Lexicon, Syntax, Semantics IIb: Modeling Meaning

Case Studies: Adjectives in Compositional DS

Eva Maria Vecchi

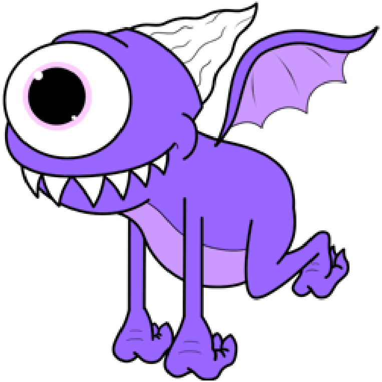
Center for Information and Language Processing
LMU Munich

June 25, 2020

Outline

Modification is fun...

One-eyed, one-horned flying purple people eater



Goals

Study the ability of Compositional Distributional Semantics (CDS) to model adjective modification

Goals

Study the ability of Compositional Distributional Semantics (CDS) to model adjective modification

1. Study how the CDS of adjective-noun phrases is able to capture linguistic phenomena concluded in previous literature
    - degree of modification of the word *white* in the phrases *white shirt*, *white wine* and *white lie*.
    - effects of adjective semantics on ordering restrictions in recursive modification

Goals

Study the ability of Compositional Distributional Semantics (CDS) to model adjective modification

1. Study how the CDS of adjective-noun phrases is able to capture linguistic phenomena concluded in previous literature
   - degree of modification of the word *white* in the phrases *white shirt*, *white wine* and *white lie*.
   - effects of adjective semantics on ordering restrictions in recursive modification
2. Investigate how CDS can provide insight to our understanding of natural language
   - nonsensicality
   - what specific semantic properties, which can be extracted from the distributional representation of phrases in a relatively painless and efficient manner, affect what we, as natural language speakers, just "understand"

Distributional semantics

Landauer and Dumais 1997, Turney and Pantel 2010, . . .

```
he curtains open and the moon shining in on the barely
ars and the cold , close moon " . And neither of the w
rough the night with the moon shining so brightly , it
made in the light of the moon . It all boils down , wr
 surely under a crescent moon , thrilled by ice-white
sun , the seasons of the moon ? Home , alone , Jay pla
m is dazzling snow , the moon has risen full and cold
un and the temple of the moon , driving out of the hug
 in the dark and now the moon rises , full and amber a
bird on the shape of the moon over the trees in front
 But I could n't see the moon or the stars , only the
rning , with a sliver of moon hanging among the stars
 they love the sun , the moon and the stars . None of
the light of an enormous moon . The plash of flowing w
man 's first step on the moon ; various exhibits , aer
 the inevitable piece of moon rock . Housing The Airsh
oud obscured part of the moon . The Allied guns behind
```

Distributional semantics

The geometry of meaning

Distributional Semantic Models (DSMs): Computational models of meaning based on their distribution: their pattern of cooccurrences within a specified context

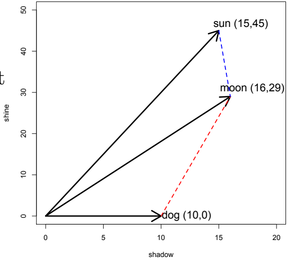
	shadow	shine	planet	night
moon	16	29	10	22
sun	15	45	14	10
dog	10	0	0	4

Distributional semantics

The geometry of meaning

Distributional Semantic Models (DSMs): Computational models of meaning based on their distribution: their pattern of cooccurrences within a specified context

	shadow	shine	planet	night
moon	16	29	10	22
sun	15	45	14	10
dog	10	0	0	4



- semantic similarity approximated by geometric distance of vectors (angle)
- successful in tasks that concern content words: detecting synonyms, lexical entailment, ... (see Turney & Pantel, 2010; Baroni & Lenci, 2010)

Compositionality in DSMs

Mitchell and Lapata 2008, 2009, 2010

	planet	night	space	color	blood	brown
red	15	3	2	24	19	20
moon	24	15	20	3	2	1

Compositionality in DSMs

Mitchell and Lapata 2008, 2009, 2010

	planet	night	space	color	blood	brown
red	15	3	2	24	19	20
moon	24	15	20	3	2	1
red moon	10	2	1	5	1	2

Compositionality in DSMs

Mitchell and Lapata 2008, 2009, 2010

	planet	night	space	color	blood	brown
red	15	3	2	24	19	20
moon	24	15	20	3	2	1
red moon	10	2	1	5	1	2
red+moon	39	18	22	27	21	21
red*moon	360	45	40	72	38	20

1. **Additive** (ADD): $\vec{p} = \vec{a} + \vec{n}$
2. **Weighted Additive** (W.ADD): $\vec{p} = \alpha\vec{a} + \beta\vec{n}$
3. **Multiplicative** (MULT): $\vec{p} = \vec{a} \odot \vec{n}$
4. **Dilation** (DL): $\vec{p} = (\vec{a} \cdot \vec{a})\vec{n} + (\lambda - 1)(\vec{a} \cdot \vec{n})\vec{a}$

Compositionality in DSMs

Guevara 2010, Baroni and Zamparelli 2010

5. **Full Additive** (F.ADD): $\vec{p} = \mathbf{W}_1\vec{a} + \mathbf{W}_2\vec{n}$

6. **Lexical Function Model** (LFM):

$$\vec{p} = \mathbf{A}\vec{n}$$

$$\begin{pmatrix} c_1 \\ c_2 \\ \ldots \\ c_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} \times \begin{pmatrix} n_1 \\ n_2 \\ \ldots \\ n_m \end{pmatrix}$$

General Setup

1. Semantic Space

Source Corpus

- about 2.8 billion tokens
  - Web-derived ukWaC corpus (`http://wacky.sslmit.unibo.it`)
  - a mid-2009 dump of the English Wikipedia (`http://en.wikipedia.org`)
  - British National Corpus (`http://www.natcorp.ox.ac.uk/`)
- tokenized, POS-tagged and lemmatized with the TreeTagger (Schmid 1995)
- co-occurrence statistics extracted at the lemma level, no inflectional information

General Setup

1. Semantic Space

Source Corpus

- about 2.8 billion tokens
  - Web-derived ukWaC corpus (`http://wacky.sslmit.unibo.it`)
  - a mid-2009 dump of the English Wikipedia (`http://en.wikipedia.org`)
  - British National Corpus (`http://www.natcorp.ox.ac.uk/`)
- tokenized, POS-tagged and lemmatized with the TreeTagger (Schmid 1995)
- co-occurrence statistics extracted at the lemma level, no inflectional information

Semantic Space Matrix

- Rows:
  - 8K most frequent Nouns
  - 4K most frequent Adjectives
  - 179K ANs with frequency > 100 in source corpus
- Contexts:
  - 10K most frequently co-occurring Adjectives, Nouns, Verbs, and Adverbs

General Setup

2. Composition model estimation

- Parameters for W.ADD, DL, F.ADD and LFM were estimated following the strategy proposed by Guevara (2010) and Baroni and Zamparelli (2010), recently extended to all composition models by Dinu et al. (2013b).

- All parameter estimations and phrase compositions were implemented using the DISSECT toolkit[1]
  - with a training set of 74,767 corpus-extracted N-AN vector pairs, ranging from 100 to over 1K items across the 663 adjectives.

[1] `http://clic.cimec.unitn.it/composes/toolkit`

(Dinu et al., 2013a)

Semantic space parameter tuning

Weighting	*Reduction*	R&G	MEN	M&L
SoA		0.82	0.69	0.43
-	-	0.77	0.72	0.36
PPMI	SVD_{300}	0.72	0.69	0.38
	SVD_{50}	0.68	0.67	0.36
	NMF_{300}	**0.81**	**0.76**	**0.40**
	NMF_{50}	0.69	0.68	**0.40**
PLMI	SVD_{300}	0.70	0.70	**0.40**
	SVD_{50}	0.54	0.55	0.28
	NMF_{300}	0.70	0.68	0.06
	NMF_{50}	0.50	0.55	0.13
PLOG	SVD_{300}	0.40	0.38	0.32
	SVD_{50}	0.39	0.38	0.30
	NMF_{300}	0.62	0.63	**0.40**
	NMF_{50}	0.46	0.51	0.31

Composed space quality evaluation

Model	ρ	*M&L*	*Parameter*
CORP	0.40	0.43	
ADD	0.34	0.37	
W.ADD	0.35	0.44	$\alpha = 0.31,\ \beta = 0.46$
MULT	0.31	0.46	
DL	0.32	0.44	$\lambda = 1.59$
F.ADD	0.35	–	
LFM	0.38	–	

Outline

Tell me something I haven't already heard...

Vecchi et al, 2011; Vecchi et al, 2017

sophisticated senator *legislative onion*

Tell me something I haven't already heard...

Vecchi et al, 2011; Vecchi et al, 2017

sophisticated senator *legislative onion*

Our Goal: Model human intuition about semantic deviance in *novel*, or unattested, attributive Adjective-Noun (AN) expressions using a number of cues in a distributional semantic space

- Applications:
  - Detecting nonsensicality as a prerequisite to metaphorical interpretation (computational/psychological, Fass, 1983, Zhou et al. 2007)
  - Language Modeling: Better probability estimates for unattested data

Criticism from Formal-Based Approaches

Seeing only what the corpus sees

Criticisms against statistical methods for meaning representation argue:

- methods based on corpus sampling are not able to generalize word combinations that have not been observed in the corpus (Chomsky, 1957)

Criticism from Formal-Based Approaches

Seeing only what the corpus sees

Criticisms against statistical methods for meaning representation argue:

- methods based on corpus sampling are not able to generalize word combinations that have not been observed in the corpus (Chomsky, 1957)
- We don't know *a priori* why things might be unattested
  - rare expressions
  - factually wrong
  - non-grammatical
  - nonsensical

Criticism from Formal-Based Approaches

Seeing only what the corpus sees

Criticisms against statistical methods for meaning representation argue:

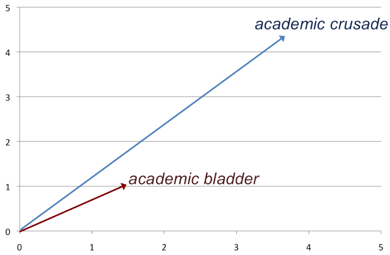
- methods based on corpus sampling are not able to generalize word combinations that have not been observed in the corpus (Chomsky, 1957)
- We don't know *a priori* why things might be unattested
  - rare expressions
  - factually wrong
  - non-grammatical
  - nonsensical
- We explore the distinction between rare vs. nonsensical expressions

wet whale	*wet literacy*
stable kidney	*angry kidney*
printed pardon	*printed gallon*
ignorant merchant	*ignorant scarf*

Measuring Semantic Deviance

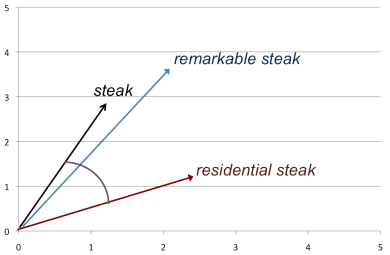
Hypothesis on detecting semantic deviance in a DSM

Vector Length



Shared contexts between component vectors ⟶ longer vectors

Cosine

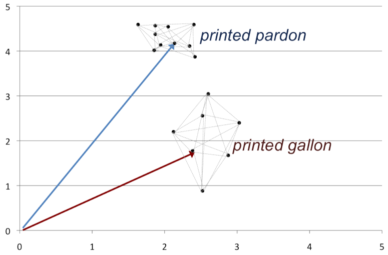


Distort meaning of component noun ⟶ greater distance

Measuring Semantic Deviance

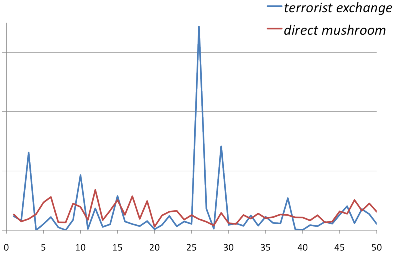
Hypothesis on detecting semantic deviance in a DSM

Neighborhood density



Shared contexts of component elements overlap greatly with attested words/expression
⟶ higher density

Entropy[1]



Dimensions containing mostly noise with a uniform distribution
⟶ higher entropy

1. (Lazaridou et al, 2013)

Evaluation Materials

- Collected an evaluation dataset of acceptability judgments on unattested ANs with a crowdsourcing experiment on CrowdFlower

  (`http://www.crowdflower.com`)

- Obtained human plausibility judgments on a set of 150K AN_x-AN_y pairs
  - set contained a random sample of 30K **unattested** ANs, composed of high frequency adjectives and nouns
  - each AN seen 5 times in position x and 5 times in position y, without repetition of pairs

Data Analysis

- Statistical analysis (logit mixed effects models) to explore the impact each measure of semantic deviance has on approximating the plausibility judgments

(Gagné and Spalding 2004, 2006, 2007, Schreuder and Baayen 1997)

Data Analysis

- Statistical analysis (logit mixed effects models) to explore the impact each measure of semantic deviance has on approximating the plausibility judgments
- **Baseline Measures:** Attempts to describe how novel compounds are processed (at a cognitive level) have shown that these measures are significant in the acceptable/unacceptable choice:

(Gagné and Spalding 2004, 2006, 2007, Schreuder and Baayen 1997)

Data Analysis

- Statistical analysis (logit mixed effects models) to explore the impact each measure of semantic deviance has on approximating the plausibility judgments
- **Baseline Measures:** Attempts to describe how novel compounds are processed (at a cognitive level) have shown that these measures are significant in the acceptable/unacceptable choice:

  1. String Length: the length of the string for the component elements

     longer string $\longrightarrow$ acceptable

(Gagné and Spalding 2004, 2006, 2007, Schreuder and Baayen 1997)

Data Analysis

- Statistical analysis (logit mixed effects models) to explore the impact each measure of semantic deviance has on approximating the plausibility judgments

- **Baseline Measures:** Attempts to describe how novel compounds are processed (at a cognitive level) have shown that these measures are significant in the acceptable/unacceptable choice:

  1. String Length: the length of the string for the component elements

     longer string $\longrightarrow$ acceptable

  2. Family Size: the number of attested ANs in which the component element occurs

     higher family size $\longrightarrow$ acceptable

(Gagné and Spalding 2004, 2006, 2007, Schreuder and Baayen 1997)

Results

1. Baseline measures

Baseline measures

Measure	*Estimate*	*Pr (> |z|)*
A_L family	-3.150e-04	< 2.2e-16 ***
A_R family	3.823e-04	< 2.2e-16 ***
N_L family	-1.803e-03	< 2.2e-16 ***
N_R family	1.875e-03	< 2.2e-16 ***
A_L length	-6.964e-02	< 2.2e-16 ***
A_R length	7.137e-02	< 2.2e-16 ***
N_L length	-1.084e-01	< 2.2e-16 ***
N_R length	1.037e-01	< 2.2e-16 ***

- Polarity of the estimate indicates the likelihood of choosing the left-hand (L, negative) or right-hand (R, positive) AN as the more acceptable AN wrt the variable.
- A larger estimate (absolute value) reflects a stronger effect on the choice of AN.

Results

2. Indices of semantic deviance in the DSM

Vector-based measures

Model	*VLENGTH*	*COSINE*	*DENSITY*	*ENTROPY*
W.ADD	***	***	**	
MULT	***	***	***	***
DL	***	***	**	***
F.ADD	***	***		***
LFM	***	***	*	***

1. Vector Length — shorter AN vector ⟶ deviant
2. Cosine from Noun — farther from noun ⟶ deviant
3. Neighborhood Density — denser neighborhood ⟶ deviant
4. Entropy — higher entropy ⟶ ??

- * All vector-based measures significantly improve the goodness of fit on the baseline measures.

Outline

Many sides to adjectival modification

red car

red meat

Many sides to adjectival modification

red car

red meat

Many sides to adjectival modification

tall boyfriend

former boyfriend

Adjective Modification

Intersective

Intersective Modifiers

- If an AN is **intersective**, then the AN is A and the AN is N

Adjective Modification

Intersective

Intersective Modifiers

- If an AN is **intersective**, then the AN is A and the AN is N
- semantic composition can be modeled by set intersection

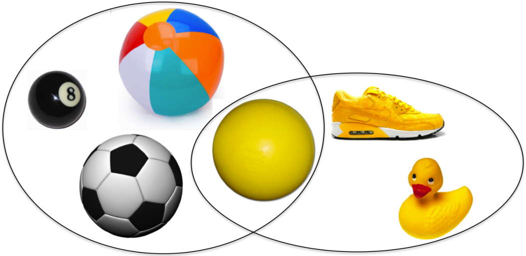
Adjective Modification

Intersective

Intersective Modifiers

- If an AN is **intersective**, then the AN is A and the AN is N
- semantic composition can be modeled by set intersection
- symmetric operation that has affinities with composition functions such as ADD and MULT

$$\| \, yellow \; ball \, \| = \| \, yellow \, \| \cap \| \, ball \, \|$$



Adjective Modification

Intersective > Subsective

Subsective Modifiers

- If an AN is **subsective**, then the AN is N but we cannot conclude that the AN is A

Adjective Modification

Intersective > Subsective

Subsective Modifiers

- If an AN is **subsective**, then the AN is N but we cannot conclude that the AN is A
- The adjective is often used to characterize a subclass of the class described by the noun
  - in the case of color adjectives, it may serve as a proxy for another property related to color (Kennedy & McNally, 2010)
  - the adjective may or may not match the literal color, e.g. *white wine* and *white lie*

Adjective Modification

Intersective > Subsective

Subsective Modifiers

- If an AN is **subsective**, then the AN is N but we cannot conclude that the AN is A
- The adjective is often used to characterize a subclass of the class described by the noun
  - in the case of color adjectives, it may serve as a proxy for another property related to color (Kennedy & McNally, 2010)
  - the adjective may or may not match the literal color, e.g. *white wine* and *white lie*
- It can be modeled by a more flexible functional approach, as in composition functions like F.ADD and LFM

white pepper

white wine

Adjective Modification

Intersective > Subsective > Intensional

Intensional Modifiers

- If an AN is **intensional**, then we cannot infer that the AN is N (or we may be able to infer that the AN is not N), nor can we conclude that the AN is A

Adjective Modification

Intersective > Subsective > Intensional

Intensional Modifiers

- If an AN is **intensional**, then we cannot infer that the AN is N (or we may be able to infer that the AN is not N), nor can we conclude that the AN is A
    - Privative: license the inference to "not N"

Dave is her former boyfriend	$\models$ *Dave is not her boyfriend now*
	$\not\models$?? *Dave is former*

Adjective Modification

Intersective > Subsective > Intensional

Intensional Modifiers

- If an AN is **intensional**, then we cannot infer that the AN is N (or we may be able to infer that the AN is not N), nor can we conclude that the AN is A
  - Privative: license the inference to "not N"

Dave is her former boyfriend
$\vDash$ *Dave is not her boyfriend now*
$\nvDash$?? *Dave is former*

  - Other intensional modifiers: license no entailments at all

John is her alleged boyfriend
$\nvDash$ *John is her boyfriend*
$\nvDash$?? *John is alleged*

Evaluation Material

Boleda et al, 2012

1. **Intersective & Subsective Color Terms**
   A random selection of very frequent color ANs annotated by two native English speaker linguists as **intersective** (239) or **subsective** (130)
   - *black, blue, brown, green, red, white, yellow* (Berlin & Kay, 1969)
   - kappa coefficient estimated on annotation is 0.86 (conf.int. 0.82-0.91)

INTERSECTIVE	*SUBSECTIVE*	*INTENSIONAL*
white towel	white wine	artificial leg
black sack	black athlete	former bassist
green coat	green politics	likely suspect
red disc	red ant	possible delay
blue square	blue state	theoretical advantage

Evaluation Material

Boleda et al, 2012

1. **Intersective & Subsective Color Terms**
   A random selection of very frequent color ANs annotated by two native English speaker linguists as **intersective** (239) or **subsective** (130)
   - *black, blue, brown, green, red, white, yellow* (Berlin & Kay, 1969)
   - kappa coefficient estimated on annotation is 0.86 (conf.int. 0.82-0.91)
2. **Intensional Modification**
   A selection of 1,200 very frequent ANs covering a pre-selected list of 10 **intensional** adjectives
   - *former, possible, future, potential, past, false, apparent, artifical, likely, theoretical*

INTERSECTIVE	*SUBSECTIVE*	*INTENSIONAL*
white towel	white wine	artificial leg
black sack	black athlete	former bassist
green coat	green politics	likely suspect
red disc	red ant	possible delay
blue square	blue state	theoretical advantage

Results: Corpus-observed Vectors

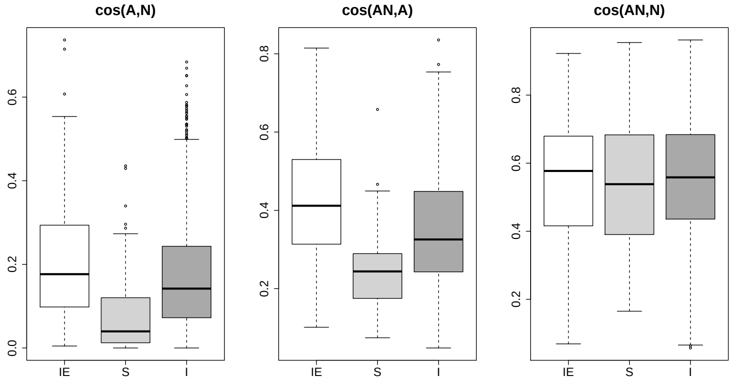
Expectations:

1. cos(**A**,**N**): Intersective (IE) $>$ Subsective (S)
2. cos(**AN**,**A**): Intersective (IE) $>$ Subsective (S) $>$ Intensional (I)
3. cos(**AN**,**N**): Intensional (I) $>$ Intersective (IE), Subsective (S)

Results: Corpus-observed Vectors

Expectations:

1. cos(**A**,**N**): Intersective (IE) > Subsective (S)
2. cos(**AN**,**A**): Intersective (IE) > Subsective (S) > Intensional (I)
3. cos(**AN**,**N**): Intensional (I) > Intersective (IE), Subsective (S)



Distribution of the cosines in the different types of AN.

Results: Model-generated Vectors

Intersective vs. Subsective

Intersective vs. Subsective Color Terms

model	Δ:AN	Δ:A	Δ:N
expectations	significant (+)	significant (+)	not significant
CORP	-	1.13 *	.08

Results: Model-generated Vectors

Intersective vs. Subsective

Intersective vs. Subsective Color Terms

model	Δ:AN	Δ:A	Δ:N
expectations	significant (+)	significant (+)	not significant
CORP	-	1.13 *	.08
ADD	.75 *	.90 *	.90 *
W.ADD	.53 *	.91 *	.89 *
MULT	.66 *	1.05 *	.62 *
DL	.19	.92 *	-.78 *
F.ADD	.50 *	.91 *	.09
LFM	.39	1.04 *	.51 *

Significances according to a t-test: * for $p < 0.001$.

Results: Model-generated Vectors

Intersective vs. Intensional

Intersective vs. Intensional Modification

model	Δ:AN	Δ:A	Δ:N
expectations	significant (+)	significant (+)	(-)
CORP	-	.51 *	-.03

Results: Model-generated Vectors

Intersective vs. Intensional

Intersective vs. Intensional Modification

model	Δ:AN		Δ:A		Δ:N	
expectations	significant (+)		significant (+)		(-)	
CORP	-		.51	*	-.03	
ADD	.28	*	.26	*	-.26	*
W.ADD	.18		.27	*	.26	*
MULT	.47	*	.34	*	.13	
DL	.01		.26	*	-.25	*
F.ADD	-.01		.30	*	.14	
LFM	-.56	*	.64	*	-.14	

Significances according to a t-test: * for $p < 0.001$.

- Intensionality was only alleged? (Boleda et al, 2013)

Outline

[Recursive] Compositionality in DSMs

	planet	night	space	color	blood	brown
red	15	3	2	24	19	20
moon	24	15	20	3	2	1
red moon	10	2	1	5	1	2
big red moon	3	1	1	4	0	1

[Recursive] Compositionality in DSMs

Given an A_xA_yN...

1. **Weighted Additive** (W.ADD):
$$\vec{p} = \alpha\vec{a}_x + \beta(\alpha\vec{a}_y + \beta\vec{n}) = \alpha\vec{a}_x + \alpha\beta\vec{a}_y + \beta^2\vec{n}$$
2. **Multiplicative** (MULT): $\vec{p} = \vec{a}_x \odot \vec{a}_y \odot \vec{n}$
3. **Full Additive** (F.ADD): $\vec{p} = \mathbf{W}_1\vec{a}_x + \mathbf{W}_2(\mathbf{W}_1\vec{a}_y + \mathbf{W}_2\vec{n})$
$$= \mathbf{W}_1\vec{a}_x + \mathbf{W}_2\mathbf{W}_1\vec{a}_y + \mathbf{W}_2^2\vec{n}$$
4. **Lexical Function Model** (LFM): $\vec{p} = \mathbf{A}_x(\mathbf{A}_y\vec{n})$

[Recursive] Compositionality in DSMs

Given an $\mathrm{A}_x\mathrm{A}_y\mathrm{N}$...

1. **Weighted Additive** (W.ADD):
$$\vec{p} = \alpha\vec{a}_x + \beta(\alpha\vec{a}_y + \beta\vec{n}) = \alpha\vec{a}_x + \alpha\beta\vec{a}_y + \beta^2\vec{n}$$
2. **Multiplicative** (MULT): $\vec{p} = \vec{a}_x \odot \vec{a}_y \odot \vec{n}$
3. **Full Additive** (F.ADD): $\vec{p} = \mathbf{W}_1\vec{a}_x + \mathbf{W}_2(\mathbf{W}_1\vec{a}_y + \mathbf{W}_2\vec{n})$
$$= \mathbf{W}_1\vec{a}_x + \mathbf{W}_2\mathbf{W}_1\vec{a}_y + \mathbf{W}_2^2\vec{n}$$
4. **Lexical Function Model** (LFM): $\vec{p} = \mathbf{A}_x(\mathbf{A}_y\vec{n})$

- How distributional composition functions behave when applied recursively
- Can we use measures extracted from the semantic space to distinguish - or even predict - adjective ordering?

Restrictions in recursive adjective modification

Vecchi et al, 2013

One-eyed, one-horned flying purple people eater

Restrictions in recursive adjective modification

Vecchi et al, 2013

One-eyed, one-horned flying purple people eater

Flexible Order (FO)

- phrases where both orders, A_xA_yN and A_yA_xN, are frequently attested

estimated total population	*total estimated population*
overall good health	*good overall health*

Restrictions in recursive adjective modification

Vecchi et al, 2013

One-eyed, one-horned flying purple people eater

Flexible Order (FO)

- phrases where both orders, $A_x A_y N$ and $A_y A_x N$, are frequently attested

estimated total population	*total estimated population*
overall good health	*good overall health*

Rigid Order (RO)

- phrases with *one* order, $A_x A_y N$, frequently attested, and $A_y A_x N$ is unattested.

ancient human remains	**human ancient remains*
fine young musician	**young fine musician*

Measures of adjective ordering

Distinguishing flexible vs. rigid order

Flexible ordering:

- the two adjectives will have a similarly strong effect on the noun
- *creative new idea* is an *idea* that is both *creative* and *new*

Measures of adjective ordering

Distinguishing flexible vs. rigid order

Flexible ordering:

- the two adjectives will have a similarly strong effect on the noun
- *creative new idea* is an *idea* that is both *creative* and *new*

Rigid ordering:

- one adjective (the one closer to the head) will dominate the meaning of the phrase, distorting the meaning of the noun
- *different architectural style* describes an *architectural style* that is *different*
- NOT a *style* that is both *architectural* and *different*

Measures of adjective ordering

Distinguishing flexible vs. rigid order

Operationalize in the distance relationship between the A_xA_yN vector and its subparts

	A_x	A_y	N
Flexible (FO):	*creative*	*new*	*idea*
Rigid (RO):	*different*	*architectural*	*style*

Distance from adjectives	$cosA_x$, $cosA_y$	FO similarly close to both component As RO systematically closer to A_y than A_x
Distance from noun	$cosN$	RO will distort the meaning of the N more than FO, i.e., farther from the N
Distance from ANs	$cosA_xN$, $cosA_yN$	FO will share properties with both ANs RO will share more with A_yN than A_xN

Properties of correct order in RO phrases

	A_x	A_y	N
Attested-order RO (A):	*rapid*	*social*	*change*
Unattested-order RO (U):	**social*	*rapid*	*change*

We expect that the fundamental property that distinguishes the orders is again the degree of modification of both component adjectives

- The modification strength of the A_y on the N results in a single concept created by the A_yN in attested-order rigid AANs, such as *social change* in *rapid social change*

Properties of correct order in RO phrases

	A_x	A_y	N
Attested-order RO (A):	*rapid*	*social*	*change*
Unattested-order RO (U):	**social*	*rapid*	*change*

We expect that the fundamental property that distinguishes the orders is again the degree of modification of both component adjectives

- The modification strength of the A_y on the N results in a single concept created by the A_yN in attested-order rigid AANs, such as *social change* in *rapid social change*
- When seen in the incorrect ordering, i.e., $^{?}$*social rapid change*, the strong modification of *social* will still dominate the meaning of the AAN

Properties of correct order in RO phrases

	A_x	A_y	N
Attested-order RO (A):	*rapid*	*social*	*change*
Unattested-order RO (U):	**social*	*rapid*	*change*

We expect that the fundamental property that distinguishes the orders is again the degree of modification of both component adjectives

- The modification strength of the A_y on the N results in a single concept created by the $\mathrm{A}_y\mathrm{N}$ in attested-order rigid AANs, such as *social change* in *rapid social change*
- When seen in the incorrect ordering, i.e., ?*social rapid change*, the strong modification of *social* will still dominate the meaning of the AAN
- Note: Analysis only possible for model-predicted vectors

Quality of corpus-extracted AAN vectors

Examples of top neighbors of gold standard

Flexible Order	Rigid Order
medieval old town	*British naval power*
fascinating town impressive cathedral medieval street	naval war British navy naval power
rural poor people	*contemporary political issue*
poor rural people rural infrastructure rural people	cultural topic contemporary debate contemporary politics
friendly helpful staff	*rapid social change*
near hotel helpful staff quick service	social conflict social transition cultural consequence
national daily newspaper	*fresh organic vegetable*
national newspaper major newspaper daily newspaper	organic vegetable organic fruit organic product
creative new idea	*last live performance*
innovative effort creative design dynamic part	final gig live dvd live release

Quality of model-generated AAN vectors

Mean cosine similarities between the corpus-extracted and model-generated gold AAN vectors

	Gold
LFM	0.655
F.ADD	0.618
W.ADD	0.565
MULT	0.424

- All composition functions are able to approximate corpus-extracted gold AAN vectors
- All average cosines are significantly above chance

Results: Corpus-observed AAN vectors

Distinguishing flexible vs. rigid order

	Measure	t	*sig.*	
	$\cos A_x$	2.478		
	$\cos A_y$	-4.348	*	RO>FO
CORP	cosN	4.656	*	FO>RO
	$\cos A_x N$	5.913	*	FO>RO
	$\cos A_y N$	1.970		

t-normalized differences for CORP vectors for the gold items.
For all significant results, $p<0.05$.

	A_x	A_y	N
Flexible (FO):	*rural*	*poor*	*people*
Rigid (RO):	*rapid*	*social*	*change*

Results: Corpus-observed AAN vectors

Distinguishing flexible vs. rigid order

	Measure	*t*	*sig.*	
	cosA_x	2.478		
	cosA_y	-4.348	*	RO>FO
CORP	cosN	4.656	*	FO>RO
	cosA_xN	5.913	*	FO>RO
	cosA_yN	1.970		

t-normalized differences for CORP vectors for the gold items.
For all significant results, $p<0.05$.

	A_x	A_y	N
Flexible (FO):	*rural*	*poor*	*people*
Rigid (RO):	*rapid*	*social*	*change*

Results: Corpus-observed AAN vectors

Distinguishing flexible vs. rigid order

	Measure	t	*sig.*	
	$\cos A_x$	2.478		
	$\cos A_y$	-4.348	*	RO>FO
CORP	cosN	4.656	*	FO>RO
	$\cos A_x N$	5.913	*	FO>RO
	$\cos A_y N$	1.970		

t-normalized differences for CORP vectors for the gold items.
For all significant results, $p<0.05$.

	A_x	A_y	N
Flexible (FO):	*rural*	*poor*	*people*
Rigid (RO):	*rapid*	*social*	*change*

Results: Model-generated AAN vectors

Distinguishing flexible vs. rigid order

	Measure	*t*	*sig.*	
W.ADD	cosA_x	4.805	*	FO>RO
	cosA_y	-1.109		
	cosN	1.140		
	cosA_xN	1.059		
	cosA_yN	0.584		
F.ADD	cosA_x	2.050		
	cosA_y	-1.451		
	cosN	**4.493**	*****	FO>RO
	cosA_xN	-0.445		
	cosA_yN	2.300		
MULT	cosA_x	3.830	*	FO>RO
	cosA_y	-0.503		
	cosN	**5.090**	*****	FO>RO
	cosA_xN	**4.435**	*****	FO>RO
	cosA_yN	3.900	*	FO>RO
LFM	cosA_x	-1.649		
	cosA_y	-1.272		
	cosN	**5.539**	*****	FO>RO
	cosA_xN	**3.336**	*****	FO>RO
	cosA_yN	4.215	*	FO>RO

Results: Properties of correct order in RO phrases

	A_x	A_y	N
Attested-order RO:	*rapid*	*social*	*change*
Unattested-order RO:	**social*	*rapid*	*change*

Results: Properties of correct order in RO phrases

	Measure	*t*	*sig.*	
W.ADD	cosA_x	**-7.840**	*	U>A
	cosA_y	**7.924**	*	A>U
	cosN	2.394		
	cosA_xN	-5.462	*	U>A
	cosA_yN	3.627	*	A>U
F.ADD	cosA_x	**-8.418**	*	U>A
	cosA_y	**6.534**	*	A>U
	cosN	-1.927		
	cosA_xN	-3.583	*	U>A
	cosA_yN	-2.185		
MULT	cosA_x	**-5.100**	*	U>A
	cosA_y	**5.100**	*	A>U
	cosN	0.000		
	cosA_xN	-0.598		
	cosA_yN	0.598		
LFM	cosA_x	**-7.498**	*	U>A
	cosA_y	**7.227**	*	A>U
	cosN	-2.172		
	cosA_xN	-5.792	*	U>A
	cosA_yN	0.774		

Results: Properties of correct order in RO phrases

	A_x	A_y	N
Attested-order RO:	*rapid*	*social*	*change*
Unattested-order RO:	**social*	*rapid*	*change*

Conclusions

- We are able to significantly model human intuitions about the semantic acceptability of novel AN phrases using simple, unsupervised cues
    - The measures and functions that model human intuition provide insight into the semantic processing and the acceptability of novel AN phrases

Conclusions

- We are able to significantly model human intuitions about the semantic acceptability of novel AN phrases using simple, unsupervised cues
    - The measures and functions that model human intuition provide insight into the semantic processing and the acceptability of novel AN phrases

- Composition functions are able to approximate corpus-extracted AANs quite well as tested on corpus-extracted vectors shown to be meaningful, semantically coherent objects
    - Distributional cues of corpus-extracted AAN vectors reflect the distinction between "flexible" AANs and "rigid" ones

Conclusions

- We are able to significantly model human intuitions about the semantic acceptability of novel AN phrases using simple, unsupervised cues
  - The measures and functions that model human intuition provide insight into the semantic processing and the acceptability of novel AN phrases

- Composition functions are able to approximate corpus-extracted AANs quite well as tested on corpus-extracted vectors shown to be meaningful, semantically coherent objects
  - Distributional cues of corpus-extracted AAN vectors reflect the distinction between "flexible" AANs and "rigid" ones

- Distributional models are able to represent the distinction between *intersective*, *subsective* and *intensional** adjectival modification (corpus-observed vectors)
  - intensionality was only alleged? (Boleda et al, 2013)

Part II: Preparing presentations in NLP

Tell a story!

1. **Take-away Message**
    - Define it: What do you want audience to have gained/learned?
    - Focus on one (only one) *central idea* for the talk

Tell a story!

1. **Take-away Message**
   - Define it: What do you want audience to have gained/learned?
   - Focus on one (only one) *central idea* for the talk
2. **Research Question**
   - Define: What question are you aiming to investigate?
   - This will be the *continuous thread* connecting all of the talk – from motivation to analysis of results
   - Do not diverge from it (unnecessarily)
   - Do not forget to address it throughout

Tell a story!

1. **Take-away Message**
    - Define it: What do you want audience to have gained/learned?
    - Focus on one (only one) *central idea* for the talk
2. **Research Question**
    - Define: What question are you aiming to investigate?
    - This will be the *continuous thread* connecting all of the talk – from motivation to analysis of results
    - Do not diverge from it (unnecessarily)
    - Do not forget to address it throughout
3. **Know your audience**
    - Consider their breadth of knowledge
    - Goal: 80% of audience should understand 80% of your talk

Elements of a talk: Introduction

- **Motivate** your research question: Why should anyone care about this?
- Shine light on relevant literature, previous work, definitions, theory behind it, etc.
- Set the stage for how you (or the work you are presenting) faced the research question

Elements of a talk: Methodology

- What tools did you need and what steps did you take?
- Approach should clearly reflect the research question – clear that this data and/or these steps should yield results relevant to hypothesis and proposed question
- Consider carefully what details should be included: what is obvious, what is not? – *reproducibility*
- **For this specific talk**: you might want to compare approach in literature with any differences in your practical study
  - What do you predict will be effect on results? Any difference in research question between two approaches?

Elements of a talk: Results and Analysis

- Provide clear, readable results for your practical study
- Graphs, tables, figures, should all be easy to read and understand based on methodology
- **Analysis of results**
  - Detailed review of results
  - What does it all *mean*? Big picture – how do these results answer your research question, did it follow your hypothesis, are there clear open issues or holes (in data, in design, etc)?

Elements of a talk: Conclusion

- Pull all elements of the story together (summary)
- What did you (we) learn?
- What can or should be done as a next step – what's not fully answered, how can these results be applied, *how can this impact our capacity to model meaning*?

Slides

- Meaningful titles
- Calculate **1-2 minutes per slide**
- Figures are key: readable and only essential information
- Explain axis of figures shown, result expected and meaning of result achieved, etc.

Prepare

- Think of what you want to say on each slide
- Practice at home or with friends, start to finish, *more than once!*
- "Memorize" first two sentences of talk – gets you going even if nervous
- Be on time

Evaluation

- Organization and Quality of talk
  - Follow logical progression
  - Proper language and content
  - Significance of topic clearly stated and explained
  - Slide quality and effectiveness
- Theory: Understanding of content
  - Ability to identify research question
  - Understanding of experimental approach and significance
  - Well-researched on relevant material
- Practical Study
  - Experimental design and implementation
  - Understanding of results, methodology and conclusions
- Ability to answer questions
  - Understanding questions
  - Integrating knowledge learned to answer questions

"It usually takes more than three weeks to prepare a good impromptu speech."

– Mark Twain

Thank you for your attention!

Questions?

http://evavecchi.com