L90: Overview of Natural Language Processing

L90: Overview of Natural Language Processing

L95
L101
R250
...

breadth over depth

L90: Overview of Natural Language Processing

L95
L101
R250
. . .

breadth over depth

L90: Overview of Natural Language Processing

L95
L101
R250
...

English, Welsh, Afrikaans, Mandarin, ...
English as a Second Language, ...
Sanskrit, ...
?dolphin language

breadth over depth

computational models

L90: Overview of Natural Language Processing

L95
L101
R250
...

English, Welsh, Afrikaans, Mandarin, ...
English as a Second Language, ...
Sanskrit, ...
?dolphin language

breadth over depth

computational models

L90: Overview of Natural Language Processing

L95
L101
R250
...

English, Welsh, Afrikaans, Mandarin, ...
English as a Second Language, ...
Sanskrit, ...
?dolphin language

Lecture 1: Overview of Overview of Natural Language Processing

1. What does it mean to know a language?
2. Form transformation
3. Why NLP is hard?
4. Scope of NLP

based on
Ann Copestake's
previous lecture

What does it mean to know a language?

universal translator

www.youtube.com/watch?v=wtAmPX1Itr0

What does it mean to *know* a language?

Some yinkish dripners blorked quastofically into the nindin with the pidibs.

the example is partly from A Carnie's *Syntax: A Generative Introduction*

What does it mean to *know* a language?

Some yinkish dripners blorked quastofically into the nindin with the pidibs.

the example is partly from A Carnie's *Syntax: A Generative Introduction*

- there was a BLORK event;

What does it mean to *know* a language?

Some yinkish dripners blorked quastofically into the nindin with the pidibs.

the example is partly from A Carnie's *Syntax: A Generative Introduction*

- there was a BLORK event;
- it happened in the PAST;

What does it mean to *know* a language?

Some yinkish dripners blorked quastofically into the nindin with the pidibs.

the example is partly from A Carnie's *Syntax: A Generative Introduction*

- there was a BLORK event;
- it happened in the PAST;
- the AGENT of BLORK is dripners;

What does it mean to *know* a language?

Some yinkish dripners blorked quastofically into the nindin with the pidibs.

the example is partly from A Carnie's *Syntax: A Generative Introduction*

- there was a BLORK event;
- it happened in the PAST;
- the AGENT of BLORK is dripners;
- the dripners were yinkish;

What does it mean to *know* a language?

Some yinkish dripners blorked quastofically into the nindin with the pidibs.

the example is partly from A Carnie's *Syntax: A Generative Introduction*

- there was a BLORK event;
- it happened in the PAST;
- the AGENT of BLORK is dripners;
- the dripners were yinkish;
- some but not all dripners blorked;

What does it mean to *know* a language?

Some yinkish dripners blorked quastofically into the nindin with the pidibs.

the example is partly from A Carnie's *Syntax: A Generative Introduction*

- there was a BLORK event;
- it happened in the PAST;
- the AGENT of BLORK is dripners;
- the dripners were yinkish;
- some but not all dripners blorked;
- *with the pidibs* may talk about *nindin* or BLORK;

Structuring a number

010101010101010101010101010101010101010101010101010101010101010...

Structuring a number

010101010101010101010101010101010101010101010101010101010101010...

011010100000010011110011001100111111100111011110011001001000...

Structuring a number

0101010101010101010101010101010101010101010101010101010101010...

011010100000100111100110011001111111001110111100110010010000...

$$\sqrt{2} - 1$$

Structuring a sentence

Some yinkish dripners blorked quastofically into the nindin with the pidibs

Structuring a sentence

Some yinkish dripners blorked quastofically into the nindin with the pidibs

dripner -s blork -ed pidib -s

Structuring a sentence

Some yinkish dripners blorked quastofically into the nindin with the pidibs

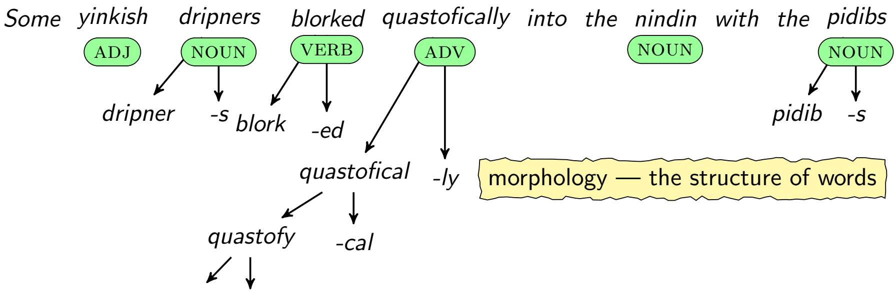
dripner *-s* *blork* *-ed* *pidib* *-s*

quastofical *-ly*

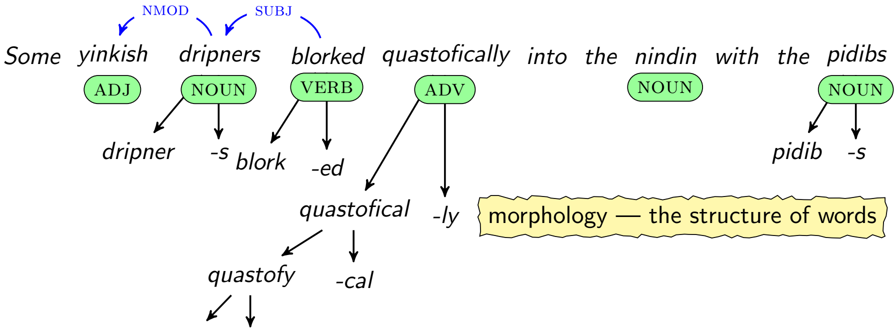
morphology — the structure of words

quastofy *-cal*

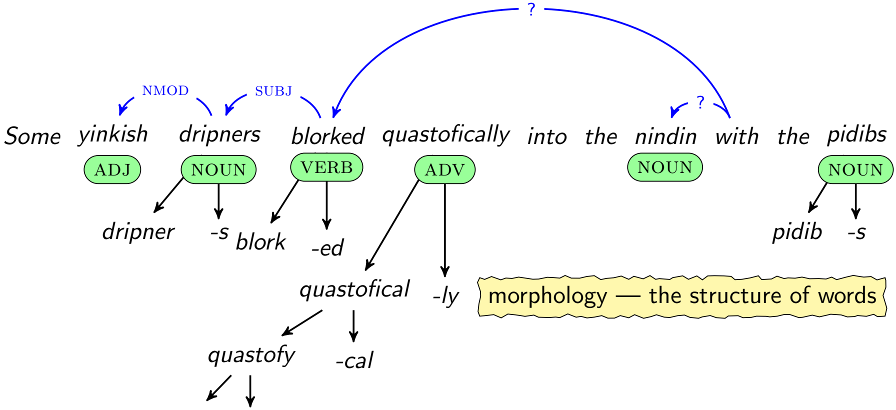
Structuring a sentence



Structuring a sentence

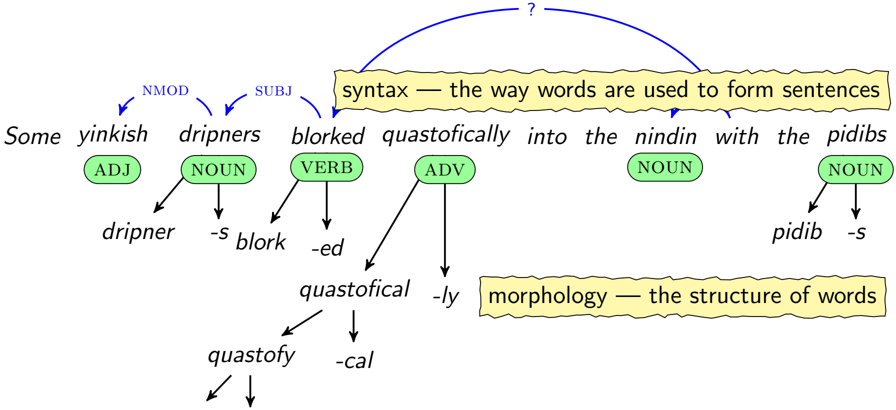


Structuring a sentence

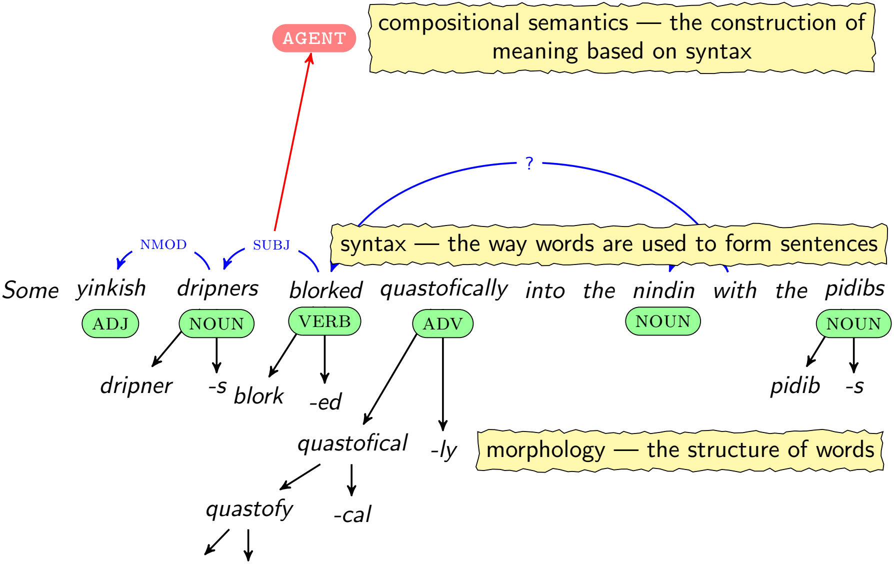


morphology — the structure of words

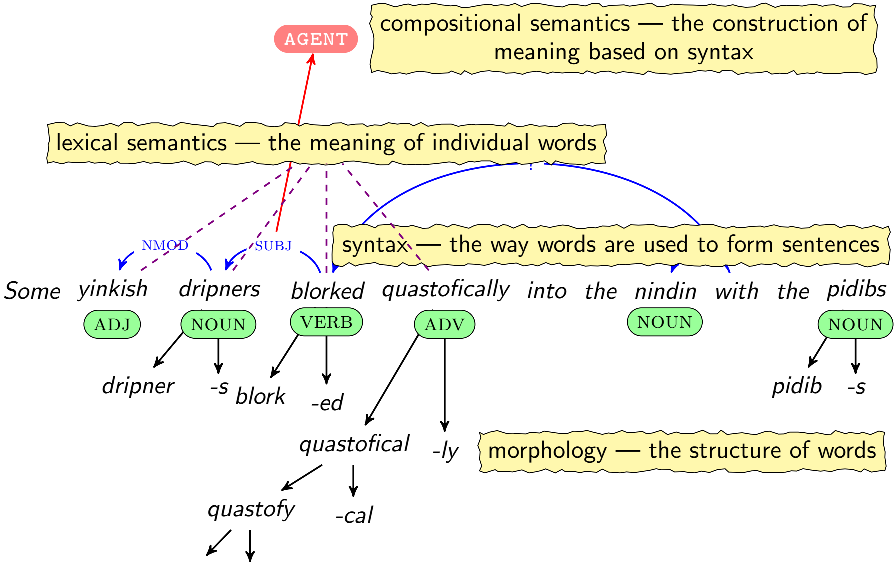
Structuring a sentence



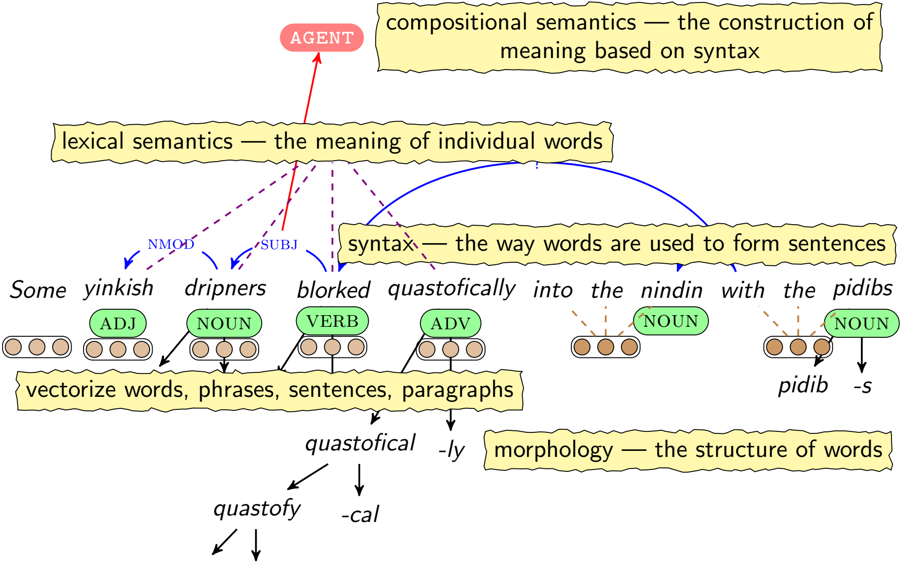
Structuring a sentence



Structuring a sentence



Structuring a sentence



NLP: the computational modelling of human language

- *Morphology* — the structure of words: lecture 2.
- *Syntax* — the way words are used to form phrases: lectures 3, 5 and 6.
- *Semantics*
    - *Compositional semantics* — the construction of meaning based on syntax: lecture 9.
    - *Lexical semantics* — the meaning of individual words: lecture 8 (sort of) and 10.
- *Pragmatics* — meaning in context: lecture 11.
- *Language generation* — lecture 12.

- *Symbolic models* — finite-state machines and context-free grammars: lecture 2 and 5.
- *Statistical models* — classification: lecture 3.
- *Neural models* — (sequential) classification: lecture 4 and 7.

Form transformation

$$\text{natural language expressions} \underset{\text{production}}{\overset{\text{comprehension}}{\rightleftarrows}} \mathcal{R} \left\{ \begin{array}{l} \text{morphological structure} \\ \text{syntactic structure} \\ \text{semantic structure} \\ \text{discourse structure} \\ \text{application-related structure} \end{array} \right.$$

Popular representations in NLP

CoNLL shared tasks

- The SIGNLL Conference on Computational Natural Language Learning
- https://www.conll.org/previous-tasks

2019	Cross-Framework Meaning Representation Parsing
2018/2017	Multilingual Parsing from Raw Text to Universal Dependencies
2018/2017	Universal Morphological Reinflection
2016/2016	(Multilingual) Shallow Discourse Parsing
2014/2013	Grammatical Error Correction
2012/2011	Modelling (Multilingual) Unrestricted Coreference in OntoNotes
2010	Hedge Detection
2009/2008	Syntactic and Semantic Dependencies in English/Multiple Languages
2007/2006	Multi-Lingual Dependency Parsing (Domain Adaptation)
2005/2004	Semantic Role Labeling
2003/2002	Language-Independent Named Entity Recognition
2001	Clause Identification
2000	Chunking
1999	NP Bracketing

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

input words

1	U.N.
2	official
3	Ekeus
4	heads
5	for
6	Baghdad
7	.

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

input words → part-of-speech tags

1	U.N.	NNP
2	official	NN
3	Ekeus	NNP
4	heads	VBZ
5	for	IN
6	Baghdad	NNP
7	.	.

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

part-of-speech tags

input words

1	U.N.	NNP	I-NP
2	official	NN	I-NP
3	Ekeus	NNP	I-NP
4	heads	VBZ	I-VP
5	for	IN	I-PP
6	Baghdad	NNP	I-NP
7	.	.	O

syntactic chunks

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

part-of-speech tags

input words

1	U.N.	NNP	I-NP
2	official	NN	I-NP
3	Ekeus	NNP	I-NP
4	heads	VBZ	I-VP
5	for	IN	I-PP
6	Baghdad	NNP	I-NP
7	.	.	O

NP

syntactic chunks

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

part-of-speech tags

input words

1	U.N.		NNP	I-NP
2	official	NP	NN	I-NP
3	Ekeus		NNP	I-NP
4	heads	VP	VBZ	I-VP
5	for	PP	IN	I-PP
6	Baghdad	NP	NNP	I-NP
7	.		.	O

syntactic chunks

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

part-of-speech tags

input words

1	U.N.	NNP	I-NP	I-ORG
2	official	NN	I-NP	O
3	Ekeus	NNP	I-NP	I-PER
4	heads	VBZ	I-VP	O
5	for	IN	I-PP	O
6	Baghdad	NNP	I-NP	I-LOC
7	.	.	O	O

syntactic chunks

named entities

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

	input words	part-of-speech tags	syntactic chunks	named entities
1	U.N.	NNP	I-NP	I-ORG
2	official	NN	I-NP	O
3	Ekeus	NNP	I-NP	I-PER
4	heads	VBZ	I-VP	O
5	for	IN	I-PP	O
6	Baghdad	NNP	I-NP	I-LOC
7	.	.	O	O

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

1	U.N.	NNP	I-NP	I-ORG	3
2	official	NN	I-NP	O	3
3	Ekeus	NNP	I-NP	I-PER	4
4	heads	VBZ	I-VP	O	0
5	for	IN	I-PP	O	6
6	Baghdad	NNP	I-NP	I-LOC	4
7	.	.	O	O	4

input words

part-of-speech tags

syntactic dependencies

syntactic chunks

named entities

CoNLL ST 1999/2000/2002/2003/2006/2007/2017/2018

input words | part-of-speech tags | syntactic chunks | named entities | syntactic dependencies

1	U.N.	NNP	I-NP	I-ORG	3
2	official	NN	I-NP	O	3
3	Ekeus	NNP	I-NP	I-PER	4
4	heads	VBZ	I-VP	O	0
5	for	IN	I-PP	O	6
6	Baghdad	NNP	I-NP	I-LOC	4
7	.	.	O	O	4

ORG NP PER VP PP LOC

SUBJ OBL

Querying a knowledge base

User query: *Has my order number 4291 been shipped yet?*

Database:

ORDER

Order number	**Date ordered**	**Date shipped**
4290	2/2/13	2/2/13
4291	2/2/13	2/2/13
4292	2/2/13	

USER: *Has my order number 4291 been shipped yet?*

DB QUERY: `order(number=4291,date_shipped=?)` $\triangleright\mathcal{R}$

RESPONSE: *Order number 4291 was shipped on 2/2/13*

Instructions

Natural language: *Go to the third junction and take a left*

Programming language:

```
(do-seq(do-n-times 3
  (move-to forward-loc
    (do-until
      (junction current-loc
        (move-to forward-loc))))
  (turn-right))
```

Many other application-based representations

P Dasigi, S Iyer, A Suhr, M Gardner and L Zettlemoyer. ACL 2018 tutorial on neural semantic parsing.
`https://github.com/allenai/acl2018-semantic-parsing-tutorial/`

Why NLP is hard?

Why is this difficult?

similar strings mean different things

(1) a. How fast is the RTX 30?

b. How fast will my RTX 30 arrive?

c. Please tell me when I can expect the RTX 30 I ordered.

different strings mean the same thing

Why is this difficult?

(2) a. Do you sell Sony laptops <u>and</u> disk drives?
b. Do you sell (Sony (laptops <u>and</u> disk drives))?
c. Do you sell (Sony laptops) <u>and</u> (disk drives)?

ambiguity

$$\boxed{2 \times (3 + 4) = 2 \times 3 + 2 \times 4} \text{ vs } \boxed{2 \times 3 + 4}$$

Wouldn't it be better if . . . ?

The properties which make natural language difficult to process are essential to human communication:

- Flexible
- Learnable but compact
- Emergent, evolving systems

Synonymy and ambiguity go along with these properties.

Natural language communication can be indefinitely precise:

- Ambiguity is mostly local (for humans)

Scope of NLP

A typical call-for-paper (1)

ACL (=Annual Meeting of the Association for Computational Linguistics) 2020 has the goal of a broad technical program. Relevant topics for the conference include, but are not limited to, the following areas:

- Theory and Formalism in NLP (Linguistic and Mathematical)
- Machine Learning for NLP
- Cognitive Modeling and Psycholinguistics

- Phonology, Morphology and Word Segmentation
- Syntax: Tagging, Chunking and Parsing
- Semantics: Lexical
- Semantics: Sentence Level
- Semantics: Textual Inference and Other Areas of Semantics
- Discourse and Pragmatics
- Generation

- Resources and Evaluation
- Interpretability and Analysis of Models for NLP

A typical call-for-paper (1)

ACL (=Annual Meeting of the Association for Computational Linguistics) 2020 has the goal of a broad technical program. Relevant topics for the conference include, but are not limited to, the following areas:

- Theory and Formalism in NLP (Linguistic and Mathematical)
- Machine Learning for NLP
- Cognitive Modeling and Psycholinguistics

- Phonology, Morphology and Word Segmentation
- Syntax: Tagging, Chunking and Parsing
- Semantics: Lexical
- Semantics: Sentence Level
- Semantics: Textual Inference and Other Areas of Semantics
- Discourse and Pragmatics
- Generation

- Resources and Evaluation
- Interpretability and Analysis of Models for NLP

A typical call-for-paper (1)

ACL (=Annual Meeting of the Association for Computational Linguistics) 2020 has the goal of a broad technical program. Relevant topics for the conference include, but are not limited to, the following areas:

- Theory and Formalism in NLP (Linguistic and Mathematical)
- Machine Learning for NLP
- Cognitive Modeling and Psycholinguistics

- Phonology, Morphology and Word Segmentation
- Syntax: Tagging, Chunking and Parsing
- Semantics: Lexical
- Semantics: Sentence Level
- Semantics: Textual Inference and Other Areas of Semantics
- Discourse and Pragmatics
- Generation

- Resources and Evaluation
- Interpretability and Analysis of Models for NLP

A typical call-for-paper (1)

ACL (=Annual Meeting of the Association for Computational Linguistics) 2020 has the goal of a broad technical program. Relevant topics for the conference include, but are not limited to, the following areas:

- Theory and Formalism in NLP (Linguistic and Mathematical)
- Machine Learning for NLP
- Cognitive Modeling and Psycholinguistics

- Phonology, Morphology and Word Segmentation
- Syntax: Tagging, Chunking and Parsing
- Semantics: Lexical
- Semantics: Sentence Level
- Semantics: Textual Inference and Other Areas of Semantics
- Discourse and Pragmatics
- Generation

- Resources and Evaluation
- Interpretability and Analysis of Models for NLP

A typical call-for-paper (2)

ACL 2020 has the goal of a broad technical program. Relevant topics for the conference include, but are not limited to, the following areas:

- Language Grounding to Vision, Robotics and Beyond
- Speech and Multimodality

- Information Extraction
- Information Retrieval and Text Mining
- Machine Translation
- Question Answering
- Dialogue and Interactive Systems
- Summarization
- Sentiment Analysis, Stylistic Analysis, and Argument Mining
- (other) NLP Applications

- Computational Social Science and Social Media
- Ethics and NLP

A typical call-for-paper (2)

ACL 2020 has the goal of a broad technical program. Relevant topics for the conference include, but are not limited to, the following areas:

- Language Grounding to Vision, Robotics and Beyond
- Speech and Multimodality

- Information Extraction
- Information Retrieval and Text Mining
- Machine Translation
- Question Answering
- Dialogue and Interactive Systems
- Summarization
- Sentiment Analysis, Stylistic Analysis, and Argument Mining
- (other) NLP Applications

- Computational Social Science and Social Media
- Ethics and NLP

A typical call-for-paper (2)

ACL 2020 has the goal of a broad technical program. Relevant topics for the conference include, but are not limited to, the following areas:

- Language Grounding to Vision, Robotics and Beyond
- Speech and Multimodality

- Information Extraction
- Information Retrieval and Text Mining
- Machine Translation
- Question Answering
- Dialogue and Interactive Systems
- Summarization
- Sentiment Analysis, Stylistic Analysis, and Argument Mining
- (other) NLP Applications

- Computational Social Science and Social Media
- Ethics and NLP

A typical call-for-paper (2)

ACL 2020 has the goal of a broad technical program. Relevant topics for the conference include, but are not limited to, the following areas:

- Language Grounding to Vision, Robotics and Beyond
- Speech and Multimodality

- Information Extraction
- Information Retrieval and Text Mining
- Machine Translation
- Question Answering
- Dialogue and Interactive Systems
- Summarization
- Sentiment Analysis, Stylistic Analysis, and Argument Mining
- (other) NLP Applications

- Computational Social Science and Social Media
- Ethics and NLP

iphone 11 review

www.cnet.com › reviews › iphone-11-2019-battery-deep-...

Apple iPhone 11 review: The best $700 iPhone Apple has ...

Apple **iPhone 11 review**: The best $700 **iPhone** Apple has ever made. The good Even faster speed, improved ...

Mar 29, 2020 - ★★★★★ Rating: 4.5 - Review by Patrick Holland

www.techradar.com › reviews › iphone-11-review ▾

iPhone 11 review | TechRadar

May 4, 2020 - The **iPhone 11** is something of a surprise - it brings more advanced technology (namely in the camera capabilities and the processing power under the hood) and yet offers it for less than the **iPhone** XR cost in 2018. It combines a large 6.1-inch display with a premium-feeling body, and comes in an array of colors too.

★★★★★ Rating: 4.5 - Review by Gareth Beavis

www.youtube.com › watch

Apple iPhone 11 review - YouTube

The Apple **iPhone 11** is the cheapest of the three new iPhones but does that mean you are giving up a lot? Our ...

Oct 3, 2019 - Uploaded by GSMArena Official

iPhone 11 review

The iPhone 11 is something of a surprise — it brings more advanced technology (namely in the camera capabilities and the processing power under the hood) and yet offers it for less than the iPhone XR cost in 2018. It combines a large 6.1-inch display with a premium-feeling body, and comes in an array of colors too.

The most eye-catching feature of the new iPhone is to the imaging capabilities: with two sensors on the rear, you can now take wider-angle snaps alongside the 'normal' main images. These sensors are 12MP each, and are raised from the rear of the phone in a square glass enclosure — which we're not enamored with visually.

The night mode is the most impressive part of the iPhone 11 imaging quality, bringing brightness and clarity to impossibly dark scenes, and the Portrait mode, defocusing the background, is improved on the new iPhone too.

from `www.techradar.com/reviews/iphone-11-review`

Opinion mining: what do they think about me?

Scan documents (webpages, tweets etc) for positive and negative opinions on people, products, etc.

Find all references to entity in some document collection: list as positive, negative (possibly with strength) or neutral.

Construct summary report plus examples (text snippets).

Fine-grained classification:
e.g., for phone, opinions about: design, performance, camera, battery life
. . .

Sentiment classification: the research task

Full task

Information retrieval, cleaning up text structure, named entity recognition, identification of relevant parts of text. Evaluation by humans.

Sentiment classification: the research task

Full task

Information retrieval, cleaning up text structure, named entity recognition, identification of relevant parts of text. Evaluation by humans.

Research task

Preclassified documents, topic known, opinion in text along with some straightforwardly extractable score.

Example

Movie review *corpus* (Pang et al, 2002): strongly positive or negative reviews from IMDb, 50:50 split, with rating score.

An American Werewolf in London (1981)	Rating: 9/10
Ooooo. Scary. *The old adage of the simplest ideas being the best is once again demonstrated in this, one of the most entertaining films of the early 80's, and almost certainly Jon Landis' best work to date. The script is light and witty, the visuals are great and the atmosphere is top class. Plus there are some great freeze-frame moments to enjoy again and again. Not forgetting, of course, the great transformation scene which still impresses to this day.* *In Summary: Top banana*	

Bag of words technique

An American Werewolf in London (1981)	Rating: 9/10
Ooooo. Scary. *The old adage of the simplest ideas being the best is once again demonstrated in this, one of the most entertaining films of the early 80's, and almost certainly Jon Landis' best work to date. The script is light and witty, the visuals are great and the atmosphere is top class. Plus there are some great freeze-frame moments to enjoy again and again. Not forgetting, of course, the great transformation scene which still impresses to this day.* *In Summary: Top banana*	

- Treat the reviews as collections of individual words.

Bag of words technique

An American Werewolf in London (1981)	Rating: 9/10
Ooooo. Scary. *The old adage of the simplest ideas being the best is once again demonstrated in this, one of the most entertaining films of the early 80's, and almost certainly Jon Landis' best work to date. The script is light and witty, the visuals are great and the atmosphere is top class. Plus there are some great freeze-frame moments to enjoy again and again. Not forgetting, of course, the great transformation scene which still impresses to this day.* *In Summary: Top banana*	

- Treat the reviews as collections of individual words.
- Classify reviews according to positive or negative words.

Bag of words technique

An American Werewolf in London (1981)	Rating: 9/10
Ooooo. Scary. *The old adage of the simplest ideas being the best is once again demonstrated in this, one of the most entertaining films of the early 80's, and almost certainly Jon Landis' best work to date. The script is light and witty, the visuals are great and the atmosphere is top class. Plus there are some great freeze-frame moments to enjoy again and again. Not forgetting, of course, the great transformation scene which still impresses to this day.* *In Summary: Top banana*	

- Treat the reviews as collections of individual words.
- Classify reviews according to positive or negative words.
- Could use word lists prepared by humans, but machine learning based on a portion of the corpus (training set) is preferable.

Bag of words technique

An American Werewolf in London (1981)	Rating: 9/10

Ooooo. Scary.
The old adage of the simplest ideas being the best is once again demonstrated in this, one of the most entertaining films of the early 80's, and almost certainly Jon Landis' best work to date. The script is light and witty, the visuals are great and the atmosphere is top class. Plus there are some great freeze-frame moments to enjoy again and again. Not forgetting, of course, the great transformation scene which still impresses to this day.
In Summary: Top banana

- Treat the reviews as collections of individual words.
- Classify reviews according to positive or negative words.
- Could use word lists prepared by humans, but machine learning based on a portion of the corpus (training set) is preferable.
- Use human rankings for training and evaluation.

Bag of words technique

An American Werewolf in London (1981)	Rating: 9/10

Ooooo. Scary.

The old adage of the simplest ideas being the best is once again demonstrated in this, one of the most entertaining films of the early 80's, and almost certainly Jon Landis' best work to date. The script is light and witty, the visuals are great and the atmosphere is top class. Plus there are some great freeze-frame moments to enjoy again and again. Not forgetting, of course, the great transformation scene which still impresses to this day.

In Summary: Top banana

- Treat the reviews as collections of individual words.
- Classify reviews according to positive or negative words.
- Could use word lists prepared by humans, but machine learning based on a portion of the corpus (training set) is preferable.
- Use human rankings for training and evaluation.
- Pang et al (2002): Chance success is 50% (corpus artificially balanced), bag-of-words gives 80%.

Some sources of errors for bag-of-words

- Negation:

  Ridley Scott has never directed a bad film.

Some sources of errors for bag-of-words

- Negation:

  Ridley Scott has never directed a bad film.

Some sources of errors for bag-of-words

- Negation:

Ridley Scott has never ⚠ directed a bad 👎 film.

Some sources of errors for bag-of-words

- Negation:

  Ridley Scott has never directed a bad film.

- Overfitting the training data:
  e.g., if training set includes a lot of films from before 2005, *Ridley* may be a strong positive indicator, but then we test on reviews for 'Kingdom of Heaven'?

Some sources of errors for bag-of-words

- Negation:

  Ridley Scott has never directed a bad film.

- Overfitting the training data:
  e.g., if training set includes a lot of films from before 2005, *Ridley* may be a strong positive indicator, but then we test on reviews for 'Kingdom of Heaven'?
- Comparisons and contrasts.

Contrasts in the discourse

This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.

Contrasts in the discourse

This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.

Contrasts in the discourse

This film should be brilliant👍. It sounds like a great👍 plot, the actors are first grade👍, and the supporting cast is good👍 as well, and Stallone is attempting to deliver a good👍 performance. However, it can't hold up.👎

Contrasts in the discourse

This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.

AN AMERICAN WEREWOLF IN PARIS is a failed attempt ... Julie Delpy is far too good for this movie. She imbues Serafine with spirit, spunk, and humanity. This isn't necessarily a good thing, since it prevents us from relaxing and enjoying AN AMERICAN WEREWOLF IN PARIS as a completely mindless, campy entertainment experience. Delpy's injection of class into an otherwise classless production raises the specter of what this film could have been with a better script and a better cast ... She was radiant, charismatic, and effective ...

Doing sentiment classification ‘properly’?

- Morphology, syntax and compositional semantics:
  who is talking about what, what terms are associated with what, tense
  ...

Doing sentiment classification ‘properly’?

- Morphology, syntax and compositional semantics:
  who is talking about what, what terms are associated with what, tense
  ...
- Lexical semantics:
  are words positive or negative in this context? Word senses (e.g., *spirit*)?

Doing sentiment classification 'properly'?

- Morphology, syntax and compositional semantics:
  who is talking about what, what terms are associated with what, tense . . .
- Lexical semantics:
  are words positive or negative in this context? Word senses (e.g., *spirit*)?
- Pragmatics and discourse structure:
  what is the topic of this section of text? Pronouns and definite references.

Doing sentiment classification 'properly'?

- Morphology, syntax and compositional semantics:
  who is talking about what, what terms are associated with what, tense ...
- Lexical semantics:
  are words positive or negative in this context? Word senses (e.g., *spirit*)?
- Pragmatics and discourse structure:
  what is the topic of this section of text? Pronouns and definite references.
- Getting all this to work well on arbitrary text is very hard.

Doing sentiment classification ‘properly’?

- Morphology, syntax and compositional semantics:
  who is talking about what, what terms are associated with what, tense
  . . .
- Lexical semantics:
  are words positive or negative in this context? Word senses (e.g., *spirit*)?
- Pragmatics and discourse structure:
  what is the topic of this section of text? Pronouns and definite references.
- Getting all this to work well on arbitrary text is very hard.
- Ultimately the problem is AI-complete, but can we do well enough for NLP to be useful?

General comments

- Even *simple* applications might need complex knowledge sources.

General comments

- Even *simple* applications might need complex knowledge sources.
- Applications cannot be 100% perfect.

General comments

- Even *simple* applications might need complex knowledge sources.
- Applications cannot be 100% perfect.
- Applications that are $< 100\%$ perfect can be useful.

General comments

- Even *simple* applications might need complex knowledge sources.
- Applications cannot be 100% perfect.
- Applications that are $< 100\%$ perfect can be useful.
- Aids to humans are easier than replacements for humans.

General comments

- Even *simple* applications might need complex knowledge sources.
- Applications cannot be 100% perfect.
- Applications that are $< 100\%$ perfect can be useful.
- Aids to humans are easier than replacements for humans.
- NLP interfaces compete with non-language approaches.

General comments

- Even *simple* applications might need complex knowledge sources.
- Applications cannot be 100% perfect.
- Applications that are $< 100\%$ perfect can be useful.
- Aids to humans are easier than replacements for humans.
- NLP interfaces compete with non-language approaches.
- Typically: shallow processing on arbitrary input or deep processing on narrow domains.

General comments

- Even *simple* applications might need complex knowledge sources.
- Applications cannot be 100% perfect.
- Applications that are $< 100\%$ perfect can be useful.
- Aids to humans are easier than replacements for humans.
- NLP interfaces compete with non-language approaches.
- Typically: shallow processing on arbitrary input or deep processing on narrow domains.
- Limited domain systems require expensive expertise to port or large amounts of (expensive) data.

General comments

- Even *simple* applications might need complex knowledge sources.
- Applications cannot be 100% perfect.
- Applications that are $< 100\%$ perfect can be useful.
- Aids to humans are easier than replacements for humans.
- NLP interfaces compete with non-language approaches.
- Typically: shallow processing on arbitrary input or deep processing on narrow domains.
- Limited domain systems require expensive expertise to port or large amounts of (expensive) data.
- External influences on NLP are very important.