



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

Кафедра прикладной математики и  
информатики

# Компьютерная лингвистика

#8

Нижний Новгород, 2019



# Лабораторная работа #9.

doc2vec

## Задача.

1. Обучить модель doc2vec на собственном корпусе
2. Построить классификатор тем
3. Добиться максимального качества



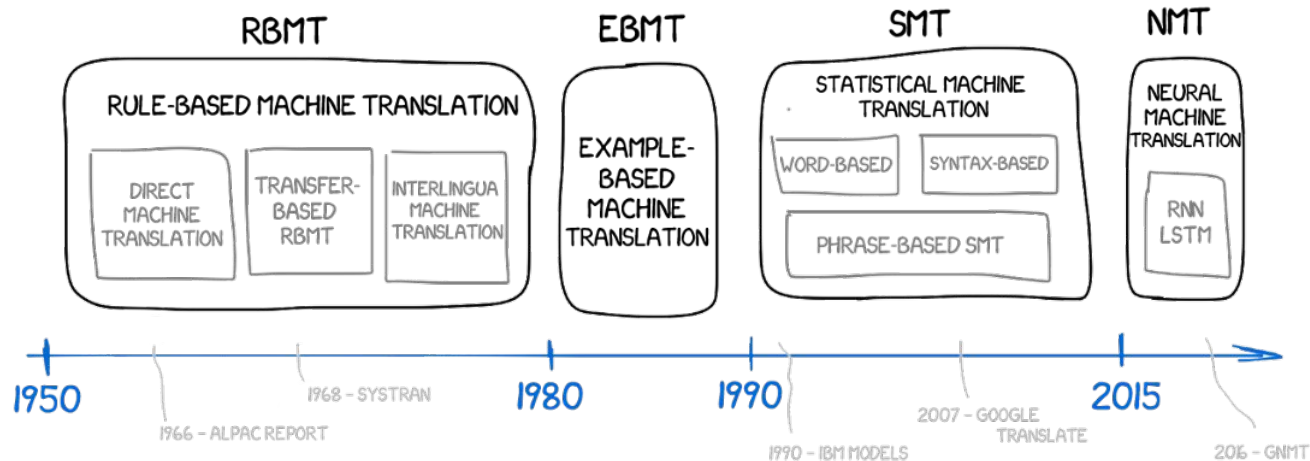
```
documents_train = [  
    gensim.models.doc2vec.TaggedDocument(tokens, [i]) for i, tokens in enumerate(corpus)  
]  
  
model = gensim.models.Doc2Vec(  
    documents_train, vector_size=100, window=5, min_count=5, workers=4  
)  
  
...  
  
vector = model.infer_vector(tokens)
```



# Лекция #8.

- Машинный перевод

## A BRIEF HISTORY OF MACHINE TRANSLATION



# Rule-based machine translation.

- Direct

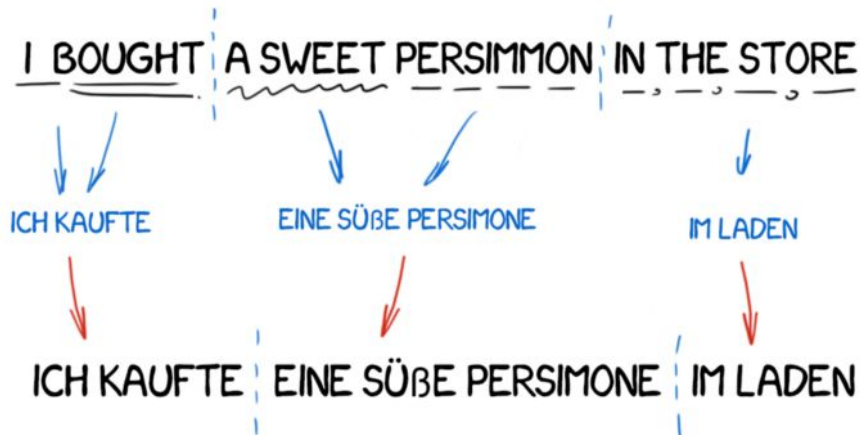
I | WANT | FORTY | KILOGRAMS OF | PERSIMMONS

↓ ↓ ↓ ↓ ↓

ICH | WOLLEN | VIERZIG | KILOGRAMM | PERSIMONEN

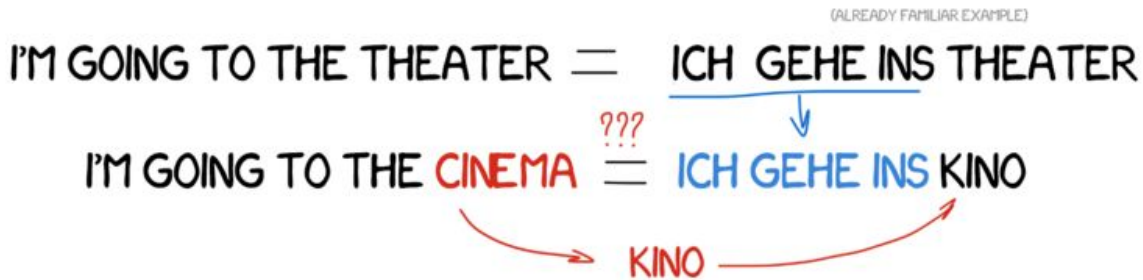
# Rule-based machine translation

- Transfer-based



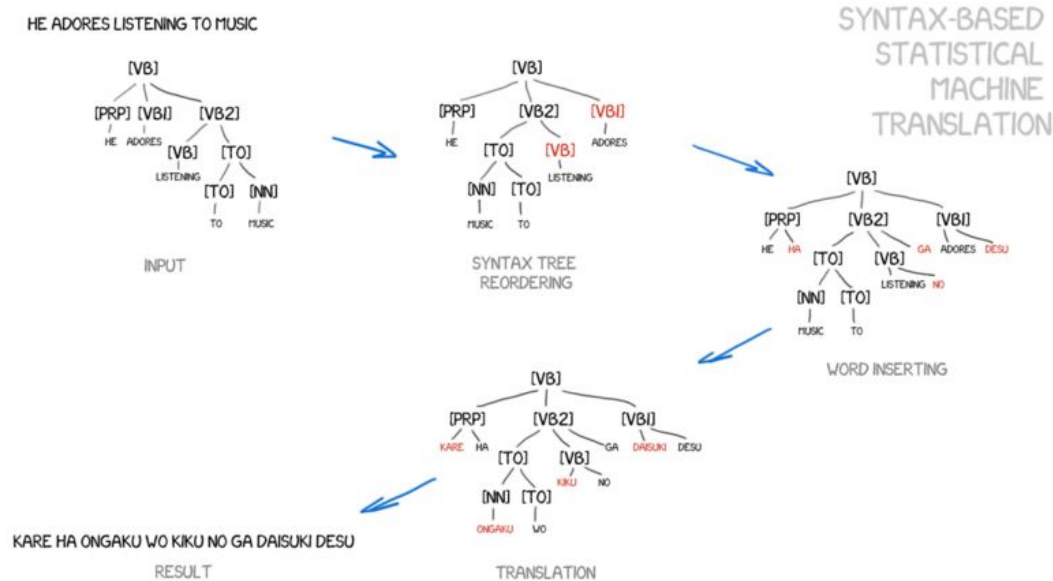


# Example-based machine translation



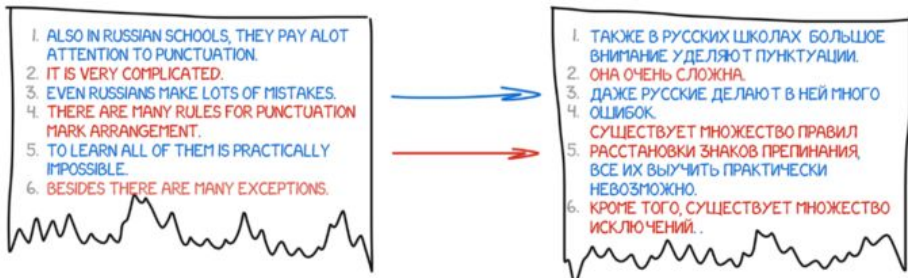


# Syntax-based machine translation

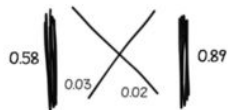


# Statistical machine translation

PARALLEL CORPUS

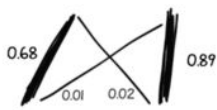


THE HOUSE



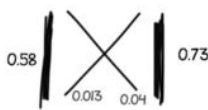
DAS HAUS

BLUE HOUSE



BLAUES HAUS

THE CAR



DAS AUTO

German English Russian Translate

verlassen ✓



Suggest an edit

Translations of leave

verb

■ verlassen	leave, abandon, exit, quit, forsake, desert
■ lassen	let, leave, allow, stop, let go, let be
■ hinterlassen	leave, leave behind, bequeath
■ abgeben	leave, dispense, deliver, submit, emit, give off
■ gehen	go, walk, leave, move, go down, quit
■ überlassen	leave, entrust, leave up to, leave with, intrust, let have
■ belassen	leave
■ abreisen	leave, depart, check out
■ stehen lassen	let stand, leave, leave behind, leave untouched, abandon
■ zurücklassen	leave, leave behind
■ austreten	escape, leave, withdraw, come out, resign, opt out
■ abfahren	leave, depart, go, ski down, wear, move off
■ liegen lassen	leave, leave out, leave lying around, leave behind
■ fortgehen	go away, leave, depart, quit
■ vergessen	forget, forget about
■ abgehen	leave, depart, come off, go off, exit, sail
■ abfliegen	depart, leave, take off, fly off, go
■ sein lassen	leave, stop, leave alone, drop
■ ausfahren	extend, come out, go for a ride, deliver
■ vermachen	bequeath, leave, will, make over
■ übrig bleiben	remain, leave, be left over
■ vererben	bequeath, pass on, leave, hand down
■ fortfahren	continue, proceed, go on, resume, leave, go away
■ sich entfernen	move away, leave, go away, recede, depart, move off
■ hinausfahren	drive out, go out, leave

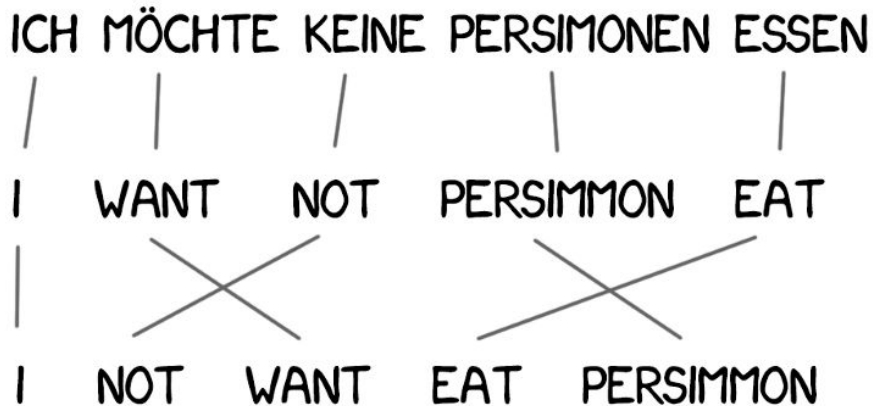


# Statistical machine translation

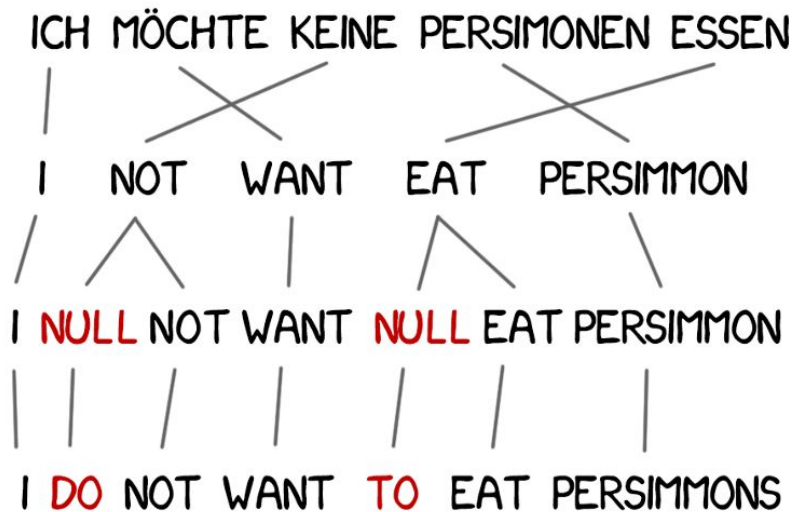
ICH MÖCHTE KEINE PERSIMONEN ESSEN  
| | | | |  
I WANT NOT PERSIMMON EAT



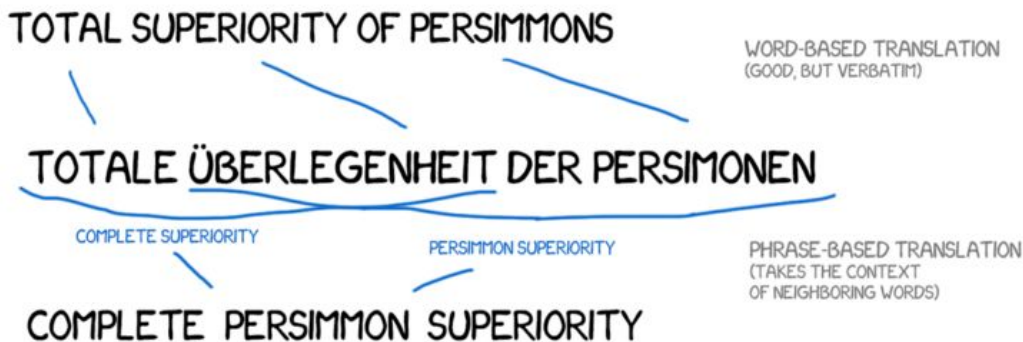
# Statistical machine translation



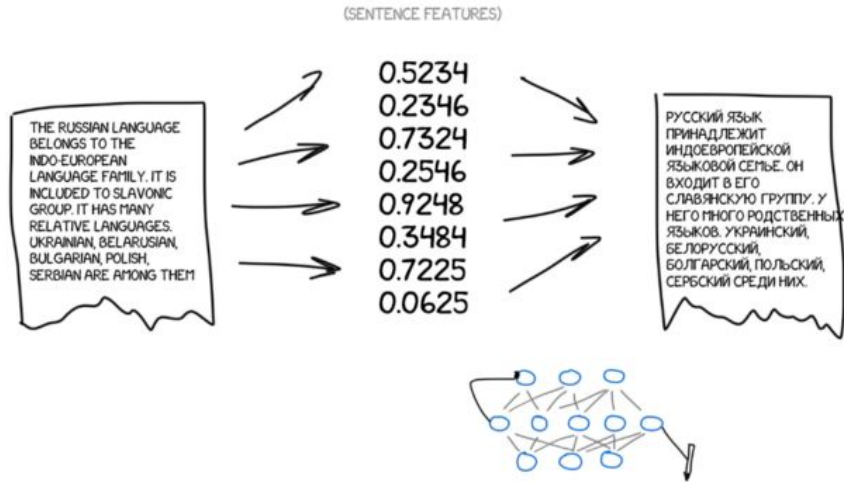
# Statistical machine translation

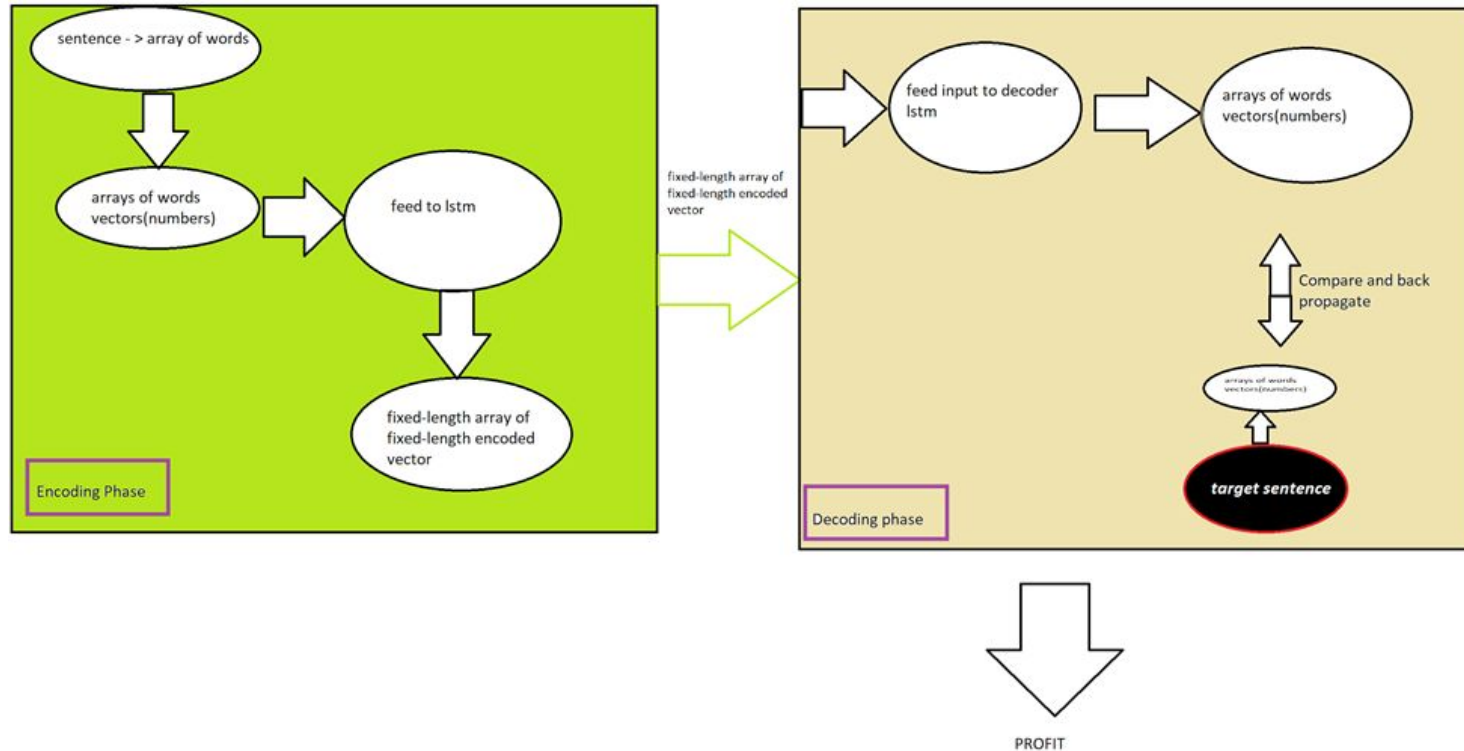


# Statistical machine translation

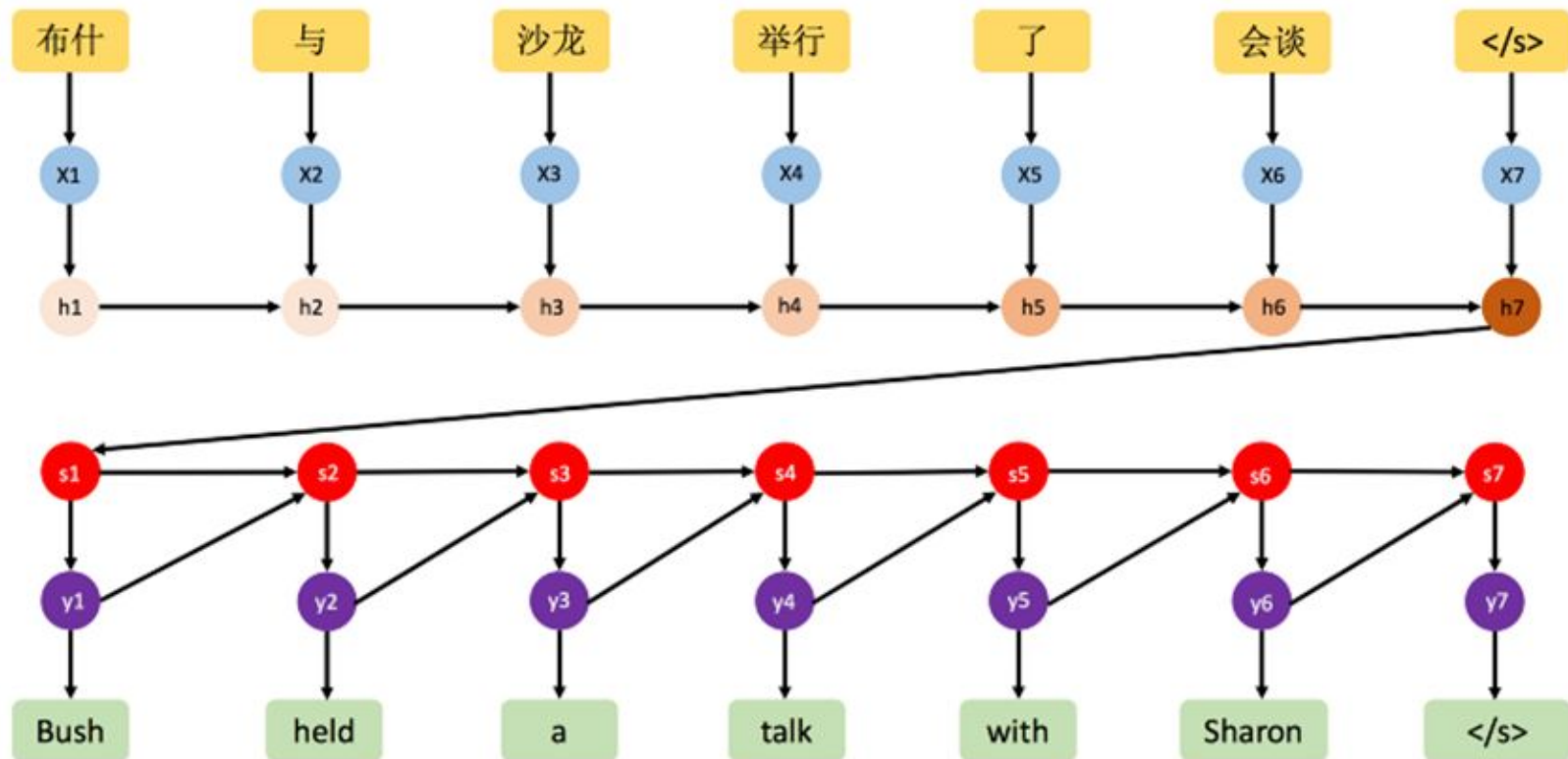


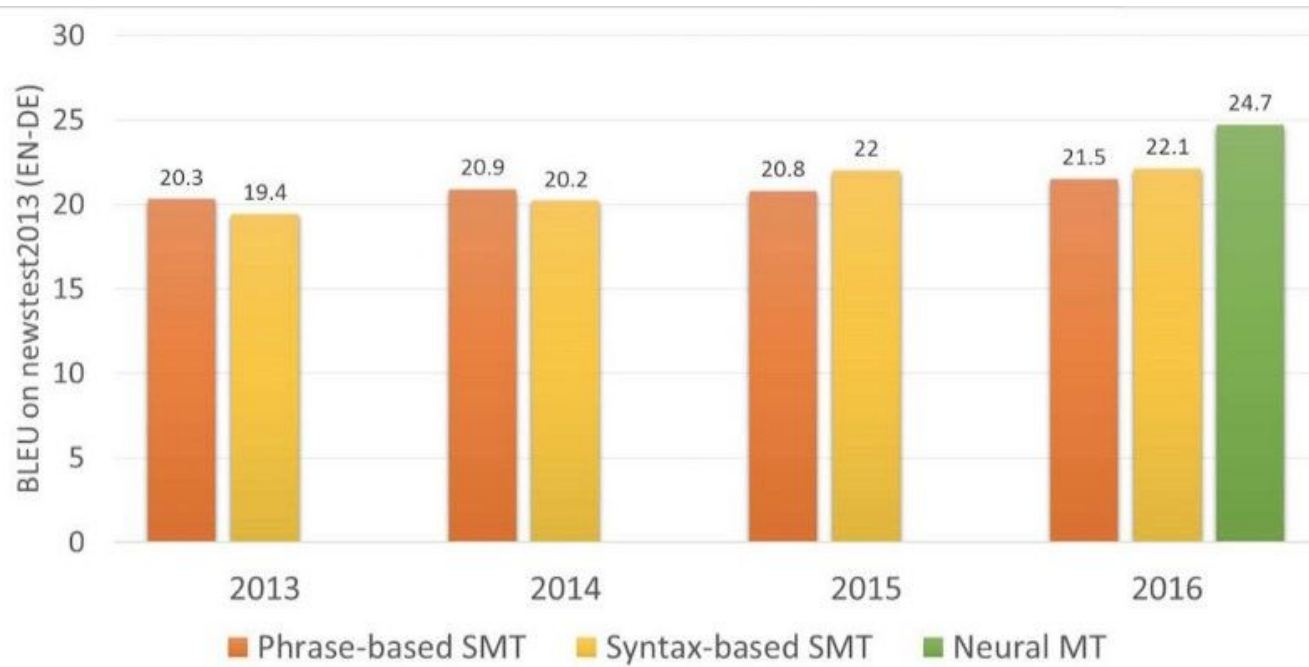
# Neural machine translation













	Neural Machine Translation	Statistical Machine Translation
Training time	More	Less
Training data	Less	More
Translation (decoding) time	More	Less
CPU usage	More	Less
Space in disk	Less	More
Mechanism	Sentence by sentence	Word by word/ phrase by phrase
	Attentional encoder-decoder networks; optimization	Statistical analysis; probability
	Train multiple features jointly	Feature engineering required
Interpretability		👍
Long distance reordering	👍	
Morphology, syntax, and agreement errors	👍	
Translation style consistency for the same word		👍
Tolerance to noisy data	👍	
Multilingual/ multi-domain translation	👍	
Vocabulary/Rare word Problem		👍



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ