A very short introduction to Language Technologies and Natural Language Processing

Jose F Quesada & Jose Luis Pro

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Language

Set of conventional spoken or written symbols used for commucation between entities.

So we can see a language as the linking between meaning (semantic side) and expression (syntantic side).

Types of languages

- **Natural languages:** Used for the communication between human beings.

- **Formal languages:** Used by computers and in mathematical areas.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Language

Set of conventional spoken or written symbols used for commucation between entities.
So we can see a language as the linking between meaning (semantic side) and expression (syntantic side).

Types of languages

- **Natural languages:** Used for the communication between human beings.

- **Formal languages:** Used by computers and in mathematical areas.

Language

Set of conventional spoken or written symbols used for commucation between entities.
So we can see a language as the linking between meaning (semantic side) and expression (syntantic side).

Types of languages

- **Natural languages:** Used for the communication between human beings.
- **Formal languages:** Used by computers and in mathematical areas.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Differences between natural and formal languages:

1. Computers don't understand natural languages, (normal) humans don't understand computer languages.

2. Formal languages shouldn't have ambiguities, but natural languages do have.

Language Technologies (LT's) are the set of technologies that aim to create software that has some kind of knowledge about natural languages.

Natural language processing (NLP) is the scientific field concerned with the interactions between computers and human by means of natural languages.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Differences between natural and formal languages:

1. Computers don't understand natural languages, (normal) humans don't understand computer languages.

2. Formal languages shouldn't have ambiguities, but natural languages do have.

Language Technologies (LT's) are the set of technologies that aim to create software that has some kind of knowledge about natural languages.

Natural language processing (NLP) is the scientific field concerned with the interactions between computers and human by means of natural languages.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Differences between natural and formal languages:

1. Computers don't understand natural languages, (normal) humans don't understand computer languages.
2. Formal languages shouldn't have ambiguities, but natural languages do have.

Language Technologies (LT's) are the set of technologies that aim to create software that has some kind of knowledge about natural languages.

Natural language processing (NLP) is the scientific field concerned with the interactions between computers and human by means of natural languages.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Differences between natural and formal languages:

1. Computers don't understand natural languages, (normal) humans don't understand computer languages.
2. Formal languages shouldn't have ambiguities, but natural languages do have.

Language Technologies (LT's) are the set of technologies that aim to create software that has some kind of knowledge about natural languages.

Natural language processing (NLP) is the scientific field concerned with the interactions between computers and human by means of natural languages.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Applications of Language Technologies:

1. Machine Translation.

2. Question - Answering.

3. Automatic Text Classification.

4. Automatic Text Summarization.

5. Social Analytics.

6. Sentiment Analysis.

7. Dialogue Systems.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Applications of Language Technologies:

1. Machine Translation.

2. Question - Answering.

3. Automatic Text Classification.

4. Automatic Text Summarization.

5. Social Analytics.

6. Sentiment Analysis.

7. Dialogue Systems.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Applications of Language Technologies:

1. Machine Translation.

2. Question - Answering.

3. Automatic Text Classification.

4. Automatic Text Summarization.

5. Social Analytics.

6. Sentiment Analysis.

7. **Dialogue Systems.**

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Applications of Language Technologies:

1. Machine Translation.

2. Question - Answering.

3. Automatic Text Classification.

4. Automatic Text Summarization.

5. Social Analytics.

6. Sentiment Analysis.

7. **Dialogue Systems.**

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Applications of Language Technologies:

1. Machine Translation.

2. Question - Answering.

3. Automatic Text Classification.

4. Automatic Text Summarization.

5. Social Analytics.

6. Sentiment Analysis.

7. **Dialogue Systems.**

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Applications of Language Technologies:

1. Machine Translation.

2. Question - Answering.

3. Automatic Text Classification.

4. Automatic Text Summarization.

5. Social Analytics.

6. Sentiment Analysis.

7. **Dialogue Systems.**

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Applications of Language Technologies:

1. Machine Translation.

2. Question - Answering.

3. Automatic Text Classification.

4. Automatic Text Summarization.

5. Social Analytics.

6. Sentiment Analysis.

7. **Dialogue Systems.**

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Dialogue Systems (written or spoken) are also known as:

- Conversational interfaces.
- Chatbots.

Applications of Dialogue Systems:

1. Information providers.

2. Transactional agents.

3. Educational and learning tutoring.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Dialogue Systems (written or spoken) are also known as:

- Conversational interfaces.
- Chatbots.

Applications of Dialogue Systems:

1 Information providers.

2 Transactional agents.

3 Educational and learning tutoring.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Dialogue Systems (written or spoken) are also known as:

- Conversational interfaces.
- Chatbots.

Applications of Dialogue Systems:

1. Information providers.
2. Transactional agents.
3. Educational and learning tutoring.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Dialogue Systems (written or spoken) are also known as:

- Conversational interfaces.
- Chatbots.

Applications of Dialogue Systems:

1. Information providers.
2. Transactional agents.
3. Educational and learning tutoring.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Dialogue systems main issue

The most difficult challenge in the design of conversational
interfaces are related with the highly
ambigous nature of spoken languages.

Example

Peter come yesterday.
Yesterday Peter come.

Two syntatic expressions $\Longleftrightarrow$ One semantic form

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Dialogue systems main issue

The most difficult challenge in the design of conversational
interfaces are related with the highly
ambigous nature of spoken languages.

Example

Peter come yesterday.
Yesterday Peter come.

Two syntatic expressions $\Longleftrightarrow$ One semantic form

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Dialogue systems main issue

The most difficult challenge in the design of conversational
interfaces are related with the highly
ambigous nature of spoken languages.

Example

Peter come yesterday.
Yesterday Peter come.

Two syntatic expressions $\Longleftrightarrow$ One semantic form

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Example

Peter said John came yesterday.

- Was it yesterday when Peter said that?
- Was it yesterday when John came?

One syntatic expression $\Longleftrightarrow$ Two semantic forms

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Example

> Peter said John came yesterday.

- Was it yesterday when Peter said that?
- Was it yesterday when John came?

One syntatic expression ⟺ Two semantic forms

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Example

> Peter said John came yesterday.

- Was it yesterday when Peter said that?
- Was it yesterday when John came?

One syntatic expression $\iff$ Two semantic forms

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Humans can deal with these ambiguities applying what is called "psicolinguistic preferences" and, of course, logic and common-sense reasoning:

Example

Peter said John will come yesterday.

From the computer point of view this sentence is such ambigous like previous one but humans know that nobody *"will come yesterday"*.
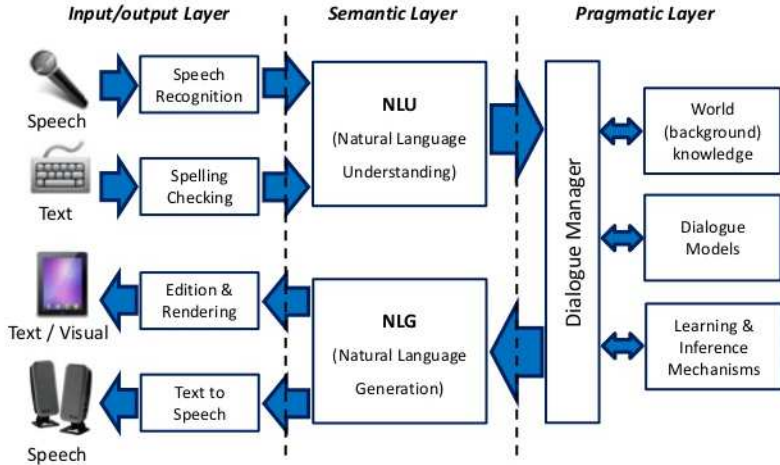
Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Humans can deal with these ambiguities applying what is called "psicolinguistic preferences" and, of course, logic and common-sense reasoning:

Example

        Peter said John will come yesterday.

From the computer point of view this sentence is such ambigous like previous one but humans know that nobody *"will come yesterday"*.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Humans can deal with these ambiguities applying what is called
"psicolinguistic preferences" and, of course, logic and
common-sense reasoning:

Example

        Peter said John will come yesterday.

From the computer point of view this sentence is such ambigous
like previous one but humans know that nobody *"will come
yesterday"*.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Type of languages
Applications
Ambiguities
Architecture

Dialogue System architecture

Picture from Institute for Infocomm Research (Singapore)

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Natural Language Understanding (NLU)

NLU main goal

The goal of NLU stage is to transform an input string (let's say *user proference*) in an abstract representation of its meaning easier for computer programs to manipulate it, in order to execute some kind of reasoning.

There is a wide variety of possible meaning representations.

- Topic maps.
- Concepts maps.
- Mind maps.
- Onthologies.
- **Feature structures.**

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Natural Language Understanding (NLU)

NLU main goal

The goal of NLU stage is to transform an input string (let's say *user proference*) in an abstract representation of its meaning easier for computer programs to manipulate it, in order to execute some kind of reasoning.

There is a wide variety of possible meaning representations.

- Topic maps.
- Concepts maps.
- Mind maps.
- Onthologies.
- **Feature structures.**

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Example

$$
\text{John came yesterday} \longrightarrow
\begin{bmatrix}
\text{SUBJECT:} & \text{John} \\
\text{ACTION:} & \text{come} \\
\text{TENSE:} & \text{past} \\
\text{OFFSETDATE:} & \text{-1 day}
\end{bmatrix}
$$

Example

$$
\text{John will talk in two days} \longrightarrow
\begin{bmatrix}
\text{SUBJECT:} & \text{John} \\
\text{ACTION:} & \text{talk} \\
\text{TENSE:} & \text{future} \\
\text{OFFSETDATE:} & \text{+2 day}
\end{bmatrix}
$$

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Feature structures

- A feature structure is a set of features.
- With no particular order between them.
- Every feature may have (but it's not required) an associated value.
- The value associated to every feature can be **atomic** or **complex**.

$$
comes \longrightarrow
\begin{bmatrix}
\text{ACTION:} & come \\
\text{TENSE:} & present \\
\text{AGREEMENT:} & \begin{bmatrix} \text{NUMBER:} & singular \\ \text{PERSON:} & 3 \end{bmatrix}
\end{bmatrix}
$$

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Feature structures

- A feature structure is a set of features.
- With no particular order between them.
- Every feature may have (but it's not required) an associated value.
- The value associated to every feature can be **atomic** or **complex**.

$$\text{comes} \longrightarrow \begin{bmatrix} \text{ACTION:} & \text{come} \\ \text{TENSE:} & \text{present} \\ \text{AGREEMENT:} & \begin{bmatrix} \text{NUMBER:} & \text{singular} \\ \text{PERSON:} & 3 \end{bmatrix} \end{bmatrix}$$

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

NLU components

Trying to convert user proference to feature structures is not trivial.
So we need to divide the process in some functional modules:

- Tokenization.

- Speller checker.

- Part Of Speech tagging (POS tagging).

- Parsing.

- Unifier.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

NLU components

Trying to convert user proference to feature structures is not trivial.
So we need to divide the process in some functional modules:

- Tokenization.
- Speller checker.
- Part Of Speech tagging (POS tagging).
- Parsing.
- Unifier.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Tokenization

Goal

Convert a sequence of characters into a sequence of tokens.

We must take into account:

- Separators: (␣-_)
- Punctuation marks: (,.;:!?)
- Special symbols: ($€%°)
- Numbers and its own separators: (1234,.)
- Alphanumeric codes: (ES772024···)

Example

	tk1		tk2		tk3		tk4		tk5		tk6	
·	The	·	dog	·	is	·	in	·	the	·	park	·
0		1		2		3		4		5		6

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Tokenization

Goal

Convert a sequence of characters into a sequence of tokens.

We must take into account:

- Separators: (␣-_)
- Punctuation marks: (,.;:!?)
- Special symbols: ($€%°)
- Numbers and its own separators: (1234,.)
- Alphanumeric codes: (ES772024···)

Example

	tk1		tk2		tk3		tk4		tk5		tk6	
·	The	·	dog	·	is	·	in	·	the	·	park	·
0		1		2		3		4		5		6	

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Tokenization

Goal

Convert a sequence of characters into a sequence of tokens.

We must take into account:

- Separators: (␣-_)
- Punctuation marks: (,.;:!?)
- Special symbols: ($€%°)
- Numbers and its own separators: (1234,.)
- Alphanumeric codes: (ES772024···)

Example

	tk1		tk2		tk3		tk4		tk5		tk6	
·	The	·	dog	·	is	·	in	·	the	·	park	·
0		1		2		3		4		5		6

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Speller checker (only in written dialogue systems)

London

- Insertion: Loondon
- Deletion: Lndon
- Substitution: Lpndon
- Switching: Lonodn
- Bad separators: Lon don

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Part Of Speech tagging (POS tagging)

Goal

To mark up lexical items with some lexical category depending on its definition and the context.

In natural language we can have some common lexical categories:

- Determiners: a, the
- Nouns: London, dog
- Pronouns: you, me
- Prepositions: to, for
- Adjectives: blue, long

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

Part Of Speech tagging (POS tagging)

Goal

To mark up lexical items with some lexical category depending on its definition and the context.

In natural language we can have some common lexical categories:

- Determiners: `a, the`
- Nouns: `London, dog`
- Pronouns: `you, me`
- Prepositions: `to, for`
- Adjectives: `blue, long`

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

POS tagging

So in the lexicon definition we can classify lexical items into categories:

- ("the", det)
- ("dog", noun)
- ("me", pronoun)
- ("to", preposition)

But in natural languages we can have several lexical categories corresponding to a single lexical item (especially in little inflectional ones, like in english):

- ("plans", noun) $\longrightarrow$ plural of plan.
- ("plans", verb) $\longrightarrow$ present of third person (singular) of verb to plan.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

POS tagging

So in the lexicon definition we can classify lexical items into categories:

- ("the", det)
- ("dog", noun)
- ("me", pronoun)
- ("to", preposition)

But in natural languages we can have several lexical categories corresponding to a single lexical item (especially in little inflectional ones, like in english):

- ("plans", noun) $\longrightarrow$ plural of plan.
- ("plans", verb) $\longrightarrow$ present of third person (singular) of verb to plan.

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

POS tagging: Garden path problem

Example

The	government	plans	to	raise	taxes	...
det	noun	**verb**	prep	verb	noun	

Example

The	government	plans	to	raise	taxes	were	defeated
det	noun	**noun**	prep	verb	noun	verb	adj

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars

POS tagging: Garden path problem

Example

The	government	plans	to	raise	taxes	...
det	noun	**verb**	prep	verb	noun	

Example

The	government	plans	to	raise	taxes	were	defeated
det	noun	**noun**	prep	verb	noun	verb	adj

Preliminaries
Natural Language Understanding
Dialogue Manager
Natural Language Generation

Goal and meaning representation
NLU parts
Grammars