Universität Stuttgart

Sabine
Schulte im Walde

Institut für
Maschinelle
Sprachverarbeitung

Aspects of Figurative Language in Noun Compound Models

MWE Workshop, Marseille – June 25, 2022

Motivation & Story

1

Noun Compounds & Compositionality

- Noun Compounds:
  - multiword expression (MWE) with some degree of idiosynchracy
  - composition of modifier constituent(s) and nominal head constituent, e.g., *climate change, crocodile tears*

Noun Compounds & Compositionality

- Noun Compounds:
  - multiword expression (MWE) with some degree of idiosynchracy
  - composition of modifier constituent(s) and nominal head constituent, e.g., *climate change, crocodile tears*
- Compositionality:
  - compositionality: meaning contributions of constituents to compound meaning $\rightarrow$ strength of semantic relatedness
  - application: linguistic theory & NLU within/across languages

Noun Compounds & Compositionality

- Noun Compounds:
  - multiword expression (MWE) with some degree of idiosynchracy
  - composition of modifier constituent(s) and nominal head constituent, e.g., *climate change, crocodile tears*
- Compositionality:
  - compositionality: meaning contributions of constituents to compound meaning $\rightarrow$ strength of semantic relatedness
  - application: linguistic theory & NLU within/across languages
- Task & Models:
  - task: predict the degree of compound compositionality as a whole/phrase and with regard to its constituents
  - models: textual/multi-modal vector-space models (VSMs); rely on cosine distance as a proxy to semantic relatedness

Examples for English

[Reddy et al., 2011]

Compounds	Mean Ratings and STDs		
	modifier	head	compound
climate change	**4.90**±0.30	**4.83**±0.38	**4.97**±0.18
polo shirt	**1.73**±1.41	**5.00**±0.00	**3.37**±1.38
search engine	**4.62**±0.96	**2.25**±1.70	**3.32**±1.16
cheat sheet	**2.30**±1.59	**4.00**±0.83	**2.89**±1.11
gilt trip	**4.71**±0.59	**0.86**±0.94	**2.19**±1.16
night owl	**4.47**±0.88	**0.50**±0.82	**1.93**±1.27
crocodile tears	**0.19**±0.47	**3.79**±1.05	**1.25**±1.09
diamond wedding	**0.78**±1.29	**3.41**±1.34	**1.70**±1.05
melting pot	**1.00**±1.15	**0.48**±0.63	**0.54**±0.63

- scale [0,5] from 0 (semantically opaque) to 5 (semantically transparent)
- annotators: 30 per compound–constituent pair
- mean values and standard deviations

Story for Today

- **Dataset creation**
  - target source, e.g., dictionary; corpus; WordNet
  - target properties:
    - empirical properties, e.g., frequency; productivity
    - lexical-semantic properties, e.g., ambiguity; concreteness
    - semantic relations between compounds and constituents

Story for Today

- **Dataset creation**
  - target source, e.g., dictionary; corpus; WordNet
  - target properties:
    - empirical properties, e.g., frequency; productivity
    - lexical-semantic properties, e.g., ambiguity; concreteness
    - semantic relations between compounds and constituents
- **Compositionality prediction models**
  - standard models: vector-space models (VSMs)
    - traditional count-based VSMs (window-/syntax-based)
    - dimensionality-reduced VSMs (including word2vec)
    - multimodal VSMs (text + images)
  - influence of target properties on prediction results

Story for Today

- **Dataset creation**
  - target source, e.g., dictionary; corpus; WordNet
  - target properties:
    - empirical properties, e.g., frequency; productivity
    - lexical-semantic properties, e.g., ambiguity; concreteness
    - semantic relations between compounds and constituents
- **Compositionality prediction models**
  - standard models: vector-space models (VSMs)
    - traditional count-based VSMs (window-/syntax-based)
    - dimensionality-reduced VSMs (including word2vec)
    - multimodal VSMs (text + images)
  - influence of target properties on prediction results
- **[Diachronic changes of compound meanings]**

Datasets

2

Reddy et al. English Dataset

[Reddy et al., 2011]

- Composition:
  - English noun compounds (with two simplex constituents)
  - compounds and heads are nouns; modifier word classes vary
- Construction:
  - four classes of modifier and head constituent combinations regarding their contribution to compound meaning
  - WordNet-based heuristic:
    a compound is considered compositional with regard to a constituent if the constituent either represents a hypernym of the compound or is used in the definition, e.g., *swimming pool*
  - semantic relations added [Bell and Schäfer, 2013]
- Dataset: 90 noun-noun compounds;
  extension to 280 noun compounds [Cordeiro et al., 2019]

Examples (Reddy et al.) – EN

[Reddy et al., 2011]

Compounds	Mean Ratings and STDs		
	modifier	head	compound
climate change	**4.90**±0.30	**4.83**±0.38	**4.97**±0.18
polo shirt	**1.73**±1.41	**5.00**±0.00	**3.37**±1.38
search engine	**4.62**±0.96	**2.25**±1.70	**3.32**±1.16
cheat sheet	**2.30**±1.59	**4.00**±0.83	**2.89**±1.11
gilt trip	**4.71**±0.59	**0.86**±0.94	**2.19**±1.16
night owl	**4.47**±0.88	**0.50**±0.82	**1.93**±1.27
crocodile tears	**0.19**±0.47	**3.79**±1.05	**1.25**±1.09
diamond wedding	**0.78**±1.29	**3.41**±1.34	**1.70**±1.05
melting pot	**1.00**±1.15	**0.48**±0.63	**0.54**±0.63

- scale [0,5] from 0 (semantically opaque) to 5 (semantically transparent)
- annotators: 30 per compound–constituent pair

Our German Datasets

- Composition:
  - German noun-noun compounds (with two simplex constituents)
  - compounds, modifiers and heads are nouns
- Example:
  - *Fliegenpilz* 'toadstool': *Fliege* 'fly/bow tie' + *Pilz* 'mushroom'
- Datasets:
  - CONCRETE-NN: 244 depictable noun-noun compounds [von der Heide and Borgwaldt, 2009, Schulte im Walde et al., 2013]
  - G$_h$oSt-NN: 868 noun-noun compounds randomly extracted from DECOW14AX [Schäfer and Bildhauer, 2012], taking compound and constituent properties into account (corpus frequencies, constituent productivity, ambiguity, semantic relations) [Schulte im Walde et al., 2016]
  - FEATURE-NN: 1,099 (≈244+868) noun-noun compounds: meaning components, compound–head hypernymy, concreteness of compounds and constituents, compound–constituent compositionality

Examples (CONCRETE-NN) – DE

[von der Heide and Borgwaldt, 2009, Schulte im Walde et al., 2013]

Compounds and Constituents			Mean Ratings and STDs		
compounds	literal constituent meanings		compound	modifier	head
Ahornblatt 'maple leaf'	maple	leaf	**6.03** ± 1.49	**5.64** ± 1.63	**5.71** ± 1.70
Postbote 'post man'	mail	messenger	**6.33** ± 0.96	**5.87** ± 1.55	**5.10** ± 1.99
Seezunge 'sole'	sea	tongue	**1.85** ± 1.28	**3.57** ± 2.42	**3.27** ± 2.32
Windlicht 'storm lamp'	wind	light	**3.52** ± 2.08	**3.07** ± 2.12	**4.27** ± 2.36
Löwenzahn 'dandelion'	lion	tooth	**1.66** ± 1.54	**2.10** ± 1.84	**2.23** ± 1.92
Maulwurf 'mole'	mouth	throw	**1.58** ± 1.43	**2.21** ± 1.68	**2.76** ± 2.10
Fliegenpilz 'toadstool'	fly/bow tie	mushroom	**2.00** ± 1.20	**1.93** ± 1.28	**6.55** ± 0.63
Flohmarkt 'flea market'	flea	market	**2.31** ± 1.65	**1.50** ± 1.22	**6.03** ± 1.50
Feuerzeug 'lighter'	fire	stuff	**4.58** ± 1.75	**5.87** ± 1.01	**1.90** ± 1.03
Fleischwolf 'meat chopper'	meat	wolf	**1.70** ± 1.05	**6.00** ± 1.44	**1.90** ± 1.42

- scale [1,7] from 1 (semantically opaque) to 7 (semantically transparent)
- annotators: 30 per compound–constituent pair

Examples (G$_h$ost-NN) – DE

[Schulte im Walde et al., 2016]

Compounds		Frequencies			Productivities		Ambiguities		Relation	Ratings	
		comp.	*mod.*	*head*	*mod.*	*head*	*mod.*	*head*		*mod.*	*head*
Sonnenenergie	*solar energy*	25,398	832,636	1,191,333	155	30	3	2	INST	4.58	5.44
Sonnenkönig	*Sun King*	2,680	832,636	494,221	155	109	3	3	LEX	1.94	5.50
Sonnenscheibe	*solar disc*	3,155	832,636	364,567	155	96	3	4	BE	4.56	3.75
Sonnenseite	*sunny side*	7,279	832,636	5,508,445	155	256	3	6	IN	4.00	4.31
Sonnenstrahl	*sunbeam*	44,612	832,636	32,182	155	27	3	3	HAVE	5.13	4.69
Sonnenuhr	*sundial*	8,407	832,636	4,507,590	155	63	3	2	INST	3.75	5.31
Kirchspiel	*parish*	6,583	1,761,187	4,122,168	319	403	3	6	LEX	4.44	3.13
Machtspiel	*power game*	4,408	806,162	4,122,168	169	403	2	6	ABOUT	4.63	3.44
Ritterspiel	*knights' tournament*	2,365	115,484	4,122,168	47	403	1	6	ACTOR	3.94	4.75
Testspiel	*tryout*	37,800	660,169	4,122,168	100	403	3	6	BE	4.25	5.19
Windspiel	*wind chimes*	2,284	551,317	4,122,168	88	403	3	6	INST	4.31	2.94
Winterspiel	*winter games*	16,067	721,552	4,122,168	207	403	1	6	IN	4.43	5.14
Würfelspiel	*game of dice*	4,408	80,371	4,122,168	14	403	2	6	INST	4.94	5.56

- **G$_h$ost-NN/S**: $20 \times 9 = 180$ compounds randomly selected but balanced for modifier productivity (low/mid/high) and head ambiguity (1/2/>2)

Examples (G$_h$ost-NN) – DE

[Schulte im Walde et al., 2016]

Compounds		Frequencies			Productivities		Ambiguities		Relation	Ratings	
		comp.	mod.	head	mod.	head	mod.	head		mod.	head
Sonnenenergie	*solar energy*	25,398	832,636	1,191,333	155	30	3	2	INST	4.58	5.44
Sonnenkönig	*Sun King*	2,680	832,636	494,221	155	109	3	3	LEX	1.94	5.50
Sonnenscheibe	*solar disc*	3,155	832,636	364,567	155	96	3	4	BE	4.56	3.75
Sonnenseite	*sunny side*	7,279	832,636	5,508,445	155	256	3	6	IN	4.00	4.31
Sonnenstrahl	*sunbeam*	44,612	832,636	32,182	155	27	3	3	HAVE	5.13	4.69
Sonnenuhr	*sundial*	8,407	832,636	4,507,590	155	63	3	2	INST	3.75	5.31
Kirchspiel	*parish*	6,583	1,761,187	4,122,168	319	403	3	6	LEX	4.44	3.13
Machtspiel	*power game*	4,408	806,162	4,122,168	169	403	2	6	ABOUT	4.63	3.44
Ritterspiel	*knights' tournament*	2,365	115,484	4,122,168	47	403	1	6	ACTOR	3.94	4.75
Testspiel	*tryout*	37,800	660,169	4,122,168	100	403	3	6	BE	4.25	5.19
Windspiel	*wind chimes*	2,284	551,317	4,122,168	88	403	3	6	INST	4.31	2.94
Winterspiel	*winter games*	16,067	721,552	4,122,168	207	403	1	6	IN	4.43	5.14
Würfelspiel	*game of dice*	4,408	80,371	4,122,168	14	403	2	6	INST	4.94	5.56

- **G$_h$ost-NN/S**: 20 $\times$ 9 = 180 compounds randomly selected but balanced for modifier productivity (low/mid/high) and head ambiguity (1/2/>2)
- **G$_h$ost-NN/XL**: 868 compounds, after adding all compounds with the same modifiers/heads as in G$_h$ost/S

Examples (G$_h$ost-NN) – DE

[Schulte im Walde et al., 2016]

Compounds		Frequencies			Productivities		Ambiguities		Relation	Ratings	
		comp.	mod.	head	mod.	head	mod.	head		mod.	head
Sonnenenergie	*solar energy*	25,398	832,636	1,191,333	155	30	3	2	INST	4.58	5.44
Sonnenkönig	*Sun King*	2,680	832,636	494,221	155	109	3	3	LEX	1.94	5.50
Sonnenscheibe	*solar disc*	3,155	832,636	364,567	155	96	3	4	BE	4.56	3.75
Sonnenseite	*sunny side*	7,279	832,636	5,508,445	155	256	3	6	IN	4.00	4.31
Sonnenstrahl	*sunbeam*	44,612	832,636	32,182	155	27	3	3	HAVE	5.13	4.69
Sonnenuhr	*sundial*	8,407	832,636	4,507,590	155	63	3	2	INST	3.75	5.31
Kirchspiel	*parish*	6,583	1,761,187	4,122,168	319	403	3	6	LEX	4.44	3.13
Machtspiel	*power game*	4,408	806,162	4,122,168	169	403	2	6	ABOUT	4.63	3.44
Ritterspiel	*knights' tournament*	2,365	115,484	4,122,168	47	403	1	6	ACTOR	3.94	4.75
Testspiel	*tryout*	37,800	660,169	4,122,168	100	403	3	6	BE	4.25	5.19
Windspiel	*wind chimes*	2,284	551,317	4,122,168	88	403	3	6	INST	4.31	2.94
Winterspiel	*winter games*	16,067	721,552	4,122,168	207	403	1	6	IN	4.43	5.14
Würfelspiel	*game of dice*	4,408	80,371	4,122,168	14	403	2	6	INST	4.94	5.56

- **G$_h$ost-NN/S**: $20 \times 9 = 180$ compounds randomly selected but balanced for modifier productivity (low/mid/high) and head ambiguity (1/2/>2)
- **G$_h$ost-NN/XL**: 868 compounds, after adding all compounds with the same modifiers/heads as in G$_h$ost/S
- scale [1,6] from 1 (semantically opaque) to 6 (semantically transparent)
- annotators: 8 per compound–constituent pair; 13 for balanced subset

Annotation++ in FEATURE-NN – DE

What's on the annotators' minds? → Ask for more!

Annotation++ in FEATURE-NN – DE

What's on the annotators' minds? → Ask for more!

- Compound meaning:
  paraphrase compound meaning in one phrase/sentence
- Constituent meaning contribution:
  provide feature(s) of constituents contributing to compound meaning
- Super-/sub-ordination:
  judge if compound *is a kind of* the head constituent on a scale 0–5
- Abstractness/concreteness: judge about concreteness of compound and constituents on a scale 0–5
- Compositionality: judge about compositionality of the compound with regard to constituents on a scale 0–5

Examples (FEATURE-NN) – DE

Compound paraphrases and constituent feature contributions:

Schriftzug	Erscheinungsbild der Schrift; Aneinanderreihung von Buchstaben; Art und Weise einer Schriftart; kurzes Stück Schrift, das meistens besonders im Fokus steht; Worte, die auf einer bestimmten Weise geschrieben wurden;
Schrift	schriftlich, handschriftlich; geschrieben; schreiben, Stift; geschrieben, festgehalten, Sprache; geschrieben
Zug	Art aufeinanderfolgend; Fortbewegungsmittel auf Schienen; lang, ziehen, zusammenhängend, Verlauf; lang, schmal

Examples (FEATURE-NN) – DE

Hypernymy ratings:

Kirchspiel	0.8	Festspiel	3.8	Abschiedsspiel	4.6
Gedankenspiel	1.6	Familienspiel	4.0	Freundschaftsspiel	4.6
Glockenspiel	2.2	Fernsehspiel	4.0	Kriegsspiel	4.6
Handspiel	2.2	Orgelspiel	4.0	Mannschaftsspiel	4.6
Liebesspiel	2.6	Passionsspiel	4.0	Punktspiel	4.6
Windspiel	2.6	Passspiel	4.0	Relegationsspiel	4.6
Lichtspiel	2.8	Ritterspiel	4.0	Testspiel	4.6
Trauerspiel	2.9	Saisonspiel	4.0	Trainingsspiel	4.6
Farbenspiel	3.0	Schauspiel	4.0	Zahlenspiel	4.6
Krippenspiel	3.0	Fingerspiel	4.2	Ballspiel	4.8
Angriffsspiel	3.2	Gewinnspiel	4.2	Computerspiel	4.8
Geduldsspiel	3.2	Jugendspiel	4.2	Endspiel	4.8
Machtspiel	3.2	Kombinationsspiel	4.2	Kampfspiel	4.8
Wortspiel	3.2	Schattenspiel	4.2	Konsolenspiel	4.8
Sprachspiel	3.4	Sommerspiel	4.2	Länderspiel	4.8
Wasserspiel	3.4	Stellungsspiel	4.2	Ligaspiel	4.8
Winterspiel	3.4	Auftaktspiel	4.4	Pokerspiel	4.8
Ferienspiel	3.6	Glücksspiel	4.4	Qualifikationsspiel	4.8
Gastspiel	3.6	Golfspiel	4.4	Sportspiel	4.8
Gitarrenspiel	3.6	Lieblingsspiel	4.4	Videospiel	4.8
Kammerspiel	3.6	Meisterschaftsspiel	4.4	Gruppenspiel	4.8
Klavierspiel	3.6	Pflichtspiel	4.4	Schachspiel	4.8
Puppenspiel	3.6	Pokalspiel	4.4	Gesellschaftsspiel	5.0
Kinderspiel	3.6	Puzzlespiel	4.4	Kartenspiel	5.0
Heimspiel	3.7	Sonntagsspiel	4.4	Würfelspiel	5.0

Examples (FEATURE-NN) – DE

Concreteness ratings (compounds, modifiers, heads):

Charakterzug	0.6	0.4	2.6
Wesenszug	1.0	1.8	2.4
Schachzug	1.5	2.8	3.3
Beutezug	1.6	3.4	2.6
Feldzug	1.6	4.4	2.8
Siegeszug	1.6	1.8	3.0
Kreuzzug	1.8	4.2	3.0
Kriegszug	2.0	0.8	3.2
Luftzug	2.4	4.4	2.6
Triumphzug	2.8	0.8	3.4
Protestzug	3.0	1.0	3.2
Trauerzug	3.0	0.4	4.0
Schriftzug	3.2	2.2	3.8
Vogelzug	3.2	4.6	2.8
Atemzug	3.4	2.0	2.2
Fackelzug	3.6	4.8	3.6
Gebirgszug	4.4	4.6	2.8
Gummizug	4.6	4.4	4.2
Reisezug	4.6	1.4	4.4
Seilzug	4.6	4.6	4.0
Dampfzug	5.0	4.0	5.0
Lastzug	5.0	3.4	4.6
Nachtzug	5.0	3.2	5.0
Personenzug	5.0	4.4	5.0
Triebzug	5.0	0.4	5.0

Summary: Datasets

- Languages: English and German
- Targets: noun compounds
- Target selection:
  - perception-based/concreteness
  - WordNet-based (hypernymy relations and definitions)
  - corpus-based (empirical)
- Annotation:
  - ratings on a scale $\rightarrow$ mean ratings and standard deviations
  - ++ paraphrases and features, hypernymy, concreteness

Standard Vector-Space Models (VSMs)

3

Distributional Hypothesis

Each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts. [Harris, 1954]

You shall know a word by the company it keeps. [Firth, 1957]

Distributional Hypothesis

Each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts. [Harris, 1954]

You shall know a word by the company it keeps. [Firth, 1957]

- The context of a linguistic unit contains indicators for its usage and its meaning.

Distributional Hypothesis

Each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts. [Harris, 1954]

You shall know a word by the company it keeps. [Firth, 1957]

- The context of a linguistic unit contains indicators for its usage and its meaning.
- Linguistic levels:
  phoneme, morpheme, syntactic, semantic, discourse, ...

Distributional Hypothesis

Each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts. [Harris, 1954]

You shall know a word by the company it keeps. [Firth, 1957]

- The context of a linguistic unit contains indicators for its usage and its meaning.
- Linguistic levels:
  phoneme, morpheme, syntactic, semantic, discourse, ...
- Comparing linguistic units on the basis of their distributional features $\rightarrow$ degree of relatedness.

Distributional Space

Vector-Space Models (VSMs) of Semantics

- Basis: represent an object through its features as a vector/point in space, e.g., $\overrightarrow{butterfly} = \langle 23, 116, 0, 0, 346 \rangle$

Distributional Space

Vector-Space Models (VSMs) of Semantics

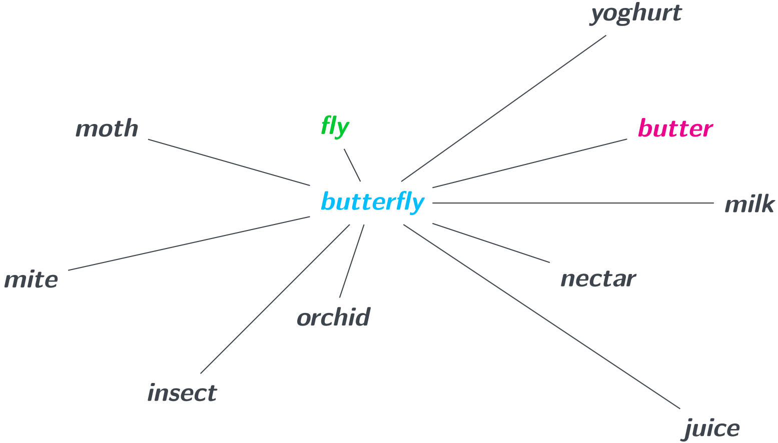
- Basis: represent an object through its features as a vector/point in space, e.g., $\overrightarrow{butterfly} = \langle 23, 116, 0, 0, 346 \rangle$
- Values of VSM objects are derived from event frequencies.
  - bag-of-word co-occurrences
    (in a window, sentence, paragraph, document)
  - syntactic dependencies
  - etc.

Distributional Space

Vector-Space Models (VSMs) of Semantics

- Basis: represent an object through its features as a vector/point in space, e.g., $\overrightarrow{butterfly} = \langle 23, 116, 0, 0, 346 \rangle$
- Values of VSM objects are derived from event frequencies.
  - bag-of-word co-occurrences
    (in a window, sentence, paragraph, document)
  - syntactic dependencies
  - etc.
- Relatedness: Points that are close together in space are semantically related, and points that are far apart are semantically distant.

Distributional Space Example



VSMs for Predicting Compositionality

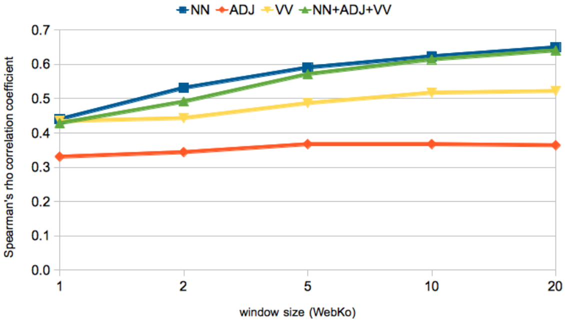
- Basis: vector representations for compounds and constituents
- Relatedness: distance between the vectors of compounds and their modifier and head constituents, relying on cosine score
  - models for German: plain cosine score
  - models for English: cosine scores for vectors weighted by pre-trained coefficients
- Compositionality: **VSM relatedness $\sim$ compositionality**
- Data:
  - CONCRETE-NN, G$_h$ost-NN, FEATURE-NN, Reddy et al.
  - web corpora: WaCky and COW
    [Baroni et al., 2009, Schäfer and Bildhauer, 2012]

Variations of VSM Representations

4

Window Models (CONCRETE-NN)

[Schulte im Walde et al., 2013]



- nouns provide the most salient features
- NN > NN+ADJ+VV > VV > ADJ (significant)
- best result: $\rho = .650$ (NN only, window size: 20)
- window-based models > syntax-based models

Vector-Space Reductions (Reddy Dataset)

[Alipoor and Schulte im Walde, 2020]

- Vector-space variants
  - all context words
  - only nouns or only verbs
  - PCA (principal components analysis)
  - word2vec
- Prediction functions
  - MOD (modifier constituent)
  - HEAD (head constituent)
  - ADD (addition)
  - MULT (multiplication)
  - COMB (addition and multiplication)
- Target subsets (3×30) [Schulte im Walde et al., 2016]
  - degree of compositionality
  - compound frequency
  - modifier productivity
  - head productivity

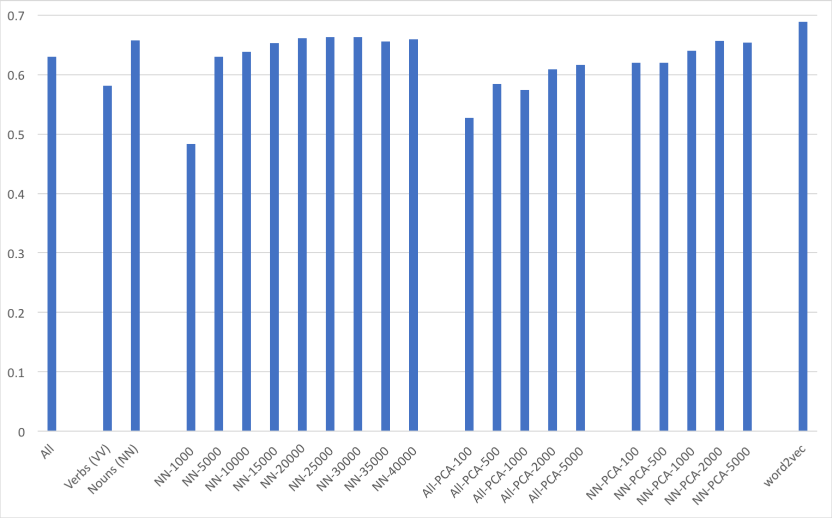
Vector-Space Reductions (Reddy Dataset)

	Vector-Space Variants	Dimensionality
All	all context words co-occurring with any of our targets	64,508
VV	all verbs	8,525
NN	all nouns	52,285
NN-1000	1,000 most frequent corpus nouns	374
NN-5000	5,000 most frequent corpus nouns	1,221
NN-10000	10,000 most frequent corpus nouns	2,392
NN-15000	15,000 most frequent corpus nouns	3,615
NN-20000	20,000 most frequent corpus nouns	4,762
NN-25000	25,000 most frequent corpus nouns	5,929
NN-30000	30,000 most frequent corpus nouns	6,970
NN-35000	35,000 most frequent corpus nouns	8,058
NN-40000	40,000 most frequent corpus nouns	9,114
All-PCA-100	PCA with 100 dimensions computed on whole matrix	100
All-PCA-500	PCA with 500 dimensions computed on whole matrix	500
All-PCA-1000	PCA with 1,000 dimensions computed on whole matrix	1,000
All-PCA-2000	PCA with 2,000 dimensions computed on whole matrix	2,000
All-PCA-5000	PCA with 5,000 dimensions computed on whole matrix	5,000
NN-PCA-100	PCA with 100 dimensions computed on noun matrix	100
NN-PCA-500	PCA with 500 dimensions computed on noun matrix	500
NN-PCA-1000	PCA with 1,000 dimensions computed on noun matrix	1,000
NN-PCA-2000	PCA with 2,000 dimensions computed on noun matrix	2,000
NN-PCA-5000	PCA with 5,000 dimensions computed on noun matrix	5,000
Word2Vec	word2vec two-layer neural network representation	300

Vector-Space Reductions (Reddy Dataset)

Vector-Space Variants	MOD	HEAD	ADD	MULT	COMB
All	0.583	0.444	**0.630**	0.626	**0.630**
Verbs (VV)	0.534	0.383	0.581	0.387	0.578
Nouns (NN)	0.634	0.433	**0.658**	**0.658**	0.655
NN-1000	0.436	0.324	0.482	0.452	0.483
NN-5000	0.614	0.377	0.630	0.592	0.630
NN-10000	0.618	0.397	0.638	0.632	0.637
NN-15000	0.631	0.429	0.653	0.648	0.652
NN-20000	0.637	0.435	0.661	0.659	0.658
NN-25000	0.640	0.438	**0.663**	**0.663**	0.662
NN-30000	0.641	0.437	0.662	**0.663**	0.662
NN-35000	0.633	0.433	0.656	0.652	0.653
NN-40000	0.635	0.432	0.657	0.659	0.656
All-PCA-100	0.456	0.321	0.527	0.487	0.504
All-PCA-500	0.510	0.357	0.584	0.573	0.577
All-PCA-1000	0.562	0.375	0.564	0.574	0.564
All-PCA-2000	0.554	0.432	0.601	0.604	0.609
All-PCA-5000	0.576	0.432	0.616	0.610	**0.616**
NN-PCA-100	0.536	0.320	0.620	0.587	0.613
NN-PCA-500	0.578	0.353	0.610	0.631	0.620
NN-PCA-1000	0.566	0.402	0.614	0.635	0.640
NN-PCA-2000	0.628	0.433	0.639	**0.657**	0.646
NN-PCA-5000	0.608	0.433	0.643	0.654	0.653
Word2Vec	0.602	0.435	0.680	0.145	**0.689**

Vector-Space Reductions (Reddy Dataset)



Multi-Modal Embeddings Model

[Köper and Schulte im Walde, 2017]

- Multi-modal features:
  - Textual modality: *word2vec* embedding vectors based on *DECOW14* (12 billion words)
  - Visual modality: *GoogLeNet* vs. *AlexNet* image embedding vectors based on `bing.de`

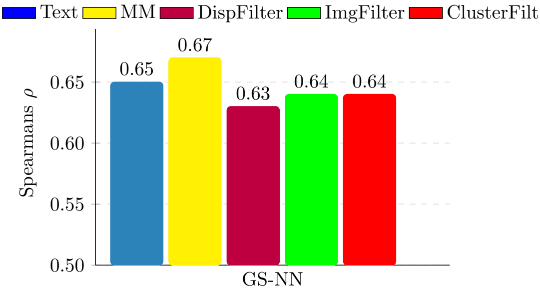
Multi-Modal Embeddings Model

[Köper and Schulte im Walde, 2017]

- Multi-modal features:
  - Textual modality: *word2vec* embedding vectors based on *DECOW14* (12 billion words)
  - Visual modality: *GoogLeNet* vs. *AlexNet* image embedding vectors based on `bing.de`

- Model parameters:
  - Mid-fusion: concatenation of L2-normalised representations
  - Unsupervised dispersion filter for perceptual information
  - Clustering filter for perceptual information
  - Affective norm filter for abstractness, imageability, valency

Multi-Modal Embeddings Model (G$_h$ost-NN/XL)



- tiny gain of multi-modal model (significant, p<0.001, Steiger's test)
- negative effects of filters

Summary: Variants of Standard VSMs

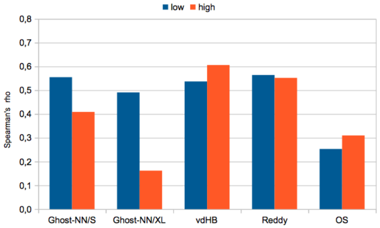
- Simple window-based textual co-occurrence models predict degree of compositionality with up to $\rho = 0.65$ on CONCRETE-NN
- Nouns provide the most salient features
- Vector-space reductions to 20,000–40,000 most frequent corpus are sufficient; but: word2vec is better ($\rho = 0.69$ on Reddy dataset)
- Adding visual information to textual information improves results; imageability, dispersion and clustering filters show negative effect (note: no sota models)

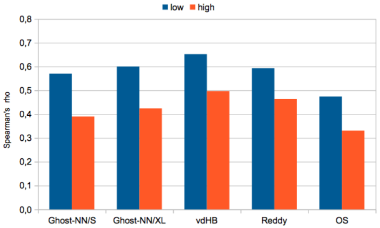
Variations of Target Properties

5

Constituent Property Models

Frequency-based results for 60 minimum (low) and 60 maximum (high) instances (left: modifiers, right: heads):

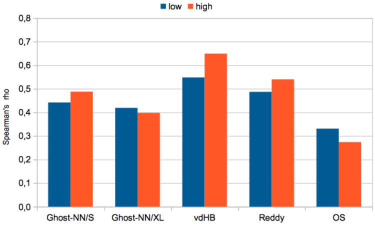


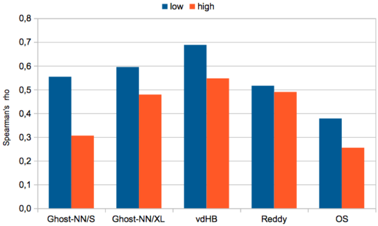


- properties of head constituents play a significant role

Constituent Property Models

Productivity-based results for 60 minimum (low) and 60 maximum (high) instances (left: modifiers, right: heads):

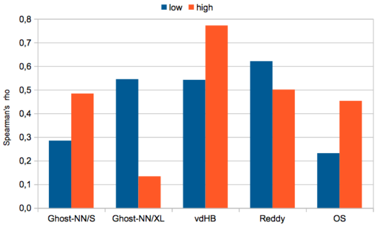


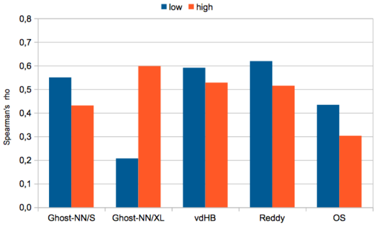


- properties of head constituents play a significant role

Constituent Property Models

Ambiguity-based results for 60 minimum (low) and 60 maximum (high) instances (left: modifiers, right: heads):

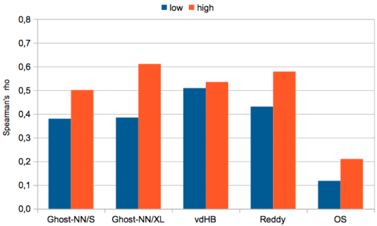




- properties of head constituents play a (significant) role

Compound Property Models

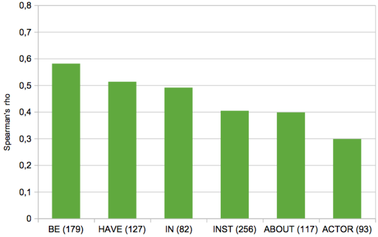
Frequency-based results for 60 minimum (low) and 60 maximum (high) instances:



- frequencies of compounds play a significant role for Ghost-NN/XL

Compound Property Models

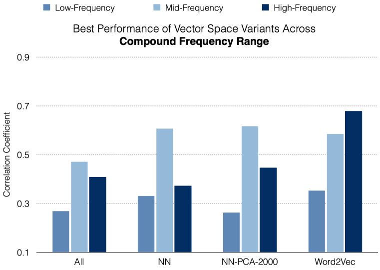
Relation-based results:

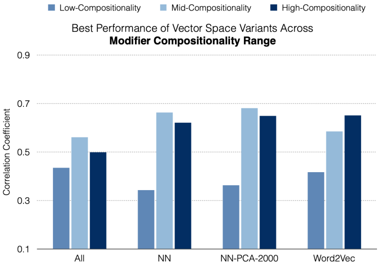


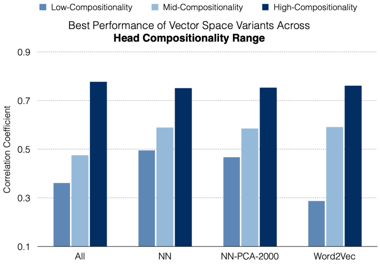
- relations between modifiers and heads in compounds play a significant role:

	HAVE	INST	ABOUT	ACTOR
BE	REDDY	G_hOST	G_hOST	G_hOST, OS
HAVE			OS	G_hOST, OS
IN				G_hOST, OS
INST				G_hOST, OS

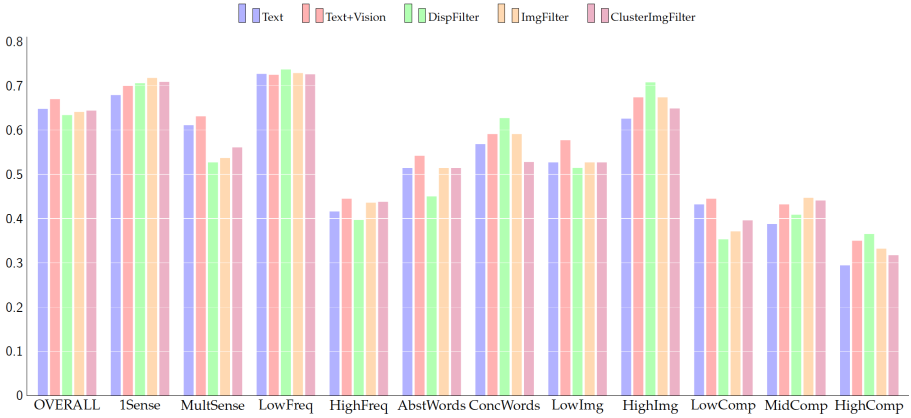
Vector-Space Reductions (Reddy Dataset)







Multi-Modal Embeddings Model ($G_h$ost-NN/XL)



- predictions for monosemous targets are better than those for ambiguous targets; ditto for low-frequency vs. high-frequency targets
- predictions strongly differ regarding the influence of target abstractness, imageability and compositionality

Summary: Variations of Target Properties

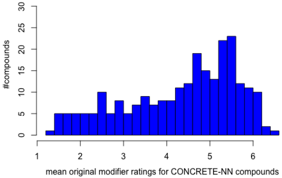
- Predictions of standard models typically report average results
- **But**: results vary strongly across compounds and constituents depending on
  - their empirical properties (frequency; productivity)
  - their semantic properties (ambiguity; semantic relations)
  - their perceptual properties (imageability; concreteness)

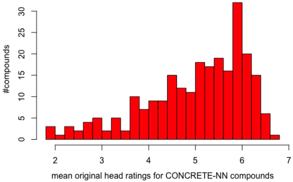
More on Distributions of Target Properties

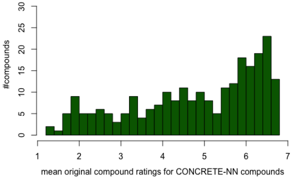
6

Rating Distributions (CONCRETE-/G$_h$ost-NN)

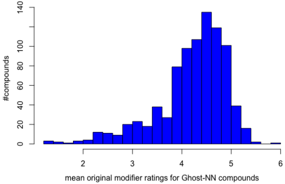
Compositionality of CONCRETE-NN compounds:

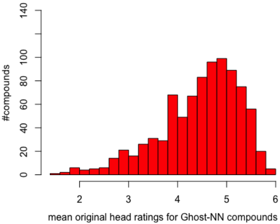






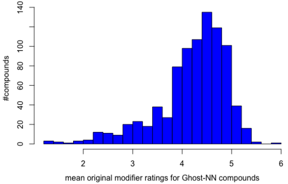
Compositionality of G$_h$ost-NN/XL compounds:

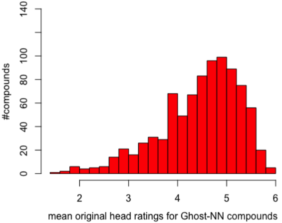




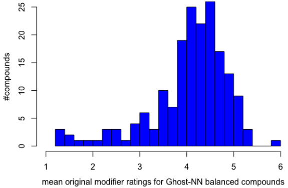
Rating Distributions (G$_h$ost-NN)

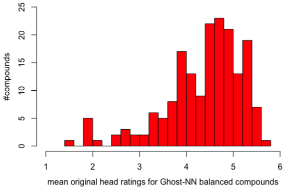
Compositionality of all G$_h$ost-NN/XL compounds:





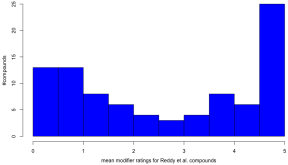
Compositionality of balanced subset of G$_h$ost-NN/S compounds:

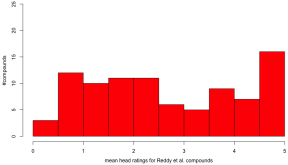


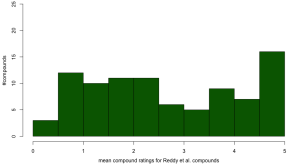


Rating Distributions (Reddy et al. & $G_h$ost-NN/S)

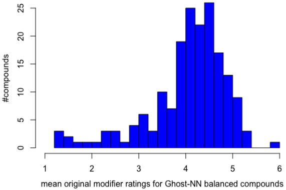
Compositionality of Reddy et al. compounds:

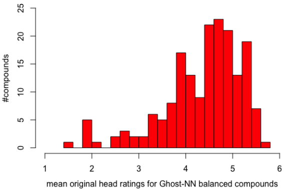






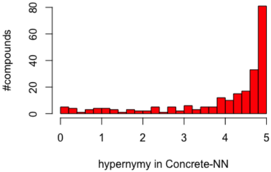
Compositionality of balanced subset of $G_h$ost-NN/S compounds:



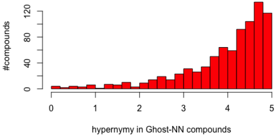


FEATURE-NN: Hypernymy Ratings

Hypernymy in CONCRETE-NN:

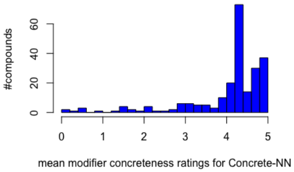


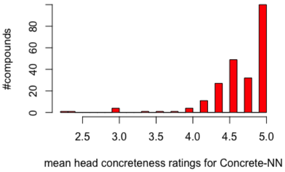
Hypernymy in G$_h$ost-NN:



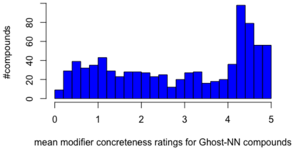
FEATURE-NN: Concreteness Ratings

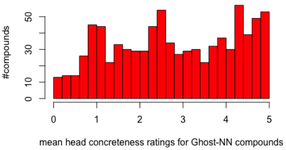
Concreteness of CONCRETE-NN compounds:





Concreteness of $\mathrm{G}_h$ost-NN compounds:





FEATURE-NN: ρ Correlations

		FEATURE-NN					
		compositionality		hypernymy	concreteness		
		mod	head		comp	mod	head
CONCRETE-NN	ORIG comp	.696	.345	.409	.124	.152	.157
	ORIG comp (mod)	.791	-.148	-.121	.046	.086	-.013
	ORIG comp (head)	-.225	.715	.678	.091	-.083	.230
	FEAT comp (mod)		.050	.052	.165	.199	.114
	FEAT comp (head)	.050		.737	.300	.047	.437
	FEAT conc (comp)	.165	.300	.066		.357	.794
	FEAT conc (mod)	.199	.047	.024	.357		.302
	FEAT conc (head)	.114	.437	.242	.794	.302	
$G_h$ost-NN/XL	ORIG comp (mod)	.654	-.084	-.061	.002	.119	-.048
	ORIG comp (head)	-.056	.683	.692	.199	.042	.179
	FEAT comp (mod)		.104	.004	.094	.225	.007
	FEAT comp (head)	.104		.637	.273	.100	.244
	FEAT conc (comp)	.094	.273	.230		.559	.893
	FEAT conc (mod)	.225	.100	.053	.559		.425
	FEAT conc (head)	.007	.244	.176	.893	.425	

Summary & Team

7

Overall Summary

- **Dataset creation**
  - target source, e.g., corpus; WordNet
  - target properties:
    - empirical properties, e.g., frequency; productivity
    - lexical-semantic properties, e.g., ambiguity; concreteness
    - semantic relations between compounds and constituents
- **Compositionality prediction models**
  - most successful standard VSMs: word2vec
  - influence of target properties on prediction results
  - interaction of target properties

Team

- Pegah Alipoor
- Stefan Bott
- Anna Hätty
- Maximilian Köper
- Stefan Müller
- Stephen Roller
- Myrto Tsigkouli

Diachronic Changes in Compositionality

8

Compound Meaning Changes over Time

[Tsigkouli, 2021]

- **Motivation**: investigate
  - compound meanings at their time of emergence
  - changes in compound and constituent meanings over time
  - changes in compositionality of compounds over time
- **Case Studies**:
  - Vector-Space Analyses in construction-grammar framework [Hilpert, 2016]
    - compound + head **spatial progression plots**
    - modifier–head **spatial density plots**
  - Semantic Density Analyses [Sagi et al., 2009]

    average pair-wise cosine score for a word's n most frequent context words in a vector space
- **Data**: English noun compounds [Reddy et al., 2011]

Historical Corpus

Corpus of Historical American English (COHA) [Davies, 2012]

- approx. 400 million words
- grouped into decades with balanced samples of various text genres
- lemmatised and part-of-speech-tagged
- cleaned version **CCOHA** [Alatrash et al., 2020]

Partitioning into **5 wider time periods** with 4 decades each:

period	decades	amount of language data
1	1810-1840	$\approx$ 37 million words
2	1850-1880	$\approx$ 70 million words
3	1890-1920	$\approx$ 88 million words
4	1930-1960	$\approx$ 95 million words
5	1970-2000	$\approx$ 103 million words

Distributional Diachronic Vector Spaces

- Vector Spaces with 250 highly frequent nouns
  - unique fixed semantic space
  - five moving semantic spaces (one for each period of time)
- Distributional Basis:
  - co-occurrence matrices: one for fixed space; five for moving
  - context words: **nouns, verbs, adverbs, adjectives** in **10-word window**, excl. 150 most frequent words and if freq<100
  - use positive point-wise mutual information (ppmi)
- Time-Specific Vector Spaces:
  - **merge** compound and constituent vectors with vector spaces
  - apply **dimensionality reduction** to transform multi-dimensional vector spaces into **2-dimensional coordinate system**: multi-dimensional scaling (MDS)

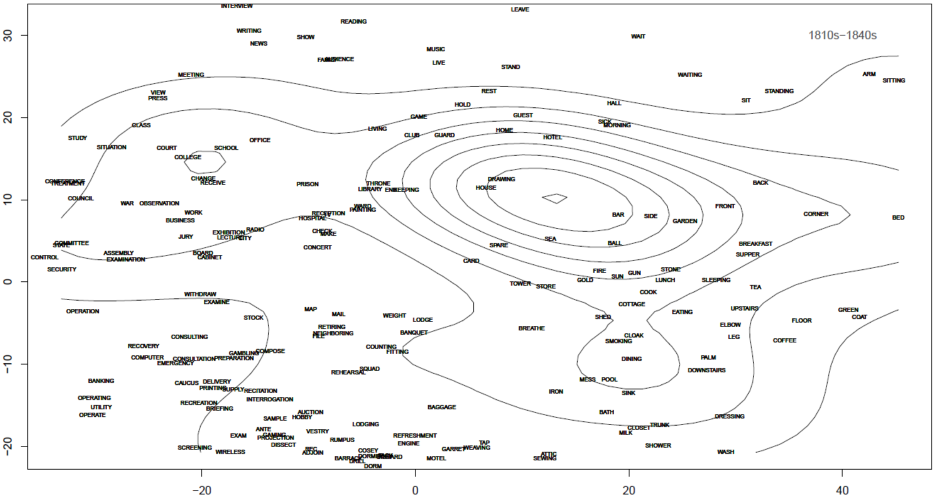
Head–Modifier Density Plots

- For each head, we collected the **100 most frequent modifiers for each time split** of the corpus → **union** of the 5 sets
  → 216 modifiers for LINE, 216 for ROOM, 169 for WALL

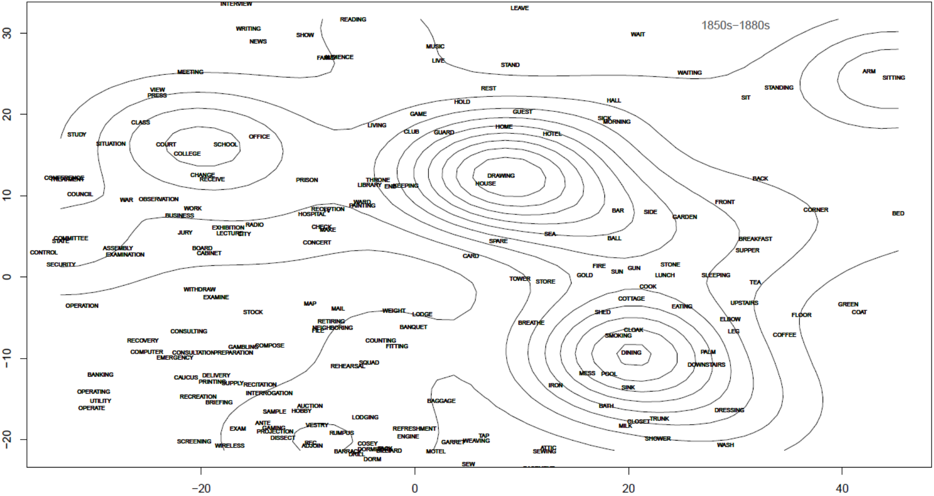
- **Co-occurrence matrix** → **PPMI** → **MDS** (as before)

- Use googleVis library (in R) to integrate frequencies for each time split into **density lines** that are added on top of the plot

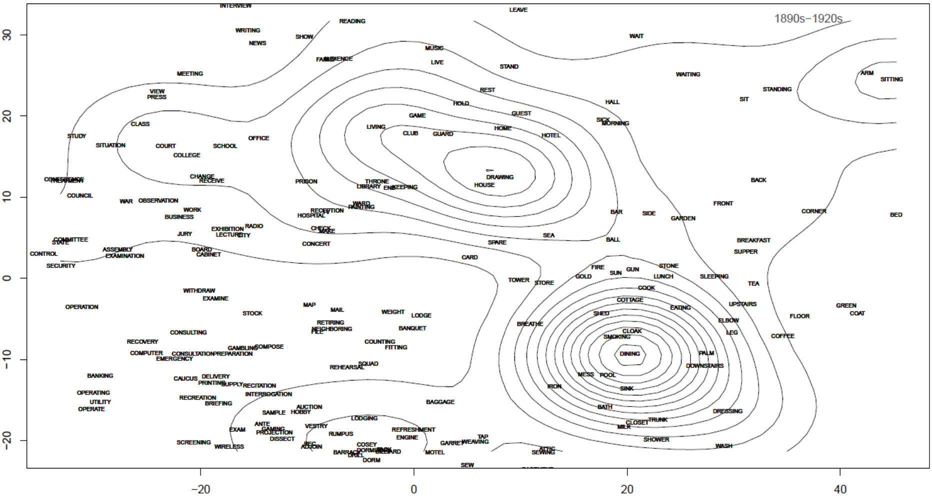
Head Density Plots: ROOM (1810–1840)



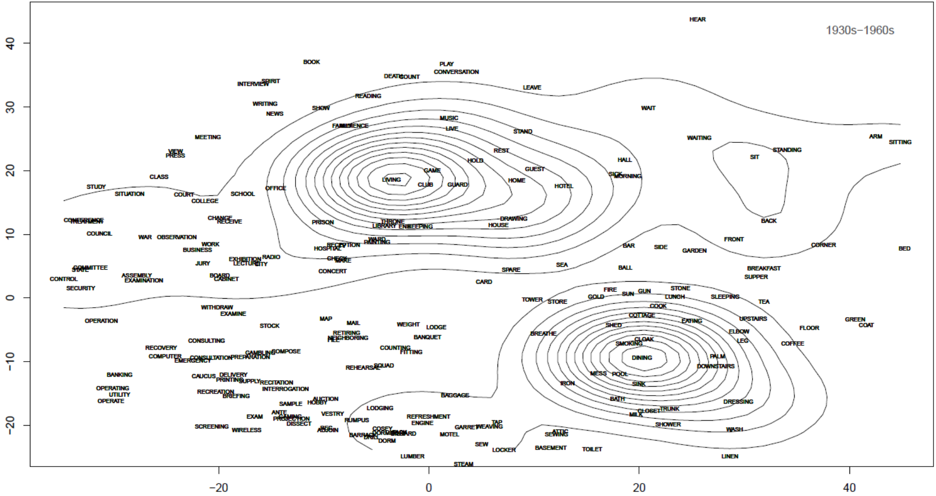
Head Density Plots: ROOM (1850–1880)



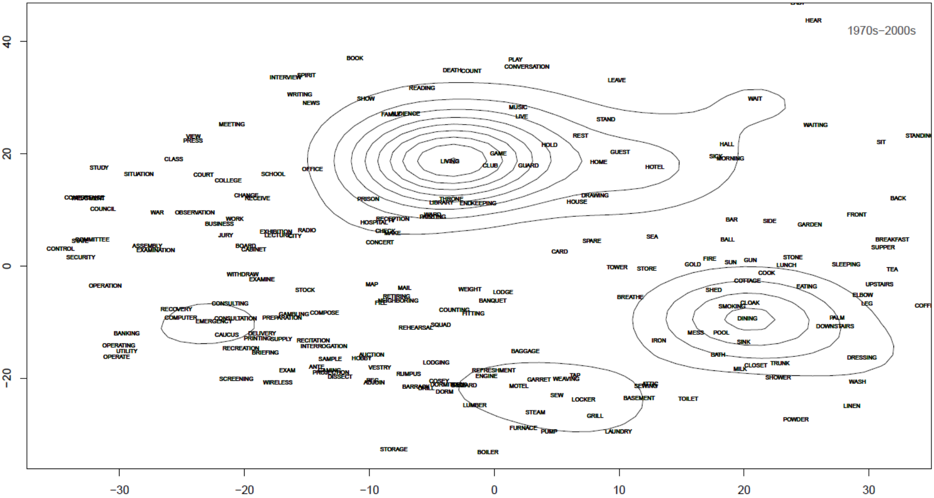
Head Density Plots: ROOM (1890–1920)



Head Density Plots: ROOM (1930–1960)



Head Density Plots: ROOM (1970–2000)



Semantic Density Analysis (SDA)

- Motivation: use SDA scores to **identify semantic shifts** with regard to meaning narrowing and meaning broadening (via polysemy)
  - compounds: rely on context words
  - heads: rely on modifiers
- Vector Spaces: five co-occurrence matrices with **nouns, verbs, adverbs, adjectives as target words and as context words**, excl. 50 most frequent words and if freq$<$100; use lmi
- Calculate the **average pair-wise cosine score between the 10 strongest context words based on LMI values** for every head and compound in every time period
- Changes in SDA score:
  - decrease in SDA score $\rightarrow$ contexts less similar to each other $\rightarrow$ semantic broadening
  - increase in SDA score $\rightarrow$ contexts more similar to each other $\rightarrow$ semantic narrowing

Semantic Density Analysis: Examples

Compound: living room

period	SDA score	frequency
1	–	0
2	0.118	99
3	0.171	1,146
4	0.269	3,459
5	0.306	5,381

Head: room

period	SDA score	frequency	#modifiers
1	0.165	11,722	162
2	0.190	37,721	328
3	0.180	53,466	587
4	0.233	62,404	958
5	0.251	65,016	1,080

Head: line

period	SDA score	frequency	#modifiers
1	0.365	8,086	136
2	0.154	19,585	362
3	0.109	34,589	768
4	0.109	36,668	1,112
5	0.075	37,280	1,305

Summary: Diachronic Distributional Models

- Head progression plots
  - **fixed semantic space** $\rightarrow$ **misleading**: sparse vectors
  - **moving semantic space** $\rightarrow$ **not very informative**: space too general
- Head-modifier density plots
  - capture semantic shifts of the heads' modifiers and meanings
  - semantic centroids **in agreement with the heads' SDA scores**
- Semantic Density Analysis
  - salient measure for **quantifying heads' polysemy**
  - salient measure for capturing semantic change in compounds
- Limitations
  - insufficient language data in the early time periods
  - moving semantic space too general to capture small semantic shifts
  - 10-word SDA sets primed for predominant meanings

References I

[Alatrash et al., 2020] Alatrash, R., Schlechtweg, D., Kuhn, J., and Schulte im Walde, S. (2020).
CCOHA: Clean Corpus of Historical American English.
In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6958–6966, Marseille, France.

[Alipoor and Schulte im Walde, 2020] Alipoor, P. and Schulte im Walde, S. (2020).
Variants of Vector Space Reductions for Predicting the Compositionality of English Noun Compounds.
In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4379–4387, Marseille, France.

[Baroni et al., 2009] Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009).
The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora.
Language Resources and Evaluation, 43(3):209–226.

[Bell and Schäfer, 2013] Bell, M. J. and Schäfer, M. (2013).
Semantic Transparency: Challenges for Distributional Semantics.
In *Proceedings of the IWCS Workshop on Formal Distributional Semantics*, pages 1–10, Potsdam, Germany.

[Cordeiro et al., 2019] Cordeiro, S., Villavicencio, A., Idiart, M., and Ramisch, C. (2019).
Unsupervised Compositionality Prediction of Nominal Compounds.
Computational Linguistics, 45(1):1–57.

[Davies, 2012] Davies, M. (2012).
Expanding Horizons in Historical Linguistics with the 400-Million Word Corpus of Historical American English.
Corpora, 7(2):121–157.

[Firth, 1957] Firth, J. R. (1957).
Papers in Linguistics 1934-51.
Longmans, London, UK.

[Harris, 1954] Harris, Z. (1954).
Distributional structure.
Word, 10(23):146–162.

[Hilpert, 2016] Hilpert, M. (2016).
Change in Modal Meanings: Another Look at the Shifting Collocates of *may*.
Constructions and Frames, 8(1):66–85.

References II

[Köper and Schulte im Walde, 2017] Köper, M. and Schulte im Walde, S. (2017).
Complex Verbs are Different: Exploring the Visual Modality in Multi-Modal Models to Predict Compositionality.
In *Proceedings of the 13th Workshop on Multiword Expressions*, pages 200–206, Valencia, Spain.

[Reddy et al., 2011] Reddy, S., McCarthy, D., and Manandhar, S. (2011).
An Empirical Study on Compositionality in Compound Nouns.
In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 210–218, Chiang Mai, Thailand.

[Sagi et al., 2009] Sagi, E., Kaufmann, S., and Clark, B. (2009).
Semantic Density Analysis: Comparing Word Meaning across Time and Phonetic Space.
In *Proceedings of the EMNLP Workshop on Geometrical Models for Natural Language Semantics*, pages 104–111, Athens, Greece.

[Schäfer and Bildhauer, 2012] Schäfer, R. and Bildhauer, F. (2012).
Building Large Corpora from the Web Using a New Efficient Tool Chain.
In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 486–493, Istanbul, Turkey.

[Schulte im Walde et al., 2016] Schulte im Walde, S., Hätty, A., and Bott, S. (2016).
The Role of Modifier and Head Properties in Predicting the Compositionality of English and German Noun-Noun Compounds: A Vector-Space Perspective.
In *Proceedings of the 5th Joint Conference on Lexical and Computational Semantics*, pages 148–158, Berlin, Germany.

[Schulte im Walde et al., 2013] Schulte im Walde, S., Müller, S., and Roller, S. (2013).
Exploring Vector Space Models to Predict the Compositionality of German Noun-Noun Compounds.
In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*, pages 255–265, Atlanta, GA, USA.

[Tsigkouli, 2021] Tsigkouli, M. (2021).
Studying the Diachronic Changes of Constituent Collocates and Meanings in English Noun Compounds.
Masterarbeit, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

[von der Heide and Borgwaldt, 2009] von der Heide, C. and Borgwaldt, S. (2009).
Assoziationen zu Unter-, Basis- und Oberbegriffen. Eine explorative Studie.
In *Proceedings of the 9th Norddeutsches Linguistisches Kolloquium*, pages 51–74.