

# Introduzione al Trattamento Automatico del Linguaggio - I

*Rachele Sprugnoli – [rachele.sprugnoli@unicatt.it](mailto:rachele.sprugnoli@unicatt.it)*

Centro Interdisciplinare di Ricerche per la Computerizzazione  
dei Segni dell'Espressione (CIRCSE)



UNIVERSITÀ  
CATTOLICA  
del Sacro Cuore



# DI COSA PARLEREMO

## PARTE I

- Linguistica Computazionale (LC) e Trattamento Automatico del Linguaggio (TAL)
- Sfide del TAL
- Come processare il linguaggio
  - esempio di pipeline

## PARTE II

- Come si sviluppa un modulo di TAL
  - sistemi rule-based
  - sistemi machine learning:
    - approcci non supervisionati e supervisionati, annotazione, valutazione



# LINGUISTICA COMPUTAZIONALE *versus* TRATTAMENTO AUTOMATICO DEL LINGUAGGIO

COMPUTATIONAL LINGUISTICS and NATURAL LANGUAGE PROCESSING [...] are sometimes used interchangeably to describe the field concerned with the processing of human language by computers

- **Computational Linguistics** is used to describe research interested in answering linguistic questions using computational methodology
- **Natural Language Processing** describes research on automatic processing of human language for practical applications

*Bender, Emily M. 2016. "Linguistic Typology in Natural Language Processing". Linguistic Typology 20(3), 645-660.*



# LINGUISTICA COMPUTAZIONALE

**Testo e Computer, 2016**

“L’obiettivo centrale della Linguistica Computazionale (LC) è quello di sviluppare modelli computazionali della lingua, cioè modelli del **funzionamento del linguaggio naturale** che possano essere tradotti in programmi eseguibili dal calcolatore e che consentano a quest’ultimo di acquisire le competenze necessarie per comunicare direttamente nella nostra lingua”



# COSA VUOL DIRE STUDIARE IL FUNZIONAMENTO DEL LINGUAGGIO?

Il computer può essere usato per la gestione e l'analisi avanzata dei dati linguistici in formato digitale studiando, ad esempio:

- le costruzioni grammaticali di una lingua
- la distribuzione delle parole in un autore o in un'opera
- i cambiamenti semantici delle parole nel tempo
- le differenze linguistiche tra vari registri/autori generi

Fonetica	Studia la produzione e la percezione dei suoni
Fonologia	Studia il sistema mentale dei suoni
Morfologia	Studia la formazione e la struttura interna delle parole
Sintassi	Studia la struttura interna delle frasi
Semantica	Studia il significato delle parole o delle frasi
Pragmatica	Studia l'uso contestuale della lingua



# COME SI STUDIA IL LINGUAGGIO?

Attraverso:

- metodi e strumenti informatici per la rappresentazione e gestione di grandi quantità di dati linguistici:
  - digitalizzazione
  - trascrizione audio
  - annotazione
- ricerche, esplorazioni, visualizzazioni del contenuto
- metodi matematici e statistici per l'analisi quantitativa

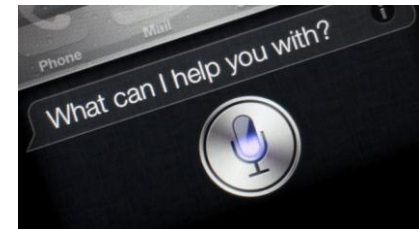


# MA...

Il computer, di per sé, **NON** conosce il linguaggio naturale!

Il **Trattamento Automatico del Linguaggio** (TAL) ha lo scopo di dotare il computer di conoscenze linguistiche, di creare macchine che capiscano (e addirittura riproducano) il linguaggio naturale, di sviluppare programmi che assistano l'essere umano in compiti (*task*) linguistici:

- riconoscimento automatico del parlato
- sintesi automatica della voce
- traduzione automatica
- analisi automatica del sentimento

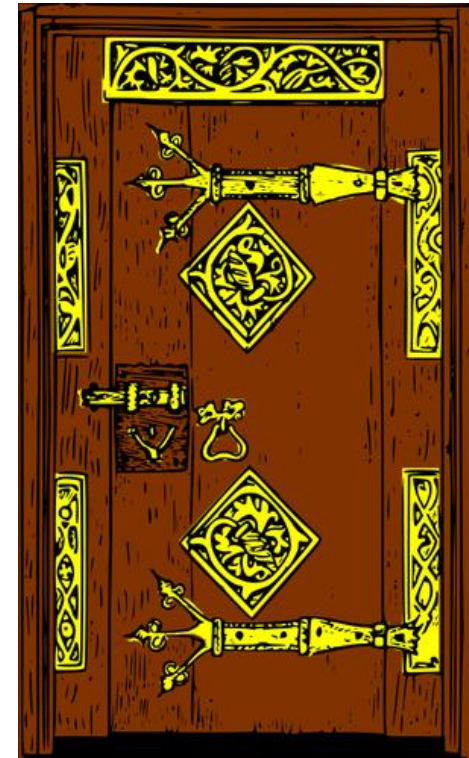


# PERCHÉ IL TAL È DIFFICILE MA ANCHE DIVERTENTE

1. Ambiguità sintattica: «una vecchia porta la sbarra»



«una **vecchia** porta la sbarra»



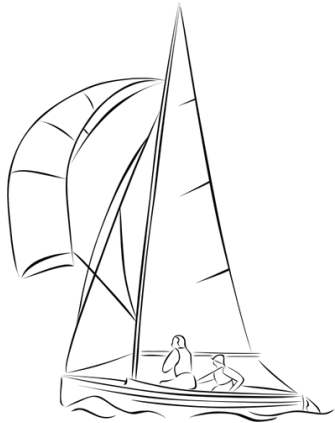
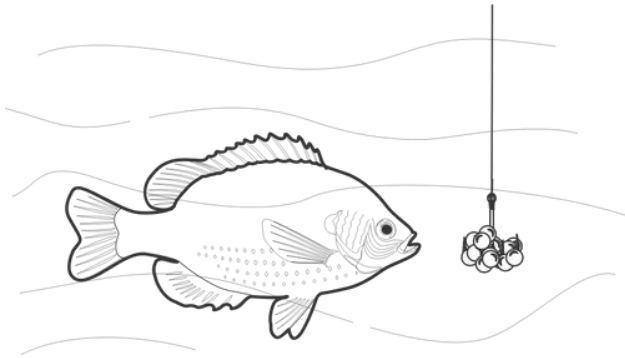
«una vecchia **porta** la sbarra (la strada)»





# PERCHÉ IL TAL È DIFFICILE MA ANCHE DIVERTENTE

## 2. Ambiguità semantica: «*amo*» / «*navigare*»



# PERCHÉ IL TAL È DIFFICILE MA ANCHE DIVERTENTE

## 3. La lingua cambia

- Lingue classiche/storiche:

*Ahi quanto a dir qual era è cosa dura  
esta selva selvaggia e aspra e forte  
che nel pensier rinova la paura!*



- Lingue non-standard:

[#SanremoFunky](#) con [@elodie](#) e qualche considerazione sulla prima serata di [#Sanremo2020](#) 🇮🇹 che sta per partire

- Neologismi: *petaloso* / *Brexit*

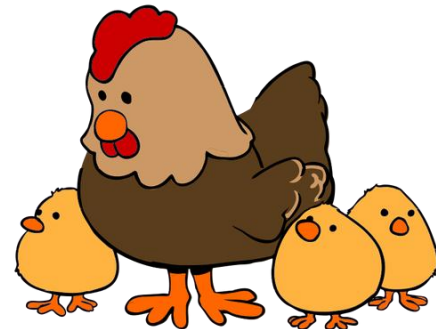


# PERCHÉ IL TAL È DIFFICILE MA ANCHE DIVERTENTE

4. Espressioni multi-parola, ovvero «2 +2 non fa sempre 4»

Il loro significato non corrisponde alla combinazione lessicale delle parole che li compongono

- espressioni metaforiche: «*parlare dietro le spalle*»
- proverbi: «*si salvi chi può*»
- espressioni idiomatiche: «*conosco i miei polli*»



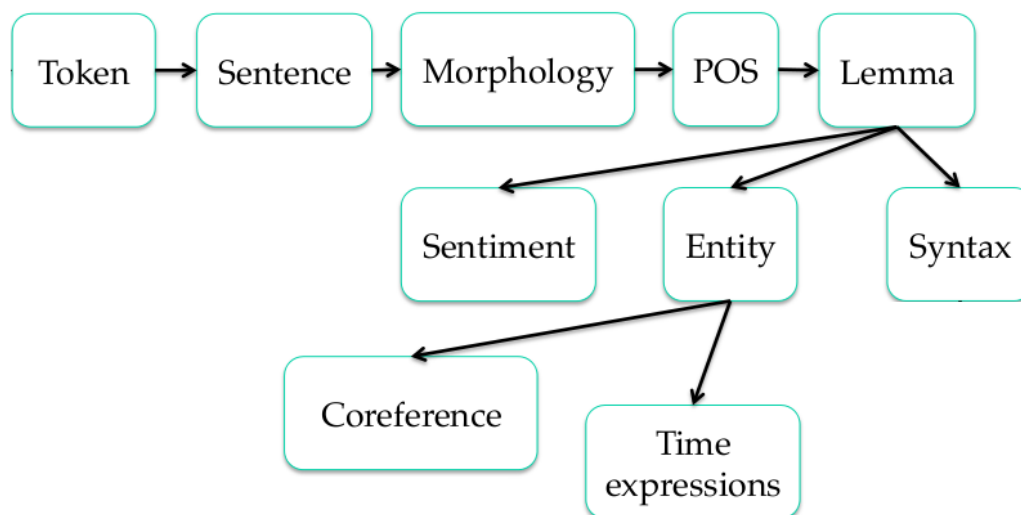
# PERCHÉ IL TAL È DIFFICILE MA ANCHE DIVERTENTE

5. Servono informazioni di contesto o di conoscenza del mondo  
*«Elsa e Anna sono sorelle»*



# COME PROCESSARE IL LINGUAGGIO

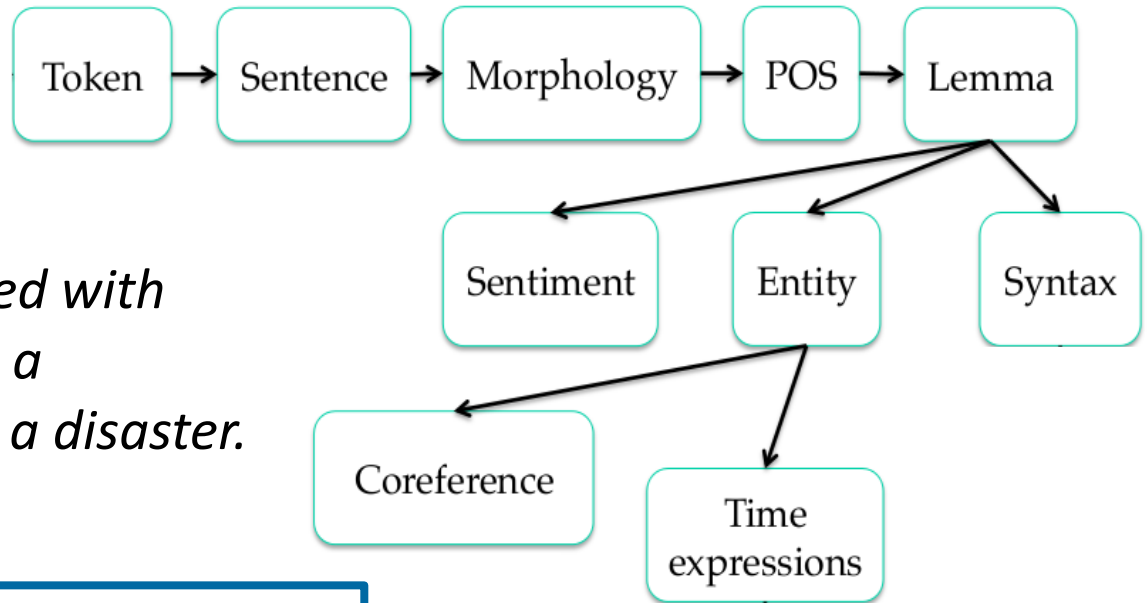
- Struttura a **PIPELINE**: catena i cui moduli descrivono ognuno un diverso livello di analisi linguistica e dove l'output di un modulo diventa l'input per il modulo successivo. Esempio:



Le analisi presentate nelle prossime slide sono l'output della pipeline di Stanford CoreNLP

- demo online: <http://corenlp.run/>





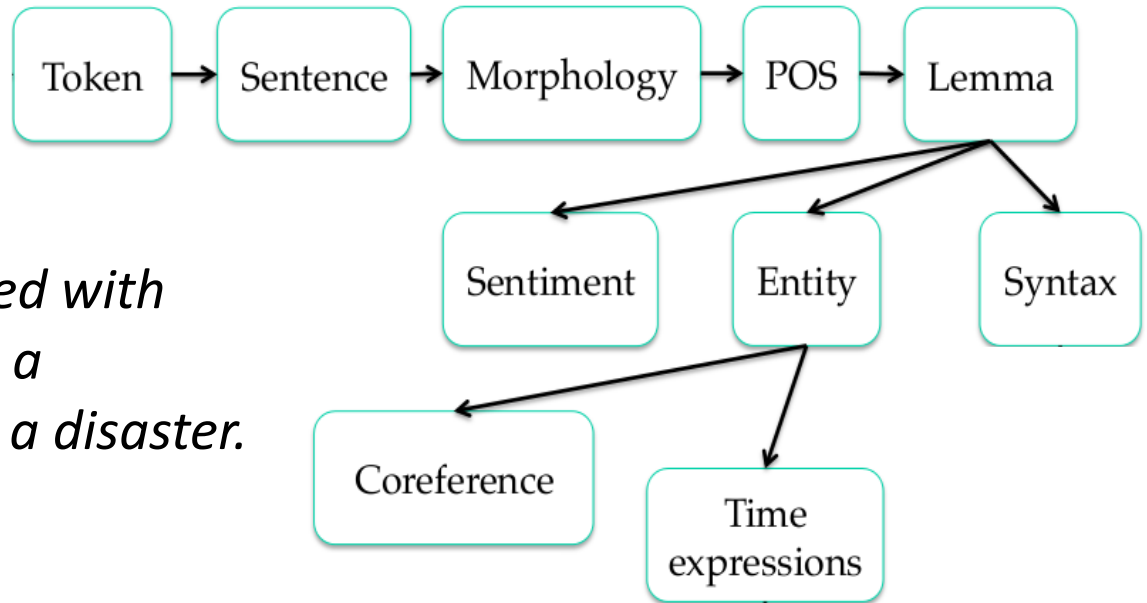
*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
Trump, 2016-08-05

## TOKEN - SENTENCE - PART OF SPEECH

1	When	you	see	what	happened	with	crooked	Hillary	today	,	it	was	a	disaster	.
2	A	disaster	.												
3	She	had	a	disaster	.										

Part of speech tags are shown above the words: WRB, PRP, VBP, WP, VBD, IN, JJ, NNP, NN, , PRP, VBD, DT, NN, .



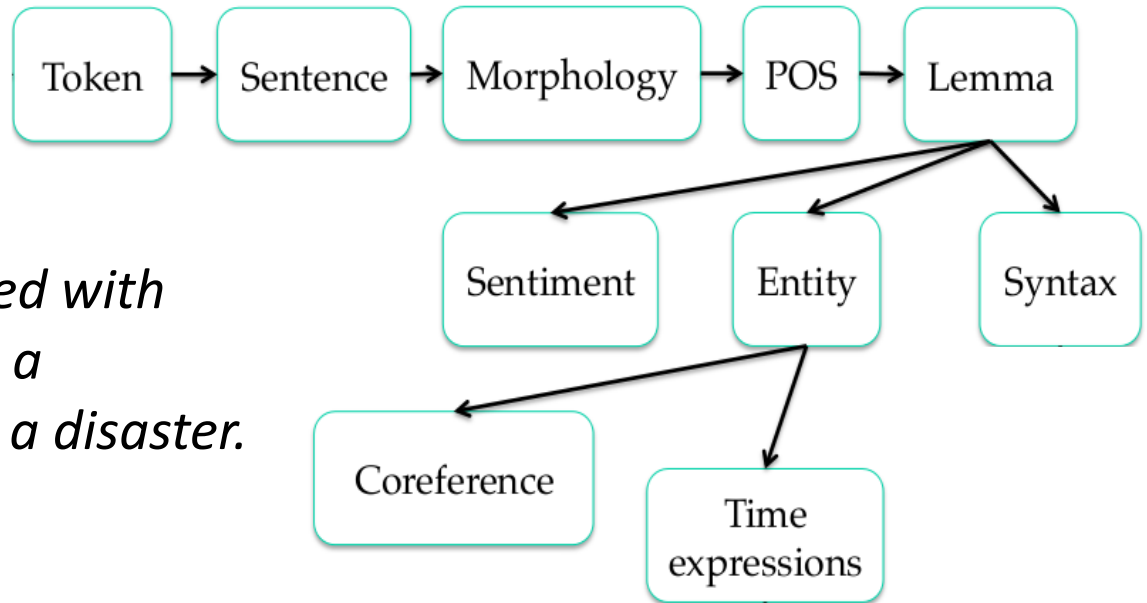


*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
 Trump, 2016-08-05

## MORPHOLOGY

when+conj	you+pron	see+v+indic+pres+no3sing	what+adj+zero	happen+v+indic+past	with+prep	crooked+adj+zero	NULL_	today+adv	NULL_
When	you	see	what	happened	with	crooked	Hillary	today	,
it+pron	be+v+indic+past	a+art	disaster+n+sing	+	punc				
it	was	a	disaster	.					
a+art	disaster+n+sing	+	punc						
A	disaster	.							
she+pron	have+v+indic+past	a+art	disaster+n+sing	+	punc				
She	had	a	disaster	.					





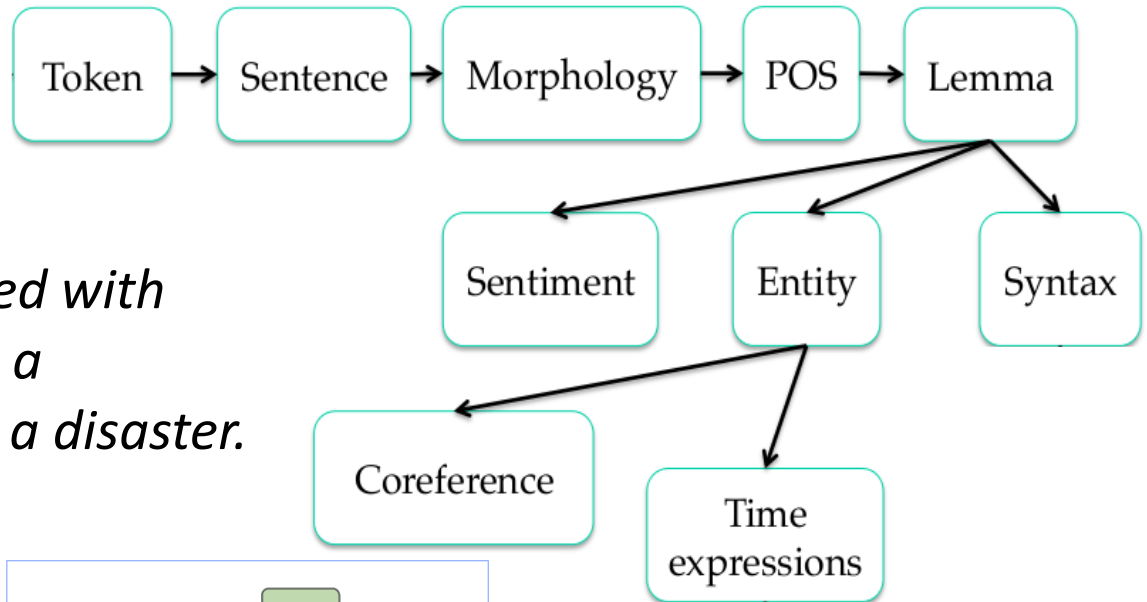
*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
Trump, 2016-08-05

## LEMMA

- 1 when you see what happen with crooked Hillary today , it be a disaster .
- 2 a disaster .
- 3 she have a disaster .



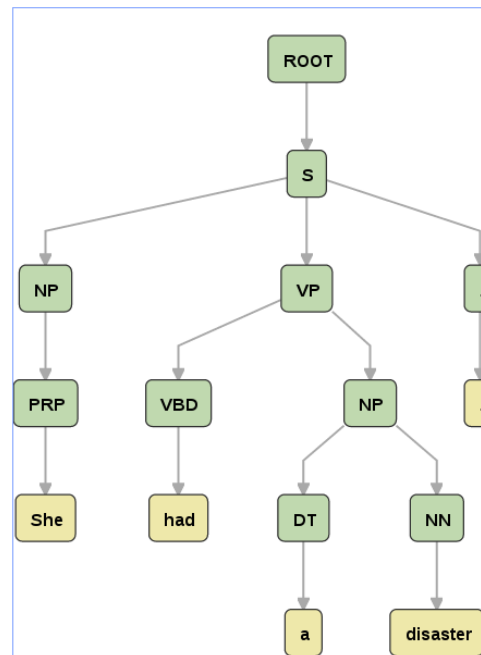


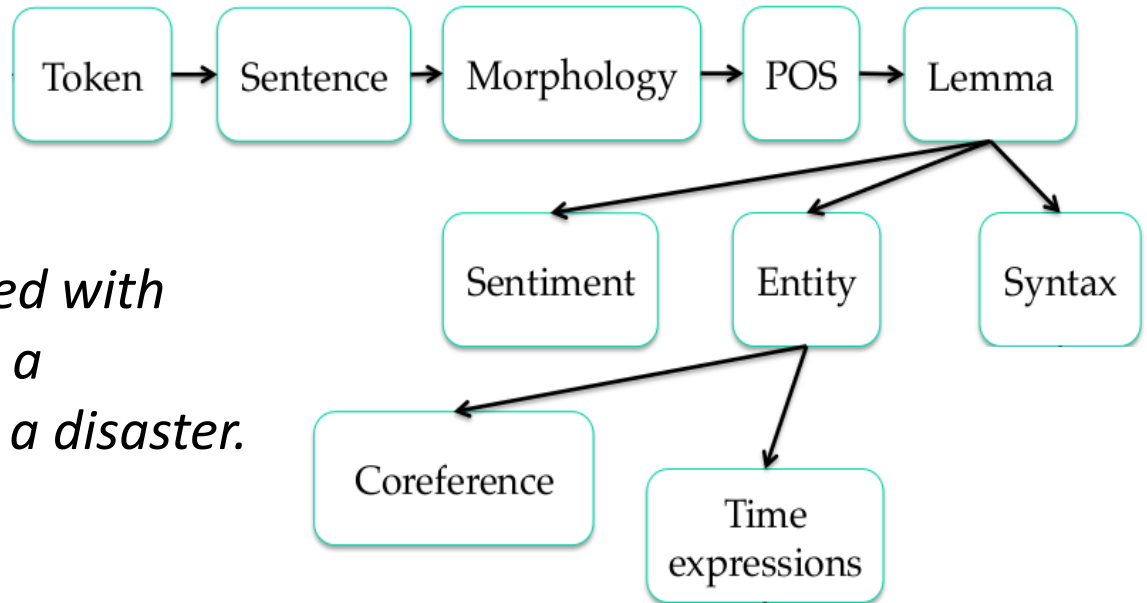


*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
Trump, 2016-08-05

## SYNTAX / PARSING

- a costituenti

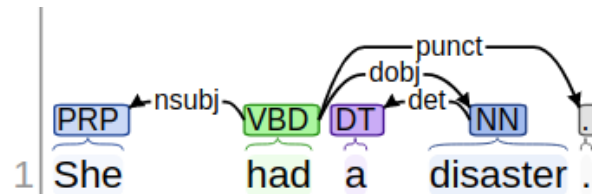


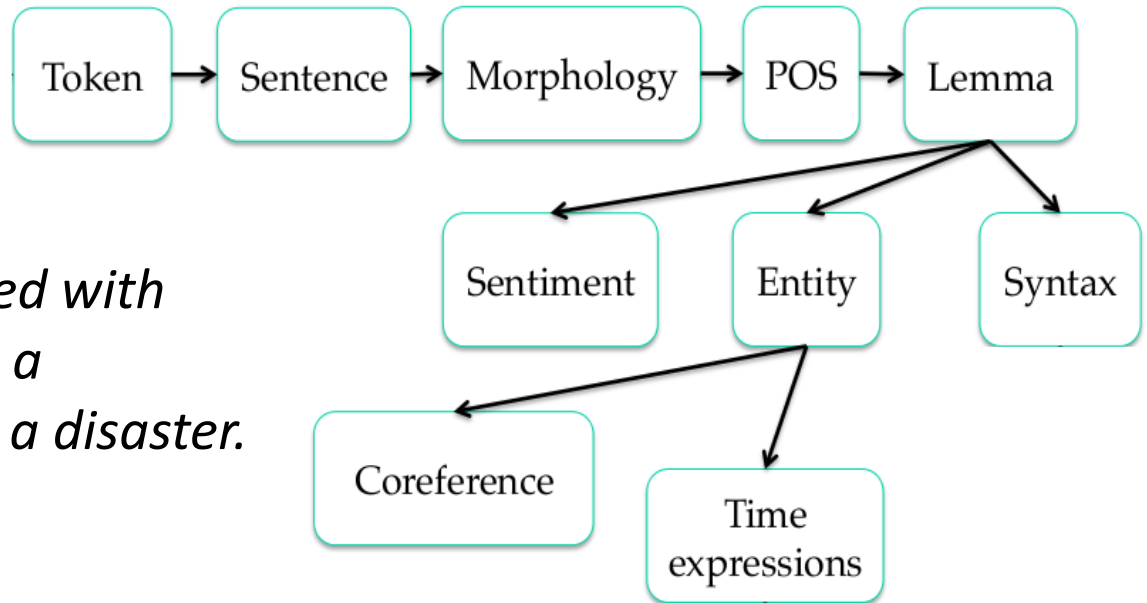


*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
Trump, 2016-08-05

## SYNTAX / PARSING

- a dipendenze



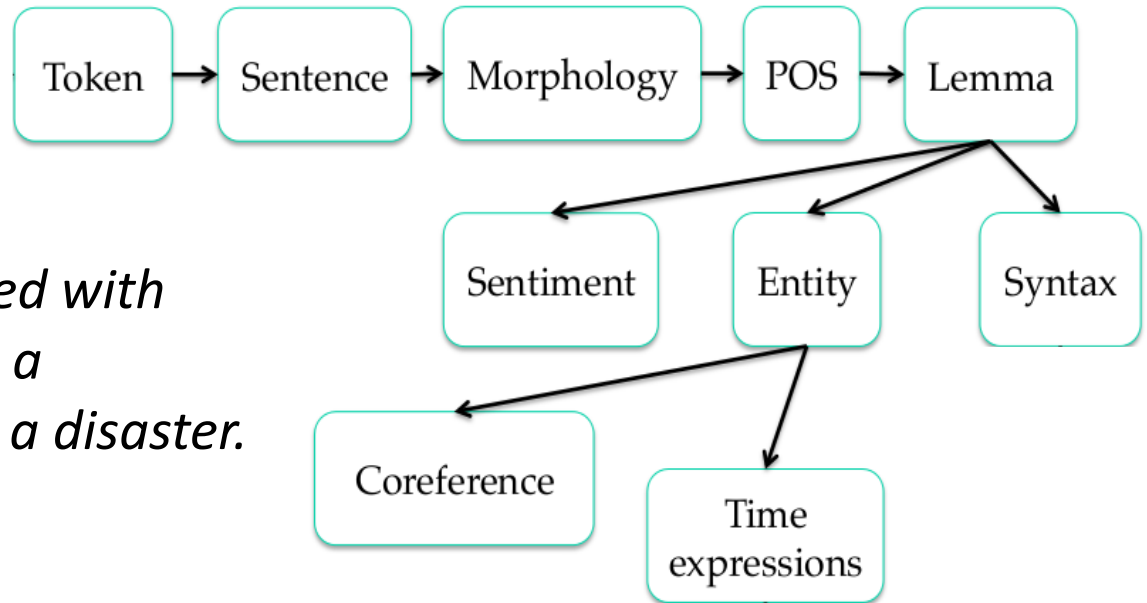


*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
Trump, 2016-08-05

## ENTITY

- 1 When you see what happened with crooked PER Hillary today , it was a disaster .
- 2 A disaster .
- 3 She had a disaster .



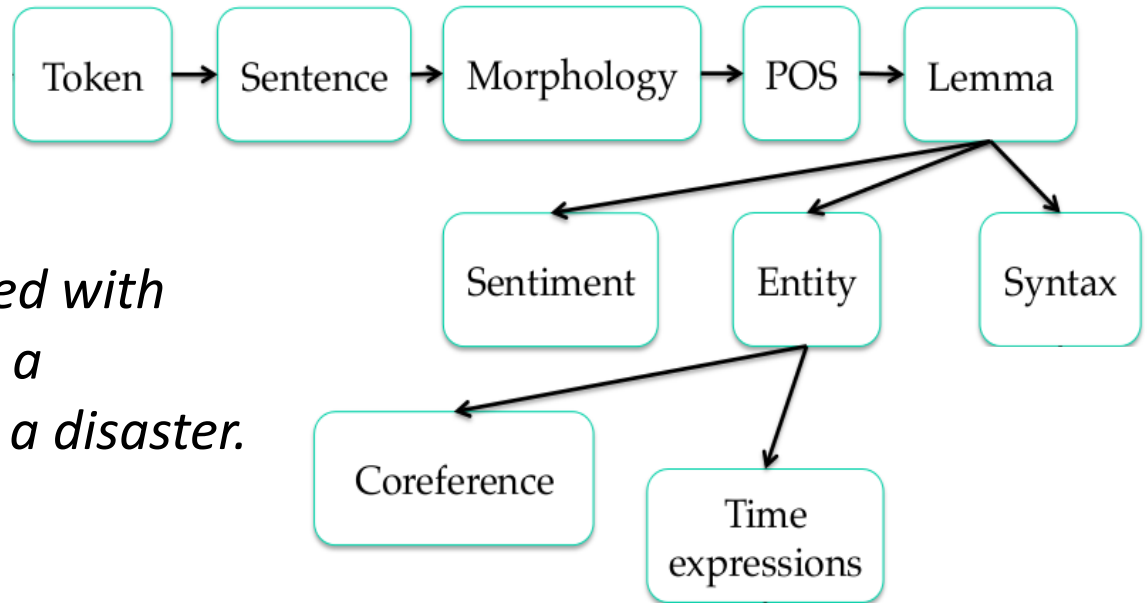


*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
Trump, 2016-08-05

## COREFERENCE

- 1 When you see what happened with crooked Hillary today , it was a disaster .  
-----coref-----
- 2 A disaster .  
-----coref-----
- 3 She had a disaster .  
-----coref-----





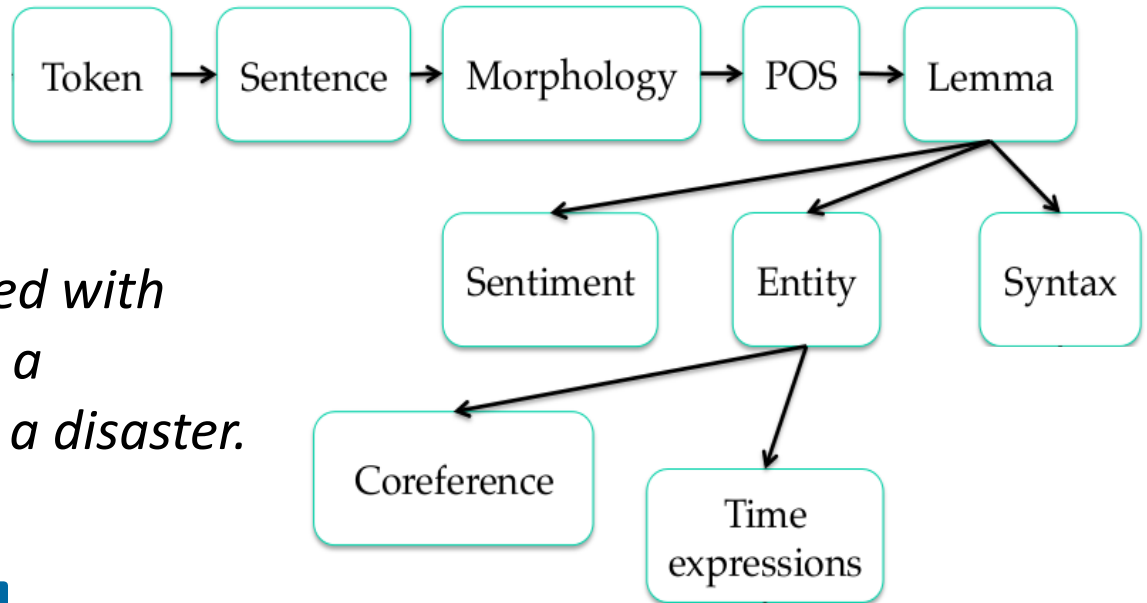
*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
Trump, 2016-08-05

## TIME EXPRESSIONS

2016-08-05

- 1 When you see what happened with crooked Hillary today , it was a disaster .
- 2 A disaster .
- 3 She had a disaster .





*When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.*  
Trump, 2016-08-05

## SENTIMENT

		NEGATIVE
1	When you see what happened with crooked Hillary today , it was a disaster .	
2	A disaster .	VERY NEGATIVE
3	She had a disaster .	NEGATIVE

