

Grammarly Under the Hood Oksana Kunikevych

November 2nd, 2017

Where to find our video

All video of speeches and webinars: 2016.utic.eu/video/

Register for new webinars and download the files of previous ones: webinar.utic.eu/

Subscribe to our YouTube channel





www.youtube.com/user/UTIConf

ПОДРОБНЕЕ...

E \$

() 0:00 / 3:49



Host

Stanislav Bogdanov COO of the UTIC



Oksana Kunikevych

Computational Linguist, Grammarly

After graduating from Lviv Polytechnic National University two years ago, Oksana joined Grammarly as Computational Linguist. At Grammarly she is working in the area of Natural Language Processing. Her current projects are focused on error correction and stylistics.

Speaker

Oksana is also passionate about coding. She is one of the organizers of DjangoGirls workshops that aim to teach women to code in Python.

What is Grammarly? Real-world NLP applications NLP at Grammarly



What Is Grammarly?

Our mission is to improve lives by improving communication.

— Grammarly

What Is Grammarly?



Al-powered assistant that makes communication:

clearmistake-freeeffective

What Is Grammarly?

Available for:



Freemium product (since Jan, 2015)



Over 10 million **active users** in Chrome Store alone





Deep intellectual property (multiple patents pending, NLP experts)

Grammarly Team



2. Real-world NLP applications

Natural Language Processing



Goal: have computers *understand* natural language in order to perform useful tasks



Types of NLP Applications:

- Analysis
- Transformation

Spam filtering



Spam filtering Sentiment analysis



Spam filtering Sentiment analysis Sarcasm detection



Spam filtering Sentiment analysis Sarcasm detection Search engines ... and many more!



NLP applications: transformation

Machine Translation

...

Translate

English	Spanish	French	German - detected	*	*=+
Die Vo Staatsv Teilgeb	lkswirtsch wissensc biet der W	haftslehr haften o /irtschaf	e (auch National der Sozialökonor tswissenschaft.	okonomie, Wirtsc nie, kurz VWL), is	haftliche × st ein
4) /					167/5000
English	Spanish	Arabic	• Translate		
The ec econor	onomics nics, eco	of econo nomics,	omics (including e economics, econ	conomics, econo omics) is a part o	omics, of economics.
☆ 匝	4)				Suggest an edit



Machine Translation

Summarization



Machine Translation

Summarization

Speech to Text / Text to Speech



Machine Translation Summarization Speech to Text / Text to Speech Question Answering



Machine Translation Summarization Speech to Text / Text to Speech Question Answering Error correction



Some statistics

Non-natives make 1 mistake every 10 words.

Most popular errors:

- spelling
- preposition choice
- missing article
- missing or redundant punctuation
- word choice

Some statistics

Grammarly:



COMPUTATIONAL LINGUIST



New feature lifecycle:

- Research the problem
- Decide on the approach
- Develop algorithm
- Test thoroughly
- Release
- Monitor feedback



Mistake: Missing article

I borrowed (an/the) excellent book from my friend.

1. Research the problem

• Grammar rules

• A or an: I left an excellent book on the

table.

• Countable or uncountable: This is the best

pizza known to man.

• Exceptions

• British English vs American English

AmE: an herb

BrE: a herb



1. Research the problem

2. Decide on the approach



1. Research the problem

2. Decide on the approach

3. Develop algorithm

I borrowed (an/the) excellent book from my friend.

ALGORITHM:

- 1. For every *noun phrase* in the sentence
- 2. Check if there is no article modifying the noun
- 3. Check if noun is singular
- 4. Check if noun is countable
- 5. Decide if noun requires "A" or "AN"
- 6. Add article to the *noun phrase*

- 1. Research the problem
- 2. Decide on the approach
- 3. Develop algorithm

How do we formalize the syntactic structure?



Example of a tree:

- constituency tree
 - every token is a part of some phrase constituent (parent node)
 - o shows relations among the constituents
 - o hierarchical structure





- 1. Research the problem
- 2. Decide on the approach
- 3. Develop algorithm

Not that easy!

This is the best pizza known to man! This is the best pizza known to **a** man! This is the best pizza known to **the** man!

- 1 💿 [countable] an adult male human
- 2 o [uncountable] humans as a group or from a particular period of history

- 1. Research the problem
- 2. Decide on the approach
- 3. Develop algorithm

N-grams

Sequences of elements and their frequencies:

- unigrams, bigrams, 3-grams, 4-grams, ... n-grams
- at different language levels
 - token ngrams:
 - ("handsome", "man"): 160,000
 - ("pretty", "man"): 5,000
 - character ngrams
 - "st": 14,000; "ct": 4,000; "str": 1,500;

1. Research the problem

2. Decide on the approach

3. Develop algorithm

Google Books Ngram Viewer

displays a graph showing how phrases have occurred in a corpus of books (e.g., "British English", "English Fiction", "French") over the selected years.

Google Books Ngram Viewer

Google Books Ngram Viewer

(click on line/label for focus)

1. Research the problem

2. Decide on the approach

3. Develop algorithm

Magic ;)

1. Research the problem

- 2. Decide on the approach
- 3. Develop algorithm
- 4. Test thoroughly

Where to get sentences to test the algorithm?

- Books
- Examples from grammar rules
- Imagination
- Corpora!

- **1.** Research the problem
- 2. Decide on the approach
- 3. Develop algorithm
- 4. Test thoroughly

Corpus is a large collection of texts.

1. Research the problem

2. Decide on the approach

- 3. Develop algorithm
- 4. Test thoroughly

Types of corpora:

- Monolingual vs bi/multilingual
- Written vs spoken
- General vs specialized
- Raw text vs tagged (lemmatized, tagged, parsed)
- Synchronic vs diachronic

Research the problem

Examples of corpora:

Corpus of Contemporary American English

- 2. Decide on the approach

1.

- 3. Develop algorithm
- 4. Test thoroughly

Chart Collocates Compare KWIC List at* book art.ALL Find matching strings Reset Sections Texts/Virtual Sort/Limit Options

Examples of corpora:

SEARCH

Corpus of Contemporary American English i 📄 🛃 🚰

1. Research the problem

2. Decide on the approach

3. Develop algorithm

4. Test thoroughly

EE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP]						
		CONTEXT	FREQ			
1		THE BOOK	34323			
2		A BOOK	19434			
3		EVERY BOOK	273			
4		NO BOOK	147			

FREQUENCY

CONTEXT

- 1. Research the problem
- 2. Decide on the approach
- 3. Develop algorithm
- 4. Test thoroughly
- 5. Release

Release

I borrowed excellent book from my friend.

an excellent or the excellent

The noun phrase **excellent book** seems to be missing a determiner before it. Consider adding an article.

MORE

× IGNORE

^ X

Takeaways

Takeaways (1):

- Use corpora to get real-world language usage examples and test language hypotheses. Links to online available corpora:
 - 1. <u>Netspeak</u> an online tool which allows you to find any word in a search phrase. It is particularly handy when you have doubts about how a phrase is formed or cannot find the right word.
 - 2. <u>Corpus of Contemporary American English (COCA)</u> the name of this resource speaks for itself. It is not only a collection of AmE texts, but also a powerful search tool.
 - **3.** <u>British National Corpus</u> 100 million word collection of samples of written and spoken language.

Takeaways (2):

2. Use <u>Google Ngrams</u> to check idiomaticity and frequency of phrases.

Thank you!

Any kwestions?

kwestions \rightarrow **questions**

Oksana Kunikevych oksana.kunikevych@grammarly.com

Join us on social media

All our social media can be found in description of every video on our YouTube channel

The next webinar

The Platinum Sponsor of UTICamp-2019

