

# NATURAL LANGUAGE PROCESSING IN A NUTSHELL

---

**Hervé Blanchon**

Laboratoire LIG

Équipe GETALP

*herve.blanchon@univ-grenoble-alpes.fr*



# Outline






---

-  Introduction
-  Natural Language Processing Applications
-  Levels of Treatment
-  Ambiguity
-  NLP Approaches in a Nutshell
-  Tools
  -  Words and Words in Context
  -  Collocation
  -  Stemming and Lemmatization
  -  POS Tagging
  -  Named-Entity Recognition
  -  Parsing
  -  Coreference Resolution
-  NLP Frameworks and Packages
-  Other Resources and Tools
-  Credits

# INTRODUCTION

# What is this Session About?

---

-  Broad overview of Natural Language Processing (NLP)
-  Vocabulary
  -  applications of NLP
  -  levels of treatment
-  Resources & Tools

# Every Natural Language...

---



...evolve over time



new vocabulary, changes of syntax



new artefacts, new concepts, new ideas,...



reading Shakespeare's writings from the sixteenth century?



reading Rabelais' writings from the fifteenth century?



reading Dante's writing from the thirteenth century?



...evolve over space



French from France, Canada, Africa









British English, American English, global English



...is ambiguous

# Natural language is...

---

-  Highly ambiguous at all levels
-  Complex and subtle use of context to convey meaning
-  Fuzzy, probabilistic
-  Involves reasoning about the world
-  A social system
  -  a key part of people interacting with other people (persuading, insulting & amusing them)

# Text Understanding is Very Hard

---



## Example




John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

# Text Understanding is Very Hard

---

## Example

 John stopped at the **donut** store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

## *What's hard about “**donut**”?*

 spare tire only intended for temporary use

 stupid individual

 deep-fried piece of dough with a hole in the center

 anything in the shape of a torus


 ...









# Text Understanding is Very Hard

---

## Example

 John stopped at the **donut store** on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.


## *What's hard about “**donut store**”?*

-  where donuts shop?
-  run by donuts?
-  which looks like a big donut?
-  made of donut?
-  which has an emptiness at its core?
-  ...

# Text Understanding is Very Hard

---

## Example

 John stopped at **the donut store on his way home from work**. He thought a coffee was good every few hours. But it turned out to be too expensive there.

 *What's hard about “**the donut store ... work**”?*


 describes where the store is?

 describes when he stopped?




# Text Understanding is Very Hard

---

## Example

 John stopped at the donut store on his way home from work. **He thought** a coffee was good every few hours. But it turned out to be too expensive there.


## *What's hard about “**He thought**”?*

-  He -> need to determine that it refers to John
-  he thought at that moment?
-  he thought habitually?

# Text Understanding is Very Hard

---

## Example


 John stopped at the donut store on his way home from work. He thought a coffee was good **every few hours**. But it turned out to be too expensive there.

## *What's hard about “**every few hours**”?*

 **he thought every few hours** that a coffee was good?

 he thought **a coffee every few hours** was good?


 he thought **a coffee stays good for every few hours**?

 *Similarly: “In America a woman has a baby every 15 minutes. Our job is to find that woman and stop her”*  
*Groucho Marx*

# Text Understanding is Very Hard

---

## Example

 John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But **it** turned out to be too expensive there.

## *What's hard about “it”?*


 stands for the coffee?

 stands for the donut store?


# Text Understanding is Very Hard


---

## Example

 John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be **too expensive** there.

## *What's hard about “**too expensive**”?*

 connect “it” to “expensive”

 too expensive for what? what are we supposed to conclude about what John did?

# Dialogue understanding is very hard

---

## Example

U: Where is **A bug's life** playing in **Mountain View**?

S: A bug's life is playing at the **Century 16 theater**.

U: When is **it** playing **there**?

S: **It's** playing at 2pm, 5pm, and 8pm.

U: I'd like 1 **adult** and 2 **children** for **the first show**. How much would **that** cost?

## Knowledge sources:

 Domain knowledge: **a bug's life**, **Mountain view**, **Century 16 theater**

 Discourse knowledge: **it**, **there**, **that**

 World knowledge: **adult**, **children**, **the first show**

# Levels of Language

---

## Phonetics/phonology/morphology

 what words (or subwords) are we dealing with?

## Syntax


 What phrases are we dealing with?

 Which words modify one another?

## Semantics

 What's the literal meaning?

## Pragmatics

 What should you conclude from the fact that I said something?

 How should you react?



# NLP APPLICATIONS

# Applications

---

## Machine Translation

 topic of the next session

## Summarization

 reduce the size of a document retaining the most important information

## Natural language generation

 produce natural language from a knowledge base, a database or a logical form

## Text classification

 assign predefined categories

 eg: automatic spam detection (binary classifier)

 eg: organize news stories by topics, ...

# LEVELS OF TREATMENT


# Phonetic/Phonological

---

 Sounds to “characters” or “words”

 problem with homophonous sequences

 fr: ils étaient très amis vs ils étaient treize amis

 en: the stuffy nose can lead to problems vs the stuff he knows  
can lead to problems

# Words


---

 Basic units for most NLP tasks

 languages with word separator: OK

 languages with no word separator: segmentation needed

 jp: 単語分割を行う (perform word segmentation)

 jp: 単語/分割/を/行う

(tango, word/bunkatsu, segmenting/particle o/okonau, to perform)

 zh: 当原子结合成分子时

 当/原子/结合/成/分子时 **(OK)**




(when/atoms/combine/molecule/the time)

 当/原子/结合/成分/子时 **(KO)**


(when/atoms/combine/ingredient/midnight)


# Words


## Construction from minimal sense units (lemma, dictionary entry)

-  study, studies, studying, studied
-  remarkable, unremarkable
-  short-sighted, fat-free, eyelashes, car park

## Lexical category (Part of Speech, POS)


 common noun (cat, cats, snow), proper noun (IBM, Italy)

 adjective (old, older, oldest), adverb (slowly)

 verb (see, count)


**Open class (lexical) words**

 ...

 number (122, 6589, one)

 conjunction (and, or) , determiner (the, some), pronoun (he, its)

 preposition (to, with), particle (off, up), interjection (Ow, Eh)

 modal (can, had)


**Closed class (functional)**

 ...


# Words

---

## Grammatical category

 tense (present, past,...), number (singular, plural,...), gender (masculine, feminine, neutral), case

## Case in German

 nominative (subject, attribute):

 **Der gute Mann** ist groß.

 The good man is big.

 accusative (complement of noun)


 Das Hemd **des guten Mannes** ist schön.

 The good man's shirt is beautiful.

 dative (indirect object)

 Ich gebe **dem guten Mann** ein Buch.

 I give the good man a book.

 genitive (direct object)

 Ich höre **den guten Mann**.










 I hear the good man.

# Phrase

---

*Syntagmatic group of words*

## Phrasal category








-  adjective phrase (very hot)
-  adverbial phrase (too slowly)
-  adpositional phrase
  -  prepositional phrase (around his desk, in the room)
  -  postpositional phrase (jp: mise ni, ie kara, hashi de)
  -  circumpositional phrase (from now on, de: Von mir aus)
-  noun phrase (the big man)
-  verb phrase (reads books)
-  ...



# Phrases

---

## Syntactic function

-  subject (**The big man** gave me a book yesterday.)
-  direct object (The big man gave me a **book** yesterday.)
-  indirect object (the big man gave **me** a book yesterday.)
-  adverbial phrase of...
  -  ...time (the big man gave me a book **yesterday**.)
  -  ...place (the big man gave me a book **at the library**.)
  -  ...

## Logico-semantic functions

-  semantic role [Fillmore 68] predicat/argument

# Phrases

---

## Logico-semantic functions

### semantic role [Fillmore 68]

Thematic Role	Definition
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of the event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	The beneficiary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

# Phrases

 Logico-semantic functions (of the predicate/verb)






 semantic role [Fillmore 68]

Thematic Role	Example
AGENT	<i><b>The waiter</b></i> spilled the soup.
EXPERIENCER	<i><b>John</b></i> has a headache.
FORCE	<i><b>The wind</b></i> blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke <i><b>the ice</b></i> ...
RESULT	The city built <i><b>a regulation-size baseball diamond</b></i> ...
CONTENT	Mona asked “ <i><b>You met Mary Ann at a supermarket?</b></i> ”
INSTRUMENT	He poached catfish, stunning them <i><b>with a shocking device</b></i> ...
BENEFICIARY	Whenever Ann Callahan makes hotel reservations <i><b>for her boss</b></i> ...
SOURCE	I flew in <i><b>from Boston</b></i> .
GOAL	I drove <i><b>to Portland</b></i> .

# Phrases

---












## Logico-semantic functions (examples)

-  John (agent) broke the window (theme).
-  John (agent) broke the window (theme) with a rock (instr.).
-  The rock (instrument) broke the window (theme).
-  The window (theme) broke.
-  The window (theme) was broken by John (agent).

# AMBIGUITY

# Ambiguity

---

-  A pervasive phenomenon in natural languages (NL)
  -  a fundamental property of linguistic expressions
-  A mean of flexibility & usability for NLs
  -  it cannot be eliminated
-  Most of the time humans do not see it
  -  we share word knowledge, “common sense”
-  On purpose, conscious uses
  -  songs, poetry, humor, jokes, advertisements
-  Accidental, unconscious uses
  -  may lead to mistakes, errors, accidents when not detected
  -  may lead to clarification sub-dialogues in conversations when detected

# Ambiguities in English

## A rough classification

### Lexical ambiguities

 Polysemy

 Homophony

 Categorical ambiguity

### Structural ambiguities

 Attachment problem

 Gap finding & filling

 Analytical ambiguity


Hirst, G. (1992) *Semantic interpretation and the resolution of ambiguity*

 any of those has implication at the semantic and pragmatic level




# English: lexical ambiguity

---

## Polysemy

-  several related “meanings” associated to a string (sequence of letters [word, term], sequence of phonemes)

## Homonymy

-  several non-related “meanings” associated to a string (sequence of letters [word, term], sequence of phonemes)
  -  written: homographs
  -  spoken: homophones

## Categorical ambiguity

-  several syntactic categories associated to a string








# Polysemy

---

Several **related** “meanings”

 The verb *open*

-  unfolding,
-  expanding,
-  revealing,
-  moving to an open position,
-  making openings in

# Homonymy

---

several non-related “meanings”

## Homographs (written)


 **row** as a noun

 a number of people or things in a more or less straight line

 a noisy acrimonious quarrel

 **bark** as a noun

 the sharp explosive cry of certain animals

 the tough, protective outer sheath of the trunk, branches, and twigs of a tree

 a sailing ship ...

## Homophones (spoken)

 **four** as a noun

 cardinal number

 **fore** as an adjective

 situated or placed in front

# Polysemy + Homonymy

---

 the word *right*

 polysemy

 senses concerning correctness & righteousness

 homonymy

 + senses concerning the right-hand side

 Linked also with metaphor

*today's metaphor may be tomorrow's polysemy or homonymy*

 a person's mouth and the mouth of a river

# Categorical ambiguity

---

several syntactic categories

 The string *sink*

 a noun

 describing a plumbing fixture

 a verb

 meaning become submerged

 *It is mainly a problem of parsing*

# Categorical ambiguity

---

several syntactic categories

 Orthogonal to the other types

 the string *respect*

 categorical and polysemous

 ... noun and verb meanings are related

 the string *sink*


 is categorical and homonymous

 ... noun and verb meanings are not related


# Structural ambiguity

---


## Attachment

-  There is more than one node to which a particular syntactic constituent may be attached

## Gap finding and filling


-  A moved constituent has to be returned to its pre-transformational starting position, and there is more than one place it might go

## Analytical ambiguity

-  The nature of the constituent is itself in doubt, that is, when there is more than one possible analysis for it

# Attachment

---

 Prepositional phrases may have more than one noun phrase available to attach it to (as well as possibly a verb)

 Example

 the door near the stairs with the “member only” sign

 the sign is one the door

 the sign is on the stairs

# Attachment

---

 A prepositional phrases may have more than one noun phrase available to attach it to ...

## Example

 I saw the man in the park

 in the park, I saw the man


 I saw the (man in the park)


 I saw the man in the park with a telescope

 in (the park with a telescope)

 in (the park with a telescope), I saw the man

 I saw the (man in (the park with a telescope))

 in (the park) ; with (a telescope)

 in (the park)<sub>location</sub>, with (a telescope)<sub>mean</sub>, I saw the man

 with (a telescope)<sub>mean</sub>, I saw the (man in (the park)<sub>location</sub>)

 I saw the (man in (the park with (a telescope)<sub>attribute</sub>)<sub>location</sub>)




# Attachment


---

 Relative clauses have similar attachment ambiguity

 Example

 The door near the stairs that had the “Members Only” sign had tempted Nadia.

 the sign is on the door

 the sign is on the stairs

# Attachment

Prepositional phrases can also be attached to an adjective phrase

## Example

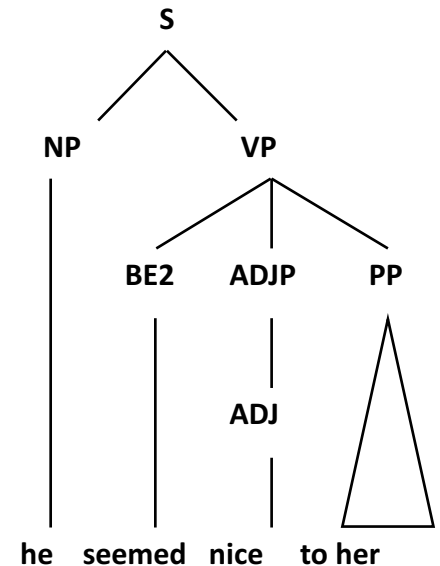
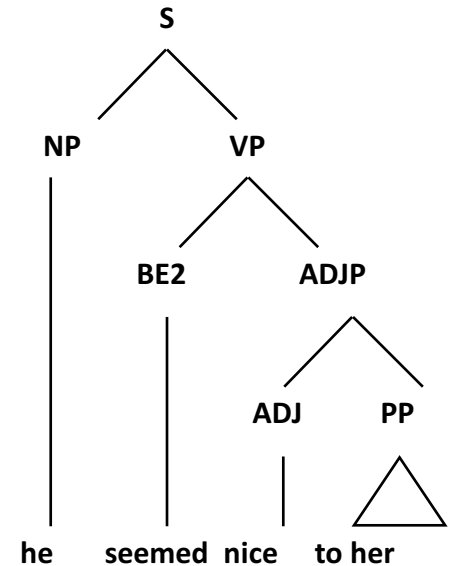
He seemed nice to her

He seemed to act nicely towards her

Attachment to the adjective phrase


He seemed to her to be nice

Attachment to the verbal phrase



# Attachment

---

 A sentence contains a subsentence, both may contain place for the attachment of a prepositional phrase or adverb

 Example







 Ross said that Nadia had taken the cleaning out yesterday

 said yesterday

 taken out yesterday

# Attachment

---

-  An adverbial may modify the sentence verb or the whole sentence
-  Example: Happily, Nadia cleaned up the mess Ross had left
  -  happily could be attached to the sentence
    -  meaning that the event was a fortunate occurrence,
    -  or it could be attached to the VP
      -  meaning that Nadia was quite happy to clean up the mess

# Attachment

---

 Adverbial placed between two clauses can be attached to the verb of either

 Examples

 The lady you met now and then came to visit us

 We were visited by the lady you met now and then

 We were visited now and then by the lady you met





 The friends you praise sometimes deserve it

 Sometimes the friend you praise deserve it

 The friends you sometimes praise deserve it


# Gap finding and filling

---


-  A moved constituent has to be returned to its pre-transformational starting position, and there is more than one place it might go
-  Example: Those are the boys that the police debated  $\triangle$  about fighting  $\triangle$ .
  -  The police debated with the boys on the topic of fighting
  -  The police debated (among themselves) about fighting the boys

# Analytical ambiguity


---

 The nature of the constituent is itself in doubt, that is, when there is more than one possible analysis for it

 Example

 “You can have the music box that’s in the closet or the one that’s on the table” said Ross. “**I want the music box  $\triangle$  on the table**” said Nadia.

 I want the music box **that is** on the table

 “I put the music box on the mantelpiece. is that okay?” asked Ross. “No,” said Nadia, “**I want the music box  $\triangle$  on the table.**”

 I want the music box **to be** on the table

# Analytical ambiguity

---

 Present participle or adjective?

 Example

 Ross and Nadia are singing madrigals

 Pen and pencils are writing implements

 Ambiguity

 They are cooking apples

 What are they doing?

 What are those apples?



# Analytical ambiguity

---

 Present participle or noun?

 Distinguishing between a present participle or a noun

 Example

 We discussed running


 We discussed the sport of running

 We discussed the possibility of our running

# Analytical ambiguity

---

 What is the subject of the supplementive?

 Participles and adjectivals at the end of a clause. A subject and an object can be the subject of a supplementative.

 Example

 We meet him leaving the room

 we were leaving the room

 he was leaving the room

 I saw him going home


 I was going home

 he was going home

# Analytical ambiguity

---

 Suppletive, restrictive relative or verb complement?

 the participle, instead of being a suppletive, could be attached to the object NP either as a reduced restrictive relative clause or as a verb complement

 Example

 The manager approached the boy smoking a cigar

 the manager is smoking (suppletive)

 the boy is smoking (relative clause)

# Analytical ambiguity

---



How is the predicate formed?



different structures that can underlies sentences of the form

NP be ADJ to V



Examples



The chicken is ready to eat



the chicken will eat



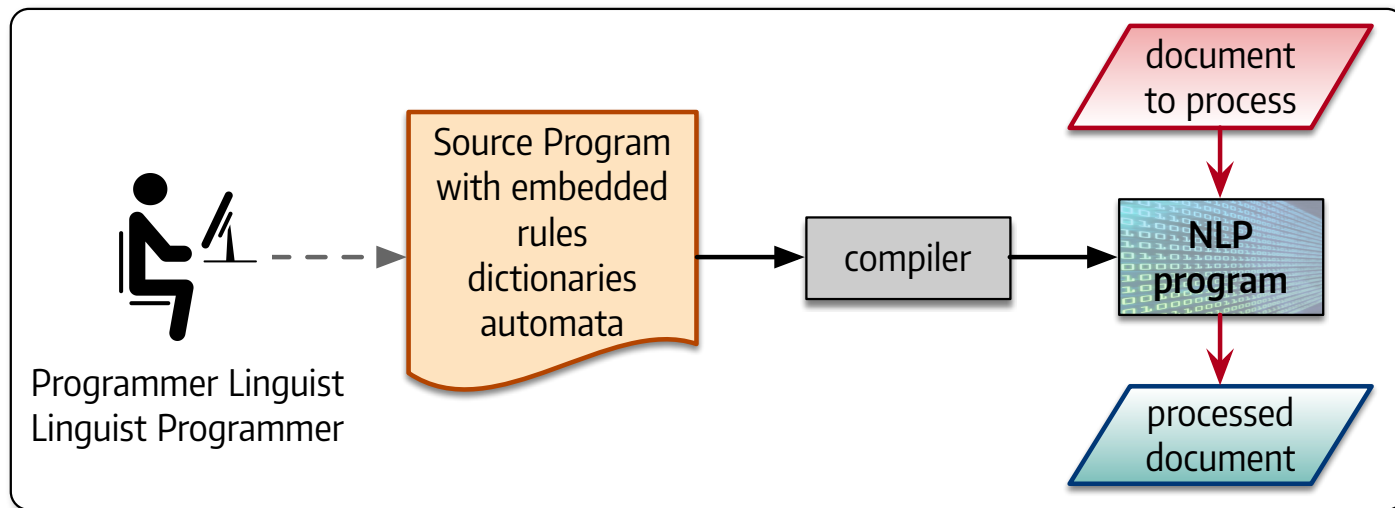
the chicken will be eaten

# NLP APPROACHES IN A NUTSHELL

# Early approach

## Linguistic hacking

-  rules, dictionaries (knowledge) are (is) encoded within the programs



# Expert approach



## Formal linguistics and compilation



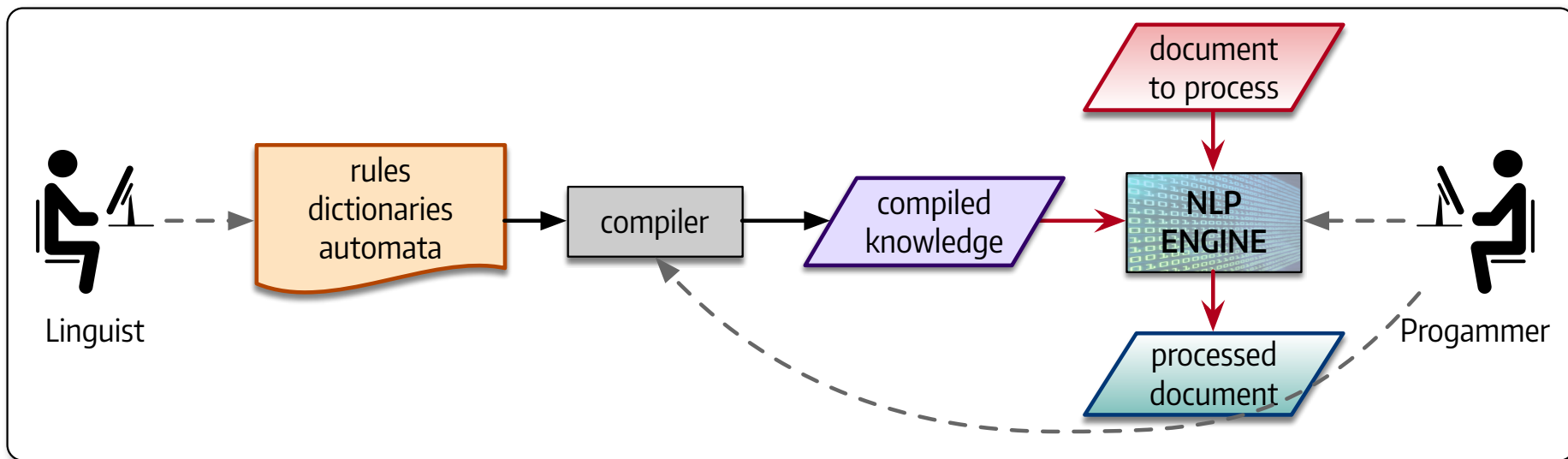
rules, dictionaries (knowledge) are (is) encoded by specialists using grammars and automata



rules (knowledge) are (is) then compiled into an internal format





the internal format act as an input for an NLP-engine



analogy with Java: source code (rules, dicos) into bytecode (compiled rules, dicos) executed by a virtual machine (NLP-engine)

# Empirical approach

## Machine learning

-  machine-learning approaches using statistical inference to learn automatically rules, knowledge, through the analysis of large, raw or annotated, corpora
-  machine-learning approaches using artificial neural networks

