Introduction to Natural Language Processing – NL Domains

Língua Natural e Sistemas Conversacionais

Luiz Faria
Adapted from lecture notes of Carlos Ramos / ISEP

Natural Language

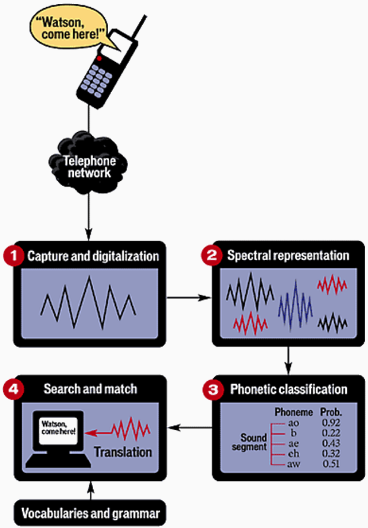
Natural Language (NL) is the area of Artificial Intelligence that focuses on textual comprehension, a simple task for humans but complicated for computer systems.

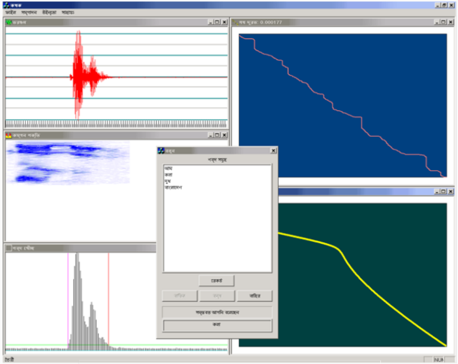
It is understood that the input of a Natural Language system is textual information, where words are already identified, for example through a string or word list. Let's look at an example:

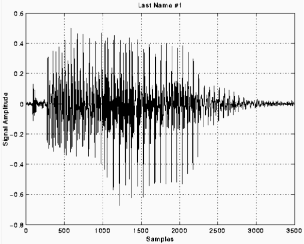
> "the boy plays the ball" or if you prefer [the,boy,plays,the,ball]

Speech Recognition can generate input to a NL program, but it corresponds to a different area, using Signal Processing techniques to proceed with the recognition of phonemes and words. It is usual in automatic subtitling and user interaction systems (mobile phones, cars, personal assistants like Alexa, Siri or Google Assistant)

Speech Recognition







Natural Language Applications

Natural Language can be used for very diverse applications such as:

- Understanding of isolated sentences
- Text comprehension
- Text Generation ⟶
- Machine Translation
- Support for user interfaces (textual or voice) such as bots
- Database Query Interface (eg SQL)
- Information Retrieval
- Text Mining

The interaction that can be established can be of the question / answer type, face to face or in free dialogue

Syntax

- In Natural Language we find two major areas: Syntax and Semantics
- In the Syntax we define the way each word is written, in practice the vocabulary, and the possible sequences of words according to their type, ie Grammar
- A Natural Language program only knows a limited vocabulary and a limited grammar. This has been one of the typical problems in Natural Language interfaces, as a misunderstanding of a sentence may be due to the limited knowledge base of the system and not due to syntax errors

Semantics

Semantics analyzes the meaning of the sentences we construct. The words in a sentence may be spelled correctly and obey a viable grammatical construct, but the sentence may not make sense

Note that the problem is very complex. There are sentences to which we can assign various meanings. When inserted in a text we can remove some ambiguities, but alone their understanding is not possible

- The boy saw the man on the hill with the telescope - who was on the hill? (the boy or the man) and who had the telescope? (the boy or the man)

Keyword Identification

One of the first user interface systems was ELIZA, developed in the 1960s by Weizenbaum [Weizenbaum-1966]. Although not very capable, for example, it did not store any representation of the sentences, this system appeared to be a minimally intelligent dialogue based on some ”tricks”

USER: Man are all alike.
ELIZA: In what way?
USER: They’re always bugging us about something or other.
ELIZA: Can you think about a specific example?
USER: Well, my boyfriend made me come here.
ELIZA: Your boyfriend made you come here?
USER: He says I’m depressed much of the time.
ELIZA: I’m sorry to hear that you are depressed.
USER: It’s true, I’m unhappy.
ELIZA: Do you think coming here will help you not to be unhappy
USER: I need some help, that much seems certain.
ELIZA: What would it mean to you if you got some help?

However, it is easy to find system problems

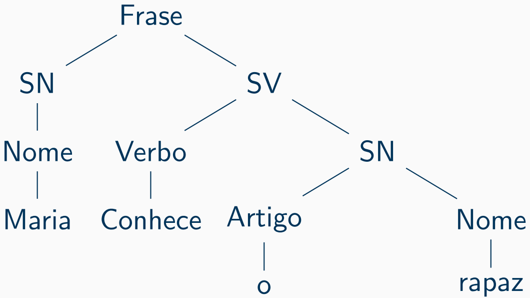
USER: My friend’s sister likes me.
ELIZA: Tell me more about your family?

Syntax Analysis

For the analysis of syntax the most common is the writing of a grammar and an interpreter, as it is usually done for the implementation of a programming language. The interpreter will assign a parsing tree to each sentence

Example in Portuguese:

$$
\begin{aligned}
&Frase \rightarrow SN \quad SV\\
&SN \rightarrow Artigo \quad Nome\\
&SN \rightarrow \quad Nome\\
&SV \rightarrow \quad Verbo \quad SN\\
&SV \rightarrow \quad Verbo\\
&Artigo \rightarrow o|a\\
&Nome \rightarrow rapaz|Maria|bola\\
&Verbo \rightarrow conhece|joga
\end{aligned}
$$



Maria knows the boy

Syntax Analysis

Example (in Portuguese) of a simple Syntax treatment done in PROLOG (Grammar and Vocabulary), we are representing Syntax by a rule-based system

```
frase --> sintagma_nominal(N), sintagma_verbal(N).
sintagma_nominal(N) --> artigo(G,N), nome(G,N).
sintagma_nominal(N) --> nome(_,N).
sintagma_verbal(N) --> verbo(N), sintagma_nominal(_,_).

artigo(m,s) --> [o].
artigo(f,p) --> [as].
nome(m,s) --> [rapaz].
nome(f,p) --> [cartas].
verbo(s) --> [joga].
verbo(p) --> [jogam].
```

Note: --> is a PROLOG operator that substitute the :- and places 2 more arguments in each term The cal would be

```
?-frase([o,rapaz,joga,as,cartas],L).
L=[ ]
```

Semantics Analysis

A possibility of semantic treatment can be done through a semantic grammar. The following illustrates a possible semantic grammar

$SENTENCE \rightarrow$ What is the $PROPERTY\_SHIP$ of ship

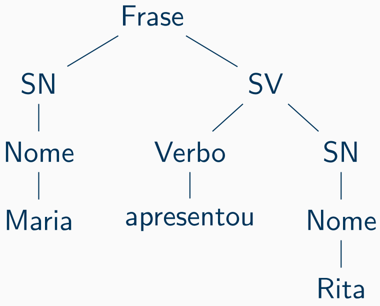
$PROPERTY\_SHIP \rightarrow ARTICLE\ PROP\_SHIP \mid PROP\_SHIP$

$PROP\_SHIP \rightarrow$ length $\mid$ width $\mid$ type

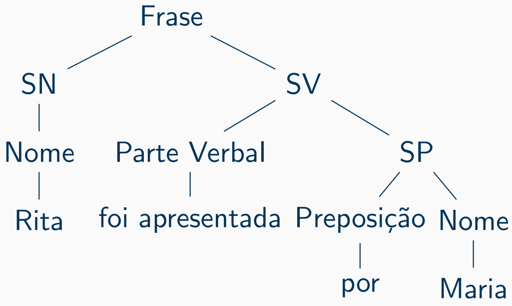
$ARTICLE \rightarrow$ the

Another possibility is the use of Case Grammars. Grammatical rules describe the syntax, but the structures created correspond to semantic relations (see next page)

Semantics Analysis



Maria presented Rita

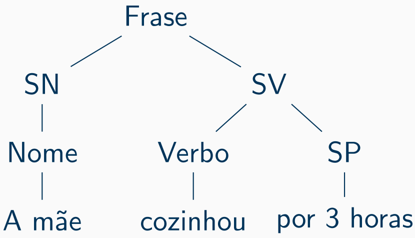


Rita was presented by Maria

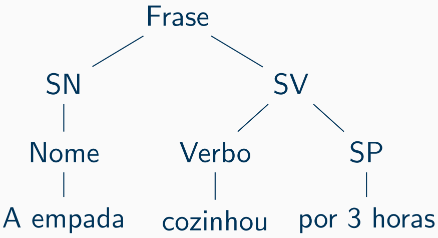
In both cases we will have a structure created of the type:

(apresentou (Atuante Maria) (Afetada Rita))
(presented (Agent Maria) (Actuated Rita))

Semantics Analysis



Mom cooked for 3 hours



The patty cooked for 3 hours

Note that, although the identification tree is identical, the semantics of the sentences is different, since in the first the subject is an active agent (the mother) while in the second the subject is the patty, which is actuated

Semantics Analysis

Another possibility is the use of Conceptual Dependence. For example, when the verb "wait" appears we can:

- Wait for something to happen
- Wait for an object
- Wait for a person

Let's look at one more example. Consider the following two sentences:

John went to the park with his girlfriend.
John went to the park with the fountain.

- In the first case "girlfriend" refers to an animated being, therefore it must be associated with John. In the second case the fountain is not an animated being, so it must be associated with the park. This example demonstrates the kind of problems we have when trying to give meaning (semantics) to a sentence

Definite Clause Grammars

Definite Clause Grammars (DCG) allow to analyse the syntax and even semantics of sentences

```
frase --> sintagma_nominal(Num), sintagma_verbal(Num).
sintagma_nominal(Num) --> artigo(Gen,Num), nome(Gen,Num).
sintagma_nominal(Num) --> nome(_,Num).
sintagma_verbal(Num) --> verbo(Num), sintagma_nominal(_).
```

MEIA
MESTRADO ENGENHARIA
INTELIGÊNCIA ARTIFICIAL

Logic Handling on Semantics

Two of the basic methods of logical handling of sentences are:

- 3 branched quantifiers (3BQ - three branched quantifiers, by Dahl and Colmerauer)
- the definite closed-world clauses (DCW - Definite Closed-World, by Pereira and Warren)

3BQ can be viewed as 3-part quantifiers (X, P, Q), where X is a variable, P is associated with the noun phrase, and Q is associated with the verbal phrase.

For example, the sentence "Haddock despises every man who does not sail." can be represented by:

every(X, and(isman(X), not(sails(X))), despises(haddock,X))

- where every is the quantifier, X is the variable, and and(isman(X), not(sails(X))) is associated with the noun phrase and despises(haddock,X) is associated with the verbal phrase
- P and Q can be true, false or undefined

Logic Handling on Semantics

In a DCW we have $H \leftarrow G1\&G2,\ldots$ corresponding in PROLOG to a rule of the type $H :- G1, G2, \ldots$

Let's see the sentence

"Is there any ocean that bathes all the countries of Europe?".

$answer(yes) \leftarrow (\exists O)(ocean(O)\&(\forall C)(country(C)\&european(C) \rightarrow bathes(O, C)))$

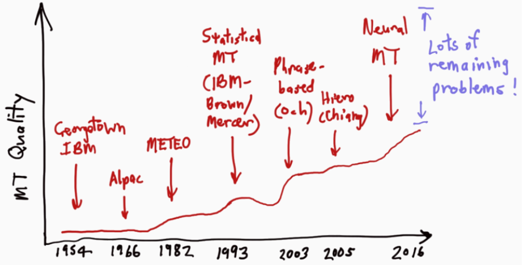
The question "What oceans bathe at least three European countries?" will be represented in the following way:

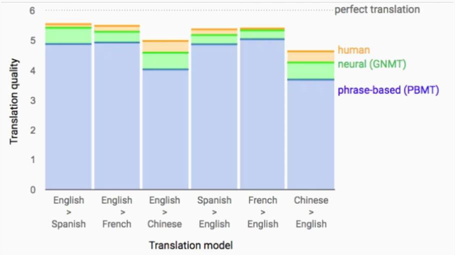
$answer(O) \leftarrow ocean(O)\&card\,C : country(C)\&european(C)\&banha(O, C) > 2$

where $card(S)$ represents the cardinal of set S

Machine Translation

Machine Translation has been one of the great challenges of AI since the early days





https://nlp.stanford.edu/projects/nmt/Luong-Cho-Manning-NMT-ACL2016-v4.pdf

Machine Translation

The Canadian METEO is a Rule-based System (1981) for translations French-English of weather forecasts

Aujourd'hui, 26 novembre	Today, 26 November
Généralement nuageux. Vents du sud-ouest de 20 km/h avec rafales à 40 devenant légers cet après-midi. Températures stables près de plus 2.	Mainly cloudy. Wind southwest 20 km/h gusting to 40 becoming light this afternoon. Temperature steady near plus 2.
Ce soir et cette nuit, 26 novembre	Tonight, 26 November
Nuageux. Neige débutant ce soir. Accumulation de 15 cm. Minimum zéro.	Cloudy. Snow beginning this evening. Amount 15 cm. Low zero.

Note: it is much simpler to translate into a specific domain that in general domains

Machine Translation

Machine translation based on rules-explicit knowledge is elegant but has major limitations:

- If understanding single sentences already raises problems of semantic analysis using more than one language raises even more problems, translation cannot be done simply by translating word by word and adapting different grammatical constructions between languages
- In most cases, both written text and (especially) voice-based text, there are errors in words and grammatical constructions

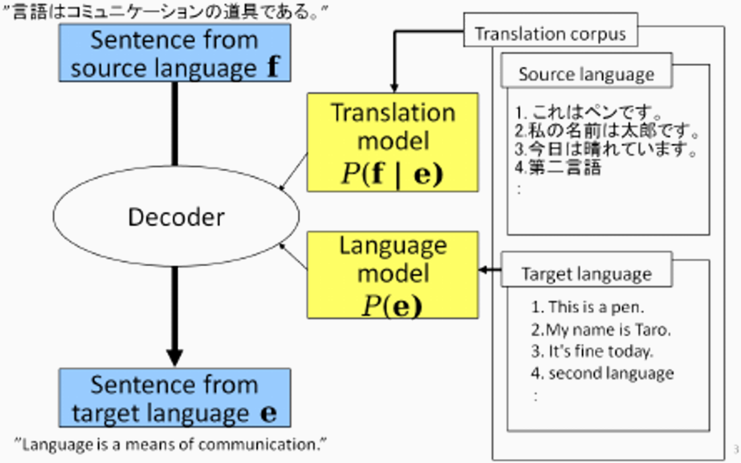
Then came the Statistical Translation systems that became popular at the beginning of this century and use a Corpus

Machine Translation

Corpus-based Translation:

- There is a set of texts in the source language and the same texts in the target language
  - For example, document translation at the United Nations was used by Google to constitute its Corpora
- Statistical Machine Translation models assume that each T sentence in one language is a translation of the S sentence in another language with a given probability $P(T|S)$ and the best translation has the highest probability value

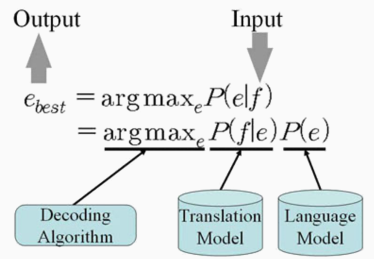
Machine Translation



https://www.researchgate.net/figure/Statistical-Machine-Translation-system_fig3_267940056

Machine Translation

- $P(f|e)$ is related to word matching
- $P(e)$ is related to the correctness of the grammatical construction in the intended language, gives us the probability that a sequence of words will be viable
- *argmax* will find the translation most likely to translate f
- Getting $argmax(P(e|f))$ is the main function of the Decoder



http://nlp.postech.ac.kr/research/previous_research/smt/

Machine Translation

- One way to deal with $P(e)$ is to use n-grams, that is, to see in a corpus the probability of appearing a given sequence of n consecutive words in a sentence
  - For example, in the phrase "I have a black dog" we have the following 3-grams: "I have a"; 'have a black"; and "a black dog"
  - 3-gram "I have a" is much more likely than "a have I"
  - Let's incorporate new words into n-gram and calculate the conditional probability
  - Word-sequence count in a Google query can be a good indicator

https://www.researchgate.net/publication/287196051_An_Overview_of_Statistical_Machine_Translation

Markov Assumption of n-order

- In a word sequence a given word usually depends on a previous word sequence, not all previous ones need to be considered

$$
\begin{aligned}
p(x_1, x_2, \ldots, x_T) &= \prod_{t=1}^{T} p(x_t | x_1, \ldots, x_{t-1}) \\
&\approx \prod_{t=1}^{T} p(x_t | x_{t-n}, \ldots, x_{t-1})
\end{aligned}
$$

Machine Translation

ON VOIT JON À LA TÉLÉVISION

	good English? P(E)	good match to French? P(F|E)
Jon appeared in TV.		✓
It back twelve saw.		
In Jon appeared TV.		✓
Jon is happy today.	✓	
Jon appeared on TV.	✓	✓
TV appeared on Jon.		✓
Jon was not happy.	✓	

https://pt.slideshare.net/hbnair080/statistical-machine-translation

Machine Translation

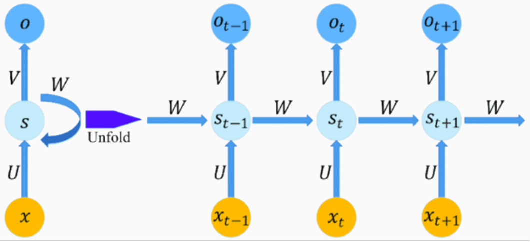
The most successful machine translators today use machine learning, in particular Deep Learnining Neural Networks, leading to what is known as Neural Machine Translation (NMT)

The most commonly used network architecture in NMT is the Encoder-Decoder architecture, consisting of 2 Recurring Neuronal Networks

Machine Translation

The NMT model represents words as densely distributed vectors that can share weights statistically between similar words.

The most commonly used network architecture in NMT is the Encoder-Decoder architecture, consisting of 2 Recurrent Neuronal Networks

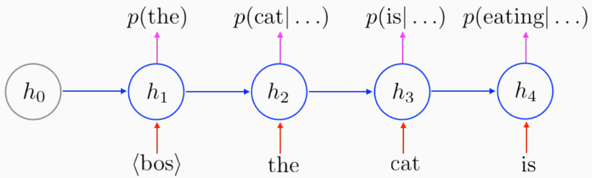


Recurrent Neural Network

Machine Translation

Recurrent Neuronal Networks (RNN) will fit the Markov chains very well, without having to consider the assumption of Markov with order n

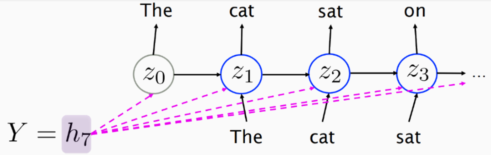
$$p(x_1, x_2, \ldots, x_T) = \prod_{t=1}^{T} p(x_t | x_1, \ldots, x_{t-1})$$



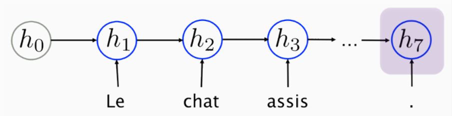
But RNN must also be trained and subject to Back-Propagation.
They are good encoders

Machine Translation

But you must have an encoder-decoder pair



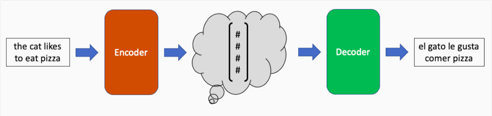
Decoder

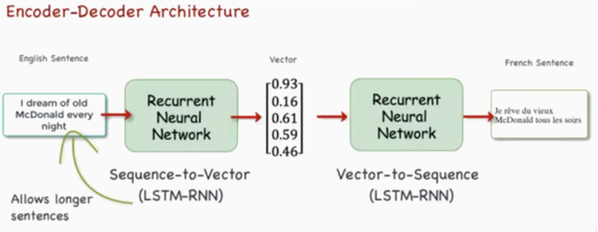


Encoder

Encoder and Decoder are both RNN
Connected in a tandem architecture

Machine Translation

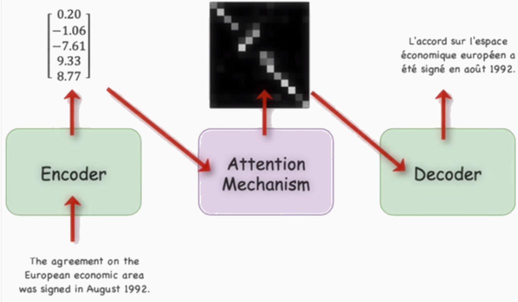




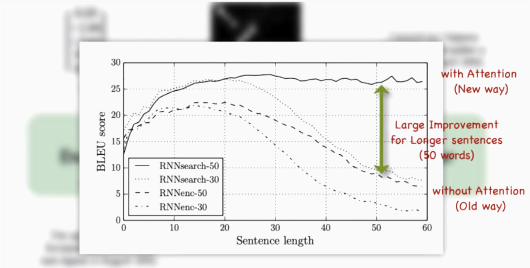
Encoder RNN converts a word sequence to a language for a vector and the decoder RNN converts the vector in a word sequence in another language

LSTM – Long Short-Term Memory, a RNN architecture

Machine Translation



The Attention Mechanism focuses on a more important part resulting from the input in this case in words that are more important



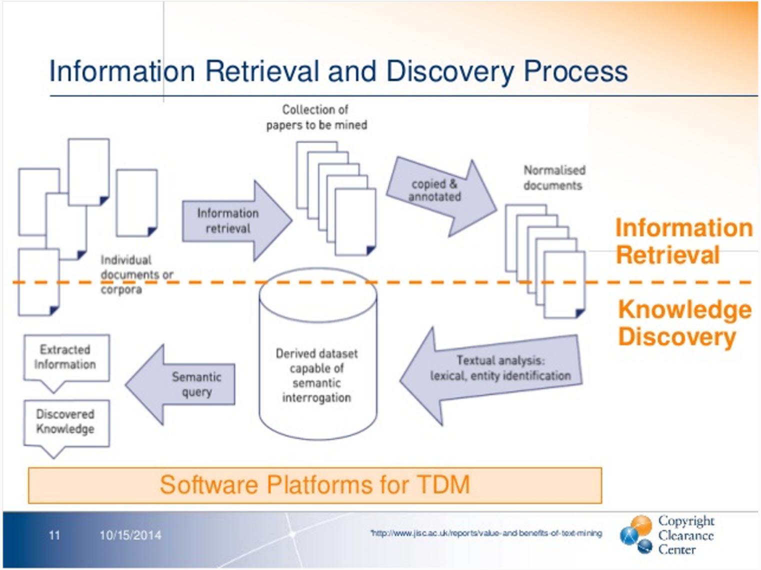
Source: Neural Machine Translation by jointly learning to align & translate (Bahdanau et al., 2016)

Information Retrieval and Text Mining

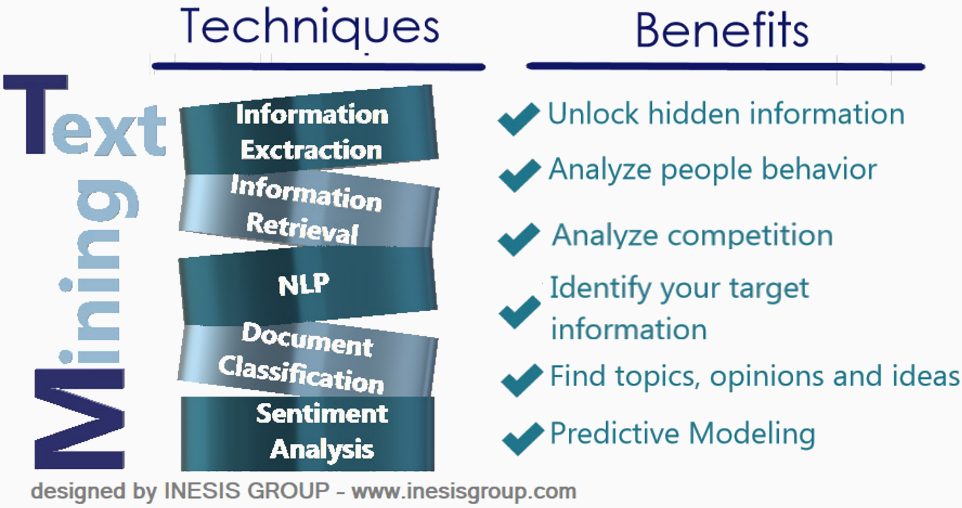
Information Retrieval - is the activity of obtaining relevant resources from an information system resource collection.

Text Mining - computer discovery of previously unknown new information automatically extracted from different written resources

Information Retrieval and Text Mining



Information Retrieval and Text Mining



Deep Learning vs. traditional NLP

Why do we need to study traditional NLP?

- Perform good enough in many tasks
  Example: sequence labeling
- Allow us not to be blinded with the hype
  Example: word2vec / distributional semantics
- Can help to further improve DL models
  Example: word alignment priors in machine translation

For sequence labeling, we can use probabilistic modeling with good results

word2vec method is inspired by some neural networks, has similar ideas as some distributional semantic methods have

Word alignments in machine translation and attention mechanism in neural networks are very similar

https://medium.com/@yoav.goldberg/an-adversarial-review-of-adversarial-generation-of-natural-language-409ac3378bd7

Deep Learning vs. traditional NLP

Why do we need to study DL in NLP?

- Provide state-of-the-art performance in many tasks
  Example: machine translation
- This is where most of research in NLP is now happening
- But many ideas in deep learning methods are really similar to something that was happening on more traditional approaches
- So it is important to study the two approaches