk_dictionary = []
        lesk_sense = ""
        context = sentence.split()

        for sense in wn.synsets(ambiguous_word):
                lesk_dictionary += sense.definition().split()
                lesk_dictionary += sense.lemma_names()

                overlaps = set(lesk_dictionary).intersection(context)

                if len(overlaps) >  max_overlaps:
                        lesk_sense = sense
                        max_overlaps = len(overlaps)

                return lesk_sense

sentence = 'I need to go to the bank to deposit money'
ambiguous_word = 'bank'
answer = lesk(sentence, bank)
```

Synset('depository_financial_institution.n.01')

Tasks: WordNet-Based Word Similarity

Given two words we want to determine how similiar they are[1].
Let's calculate the similarity of bank and financial institution

Path

$*ROOT* * .n.01 < entity.n.01 < abstraction.n.06 < group.n.01 < social\_group.n.01 <$
$organization.n.01 < institution.n.01 < financial\_institution.n.01 < bank.n.02$

- **Path-based**: $score = 1/path\_length = 0.5$ (# nodes)
- **LESK**: counting overlaps between definitions and related definitions; score = 746 (word count)
- Wu and Palmer **WUP**: Least Common Subsumer (LCS); $wup(s1, s2) = \frac{2*depth(LCS(s1,s2)}{depth(s1)+depth(s2)}$ which is $(2 * 8)/(9 + 8) = 0.9412$

[1] See `http://ws4jdemo.appspot.com` for an overview.

Problem with Lexical Resource Representation

- connotations of synonyms are missing (e.g. interested, questioning, nosy)
- difficult to keep up to date with new words introduced
- certain bias of created
- expensive to generate since it requires a lot of manual labor

Pragmatics

Pragmatics focuses on the meaning of language in interactional context, that is, language in use and conversational implicature (speaker implies and listener infers - negotiation between speaker and listener). One of the most important founders: Paul Grice.

- deixis (e.g. “here” cannot be understood without context),
- turn taking in a conversation (interruption is power),
- presupposition (implicit assumption about world knowledge),
- many more

A: “Do you want to go for dinner tomorrow?”
B: “Tomorrow is Tuesday.”
(Explanation: B plays tennis every Tuesday and presumes that A knows this. Only with this additional information does the reply make sense).

Grice's Meaning

- **literal meaning**: this is the type of meaning semantics is concerned with more than pragmatics, it is the semantic information that you have from your knowledge of English
- **explicature**: basic interpretation of an utterance, using contextual information and world knowledge; they go beyond the literal meaning; e.g. "This was the last bus." (scheduled that day? the previous bus would have been the right one to go to Monterey?)
- **implicature**: act of meaning or implying one thing by saying something else; used in metaphors, irony, to mislead without lying, verbal efficiency, or to maintain good social relations (example from previous slide: B implies to not have time for dinner the next day - however, what s/he means is different from what s/he says)

Performative Utterances

Within the theory of speech acts, John L. Austin defined performative utterances, which is statements that change the social reality they are describing. This is just one example of pragmatic theories.

Uttering a performative is, or is in part, doing an action.

I hereby christen this ship Argo.
I now pronounce you husband and wife.
You are under arrest.
I accept your apology.

Review of Lecture 1

- What is Semantic Computing? Which problems does it address?
- What is an example of a theory of meaning? What is an example of linguistic meaning?
- What is Linguistics?
- Which subfields of linguistics do you know? By structure?
- What is linguistic competence?
- What is the difference between lexical and structural ambiguity?
- Which types of lexical-semantic relations do you know?
- What is implicature?