Artificial Intelligence
EDA132
Lecture 13.1: An Overview of Language Processing

Pierre Nugues

Lund University
Pierre.Nugues@cs.lth.se
http://cs.lth.se/pierre_nugues/

March 1st, 2017

Applications of Language Processing

- Spelling and grammatical checkers: *MS Word*
- Text indexing and information retrieval on the Internet: *Google*, *Microsoft Bing*, *Yahoo*
- Telephone information that understands some spoken questions: *SJ* (trains in Sweden) or *Tellme.com* in the United States
- Speech dictation of letters or reports: *IBM ViaVoice*, *Windows Vista*
- Translation: *Google Translate*, *SYSTRAN*

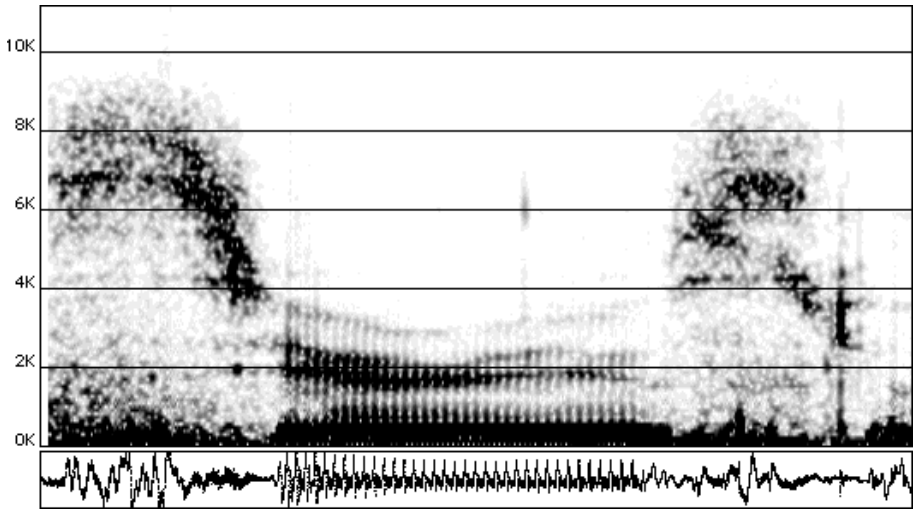
Applications of Language Processing (ctn'd)

- Direct translation from spoken English to spoken Swedish in a restricted domain: *SRI* and *SICS*
- Voice control of domestic devices such as tape recorders: *Philips* or disc changers: *MS Persona*
- Conversational agents able to dialogue and to plan: *TRAINS*
- Spoken navigation in virtual worlds: *Ulysse*, *Higgins*
- Generation of 3D scenes from text: *Carsim*
- Question answering systems: *IBM Watson*

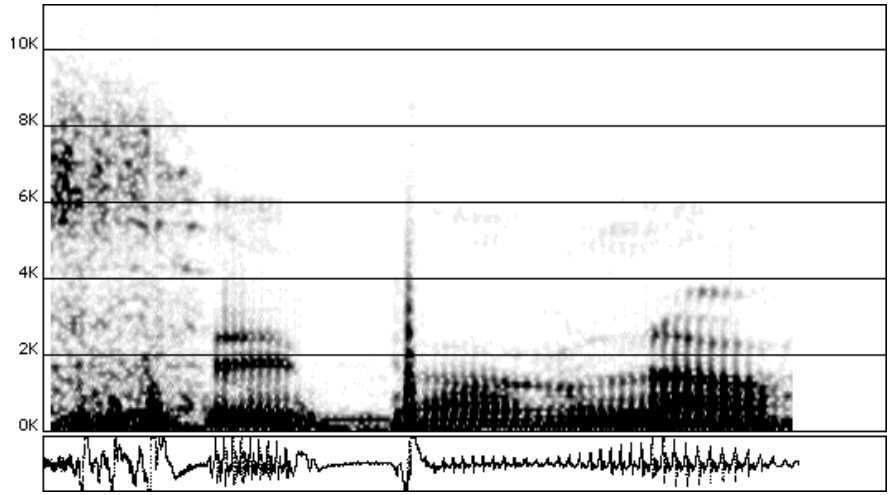
Linguistics Layers

- Sounds
- Phonemes
- Words and morphology
- Syntax and functions
- Semantics
- Dialogue

Sounds and Phonemes



Serious



C'est par là 'It is that way'

Lexicon and Parts of Speech

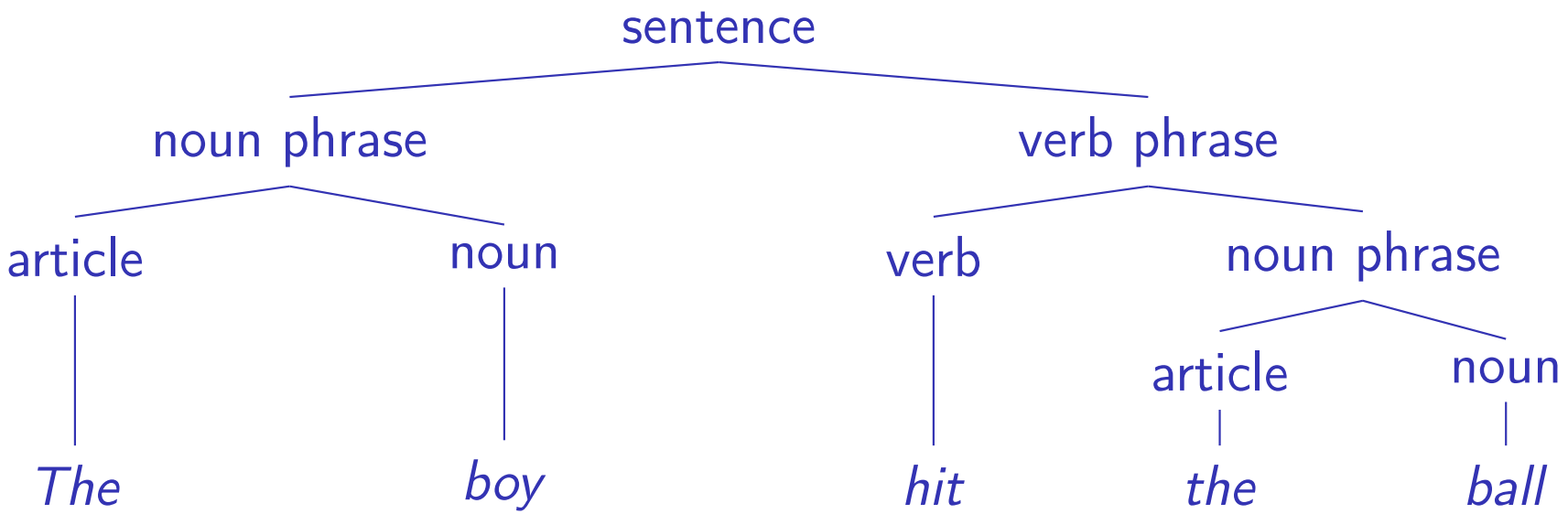
The big cat ate the gray mouse

The/article *big*/adjective *cat*/noun *ate*/verb *the*/article *gray*/adjective *mouse*/noun
Le/article *gros*/adjectif *chat*/nom *mange*/verbe *la*/article *souris*/nom *grise*/adjectif
Die/Artikel *große*/Adjektiv *Katze*/Substantiv *ißt*/Verb *die*/Artikel *graue*/Adjektiv *Maus*/Substantiv

Morphology

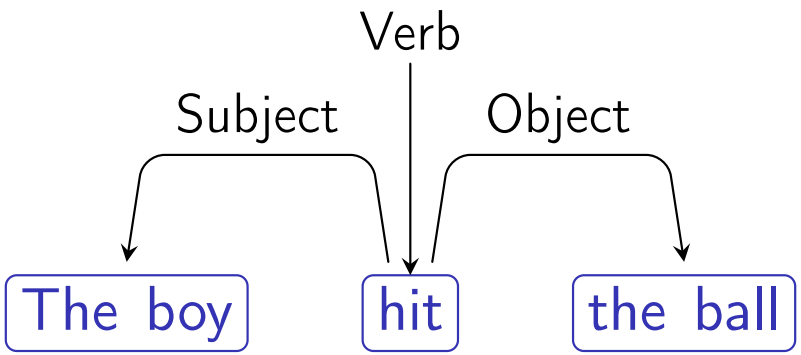
Word	Root form
worked	*to work* + verb + preterit
travaillé	*travailler* + verb + past participle
gearbeitet	*arbeiten* + verb + past participle

Syntactic Tree



Syntax: A Classical View

A graph of dependencies and functions



Semantics

As opposed to syntax:

1. Colorless green ideas sleep furiously.
2. *Furiously sleep ideas green colorless.

Determining the logical form:

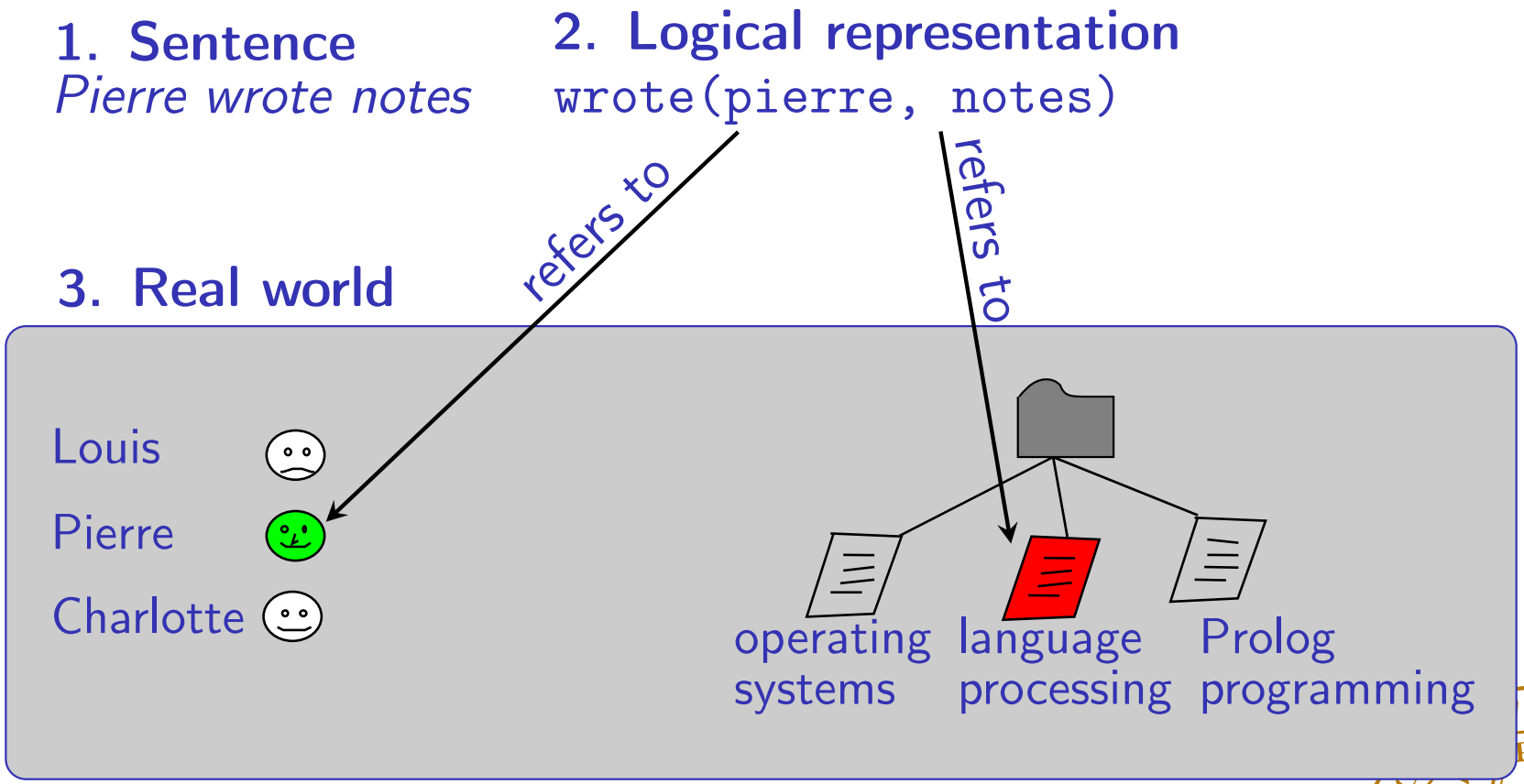
Sentence	Logical representation
Frank is writing notes	`writing(Frank, notes).`
François écrit des notes	`écrit(François, notes).`
Franz schreibt Notizen	`schreibt(Franz, Notizen).`

Lexical Semantics

Word senses:

1. **note** (*noun*) short piece of writing;
2. **note** (*noun*) a single sound at a particular level;
3. **note** (*noun*) a piece of paper money;
4. **note** (*verb*) to take notice of;
5. **note** (*noun*) of note: of importance.

Reference


1. Sentence
Pierre wrote notes
2. Logical representation
wrote(pierre, notes)
refers to
refers to
3. Real world
Louis
Pierre
Charlotte
operating systems
language processing
Prolog programming

Communication

Exchange of information between two parties
A dialogue is a set of linguistic interactions to carry out this exchange for instance to ask, inform, command, accept, etc.
It involves the generation of phrases/sentences by the speaker and their analysis by the hearer
Generation can be modeled as logical terms and then converted into sentences
Analysis involves the perception of the message, its syntactic and semantic parsing, and a pragmatic interpretation

Ambiguity

Many analyses are ambiguous. It makes language processing difficult.
Ambiguity occurs in any layer: speech recognition, part-of-speech tagging, parsing, etc.
Example of an ambiguous phonetic transcription:
The boys eat the sandwiches
That may correspond to:
The boy seat the sandwiches; *the boy seat this and which is*; *the buoys eat the sand which is*

Models and Tools

Linguistics has produced an impressive set of theories and models
Language processing requires significant resources
Models and tools have matured. Resources are available.
Tools involve notably finite-state automata, regular expressions, rewriting rules, logic, statistics and machine learning.

The Carsim System: A Text-to-Scene Converter

Texts	XML Templates	3D Animation

Véhicule B venant de ma gauche, je me trouve dans le carrefour, à faible vitesse environ 40 km/h, quand le véhicule B, percute mon véhicule, et me refuse la priorité à droite. Le premier choc atteint mon aile arrière gauche,

$\Longrightarrow$

NLP engine

```
// Static Objects
STATIC [
ROAD
TREE
]
// Dynamic Objects
DYNAMIC [
VEHICLE [
ID = vehicule_b;
INITDIRECTION = east;
```

$\Longrightarrow$

Java 3D animation program