Decomposing Generalization

MODELS OF GENERIC, HABITUAL AND EPISODIC STATEMENTS

Venkata S Govindarajan, Benjamin Van Durme, Aaron Steven White

Transactions of the ACL
Volume 7, 2019 p.501-517

① **The service** at that restaurant was good

How to capture linguistic generalization like in the above in a framework for research and annotation?

The ability to capture different modes of generalization is key to building systems with robust **commonsense reasoning**. (Zhang, Rudinger, Duh, et al. 2017, Bauer et al. 2018, McCarthy 1960, 1980, Minsky 1974, Hobbs et al. 1987)

OUR CLAIM

Linguistic generalizations should be captured in a **continuous multi-label system**, using simple real-valued referential properties.

Our framework is based on **Decompositional Semantics**. (White et al. 2016)

BACKGROUND

STANDARD CLASSIFICATION

②	Mary	ate	lunch.
	individual	episodic	
③	Mary	eats	oatmeal for breakfast.
	individual	habitual	
④	The lion	is	in the cage.
	individual	stative	
⑤	The lion	disappeared from Asia.	
	kind	episodic	
⑥	Lions	eat	meat.
	kind	generic	

G. N. Carlson et al. 1995, Carlson 2005

PROBLEMS

Arguments and Predicates do not always fall under such well defined categories as described.

⑦ Taxonomic Reference (G. N. Carlson et al. 1995)

a. **One whale**, namely the blue whale, is nearly extinct.

b. That vintner makes **three different wines.**

⑧ Abstract Reference (Grimm 2014, 2016)

a. Know where **crimes** usually happen, and be safe .

b. **The atmosphere** may not be for everyone.

⑨ Indefinite definites (G. Carlson et al. 2006)

a. Open **the window**, will you please?

b. That bureaucrat takes **the 90 bus** to work.

CURRENT CORPORA

The **ACE-2** program (Doddington et al. 2004, Reiter et al. 2010) associated entity mentions with two classes - specific and generic.

The **ACE-2005** (Walker et al. 2006) corpus adds data and provides two additional classes - neg (empty sets), and usp (underspecified).

The **EventCorefBank**(ECB) (Bejan et al. 2010, Lee et al. 2012) annotates event and entity mentions with a generic class.

SitEnt – the Situational Entities Corpus (Friedrich et al. 2016, 2015, 2014) annotates NPs and clauses separately for their genericity, habituality, and lexical aspectual class of main verb.

They fail to deal with taxonomic reference, abstract reference and indefinite definites.

All of these frameworks employ multi-class annotation schemes.

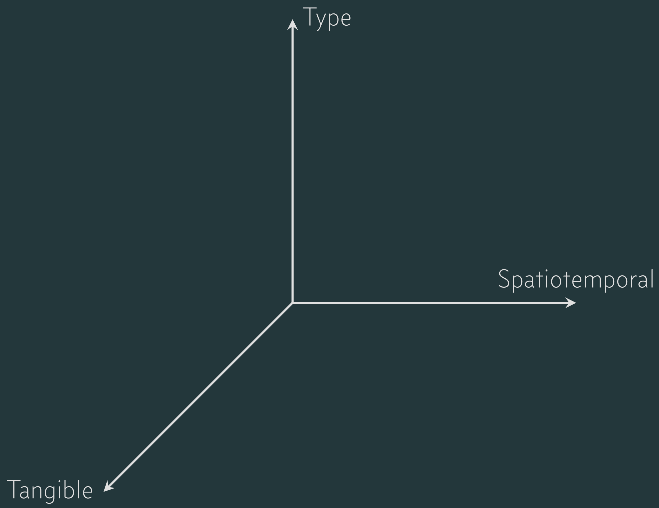
ANNOTATION FRAMEWORK AND DATA COLLECTION

Decompose arguments and predicates into simple referential properties.

Collect annotations for argument and predicate properties separately, with confidence ratings for each annotation.

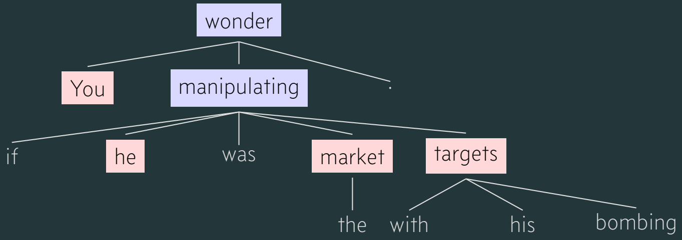
Multiple properties can be true of a predicate/argument – multi-label annotation schema.

AXES OF REFERENCE



You wonder if he was manipulating the market with his bombing targets .

Universal Dependencies (Bies et al. 2012)



PredPatt(Zhang, Rudinger & Durme 2017) extracts Arguments & Predicates

Filtering

wonder, manipulating, you, market, targets

Annotation on Mechanical Turk

(True,4), (False, 3), (True,2), ...

ARGUMENT ANNOTATION

You wonder if he was manipulating the market with his bombing targets .

The noun You does not refer to a particular thing in this sentence and I am totally confident about my choice.

The noun You does refer to a type of thing in this sentence and I am totally confident about my choice.

The noun You does refer to an abstract concept in this sentence and I am about my choice.

✓ totally confident
very confident
somewhat confident
not very confident
not at all confident

PREDICATE ANNOTATION

You wonder if he was manipulating the market with his bombing targets .

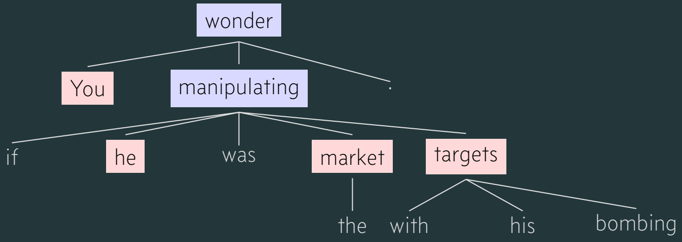
The situation referred to by wonder ------- hypothetical and I am totally confident about my choice.

The situation referred to by wonder ------- a particular situation or a group of particular situations and I am totally confident about my choice.

The situation referred to by wonde ✓ ------- is is not dynamic and I am totally confident about my choice.

You wonder if he was manipulating the market with his bombing targets .

Universal Dependencies (Bies et al. 2012)



PredPatt(Zhang, Rudinger & Durme 2017) extracts Arguments & Predicates

Filtering

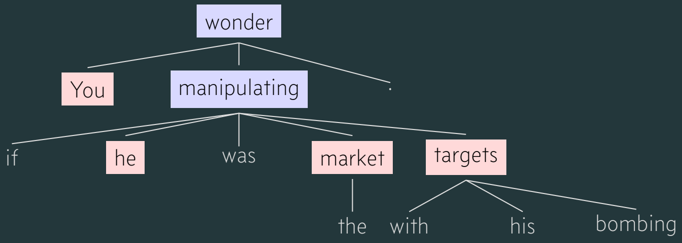
wonder, manipulating, you, market, targets

Annotation on Mechanical Turk

(True,4), (False, 3), (True,2), ...

You wonder if he was manipulating the market with his bombing targets .

Universal Dependencies (Bies et al. 2012)



PredPatt(Zhang, Rudinger & Durme 2017) extracts Arguments & Predicates

Filtering

wonder, manipulating, you, market, targets

Annotation on Mechanical Turk

(True,4), (False, 3), (True,2), ...

Normalization

The need to adjust annotation bias has long been recognized in psycholinguistics literature(Baayen 2008). We employ such procedures to arrive at a **single real-valued score**.

Confidence Normalization

To adjust for annotator bias while using confidence scales, we use **ridit scoring** (Agresti 2003). It reweights confidences based on frequency.

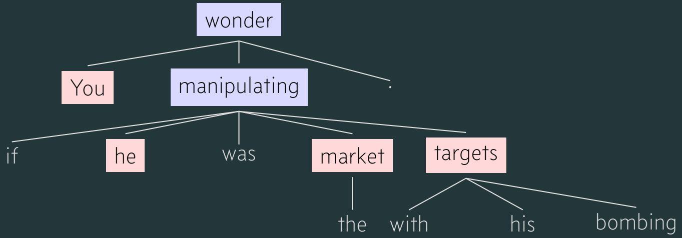
Binary Normalization

To adjust for annotator bias while assigning labels to properties, we use a **mixed effects logistic model** (Gelman et al. 2014)

We thus estimate a real-valued score for each property and each token based on the **average annotator**.

You wonder if he was manipulating the market with his bombing targets .

Universal Dependencies (Bies et al. 2012)



PredPatt(Zhang, Rudinger & Durme 2017) extracts Arguments & Predicates

Filtering

wonder, manipulating, you, market, targets

Annotation on Mechanical Turk

(True,4), (False, 3), (True,2), ...

Normalization

3.2, -2.3, 1.1, ...

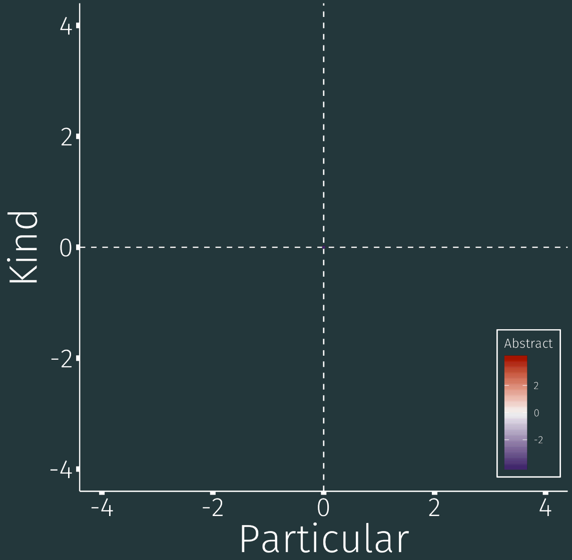
Universal Decompositional Semantics-Genericity (UDS-G) dataset:

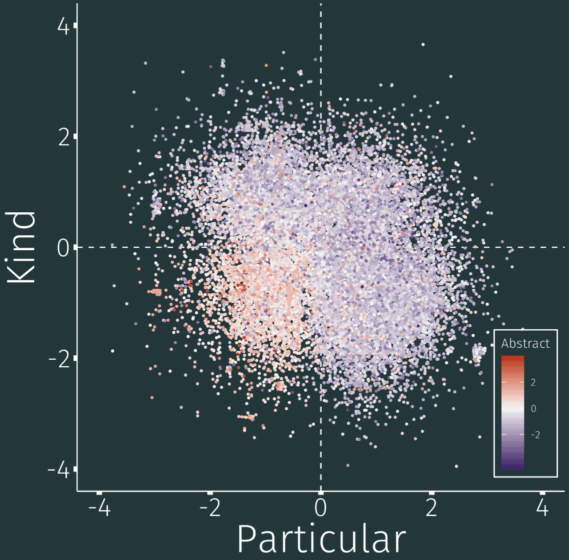
37,146 Arguments, **33,114 Predicates**

Data (and code) available at decomp.io

PRELIMINARY ANALYSIS

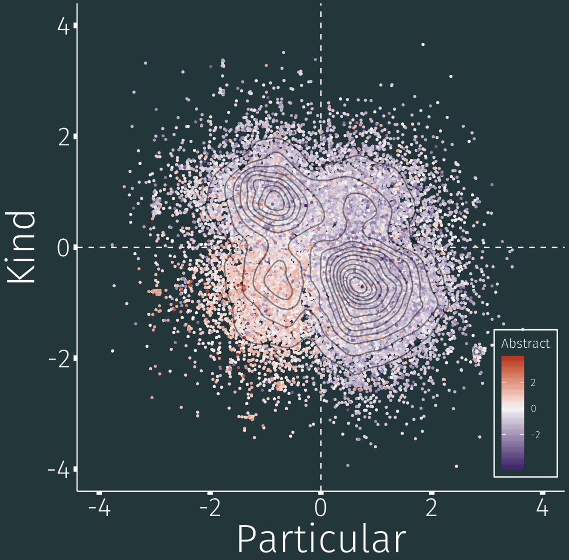
ARGUMENT NORMALIZED DISTRIBUTION



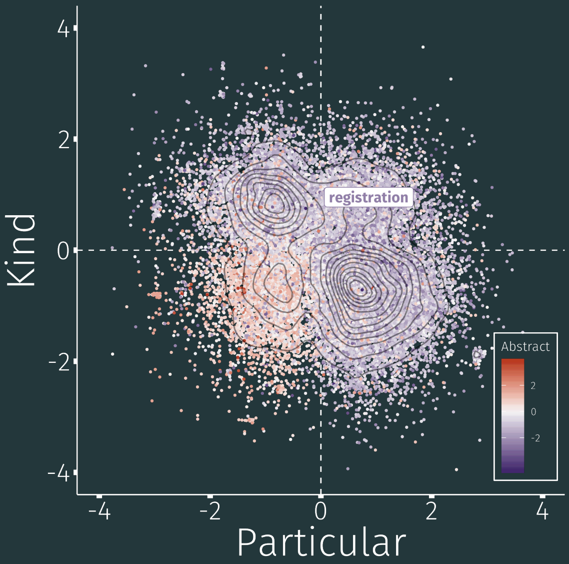


4
2
0
-2
-4
Kind
-4
-2
0
2
4
Particular
Abstract
2
0
-2

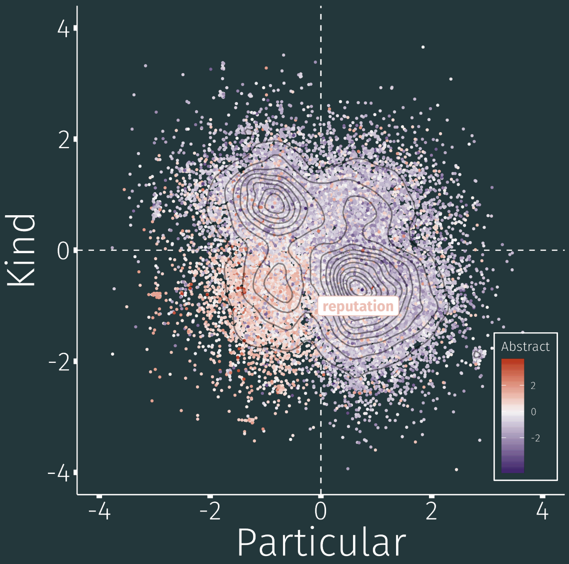
ARGUMENT NORMALIZED DISTRIBUTION



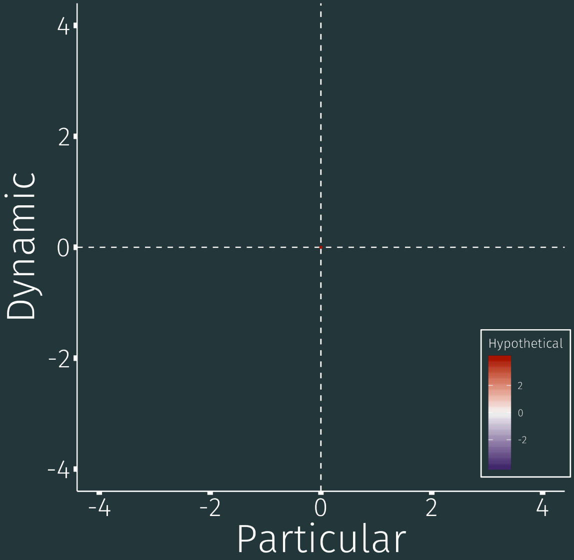
⑩ Some places do the **registration** right at the hospital...



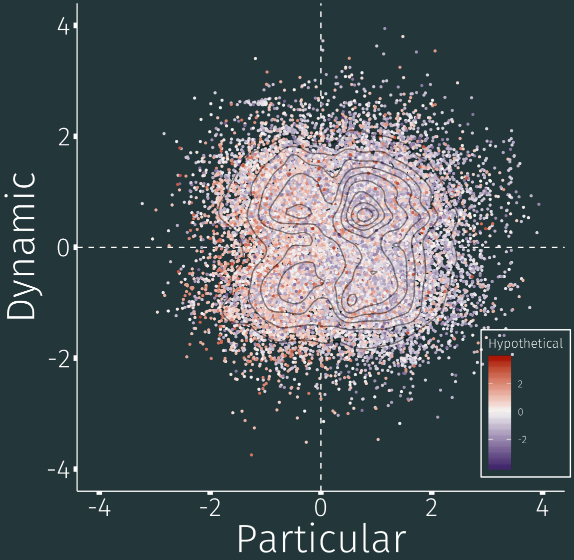
11 Meanwhile, his **reputation** seems to be improving...



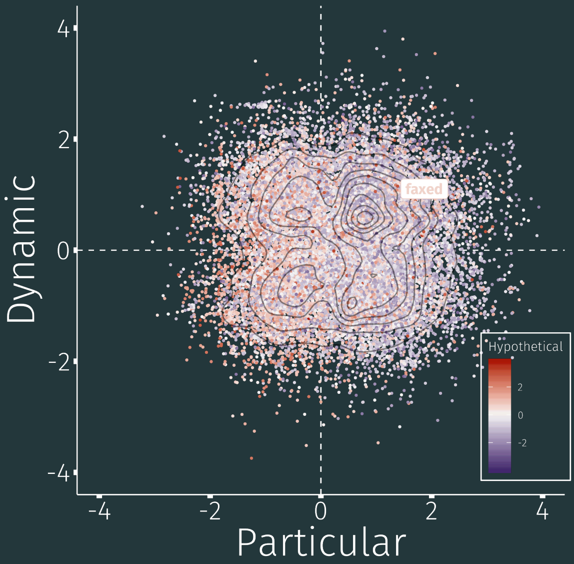
PREDICATE NORMALIZED DISTRIBUTION



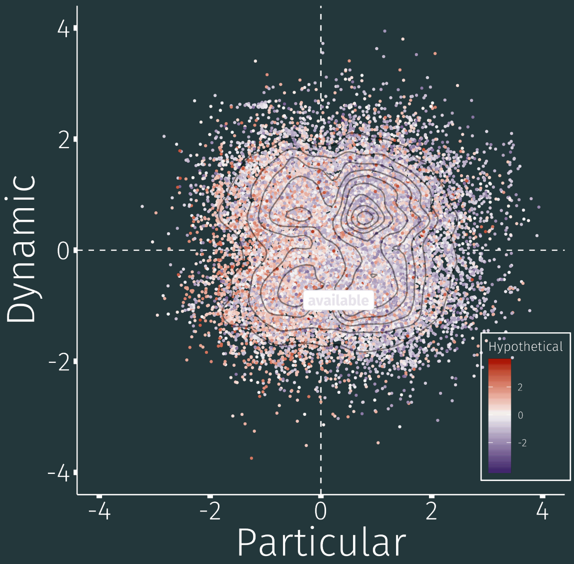
PREDICATE NORMALIZED DISTRIBUTION



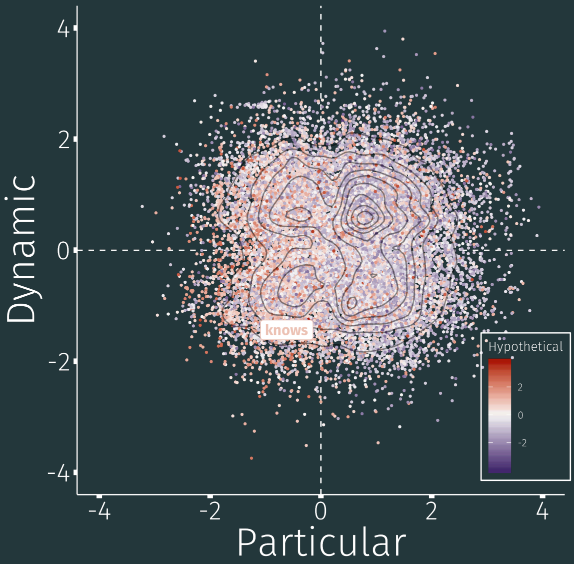
⑫ I have **faxed** to you the form of Bond...



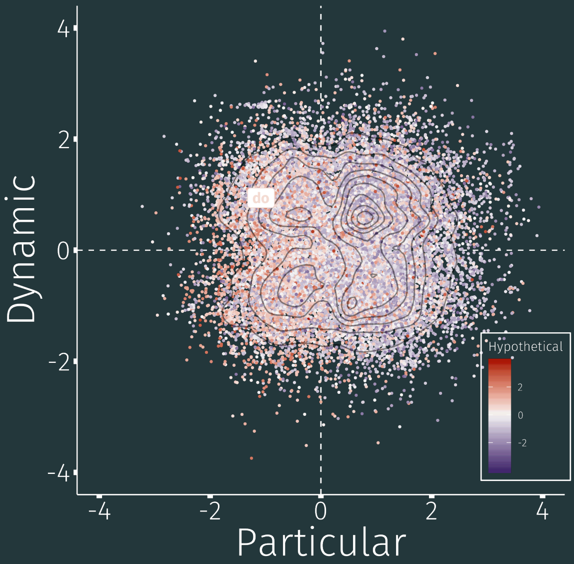
⑬ Is gare montparnasse storage still **available** ?



14 Who **knows** what the future might hold, and it might still be expensive?



15 I have tryed to give him water but he wont take it..what should i **do**?



MODELING

FEATURE REPRESENTATIONS

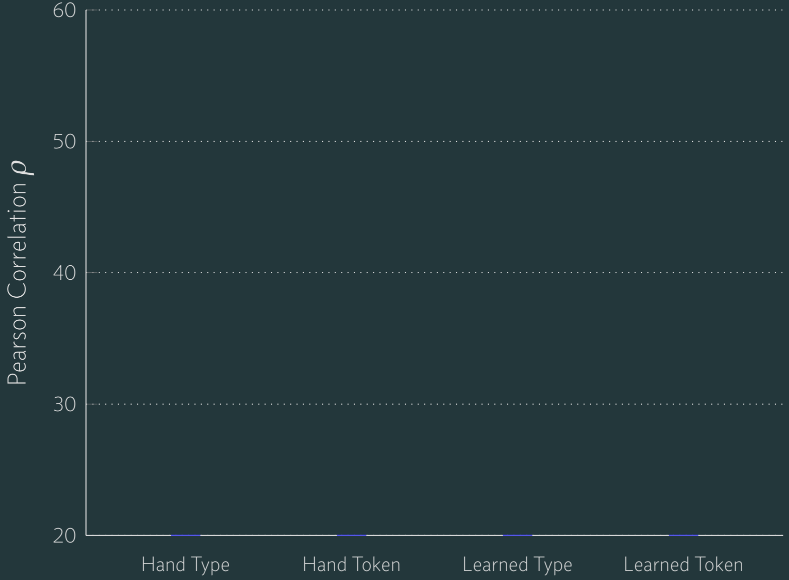
To predict the real-valued properties using a computational model, arguments and predicates need rich feature representations.

- Hand engineered:
  - **Type level** VerbNet classes, FrameNet frames, WordNet supersenses, Concreteness ratings (Brysbaert et al. 2014)
  - **Token level** Part-of-Speech tags, Inflectional features, Syntactic Relations
- Learned (word embeddings):
  - **Type level** GloVe static embeddings (Pennington et al. 2014)
  - **Token level** ELMO contextual embeddings (Peters et al. 2018)

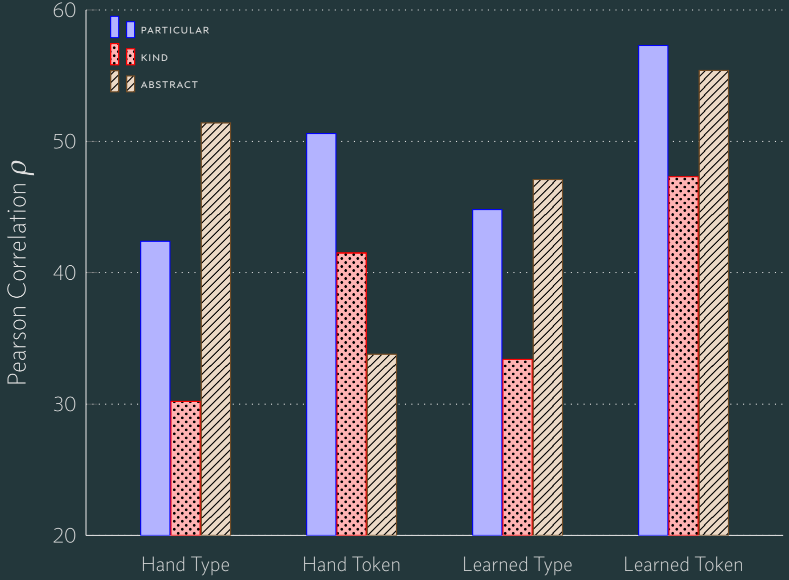
LABELLING MODEL

Multi-Layer Neural Network that takes as input one (or more) of the feature representations of the argument/predicate token that was annotated, and outputs 3 real values corresponding to the 3 properties.

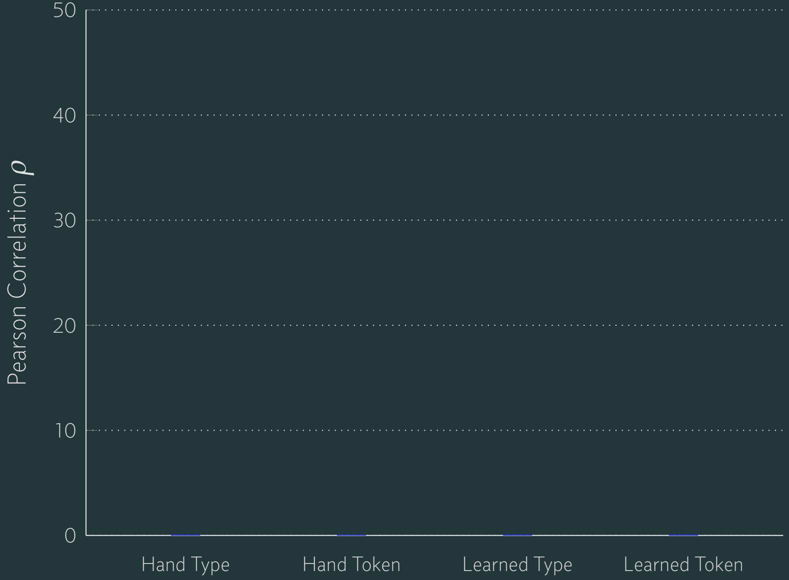
RESULTS - ARGUMENT



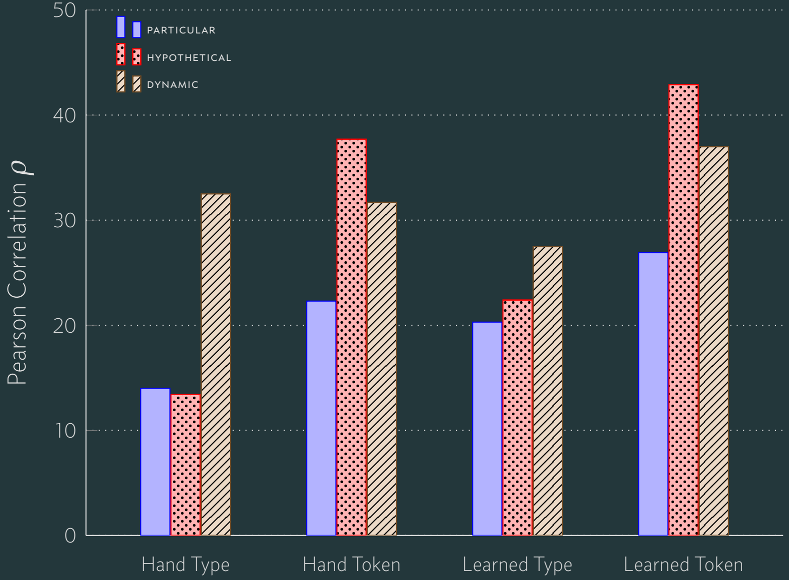
RESULTS - ARGUMENT



RESULTS - PREDICATE



RESULTS - PREDICATE



CONCLUSION

Framework We have proposed a novel semantic framework for modeling linguistic expressions of generalization as combinations of real-valued referential properties of predicates and arguments.

Dataset We used this framework to construct a large-scale dataset covering the entirety of the Universal Dependencies English Web Treebank.

Modeling We have built baseline models to probe the efficacy of hand-engineered and learned type and token level features.

ACKNOWLEDGEMENTS

I'm grateful to Aaron White and Benjamin Van Durme for guiding me through the entirety of my work on this project. I would also like to thank Scott Grimm for his useful comments.

Thanks also to Siddharth Vashistha, YoungEun An, Elizabeth Lee, and Gene Kim for their comments and help.

Special thanks to Ellise Moon for helping me perfect this presentation.

REFERENCES

REFERENCES I

Agresti, Alan. 2003. Categorical Data Analysis. Vol. 482. John Wiley & Sons.

Baayen, RH. 2008. Analyzing Linguistic Data: A Practical Introduction to Statistics using R. Cambridge: Cambridge University Press.

Bauer, Lisa, Yicheng Wang & Mohit Bansal. 2018. Commonsense for Generative Multi-Hop Question Answering Tasks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4220–4230. Brussels, Belgium: Association for Computational Linguistics.

Bejan, Cosmin Adrian & Sanda Harabagiu. 2010. Unsupervised Event Coreference Resolution with Rich Linguistic Features. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1412–1422. Uppsala Sweden.

Bies, Ann, Justin Mott, Colin Warner & Seth Kulick. 2012. English Web Treebank LDC2012T13. Linguistic Data Consortium, Philadelphia, PA.

Brysbaert, Marc, Amy Beth Warriner & Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46(3). 904–911.

Carlson, Greg, Rachel Sussman, Natalie Klein & Michael Tanenhaus. 2006. Weak Definite Noun Phrases. In Christopher Davis, Amy Rose Deal & Youri Zabbal (eds.), *Proceedings of NELS 36*, 179–196. Amherst, MA: GLSA.

Carlson, Gregory N. 2005. Generics, Habituals and Iteratives. In Alex Barber (ed.), *Encyclopedia of Language and Linguistics*. Elsevier.

Carlson, Gregory N. & Francis Jeffrey Pelletier. 1995. The Generic Book. The University of Chicago Press, Chicago. 488 pp.

Doddington, George R., Alexis Mitchell, Mark A. Przybocki, Lance A. Ramshaw, Stephanie Strassel & Ralph M. Weischedel. 2004. The Automatic Content Extraction (ACE) Program - Tasks, Data, and Evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Lisbon, Portugal: European Language Resources Association (ELRA).

Friedrich, Annemarie & Alexis Palmer. 2014. Automatic prediction of aspectual class of verbs in context. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 517–523. Baltimore, Maryland: Association for Computational Linguistics.

Friedrich, Annemarie, Alexis Palmer, Melissa Peate Sørensen & Manfred Pinkal. 2015. Annotating genericity: a survey, a scheme, and a corpus. In *Proceedings of The 9th Linguistic Annotation Workshop*, 21–30. Denver, Colorado: Association for Computational Linguistics.

Friedrich, Annemarie, Alexis Palmer & Manfred Pinkal. 2016. Situation entity types: automatic classification of clause-level aspect. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1757–1768. Berlin, Germany: Association for Computational Linguistics.

Gelman, Andrew & Jennifer Hill. 2014. Data Analysis using Regression and Multilevel-Hierarchical Models. New York City: Cambridge University Press.

Grimm, Scott. 2014. Individuating the Abstract. In Urtzi Etxeberria, Anamaria Fălăuş, Aritz Irurtzun & Bryan Leferman (eds.), *Proceedings of Sinn und Bedeutung 18*, 182–200. Bayonne & Vitoria-Gasteiz.

Grimm, Scott. 2016. Crime Investigations: The Countability Profile of a Delinquent Noun. *Baltic International Yearbook of Cognition, Logic and Communication* 11.

REFERENCES III

Hobbs, Jerry R., William Croft, Todd Davies, Douglas Edwards & Kenneth Laws. 1987. Commonsense Metaphysics and Lexical Semantics. *Computational Linguistics* 13(3-4). 241–250.

Lee, Heeyoung, Marta Recasens, Angel Chang, Mihai Surdeanu & Dan Jurafsky. 2012. Joint Entity and Event Coreference Resolution across Documents. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 489–500. Jeju Island, Korea: Association for Computational Linguistics.

McCarthy, John. 1960. Programs with Common Sense. RLE & MIT computation center.

McCarthy, John. 1980. Circumscription—A Form of Nonmonotonic Reasoning. *Artificial Intelligence* 13(1-2). 27–39.

Minsky, Marvin. 1974. A Framework for Representing Knowledge. *MIT-AI Laboratory Memo 306*.

Pennington, Jeffrey, Richard Socher & Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Doha, Qatar: Association for Computational Linguistics.

Peters, Matthew, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee & Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2227–2237. New Orleans, Louisiana: Association for Computational Linguistics.

Reiter, Nils & Anette Frank. 2010. Identifying Generic Noun Phrases. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (ACL '10), 40–49. Uppsala, Sweden: Association for Computational Linguistics.

REFERENCES IV

Walker, Christopher, Stephanie Strassel, Julie Medero & Kazuaki Maeda. 2006. ACE 2005 Multilingual Training Corpus LDC2006T06. Linguistic Data Consortium, Philadelphia, PA.

White, Aaron Steven, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins & Benjamin Van Durme. 2016. Universal Decompositional Semantics on Universal Dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1713–1723. Austin, TX: Association for Computational Linguistics.

Zhang, Sheng, Rachel Rudinger, Kevin Duh & Benjamin Van Durme. 2017. Ordinal Common-sense Inference. *Transactions of the Association for Computational Linguistics* 5. 379–395.

Zhang, Sheng, Rachel Rudinger & Benjamin Van Durme. 2017. An Evaluation of PredPatt and Open IE via Stage 1 Semantic Role Labeling. In *IWCS 2017 — 12th International Conference on Computational Semantics — Short papers*. Montpellier, France.

APPENDIX

ANALYZING ARGUMENTS

Proper Nouns

1.
   a. **The US Marines** took most of Wednesday, but still face…
   b. I'm writing an essay…and I need to know if **the iPhone** was the first Smart Phone.

Pronouns

2.
   a. I like Hayes Street Grill….another plus, it's right by Civic Center, so **you** can take a romantic walk.
   b. What would happen if **you** flew the flag of South Vietnam in Modern day Vietnam?

ANALYZING PREDICATES

Hypothetical and Particular

3. a. **Read** the entire article; there 's a punchline…
   b. it **s illegal** to sell stolen property, even if you don't know its stolen.

Dynamic and Particular

4. a. library **is closed**
   b. I have a new born daughter and she **helped** me with a lot.

RESULTS - ALL ABLATIONS

	Feature sets				Is.Particular		Is.Kind		Is.Abstract		All
	Type	Token	GloVe	ELMO	ρ	R1	ρ	R1	ρ	R1	wR1
ARGUMENT	+	–	–	–	42.4	7.4	30.2	4.9	51.4	11.7	8.1
	–	+	–	–	50.6	13.0	41.5	8.8	33.8	4.8	8.7
	–	–	+	–	44.8	10.5	33.4	3.9	47.1	9.9	8.2
	–	–	–	+	57.3	16.5	47.3	12.8	55.4	15.3	14.9
	+	+	–	–	55.3	14.1	46.2	11.6	52.6	13.0	12.9
	–	+	–	+	57.6	**17.2**	48.3	13.0	55.6	15.5	15.3
	+	+	–	+	57.8	16.7	47.8	13.1	**56.2**	**15.7**	15.2
	+	+	+	+	**58.0**	17.0	**48.4**	**13.5**	55.4	15.5	**15.4**
					Is.Particular		**Is.Hypothetical**		**Is.Dynamic**		
PREDICATE	+	–	–	–	14.0	0.8	13.4	0.0	32.5	5.6	2.0
	–	+	–	–	22.3	2.8	37.7	7.3	31.7	5.1	5.1
	–	–	+	–	20.3	2.4	22.4	1.5	27.5	3.6	2.5
	–	–	–	+	26.9	3.9	42.9	9.9	37.0	7.2	7.0
	–	–	+	+	26.2	3.8	42.6	10.0	37.3	7.3	7.0
	+	+	–	–	24.0	3.3	37.9	7.6	37.1	7.6	6.1
	–	+	–	+	26.9	4.0	**45.5**	**11.8**	**38.0**	**7.4**	**7.7**
	+	–	–	+	**28.2**	**4.3**	44.4	10.5	36.6	7.0	7.3
	+	+	+	+	26.1	3.5	43.8	10.4	37.3	7.3	7.0

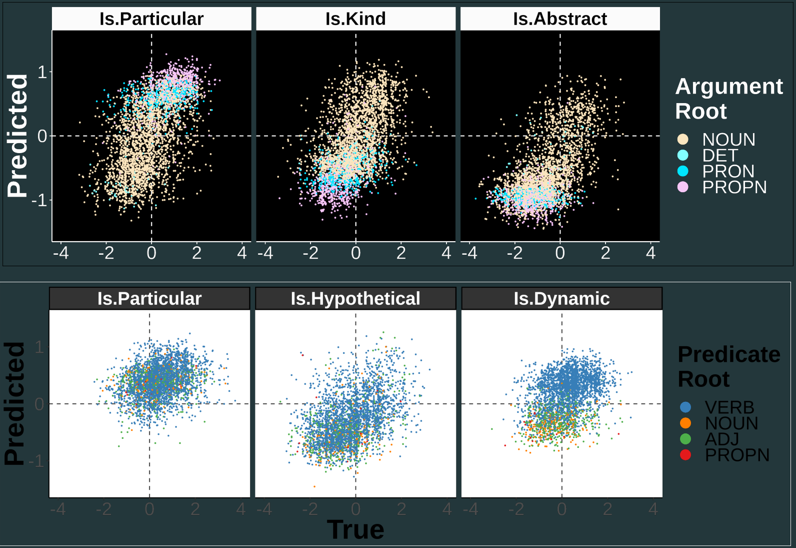
CORPUSES

Corpus	Level	Scheme	Size
ACE-2 ACE-2005	NP	multi-class	40,106
ECB+	Arg. Pred.	multi-class multi-class	12,540 14,884
CFD	NP	multi-class	3,422
Matthew et al	clause	multi-class	1,052
ARRAU	NP	multi-class	91,933
SitEnt	Topic Clause	multi-class multi-class	40,940
RED	Arg. Pred.	multi-class multi-class	10,319 8,731
UDS-G	**Arg.** **Pred.**	**multi-label** **multi-label**	**37,146** **33,114**

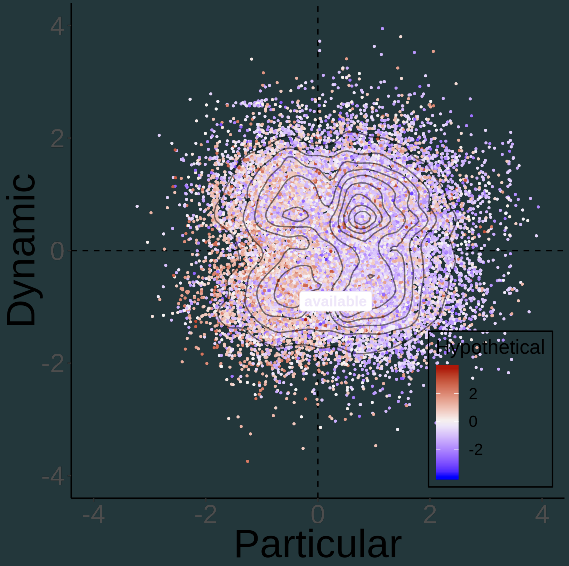
PRELIMINARY ANALYSIS - SPR

Property	Is Part	Is Kind	Is Abs
awareness	0.16	-0.1	-0.15
volition	0.16	-0.11	-0.15
sentient	0.16	-0.08	-0.16
instigation	0.10	-0.08	-0.09
existed before	0.16	-0.04	-0.17
existed during	0.10	-0.02	-0.07
existed after	0.15	-0.06	-0.14
was for benefit	0.11	-0.08	-0.11
change of location	0.07	0.06	-0.17
change of state	-0.02	0.03	-0.03
was used	0.08	-0.03	-0.09
change of possession	-0.04	0.11	-0.04
partitive	-0.02	0.04	-0.06

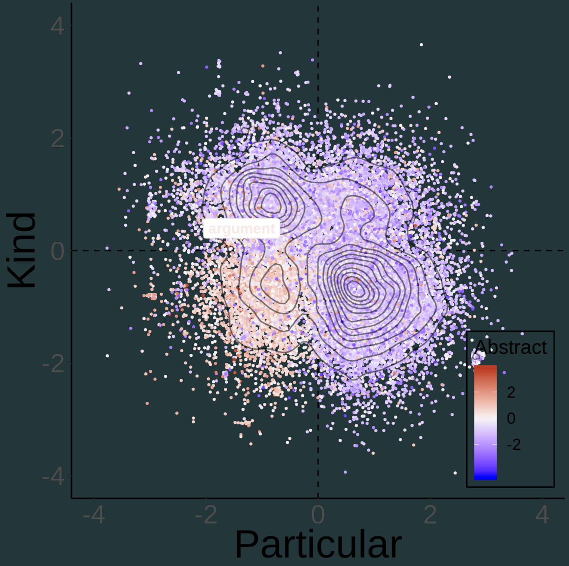
ANALYSIS - TRUE VS PREDICTED DISTRIBUTION



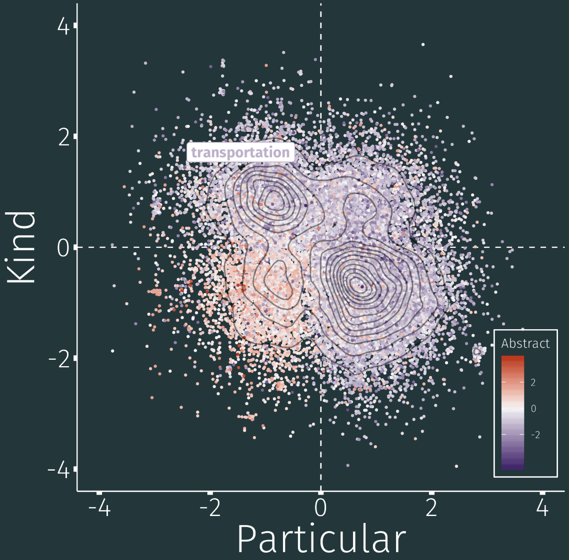
5 is gare montparnasse storage still **available**?



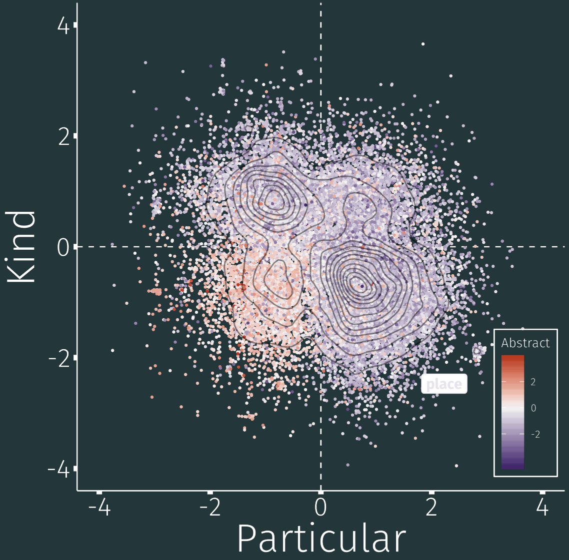
6 The Pew researchers tried to transcend the economic **argument**.



7 What made it perfect was that they offered **transportation** so that I would not have to wait...



8 I think this **place** is probably really great especially...



9 **Power** be where power lies.