Language Technologies Institute

Carnegie Mellon University

Multimodal Affective Computing

Lecture 6: Verbal Messages

Louis-Philippe Morency
Jeffrey Girard

Originally developed with help from
Stefan Scherer and Tadas Baltrušaitis

From Nonverbal to Verbal…

Internal Monologue (alt-text)

Oh right, eye contact. Ok, good, holding the eye contact ... holding ... still holding ... ok, too long! Getting weird! Quick, look thoughtfully into space and nod. Oh, dammit, said 'yeah' again!

Lecture Outline

- Linguistics and the study of language
- Word and lexical representations
  - Sentiment and topic analysis
  - LIWC and lexicons
  - Word2vec and word embeddings
- Language structure
  - Grammar, syntax and language models
- Discourse and dialogue analysis
  - Adjacency pairs, common ground
  - Speech and dialogue acts
- Turn-taking and conversation dynamics
  - Overlaps, interruptions, Backchannel, Disfluencies
  - Multi-party floor management
- Practical tools for automatic annotation

Natural Language and Verbal Messages

Communication, Natural Language and Machine Learning

- What is Natural Language?
- Natural Language is of obvious importance in human communication
- How can machines deal with natural language?

Levels of Analysis

Last week

sounds

- Phonetics
  - Articulation and perception of the sounds of human language
- Phonology
  - Sound patterns in a human language
- Morphology
  - Structure of words
- Syntax
  - Words are combined in a hierarchical structure
- Semantics
  - Meaning of symbols and their structure

meaning

- Pragmatics
  - Speaker’s intended meaning, goal
- Discourse analysis

Syntax, Semantic, Pragmatics

- **Syntax** concerns the proper ordering of words and its affect on meaning.
  - The dog bit the boy.
  - The boy bit the dog.
  - * Bit boy dog the the.
- **Semantics** concerns the (literal) meaning of words, phrases, and sentences.
  - “plant” as a photosynthetic organism
  - “plant” as a manufacturing facility
  - “plant” as the act of sowing
- **Pragmatics** concerns the overall communicative and social context and its effect on interpretation.
  - The ham sandwich wants another beer. (co-reference, anaphora)
  - John thinks vanilla. (ellipsis)

Semantics vs. **Pragmatics**

Semantics deals with the meaning or the meaning potential of words, phrases and sentences out of context (context-invariant, speaker-independent meaning) – propositional meaning

Pragmatics deals with the meaning of expressions (mainly utterances) in a particular context (with context-sensitive, speaker-dependent meaning) – "the interesting stuff about language."

(Lakoff 1993: 367)

(cp. Kortmann 2005: 14; chapters 6 and 7)

Sentence vs (Verbal) Utterance

- An utterance is any stretch of talk, by one person, before and after which there is silence on the part of that person.
- An utterance is the USE by a particular speaker, on a particular occasion, of a piece of language, such as a sequence of sentences, or a single phrase, or even a single word.

Sentence vs (Verbal) Utterance

- A SENTENCE is neither a physical event nor a physical object. It is, conceived abstractly, a string of words put together by the grammatical rules of a language. A sentence can be thought of as the IDEAL string of words behind various realizations in utterances and inscriptions.

Sentiment, Lexicons, and Topic Analysis

What is subjectivity?

- The **linguistic** expression of somebody's **opinions**, **sentiments**, emotions, evaluations, beliefs, speculations ***(private states)***
- **Private state:** state that is not open to objective observation or verification

  Quirk, Greenbaum, Leech, Svartvik (1985). *A Comprehensive Grammar of the English Language*.
- Subjectivity analysis classifies content in **objective** or **subjective**

Components of an opinion

- Basic components of an opinion:
  - Opinion holder: The person or organization that holds a specific opinion on a particular object.
  - Object: on which an opinion is expressed
  - Opinion: a view, attitude, or appraisal on an object from an opinion holder.

What is sentiment analysis?

- Also known as opinion mining
- Attempts to identify the opinion/sentiment that a person may hold towards an object
- It is a finer grain analysis compared to subjectivity analysis

Sentiment Analysis	Subjectivity analysis
Positive	Subjective
Negative	
Neutral	Objective

Example: Sentiment analysis of movie reviews

Traditional/Naive sentiment analysis = bag-of-words

Best movie of the year 

Slick and entertaining, despite a weak script 

Fun and sweet but ultimately unsatisfying 

- A count-based method would completely fail in predicting the last example

- More generally, a bag-of-words approach does not take into account sentence structure and semantic composition

Newer approach (neural network!) : http://nlp.stanford.edu/sentiment/

Main resources

- Lexicons
  - General Inquirer (Stone et al., 1966)
  - OpinionFinder lexicon (Wiebe & Riloff, 2005)
  - SentiWordNet (Esuli & Sebastiani, 2006)
  - LIWC (Pennebaker)

- Tools
  - LightSIDE
  - Stanford NLP toolbox
  - IBM Watson Tone Analyzer
  - Google Cloud Natural Language
  - Microsoft Azure Text Analytics

Topic Clustering - Example from a Science Magazine papers collection

universe	0.0439
galaxies	0.0375
clusters	0.0279
matter	0.0233
galaxy	0.0232
cluster	0.0214
cosmic	0.0137
dark	0.0131
light	0.0109
density	0.01

drug	0.0672
patients	0.0493
drugs	0.0444
clinical	0.0346
treatment	0.028
trials	0.0277
therapy	0.0213
trial	0.0164
disease	0.0157
medical	0.00997

cells	0.0675
stem	0.0478
human	0.0421
cell	0.0309
gene	0.025
tissue	0.0185
cloning	0.0169
transfer	0.0155
blood	0.0113
embryos	0.0111

sequence	0.0818
sequences	0.0493
genome	0.033
dna	0.0257
sequencing	0.0172
map	0.0123
genes	0.0122
chromosome	0.0119
regions	0.0119
human	0.0111

years	0.156
million	0.0556
ago	0.045
time	0.0317
age	0.0243
year	0.024
record	0.0238
early	0.0233
billion	0.0177
history	0.0148

bacteria	0.0983
bacterial	0.0561
resistance	0.0431
coli	0.0381
strains	0.025
microbiol	0.0214
microbial	0.0196
strain	0.0165
salmonella	0.0163
resistant	0.0145

male	0.0558
females	0.0541
female	0.0529
males	0.0477
sex	0.0339
reproductive	0.0172
offspring	0.0168
sexual	0.0166
reproduction	0.0143
eggs	0.0138

theory	0.0811
physics	0.0782
physicists	0.0146
einstein	0.0142
university	0.013
gravity	0.013
black	0.0127
theories	0.01
aps	0.00987
matter	0.00954

immune	0.0909
response	0.0375
system	0.0358
responses	0.0322
antigen	0.0263
antigens	0.0184
immunity	0.0176
immunology	0.0145
antibody	0.014
autoimmune	0.0128

stars	0.0524
star	0.0458
astrophys	0.0237
mass	0.021
disk	0.0173
black	0.0161
gas	0.0149
stellar	0.0127
astron	0.0125
hole	0.00824

Topic Clustering: Term-document matrix

Consider the frequencies of words in certain documents. From this information we can construct a vector for each word reflecting the corresponding frequencies:

	Document 1	Document 2	Document 3
bank	0	0	4 / 1
bass	2 / 0.447	4	0
cream	1 / 2	0	0
guitar	1	0	0
fisherman	0	3 / 1	0
money	0	1 / 0.447	2 / 0.894

Document 1 is about music instruments, document 2 about fishermen, and document 3 about financial institutions.

Topic Clustering: Latent Semantic Analysis (LSA)

- LSA aims to discover something about the meaning behind the words; about the topics in the documents.
- What is the difference between topics and words?
  - Words are observable
  - Topics are not. They are latent.
- Finding topics from the words in an automatic way
  - We can imagine them as a compression of words
  - A combination of words
  - LSA is one approach to do this

Toolbox: MALLET (document classification, topic modeling)
http://mallet.cs.umass.edu/

LIWC: Language Inquiry & Word Count

A large lexicon with an extensive list of topics and categories:

- Function words: *pronouns, preposition, negation…*
- Affect words: *positive, negative emotions*
- Social words: *family, friends, referents*
- Cognitive processes: *Insight, cause, …*
- Perceptual processes: *Seeing, hearing, feeling*
- Biological processes: *Body, health/illness,…*
- Drives and needs*: Affiliation, achievement, …*
- Time orientation: *past, present, future*
- Relativity: *motion, space, time*
- Personal concerns: *work, leisure, money, religion …*
- Informal speech: *swear words, fillers, assent,…*

URL: https://liwc.wpengine.com/

Word Representations

Possible ways of representing words

Given a text corpus containing 100,000 unique words

- Classic binary word representation: [0; 0; 0; 0;….; 0; 0; 1; 0;…; 0; 0]

  100,000d vector

  - Only non-zero at the index of the word

- Classic word feature representation: [5; 1; 0; 0;….; 0; 20; 1; 0;…; 3; 0]

  300d vector

  - Manually define 300 "good" features (e.g., ends on –ing)

- Learned word representation: [0,1; 0,0003; 0;….; 0,02; 0.08; 0,05]

  300d vector

  - Learn 300 features using an algorithm

What is the meaning of “bardiwac”?

- He handed her her glass of bardiwac.
- Beef dishes are made to complement the bardiwacs.
- Nigel staggered to his feet, face flushed from too much bardiwac.
- Malbec, one of the lesser-known bardiwac grapes, responds well to Australia’s sunshine.
- I dined off bread and cheese and this excellent bardiwac.
- The drinks were delicious: blood-red bardiwac as well as light, sweet Rhenish.

⇒ bardiwac is a heavy red alcoholic beverage made from grapes

How to learn (word) features/representations?

- **Distribution hypothesis:** Approximate the word meaning by its surrounding words

- Words used in a similar context will lie close together

He was $\boxed{\text{walking}}$ away because …
He was $\boxed{\text{running}}$ away because …

- **Instead of capturing co-occurrence counts directly, predict surrounding words of every word**

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c,j\neq 0}\log p(w_{t+j}|w_t)$$

Geometric interpretation

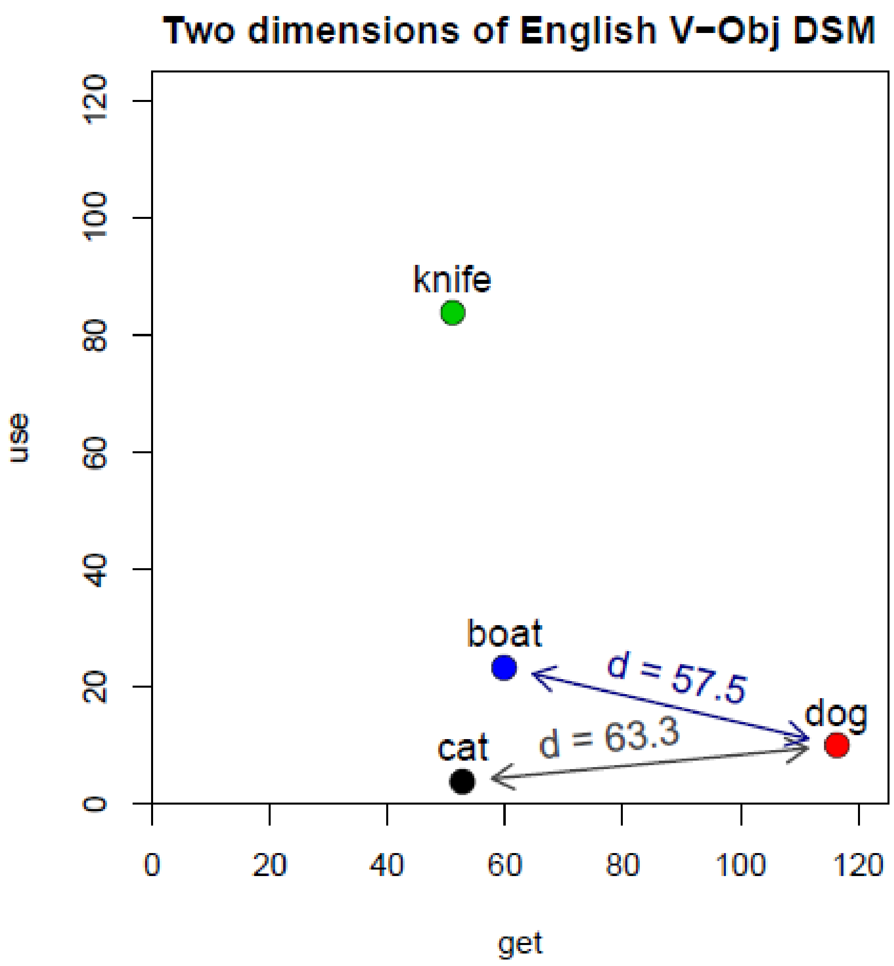
- row vector $\mathbf{x}_{dog}$ describes usage of word *dog* in the corpus
- can be seen as coordinates of point in *n*-dimensional Euclidean space R^n

	get	see	use	hear	eat	kill
knife	51	20	84	0	3	0
cat	52	58	4	4	6	26
dog	115	83	10	42	33	17
boat	59	39	23	4	0	0
cup	98	14	6	2	1	0
pig	12	17	3	2	9	27
banana	11	2	2	0	18	0

co-occurrence matrix **M**

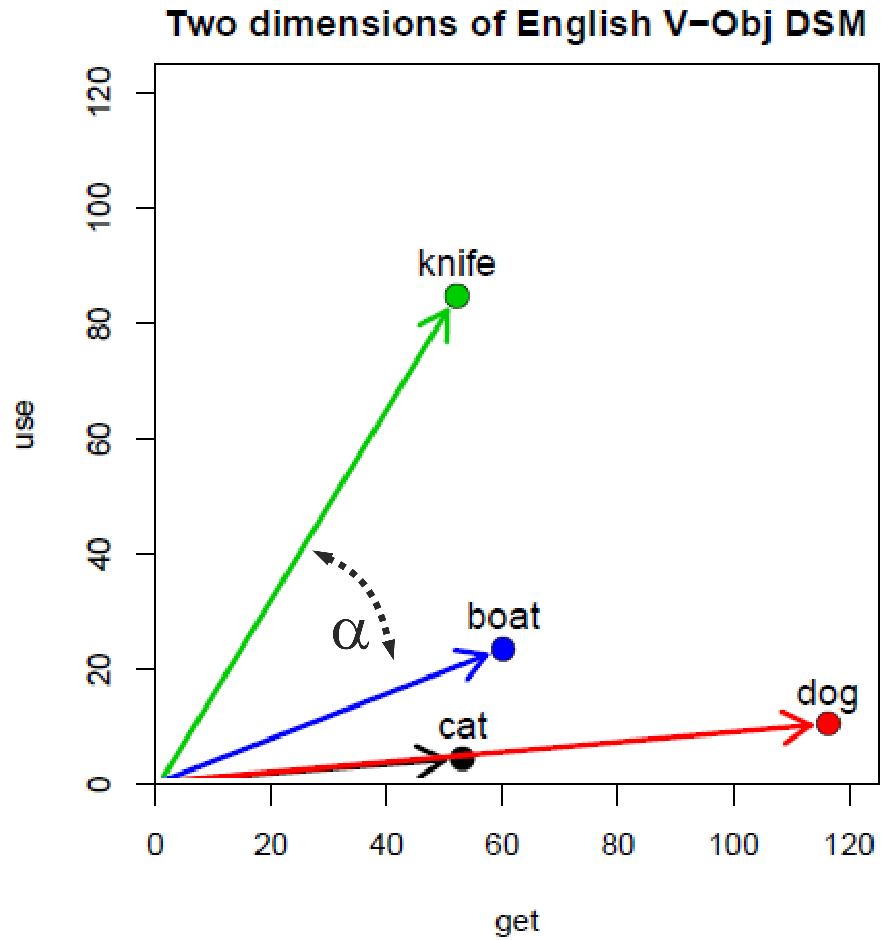
Distance and similarity

- illustrated for two dimensions: *get* and *use*: $\mathbf{x}_{dog} = (115, 10)$
- similarity = spatial proximity (Euclidean distance)
- location depends on frequency of noun ($f_{dog} \approx 2.7 \cdot f_{cat}$)

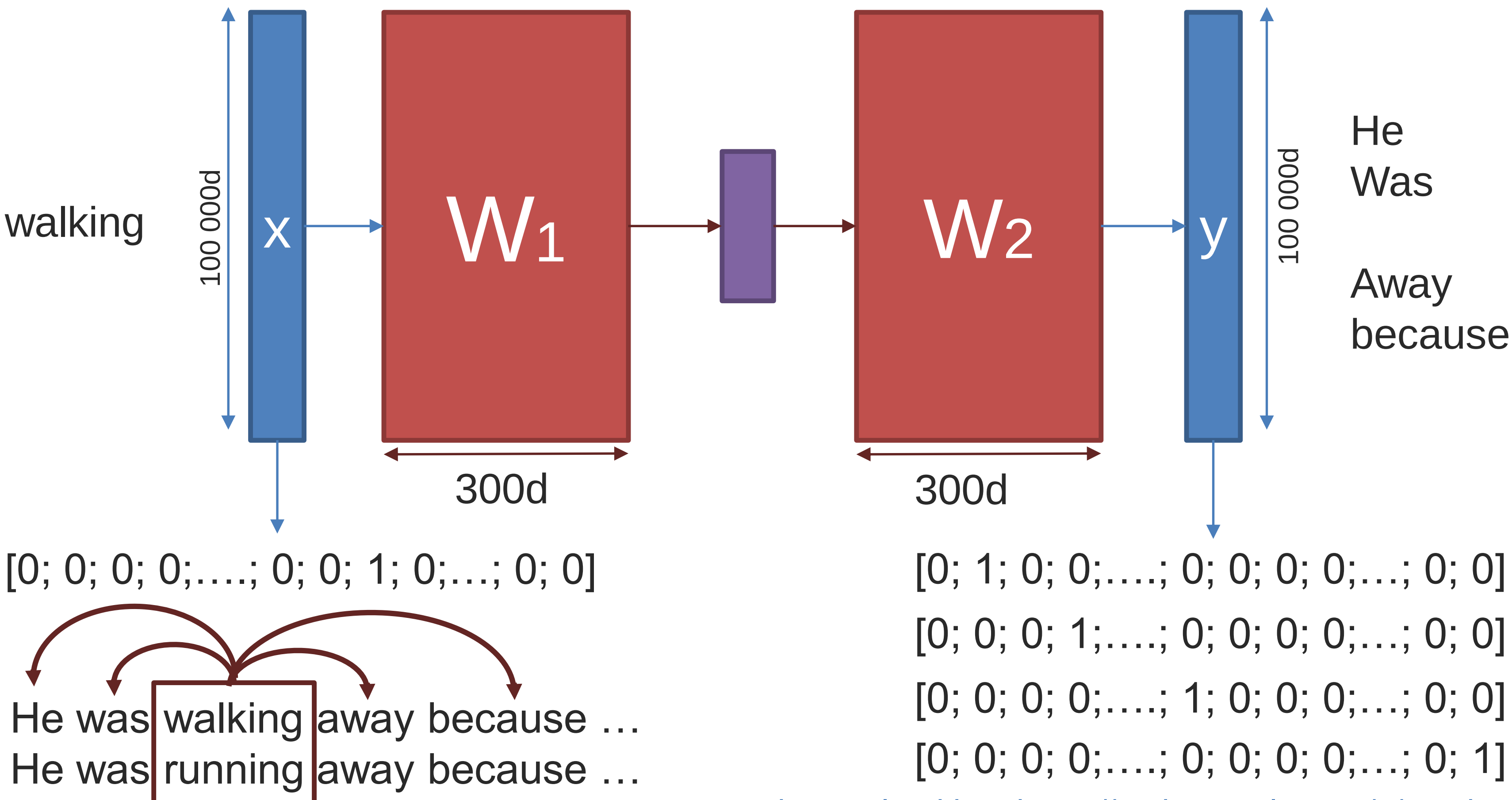


Angle and similarity

- direction more important than location
- normalise "length" $||\mathbf{x}_{dog}||$ of vector
- or use angle α as distance measure



How to learn (word) features/representations?



Word2vec algorithm: https://code.google.com/p/word2vec/

How to use these word representations

If we would have a vocabulary of 100 000 words:

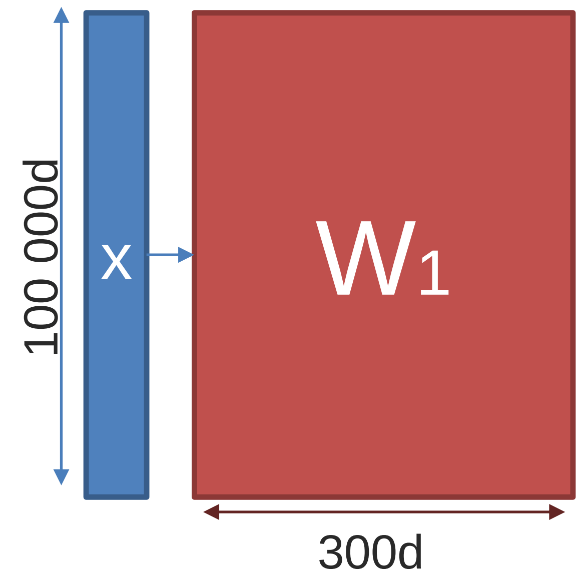
Classic NLP: 100 000 dimensional vector

Walking: [0; 0; 0; 0;….; 0; 0; 1; 0;…; 0; 0]

Running: [0; 0; 0; 0;….; 0; 0; 0; 0;…; 1; 0]

Similarity = 0.0

Transform: x'=x*W



Goal: 300 dimensional vector

Walking: [0,1; 0,0003; 0;….; 0,02; 0.08; 0,05]

Running: [0,1; 0,0004; 0;….; 0,01; 0.09; 0,05]

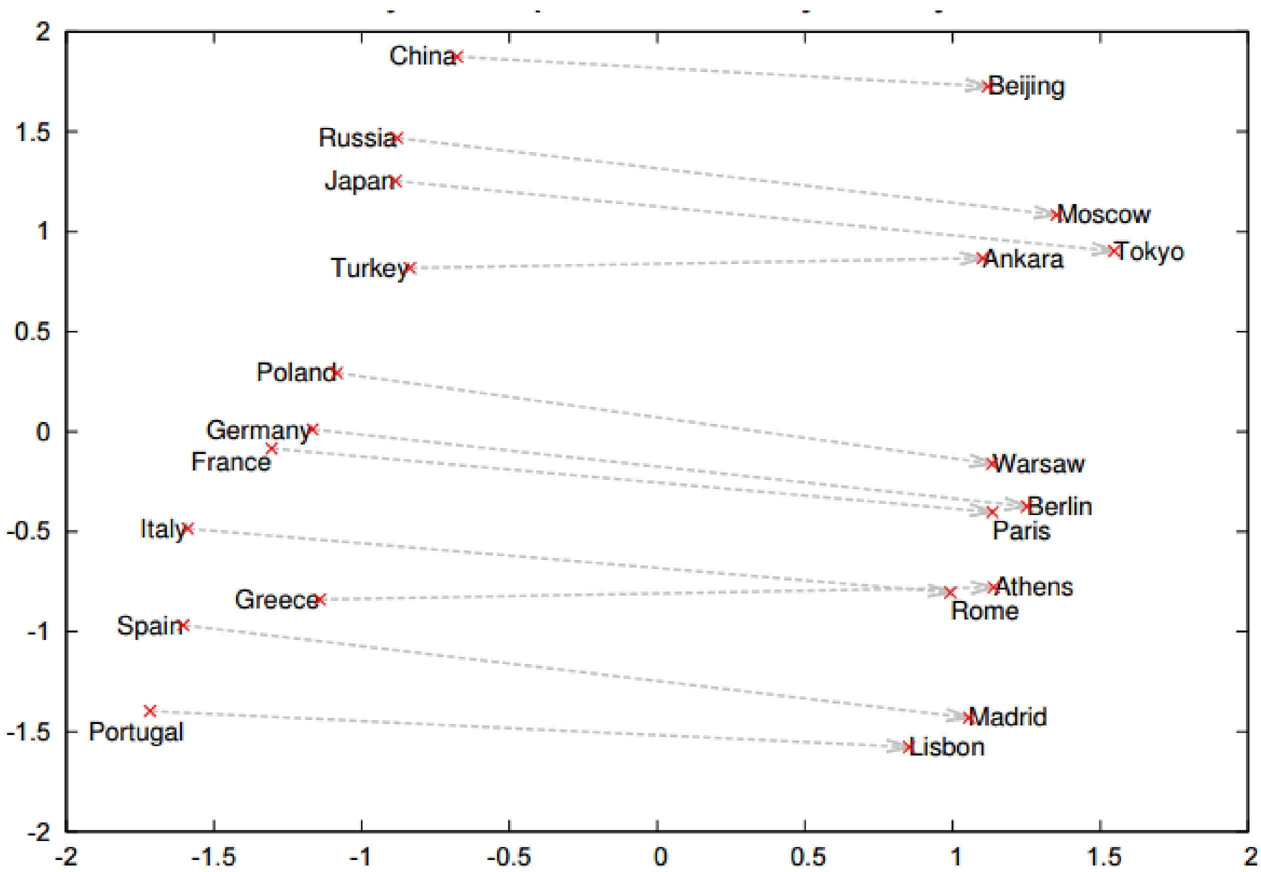
Similarity = 0.9

Vector space models of words

- While learning these word representations, we are actually building a vector space in which all words reside with certain relationships between them
- Encodes both syntactic and semantic relationships
- This vector space allows for algebraic operations:

$$\text{Vec(king)} - \text{vec(man)} + \text{vec(woman)} \approx \text{vec(queen)}$$

Vector space models of words: semantic relationships



Trained on the Google news corpus with over 300 billion words

Resources

- BERT (Google, 2018)
  - https://github.com/google-research/bert
- ELMO (Allen Institute for AI, 2018)
  - https://allennlp.org/elmo
- FastText (Facebook, 2017)
  - https://fasttext.cc/
- Glove (Stanford, 2014)
  - https://nlp.stanford.edu/projects/glove/
- Word2Vec (Google, 2013)
  - https://code.google.com/archive/p/word2vec/

Language Models

How Likely are These Utterances?

- The facts in that complex case is questionable.
- I will eat fish for dinner and drank milk with my dinner.
- I don't want no pudding.
- He is strong and a tough competitor.
- He is strong and competitive.

p(sentence) = ***?***

p(word1,word2,word3, word4, …)

N-Gram Language Models

- Estimate probability of each word given prior context.
  - P(competitive | he is strong and)
- An N-gram model uses only N–1 words of prior context.
  - Unigram: P(competitive)
  - Bigram: P(competitive | and)
  - Trigram: P(competitive | strong and)
- The ***Markov assumption*** is the presumption that the future behavior of a dynamical system only depends on its recent history.
- Language models can also be used for prediction. For example, predicting the next word to be spoken.

https://www.cs.utexas.edu/~mooney/

N-Gram Model Formulas

- Word sequences

$$w_1^n = w_1 \ldots w_n$$

- Chain rule of probability

$$P(w_1^n) = P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1^2)\ldots P(w_n \mid w_1^{n-1}) = \prod_{k=1}^{n} P(w_k \mid w_1^{k-1})$$

- Bigram approximation

$$P(w_1^n) = \prod_{k=1}^{n} P(w_k \mid w_{k-1})$$

- N-gram approximation

$$P(w_1^n) = \prod_{k=1}^{n} P(w_k \mid w_{k-N+1}^{k-1})$$

https://www.cs.utexas.edu/~mooney/

Estimating Probabilities

- N-gram conditional probabilities can be estimated from raw text based on the *relative frequency* of word sequences.

Bigram: $$P(w_n \mid w_{n-1}) = \frac{C(w_{n-1}w_n)}{C(w_{n-1})}$$

N-gram: $$P(w_n \mid w_{n-N+1}^{n-1}) = \frac{C(w_{n-N+1}^{n-1}w_n)}{C(w_{n-N+1}^{n-1})}$$

https://www.cs.utexas.edu/~mooney/

Application: Speech Recognition

$$\arg\max_{wordsequence} P(wordsequence \mid acoustics) =$$

$$\arg\max_{wordsequence} \frac{P(acoustics \mid wordsequence) \times P(wordsequence)}{P(acoustics)}$$

$$\arg\max_{wordsequence} P(acoustics \mid wordsequence) \times \underbrace{P(wordsequence)}_{\text{Language model}}$$

Resources

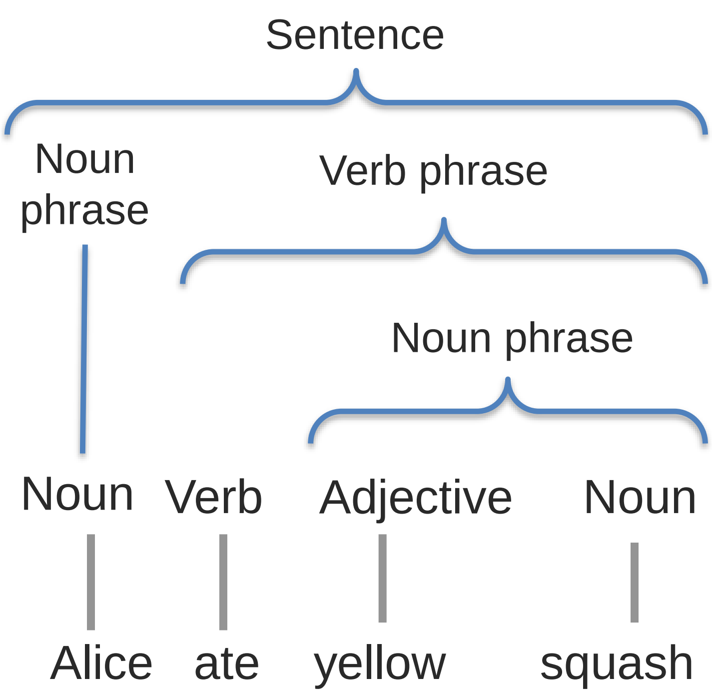
Many pre-trained models exist for language models using neural networks. Most of them are using corpus with written text. Some examples:

- Using TensorFlow (& 1 billion words):
  https://github.com/tensorflow/models/tree/master/research/lm_1b
- PyTorch tutorial
  https://github.com/pytorch/examples/tree/master/word_language_model
- Character-based level, DeepSpeech
  https://github.com/mozilla/DeepSpeech/
- SRI Language Modeling Toolkit (language models)
  http://www.speech.sri.com/projects/srilm/

Syntactic Analysis

Syntax and Language Structure

What can you tell about this sentence?



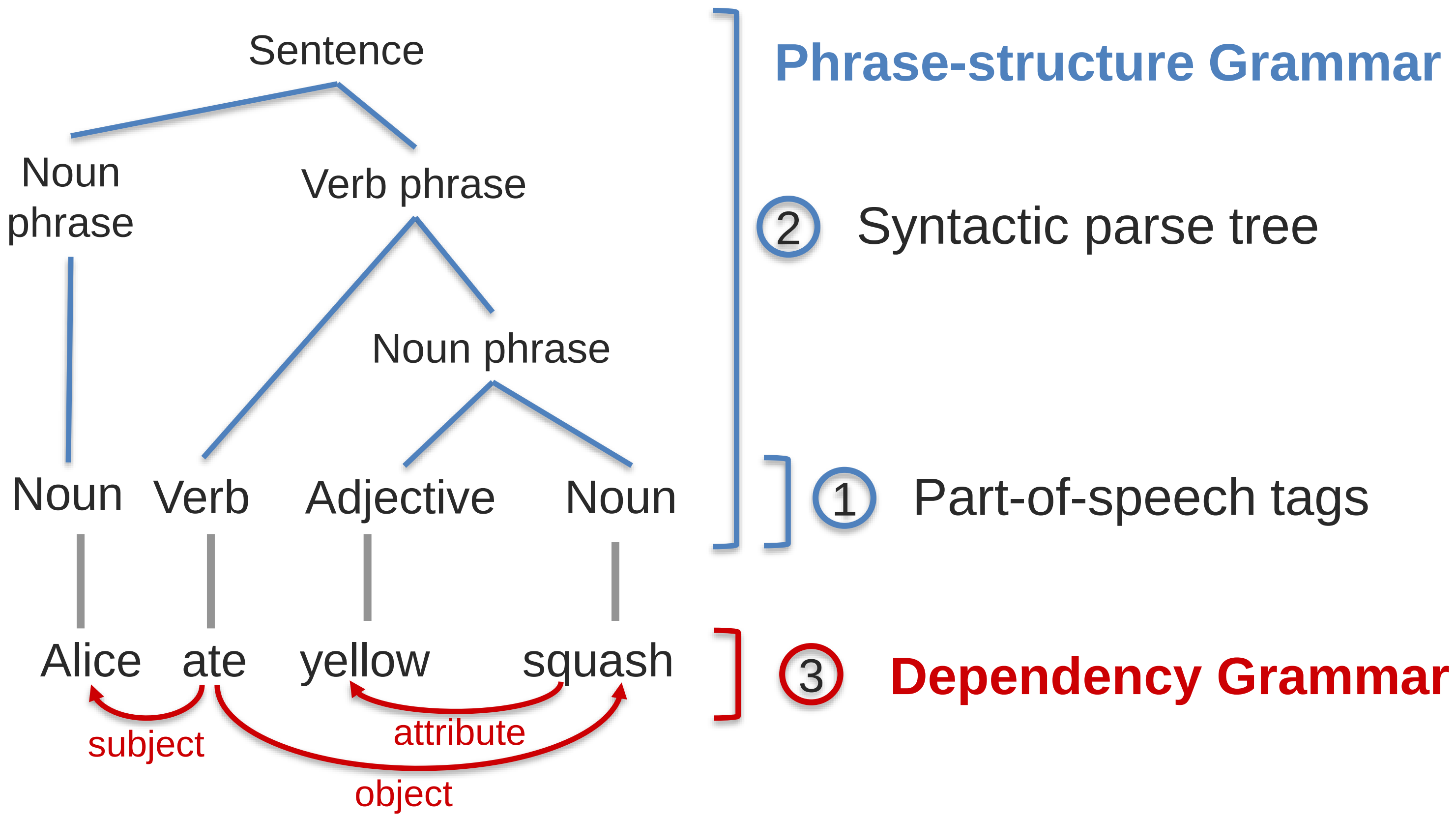
Phrase-structure Grammar

2 Syntactic parse tree

1 Part-of-speech tags

Syntax and Language Structure

What can you tell about this sentence?



Ambiguity in Syntactic Parsing

- **“Like” can be a verb or a preposition**
  - I like/VBP candy.
  - Time flies like/IN an arrow.
- **“Around” can be a preposition, particle, or adverb**
  - I bought it at the shop around/IN the corner.
  - I never got around/RP to getting a car.
  - A new Prius costs around/RB $25K.

The Importance of Parsing

What does "fake" modify?

In the hotel fake property was sold to tourists.

What does "In the hotel" modify?

Resources

- Stanford NLP software

https://nlp.stanford.edu/software/

  - Stanford Parser
  - Stanford POS Tagger

- UC Berkeley Parser

https://github.com/slavpetrov/berkeleyparser

- Parsers by Kenji Sagae (syntactic parsers) http://www.sagae.org/software.html

Conversation Annotations

Speech Acts (Austin, 1962)

Speech acts can be analyzed on three levels:

- Locutionary act: the performance of an utterance
  - The verbal, syntactic and semantic aspects of any meaningful utterance
- Illocutionary act: the pragmatic 'illocutionary force' of the utterance
  - the act of asking, answering, promising, etc., in uttering a sentence.
- Perlocutionary act: its actual effect
  - the (often intentional) production of certain **effects** upon the thoughts, feelings, or actions of addressee in uttering a sentence.

The 3 levels of act revisited

	Locutionary Force	Illocutionary Force	Perlocutionary Force
Can I have the rest of your sandwich? Or Are you going to finish that?	Question	Request	Effect: You give me sandwich (or you are amused by my quoting from “Diner”) (or etc)
I want the rest of your sandwich	Declarative	Request	Effect: as above
Give me your sandwich!	Imperative	Request	Effect: as above.

Locutionary and illocutionary

- "You can't do that!"
- Illocutionary force:
  - Protesting
- Perlocutionary force:
  - **Effect** of annoying addressee
  - **Effect** of stopping addressee from doing something

Extra reference: Illocutionary Speech acts (Searle, 1975)

Dialogue acts

- Also called "conversational moves"
- An act with (internal) structure related specifically to its dialogue function
- Incorporates ideas of grounding
- Incorporates other dialogue and conversational functions that Austin and Searle didn't seem interested in

Adjacency pairs

An adjacency pair is a unit of conversation that contains an exchange of one turn each by two speakers. The turns are functionally related to each other in such a fashion that the first turn requires a certain type or range of types of second turn.

- Question→ answer
  - "*What does this big red button do?*" → "*It causes two-thirds of the universe to implode*"
- Greeting → greeting
  - "*Heya!*" → "*Oh, hi!*"
- Request → acceptance/rejection
  - "*Is it OK if I borrow this book?*" → "*I'd rather you didn't, it's due back at the library tomorrow*"

Dialogue Levels & Dialogue Acts (Traum)

- Contact (make,break)
- Attention (show, request, accept)
- Conversation (begin, join, leave, end)
  - Turn-taking (take, hold, release, assign)
  - Initiative (take, assign, release)
  - Utterance
    - Main Function (assert, request, suggest, order, offer, promise, info-request,…)
    - Relational (answer, accept, reject, avoid, hold,…)
    - Features: speaker, addressee, overhearer, referent, content
      - Polarity (positive, negative)
  - Grounding (initiate, continue, acknowledge, repair, request repair…)
  - Topic (set topic, set subtopic, close topic)
- Social
  - Obligations & Commitments
  - Relationships (Face, Status, Affilliation, Trust)
  - Social Roles

Example Transcription

Speaker	Dialogue Act	Utterance
A		So do you go to college right now?
A		Are yo-,
B		*Yeah,*
B		*it's my last year* [laughter].
A		You're a, so you're a senior now.
B		*Yeah,*
B		*I'm working on my projects trying to graduate* [laughter].
A		Oh, good for you.
B		*Yeah.*
A		That's great,
A		um, is, is N C University is that, uh, State,
B		*N C State.*
A		What did you say?
B		*N C State.*

References - Annotation schemes

- Speech Acts
  - Austin, 1962
- Illocutionary Speech acts
  - Searle, 1975
- Cooperative Principle
  - 4 Gricean Maxims - 1975
- Grounding Acts
  - Traum & Hinkelman92, Traum 94
- DAMSL annotation scheme
  - https://www.cs.rochester.edu/research/speech/damsl/RevisedManual/

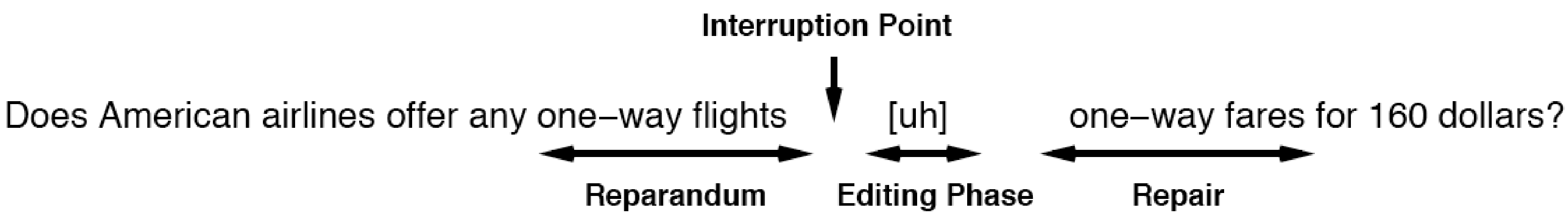
Disfluencies and Turn-taking

Disfluencies and Pause Fillers

the . [exhale] . . . [inhale] . . [uh] does American airlines . offer any . one way flights . [uh] one way fares, for one hundred and sixty one dollars
[mm] i'd like to leave i guess between [um] . [smack] . five o'clock no, five o'clock and [uh], seven o'clock . P M
around, four, P M
all right, [throat_clear] . . i'd like to know the . give me the flight . times . in the morning . for September twentieth . nineteen ninety one
[uh] one way
[uh] seven fifteen, please
on United airlines . . give me, the . . time . . from New York . [smack] . to Boise-, to . I'm sorry . on United airlines . [uh] give me the flight, numbers, the flight times from . [uh] Boston . to Dallas

Figure 9.5 Some sample spoken utterances from users interacting with the ATIS system.

Disfluencies: standard terminology (Levelt)



- Reparandum: thing repaired
- Interruption point (IP): where speaker breaks off
- Editing phase (edit terms): uh, I mean, you know
- Repair: fluent continuation

Turn-Taking Key Definitions

(Conversational) Floor

“The ***right*** to address an assembly” (Merriam-Webster)

The interactional ***state*** that describes which participant in a dialog has the ***right*** to provide or request information at any point.

Turn-Taking

The process by which participants in a conversation ***alternately own*** the conversational floor.

Floor, Intentions and Beliefs

The floor is ***not*** an ***observable*** state.

Rather, participants have:

- ***intentions*** to claim the floor or not
- ***beliefs*** over whether others are claiming it

Participants ***negotiate*** the floor to limit ***gaps*** and ***overlaps***. [Sacks et al 1974, Clark 1996]

Overlaps and Interruptions

- Overlaps:
  - Coates (2003) : instances of slight over-anticipation by the next speaker.
  - Over-anticipation does not necessarily force the first speaker to finish his / her turn.
  - Continues the previous turn of the other (Person A continues)
  - Occurs at a TRP (Transition-Relevant Place)

Overlaps and Interruptions

- Interruptions:
  - have the potential to disrupt a speaker´s turn and disorganize ongoing construction of the conversational topic of the first speaker;
  - Occurs NOT at a TRP
  - regarded as a hostile act

Turn-taking rules, Sacks et al. (1974)

- At each transition-relevance place of each turn:
  a) If during this turn the current speaker has selected B as the next speaker then B must speak next.
  b) If the current speaker does not select the next speaker, any other speaker may take the next turn.
  c) If no one else takes the next turn, the current speaker may take the next turn.

Violation of the turn-taking model

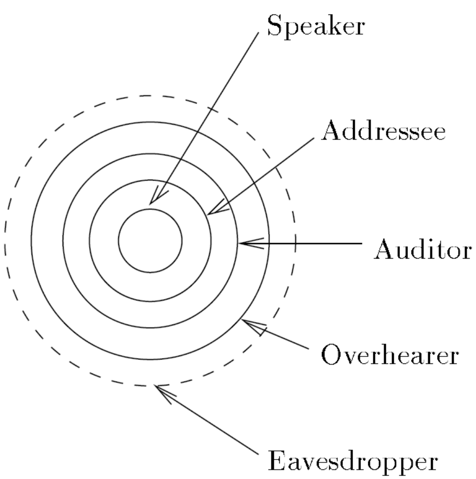
- grabbing the floor
- hogging the floor (taking the floor although other speaker was selected)
- not responding (silence)

Turn-Constructional Unit

- TCU are turns at talk, e.g. in sentences, clauses, single words or phrases
  - TCUs have the property of projectability: it is possible for participants to project, in the course of TCU, what sort of unit it is and at what point it is likely to end.
  - TCUs have transition-relevance places (TRPs) at their boundaries: at the end of each TCU there is the possibility for legitimate transition between speakers (example)

Multiparty Floor Management [Bell, 1984]

Role	Description
Addressee	participant that utterance is addressed to
Side participant	participant that utterance is not addressed to
Overhearer	others known to the speaker who are not participants in conversation but will hear the utterance
Eavesdropper	others not known to the speaker who are not participants in the conversation but will hear the utterance



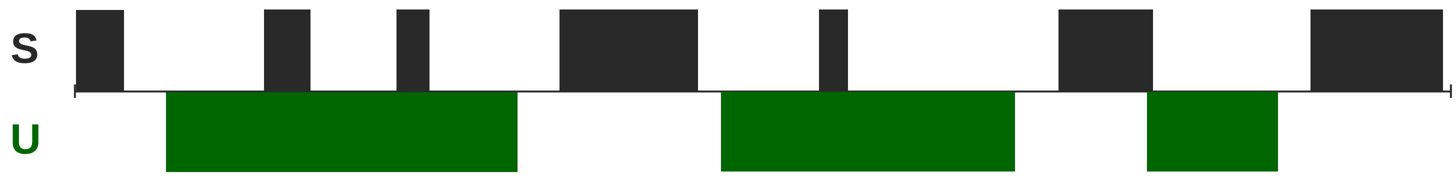
	Known	Ratified	Addressed
Addressee	+	+	+
Auditor	+	+	−
Overhearer	+	−	−
Eavesdropper	−	−	−

Multiparty Floor Management: Example

1	C:	Ann,
2		what's playing at the theater next week?
3	A:	Sorry,
4		I don't know.
5		But Barbara does.
6	B:	"Much Ado about Nothing".

#	Speaker	Addressee	Auditor
1	C	A,B	—
2	C	A	B
3	A	C	B
4	A	C	B
5	A	C	B
6	B	A	C

Dialogue Systems: Uncertainty over the Floor



Uncertainty over the floor leads to ***breakdowns in turn-taking***:

- Cut-ins
- Latency
- Self interruptions

Resources

- J. Hough and D. Schlangen. 2017. Joint, Incremental Disfluency Detection and Utterance Segmentation from Speech. In *Proceedings of the 15th International Conference of the European Chapter of the Association for Computational Linguistics*

  https://github.com/dsg-bielefeld/deep_disfluency