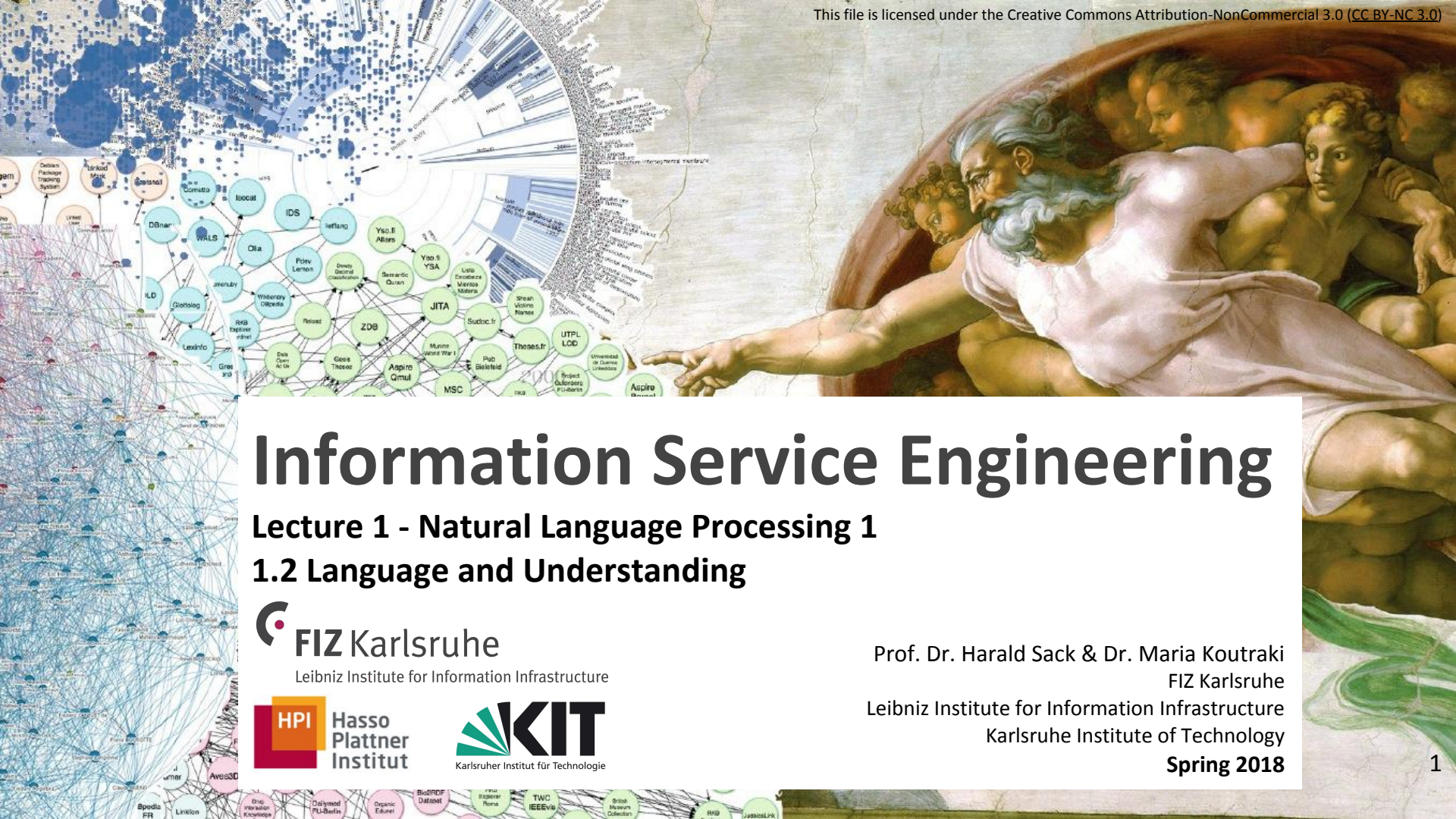




Information Service Engineering

Lecture 1 - Natural Language Processing 1

1.2 Language and Understanding



Leibniz Institute for Information Infrastructure



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

Leibniz Institute for Information Infrastructure
Karlsruhe Institute of Technology
Spring 2018

Information Service Engineering

Lecture 1: Natural Language Processing

1.1 Information and Knowledge

1.2 Language and Understanding

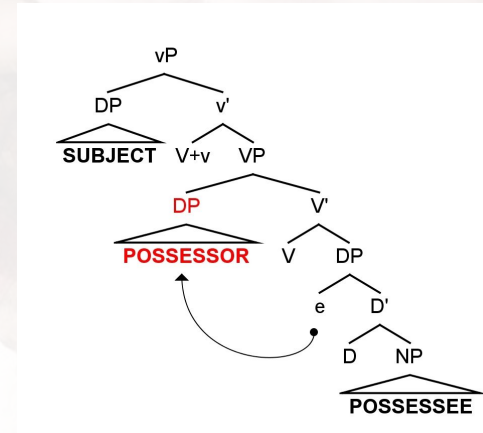
1.3 NLP Techniques and Challenges

1.4 Linguistic Basics

1.5 Morphology

1.6 Evaluation, Precision and Recall

1.7 Regular Expressions



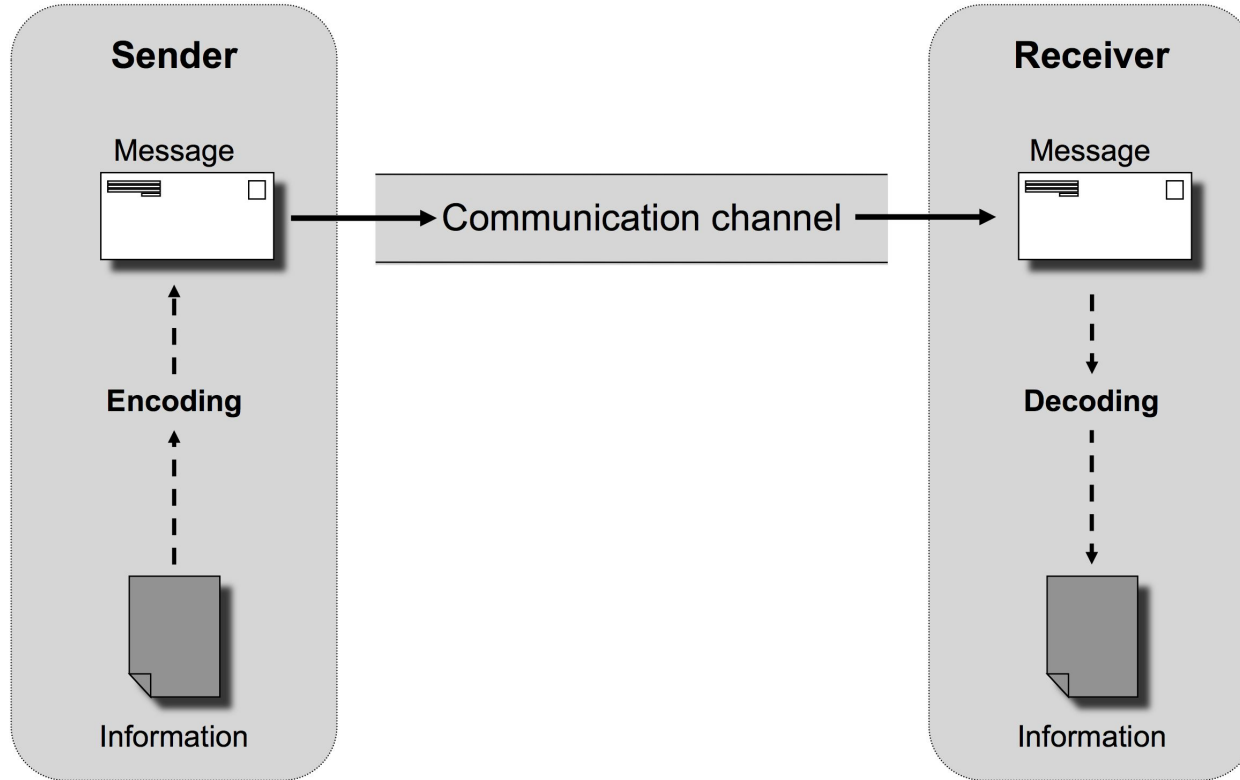
What is Communication?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- **Communication** is a process by which **information** is **exchanged** between individuals through a **common system of symbols, signs, or behavior**.
[Merriam-Webster]
- **Communication** is the act of **conveying intended meanings** from one entity or group to another through the use of **mutually understood signs and semiotic rules**. [wikipedia]

The Communication Process

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding



Ch. Meinel, H. Sack:
Digital Communication -
Communication Multimedia,
Security, Springer, 2014.

What is Language?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

Language, a system of conventional **spoken, manual, or written symbols** by means of which human beings, as members of a social group and participants in its culture, **express** themselves.

The functions of language include **communication**, the expression of identity, play, imaginative expression, and emotional release.

[Encyclopaedia Britannica]

Natural Language

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

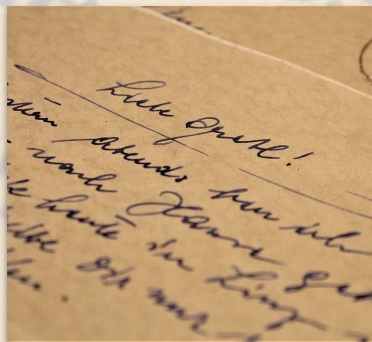
How do we communicate Natural Language?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding



Speech

- Listening
- Speaking

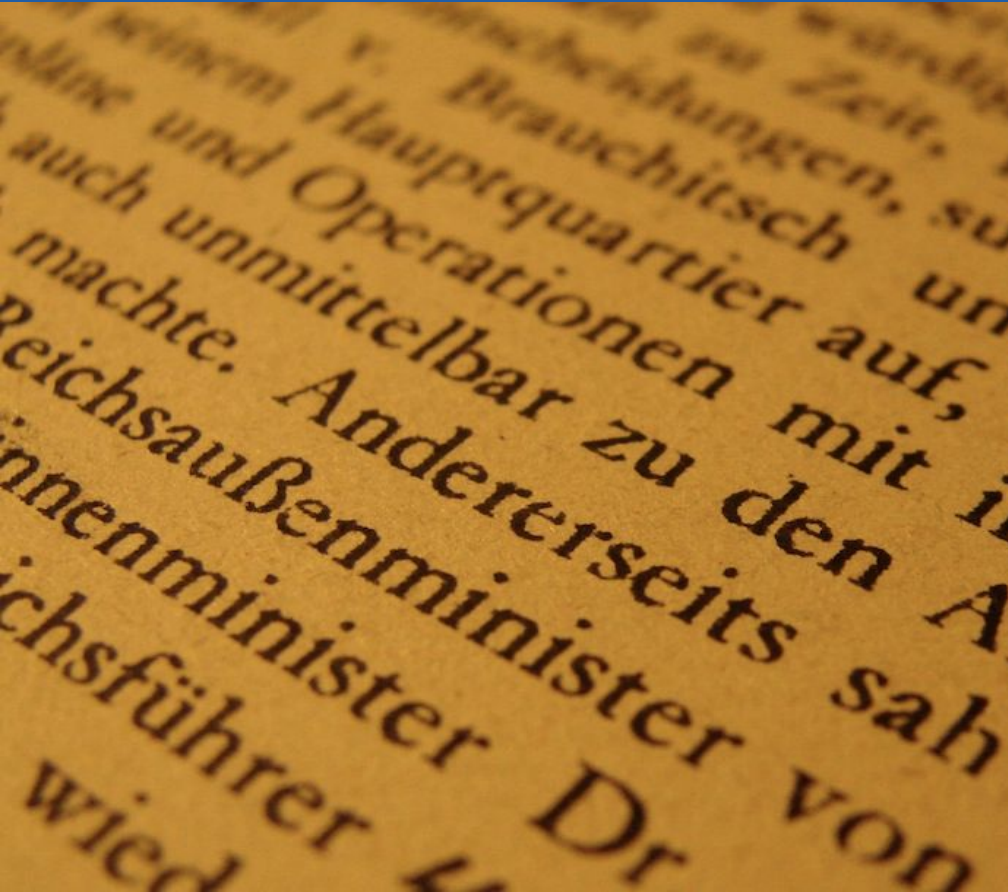


Text

- Reading
- Writing

How do we Encode Textual Information?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

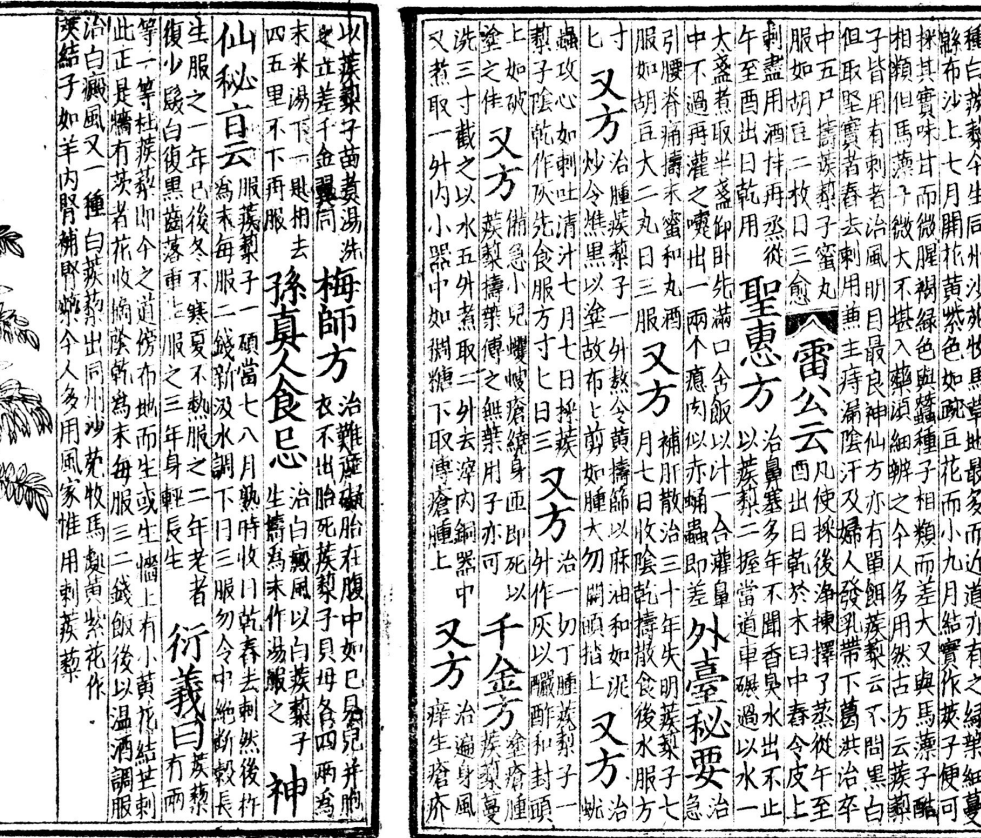


Alphabetic Writing Systems

- Single character refers to a single sound (phonemic alphabet)
- Consonant alphabets (abjads), vowels can be deduced from context

How do we Encode Textual Information?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding



Syllabic and Logographic Writing System

- Syllabic systems also involve a mapping between characters and sound, (but refer on larger units)
- A logograph is a symbol that represents a unit of meaning

Artificial Language

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

```
public static void main(String[] args) throws FileNotFoundException {

    String filePath = "/Users/hsa/Documents/Workspace/Scientometrics2016/testdata_input";
    File fout = new File("/Users/hsa/Documents/Workspace/Scientometrics2016/SemanticSim_testdata_10000");
    FileOutputStream fos = new FileOutputStream(fout);

    try {
        BufferedReader lineReader = new BufferedReader(new FileReader(filePath));
        BufferedWriter bw = new BufferedWriter(new OutputStreamWriter(fos));
        String lineText = null;
        int nr=0;

        while ((lineText = lineReader.readLine()) != null) {

            String[] splitStr = lineText.split("\\s+");
            nr++;

            // DBpedia settings - without giving prefix list, using property patterns instead
            // it will filter properties which start with "http://dbpedia.org/ontology"
            ResourceSimilarityMeasure rsmForDBpedia = new ResourceSimilarityMeasure(
                PropertyRestriction.SamePropertyPath,
                "http://dbpedia.org/sparql",
                null,
                "http://dbpedia.org/ontology/",
                null, null, null, null);

            System.out.println(
                rsmForDBpedia.getSimilarity("<http://dbpedia.org/resource/Albert_Einstein>", "<http://dbpedia.org/re
```

What is Information?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- **Information** is that which informs.
- **Information** is conveyed as the **content of a message**.
- **Information** can be **encoded** into various forms for **transmission** and **interpretation**
- **Information** is any **propagation of cause and effect** within a system.
- **Information's existence** is not necessarily coupled to an **observer**.
- **Information reduces uncertainty (Information Theory)**

What Does it Mean to Understand?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- **Understanding** is the ability to grasp the **meaning of information**.

What Does it Mean to Understand?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- **Understanding** is the ability to grasp the **meaning of information**.



- Information is conveyed in a **message** using a **specific language**.

What Does it Mean to Understand?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- **Understanding** is the ability to grasp the **meaning of information**.



- Information is conveyed in a **message** using a **specific language**.



- **Information is understood** by the receiver of a message, **if the receiver interprets the information correctly**.

What Does it Mean to Understand?

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- **Correct Interpretation** depends on
 - Syntax,
 - Semantics,
 - Context,
 - Pragmatics, and
 - Experience

Syntax

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- =[greek] Arrangement, Ordering
- In **grammatics** syntax denotes the study of the principles and processes by which sentences are constructed in particular languages.
- In **formal languages**, syntax is just a set of rules, by which well formed expressions can be created from a fundamental set of symbols (alphabet).
- In **computer science**, syntax defines the normative structure of data.

Semantics

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- =[greek] pertains to the character, the study of meaning
- is **part of the linguistics** focussed on **Sense and Meaning** of language or symbols of language.
- is the **study of interpretation of signs or symbols** as used by agents or communities within **particular circumstances** and **contexts**.
- Semantics asks, how sense and meaning of complex concepts can be derived from simple concepts based on the **rules of syntax**.
- The semantics of a message depends on **context** and **pragmatics**.

Context

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- [lat.] contextus = interweaved
- denotes the **surrounding of a symbol** in an expression respective its **relationship with surrounding expressions** and further related elements,
- Contexts denotes **all elements of any sort of communication that define the interpretation of the communicated content.**

Pragmatics

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- =[greek] action
- reflects the **intention by which the language is used** to communicate a message.
- In **linguistics** pragmatics denotes the **study of applying language in different situations**.
- It also **denotes the intended purpose** of the speaker.
- Pragmatics studies **the ways in which context contributes to meaning**.

Experience

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- **Experience** considers all information that you have learned and put in context with the world you are living in.
- Experience in this sense also often is referred to as **common sense knowledge** or **world knowledge**.

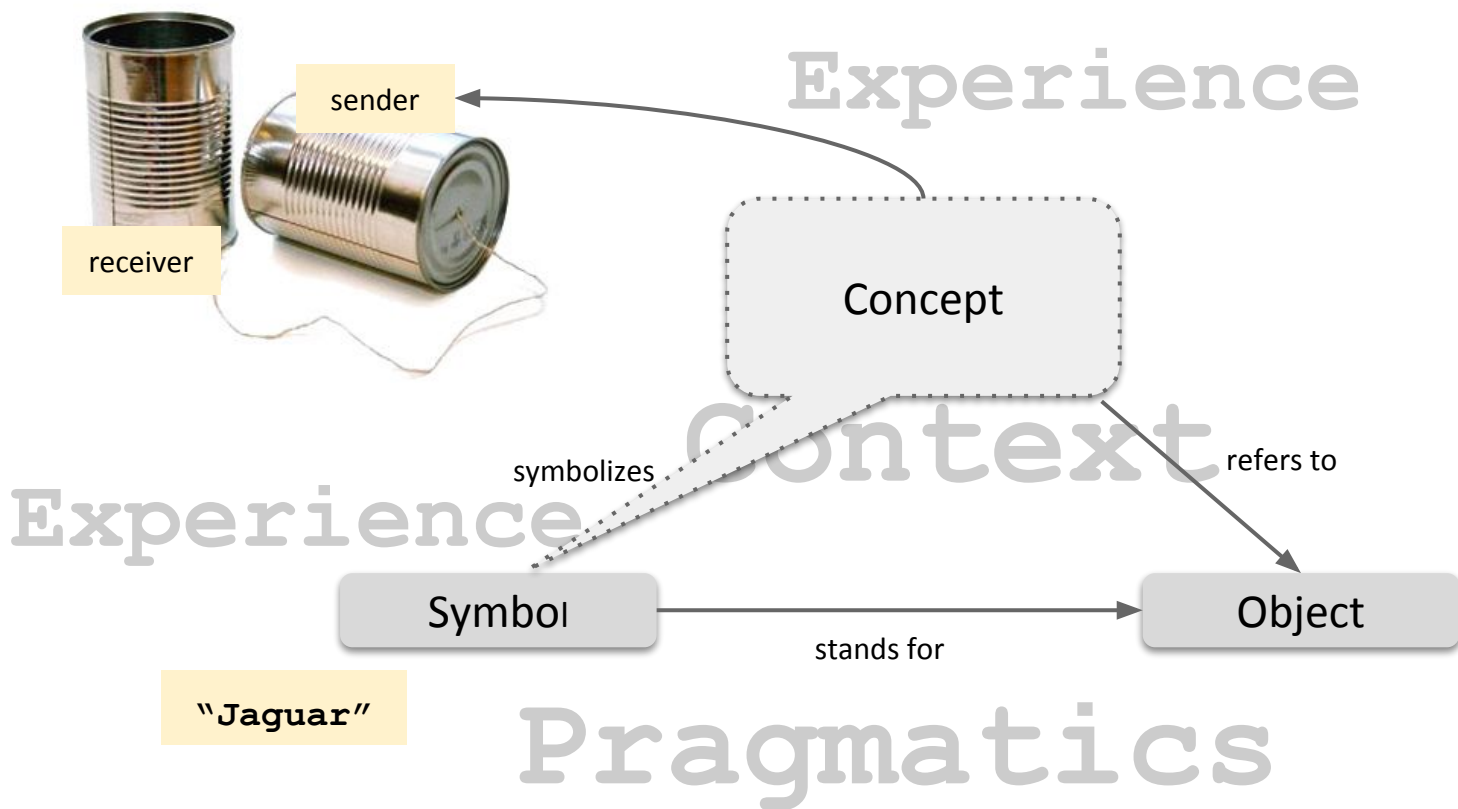
Successful Communication

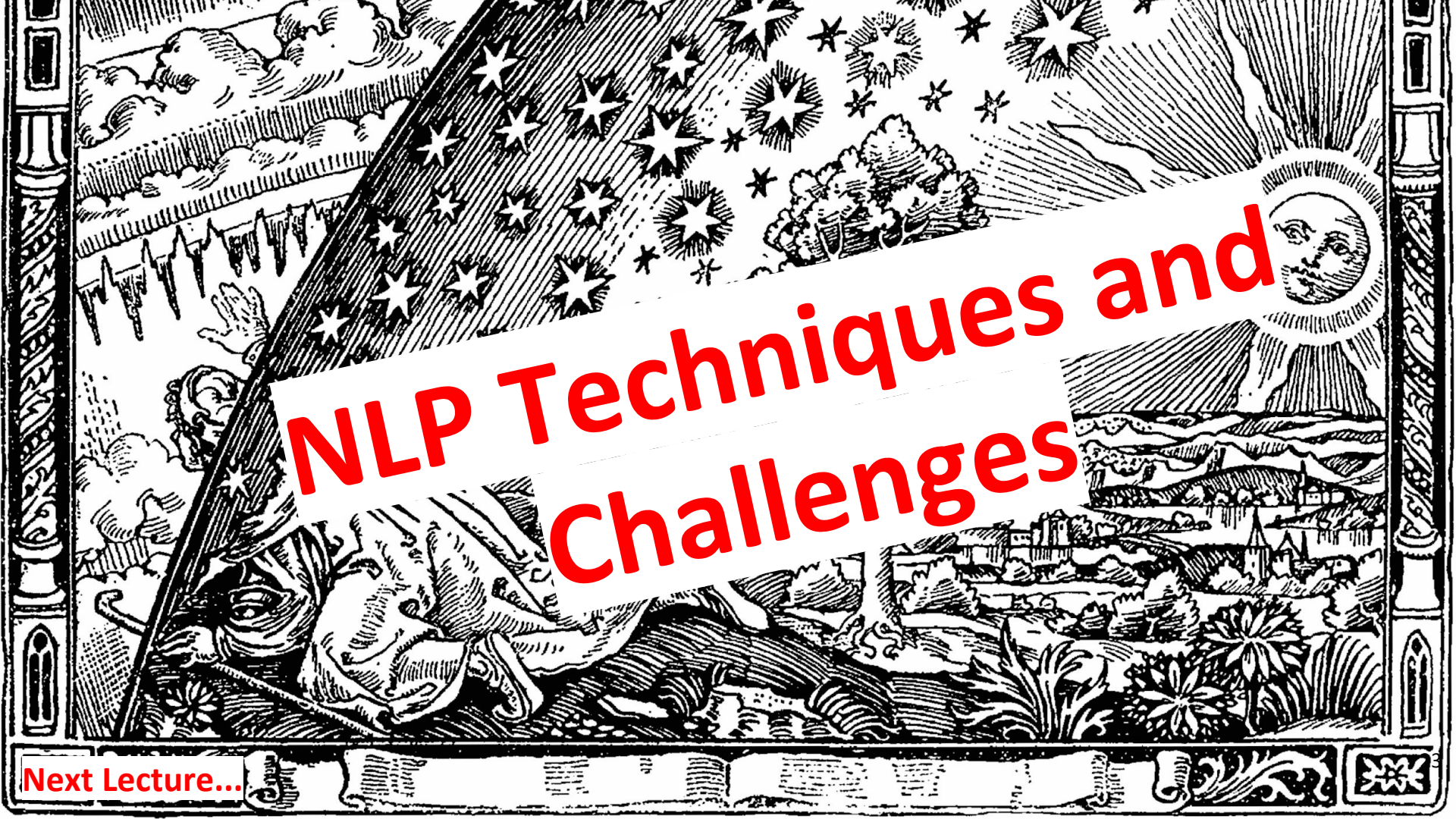
Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding

- For **successful communication**,
 - information has to be correctly transmitted (**Syntax**)
 - the meaning (**Semantics**) of the transmitted information must be interpreted correctly (= **understanding**)
- **Understanding** depends on
 - the **context** of both sender and receiver and
 - the **pragmatics** of the sender
- (Personal) **experience** determines
 - how sender and receiver **interpret the semantics, context, and pragmatics** of a message, and thus its intended meaning

Successful Communication

Information Service Engineering - Lecture 1 Natural Language Processing - 1.2 Language and Understanding



A detailed woodcut-style illustration of a landscape. In the upper right, a sun with a human-like face and radiating rays shines over a town. The sky is filled with numerous stars of varying sizes. In the lower left, a figure is shown in a dynamic, possibly dancing or falling, pose. The entire scene is framed by an ornate border. A large, white, tilted banner with red text is superimposed over the center of the image.

NLP Techniques and Challenges

Next Lecture...

Picture References:

- P.6: <http://expertenough.com/2392/german-language-hacks>
- P.7: <http://expertenough.com/2392/german-language-hacks>
Real-time MRI - Speaking (English), [https://en.wikipedia.org/wiki/File:Real-time_MRI_-_Speaking_\(English\).ogv](https://en.wikipedia.org/wiki/File:Real-time_MRI_-_Speaking_(English).ogv)
<https://pixabay.com/en/letters-old-antique-handwriting-2794215/>
- P.9: https://en.wikipedia.org/wiki/Hua_Sui#/media/File:Pen_ts%27ao,_woodblock_book_1249-ce.png
- P.20: Jaguar Car, <https://pixabay.com/en/car-jaguar-vehicle-automobile-161046/>
A leopard sitting on the ground. Engraving by M De Bye after P Potter.
https://commons.wikimedia.org/wiki/File:A_leopard_sitting_on_the_ground._Engraving_by_M_De_Bye_after_Wellcome_V0021498.jpg
Default Jaguar Desktop, <https://www.flickr.com/photos/splorp/75563891/>
- P.21: Camille Flammarion, L'atmosphère: météorologie populaire (1888), <https://commons.wikimedia.org/wiki/File:FlammarionWoodcut.jpg>