Ease of learning explains semantic universals

Shane Steinert-Threlkeld Jakub Szymanik

Overview

Universals in Linguistic Theory

Universals in Linguistic Theory

Question

What is the range of variation in human languages? That is: which out of all of the logically possible languages that humans could speak, do they in fact speak?

Explaining Universals

Natural Question

Why do universals hold?

Explaining Universals

Natural Question

Why do universals hold?

- Answer 1: *learnability*.
  (Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)

Explaining Universals

Natural Question

Why do universals hold?

- Answer 1: *learnability*.
  (Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)
- The universals greatly restrict the search space that a language learner must explore when learning the meanings of expressions. This makes it easier (possible?) for them to learn such meanings from relatively small input.
  [Compare: Poverty of the Stimulus argument for UG. (Chomsky 1980; Pullum and Scholz 2002)]

Explaining Universals

Natural Question

Why do universals hold?

- Answer 1: *learnability*.
  (Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)
- The universals greatly restrict the search space that a language learner must explore when learning the meanings of expressions. This makes it easier (possible?) for them to learn such meanings from relatively small input.
  [Compare: Poverty of the Stimulus argument for UG. (Chomsky 1980; Pullum and Scholz 2002)]
- In a sense must be true, but:
  - May not help much (Piantadosi 2013)
  - Does not explain *which* universals are attested.

Explaining Universals

Natural Question

Why do universals hold?

- Answer 2: *learnability*. (hints in Peters and Westerståhl 2006)

Explaining Universals

Natural Question

Why do universals hold?

- Answer 2: *learnability*. (hints in Peters and Westerståhl 2006)
- Universals aid learnability because expressions satisfying the universals are *easier* to learn than those that do not.

Explaining Universals

Natural Question

Why do universals hold?

- Answer 2: *learnability*. (hints in Peters and Westerståhl 2006)
- Universals aid learnability because expressions satisfying the universals are *easier* to learn than those that do not.
- Our goal: make good on this claim by providing a single model of learning and using it to explain **semantic** universals from several different domains.

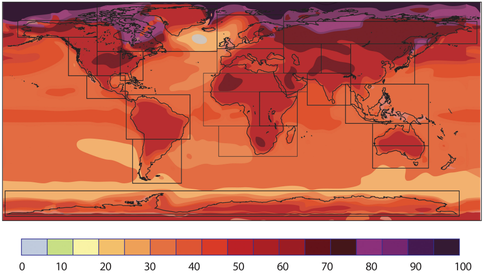
Explaining Universals

Natural Question

Why do universals hold?

- Answer 2: *learnability*. (hints in Peters and Westerståhl 2006)
- Universals aid learnability because expressions satisfying the universals are *easier* to learn than those that do not.
- Our goal: make good on this claim by providing a single model of learning and using it to explain **semantic** universals from several different domains.
- In particular, we train *artificial neural networks* to learn the meanings of different kinds of expressions. Within each kind, we will compare expressions satisfying proposed universals to those that do not.

Heat-map



Overview

1. Main Question
2. (Machine) Learning
3. Color Terms
4. Quantifiers
5. Responsive Verbs
6. Conclusion

Learnability and Semantic Universals

- Our goal: argue that universals arise because expressions satisfying them are *easier to learn*.

Learnability and Semantic Universals

- Our goal: argue that universals arise because expressions satisfying them are *easier to learn*.
- An innovation: using *artificial neural networks* as a model of learning.

Learnability and Semantic Universals

- Our goal: argue that universals arise because expressions satisfying them are *easier to learn*.
- An innovation: using *artificial neural networks* as a model of learning.
- Allows us to test many domains quickly, in a roughly biologically plausible fashion.

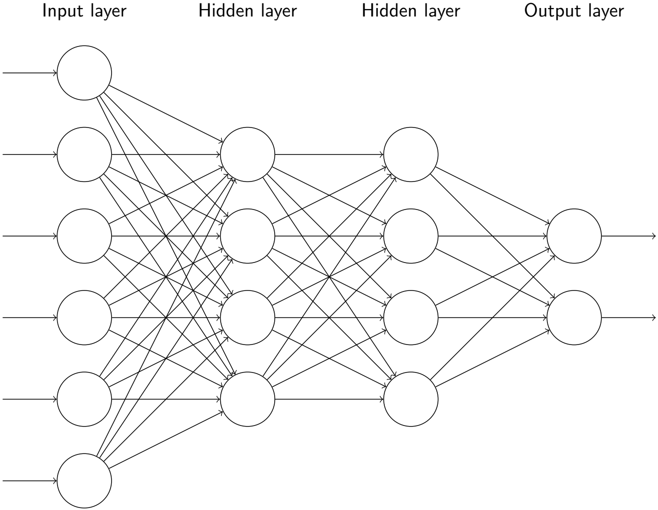
Existing Study with Children


(a)


(b)

From: Hunter and Lidz 2013

Artificial Neural Network



Nielsen 2015; Goodfellow, Bengio, and Courville 2016

http://www.3blue1brown.com/neural-networks

How to Train a Neural Network

- The paradigm method is called *supervised learning*.

How to Train a Neural Network

- The paradigm method is called *supervised learning*.
- You give the network a whole bunch of *labelled examples*, i.e. a bunch of true/correct input-output pairs.

How to Train a Neural Network

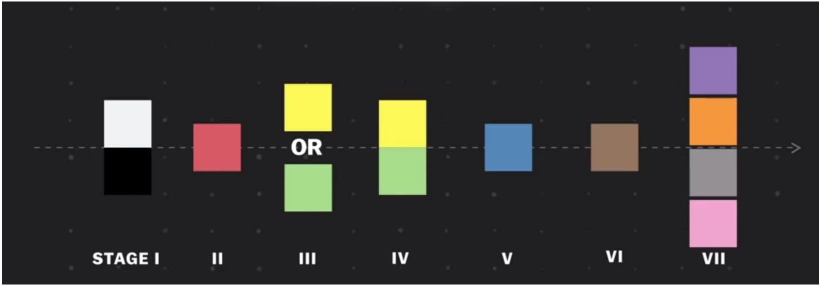
- The paradigm method is called *supervised learning*.
- You give the network a whole bunch of *labelled examples*, i.e. a bunch of true/correct input-output pairs.
- After it processes these examples, it lightly adjusts the weights and biases in the network so as to *make its future guesses* better. It tries to *minimize a loss function* between the true output and the network's output.

How to Train a Neural Network

- The paradigm method is called *supervised learning*.
- You give the network a whole bunch of *labelled examples*, i.e. a bunch of true/correct input-output pairs.
- After it processes these examples, it lightly adjusts the weights and biases in the network so as to *make its future guesses* better. It tries to *minimize a loss function* between the true output and the network's output.
- This is called (stochastic) gradient descent; there are fancier variations now.

Overview

The Order of Color Terms



Berlin and Kay 1969; Regier, Kay, and Khetarpal 2007; Gibson et al. 2017

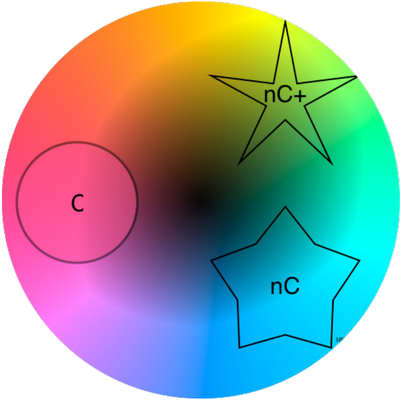
https://www.vox.com/videos/2017/5/16/15646500/color-pattern-language

Convexity

While natural languages vary in how many color terms they have and which specific colors are denoted, it seems that all color terms denote very 'well-behaved' regions of color space.

- X is *convex* just in case if $x, y \in X$, then for every $t \in (0, 1)$,

$$tx + (1 - t)y \in X$$



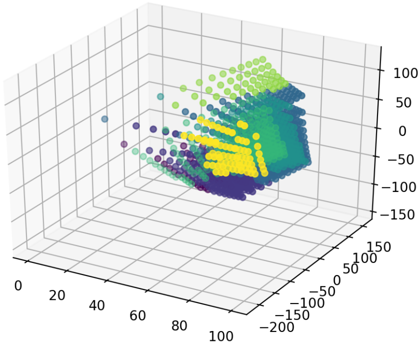
Convexity universal

Convexity Universal (Gärdenfors 2014; Jäger 2010)

All color terms denote convex regions of color space.

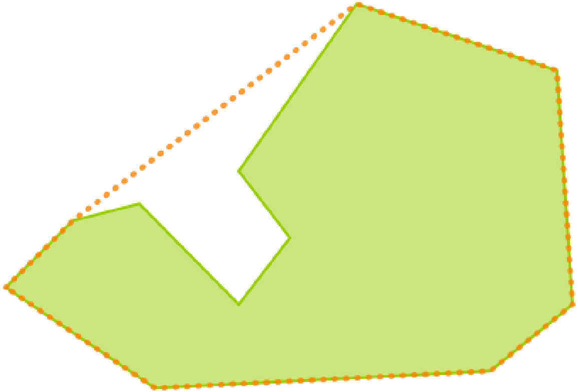
Partitioning CIE-L*a*b* Space

We generated 300 artificial color-naming systems by partitioning the CIELab color space into distinct categories. CIELab approximates human color vision. It is perceptually uniform, meaning that the distance in the space corresponds well with the visually perceived color change.

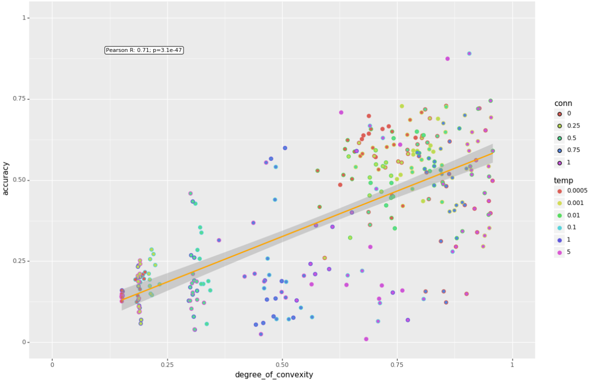


Degree of convexity

We varied the degree of convexity, measured as the average area of the convex hull of each color that is covered by that color.



Convexity: Results



Steinert-Threlkeld and Szymanik 2018a

Overview

Determiners

- Meaning (semantics):
- If languages have syntactic constituents (NPs), then their semantic function is to express generalized quantifiers. (Barwise and Cooper 1981)
- Determiners:
  - Simple: *every*, *some*, *few*, *most*, *five*, . . .
  - Complex: *all but five*, *fewer than three*, *at least eight or fewer than five*, . . .

Determiners

- Meaning (semantics):
- If languages have syntactic constituents (NPs), then their semantic function is to express generalized quantifiers. (Barwise and Cooper 1981)
- Determiners:
    - Simple: *every*, *some*, *few*, *most*, *five*, . . .
    - Complex: *all but five*, *fewer than three*, *at least eight or fewer than five*, . . .
- Denote type $\langle 1, 1\rangle$ generalized quantifiers: sets of models of the form $\langle M, A, B\rangle$ with $A, B \subseteq M$

Determiners

- Meaning (semantics):
- If languages have syntactic constituents (NPs), then their semantic function is to express generalized quantifiers. (Barwise and Cooper 1981)
- Determiners:
  - Simple: *every*, *some*, *few*, *most*, *five*, ...
  - Complex: *all but five*, *fewer than three*, *at least eight or fewer than five*, ...
- Denote type $\langle 1, 1\rangle$ generalized quantifiers: sets of models of the form $\langle M, A, B\rangle$ with $A, B \subseteq M$
- For example:

$$[\![every]\!] = \{\langle M, A, B\rangle : A \subseteq B\}$$
$$[\![three]\!] = \{\langle M, A, B\rangle : |A \cap B| \geq 3\}$$
$$[\![most]\!] = \{\langle M, A, B\rangle : |A \cap B| > |A \setminus B|\}$$

Monotonicity

- Many French people smoke cigarettes
  $\Rightarrow$ Many French people smoke

Monotonicity

- Many French people smoke cigarettes
  $\Rightarrow$ Many French people smoke

So: ‘many’ is *upward monotone*.

Monotonicity

- Many French people smoke cigarettes
  ⇒ Many French people smoke

So: ‘many’ is *upward monotone*.

- Few French people smoke
  ⇒ Few French people smoke cigarettes

Monotonicity

- Many French people smoke cigarettes
  ⇒ Many French people smoke

So: ‘many’ is *upward monotone*.

- Few French people smoke
  ⇒ Few French people smoke cigarettes

So: ‘few’ is *downward monotone*.

Monotonicity

- Many French people smoke cigarettes
  $\Rightarrow$ Many French people smoke

So: 'many' is *upward monotone*.

- Few French people smoke
  $\Rightarrow$ Few French people smoke cigarettes

So: 'few' is *downward monotone*.

- At least 6 or at most 2 French people smoke cigarettes.
  $\not\Rightarrow$ (and $\not\Leftarrow$) At least 6 or at most 2 French people smoke.

Monotonicity

- Many French people smoke cigarettes
  $\Rightarrow$ Many French people smoke

So: 'many' is *upward monotone*.

- Few French people smoke
  $\Rightarrow$ Few French people smoke cigarettes

So: 'few' is *downward monotone*.

- At least 6 or at most 2 French people smoke cigarettes.
  $\not\Rightarrow$ (and $\not\Leftarrow$) At least 6 or at most 2 French people smoke.

So: 'at least 6 or at most 2' is not monotone.

Monotonicity Universal

- Q is upward monotone:
  if $\langle M, A, B \rangle \in \mathsf{Q}$ and $B \subseteq B'$, then $\langle M, A, B' \rangle \in \mathsf{Q}$

Monotonicity Universal

- Q is upward monotone:
  if $\langle M, A, B \rangle \in \mathsf{Q}$ and $B \subseteq B'$, then $\langle M, A, B' \rangle \in \mathsf{Q}$
- Q is downward monotone:
  if $\langle M, A, B \rangle \in \mathsf{Q}$ and $B' \subseteq B$, then $\langle M, A, B' \rangle \in \mathsf{Q}$

Monotonicity Universal

- Q is upward monotone:
  if $\langle M, A, B \rangle \in \mathsf{Q}$ and $B \subseteq B'$, then $\langle M, A, B' \rangle \in \mathsf{Q}$
- Q is downward monotone:
  if $\langle M, A, B \rangle \in \mathsf{Q}$ and $B' \subseteq B$, then $\langle M, A, B' \rangle \in \mathsf{Q}$

Monotonicity Universal (Barwise and Cooper 1981)

All simple determiners are monotone.

Quantity

- At least three houses on Cambridge Ave are blue.
  There are exactly as many blue and non-blue houses on El Camino Real as on Cambridge Ave.
  $\Rightarrow$ At least three houses on El Camino Real are blue.

Quantity

- At least three houses on Cambridge Ave are blue.
  There are exactly as many blue and non-blue houses on El Camino Real as on Cambridge Ave.
  $\Rightarrow$ At least three houses on El Camino Real are blue.

So: ‘at least three’ is *quantitative*.

Quantity

- At least three houses on Cambridge Ave are blue.
  There are exactly as many blue and non-blue houses on El Camino Real as on Cambridge Ave.
  $\Rightarrow$ At least three houses on El Camino Real are blue.

So: 'at least three' is *quantitative*.

- The first three houses on Cambridge Ave are blue.
  There are exactly as many blue and non-blue houses on El Camino Real as on Cambridge Ave.
  $\not\Rightarrow$ The first three houses on El Camino Real are blue.

Quantity

- At least three houses on Cambridge Ave are blue.
  There are exactly as many blue and non-blue houses on El Camino Real as on Cambridge Ave.
  $\Rightarrow$ At least three houses on El Camino Real are blue.

So: 'at least three' is *quantitative*.

- The first three houses on Cambridge Ave are blue.
  There are exactly as many blue and non-blue houses on El Camino Real as on Cambridge Ave.
  $\not\Rightarrow$ The first three houses on El Camino Real are blue.

So: 'the first three' is not quantitative.

Quantity Universal

- Q is *quantitative*:
  if $\langle M, A, B, \ldots \rangle \in$ Q and $A \cap B, A \setminus B, B \setminus A, M \setminus (A \cup B)$ have the same cardinality (size) as their primed-counterparts, then $\langle M', A', B', \ldots \rangle \in$ Q

Quantity Universal

- Q is *quantitative*:
  if $\langle M, A, B, \ldots \rangle \in \mathrm{Q}$ and $A \cap B, A \setminus B, B \setminus A, M \setminus (A \cup B)$ have the same cardinality (size) as their primed-counterparts, then $\langle M', A', B', \ldots \rangle \in \mathrm{Q}$

Quantity Universal (Keenan and Stavi 1986; Peters and Westerståhl 2006)

All simple determiners are quantitative.

Conservativity

- Many French people smoke cigarettes
  $\equiv$ Many French people are French people who smoke cigarettes

Conservativity

- Many French people smoke cigarettes
  $\equiv$ Many French people are French people who smoke cigarettes

So: ‘many’ is *conservative*.

- Only French people smoke cigarettes
  $\not\equiv$ Only French people are French people who smoke cigarettes

Conservativity

- Many French people smoke cigarettes
  $\equiv$ Many French people are French people who smoke cigarettes

So: ‘many’ is *conservative*.

- Only French people smoke cigarettes
  $\not\equiv$ Only French people are French people who smoke cigarettes

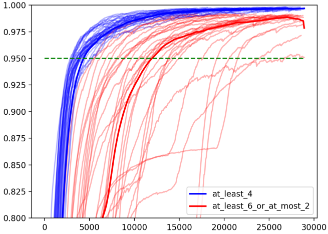
So: ‘only’ is not conservative.

Conservativity Universal

Conservativity Universal (Barwise and Cooper 1981; Keenan and Stavi 1986)

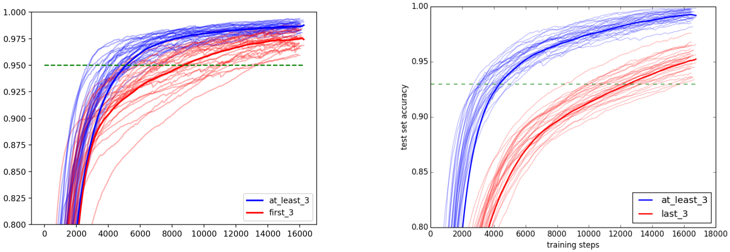
All simple determiners are conservative.

Monotonicity: Results



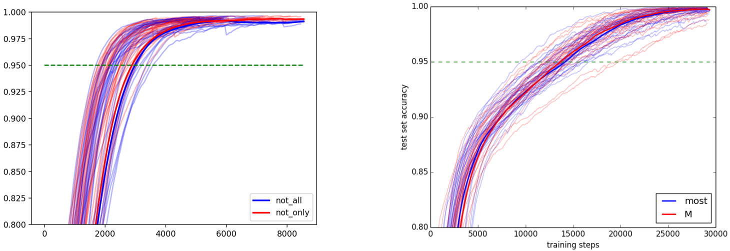
Steinert-Threlkeld and Szymanik 2018b

Quantity: Results



Steinert-Threlkeld and Szymanik 2018b

Conservativity: Results



Steinert-Threlkeld and Szymanik 2018b

Conservativity: Discussion

- No way of 'breaking the symmetry' between $A \setminus B$ and $B \setminus A$

Conservativity: Discussion

- No way of ‘breaking the symmetry’ between $A \setminus B$ and $B \setminus A$
- Cons as a syntactic/structural constraint, not a semantic universal [See Fox 2002; Sportiche 2005; Romoli 2015]

Overview

Types of Verbs

-
  - Jacopo believes that Amsterdam is the capital of the Netherlands.
  - # Jacopo believes where Amsterdam is.

Types of Verbs

- - Jacopo believes that Amsterdam is the capital of the Netherlands.
  - # Jacopo believes where Amsterdam is.
- - # Jacopo wonders that Amsterdam is the capital of the Netherlands.
  - Jacopo wonders where Amsterdam is.

Types of Verbs

- - Jacopo believes that Amsterdam is the capital of the Netherlands.
  - # Jacopo believes where Amsterdam is.
- - # Jacopo wonders that Amsterdam is the capital of the Netherlands.
  - Jacopo wonders where Amsterdam is.
- - Jacopo knows that Amsterdam is the capital of the Netherlands.
  - Jacopo knows where Amsterdam is.

Types of Verbs

type	declarative	interrogative	example
rogative	x	✓	‘wonder’
anti-rogative	✓	x	‘believe’
responsive	✓	✓	‘know’

Lahiri 2002; Theiler, Roelofsen, and Aloni 2018; Uegaki 2018

Veridicality

- Meica knows that Carlos won the race.
  $\rightsquigarrow$ Carlos won the race.

Veridicality

- Meica knows that Carlos won the race.
  $\rightsquigarrow$ Carlos won the race.

So: ‘know‘ is *veridical with respect to declarative complements*.

Veridicality

- Meica knows that Carlos won the race.
  $\rightsquigarrow$ Carlos won the race.

So: ‘know‘ is *veridical with respect to declarative complements*.

- Meica knows who won the race.
  Carlos won the race.
  $\rightsquigarrow$ Meica knows that Carlos won the race.

Veridicality

- Meica knows that Carlos won the race.
  $\rightsquigarrow$ Carlos won the race.

So: ‘know‘ is *veridical with respect to declarative complements*.

- Meica knows who won the race.
  Carlos won the race.
  $\rightsquigarrow$ Meica knows that Carlos won the race.

So: ‘know‘ is *veridical with respect to interrogative complements*.

Veridicality

- Meica knows that Carlos won the race.
  $\rightsquigarrow$ Carlos won the race.

So: ‘know‘ is *veridical with respect to declarative complements*.

- Meica knows who won the race.
  Carlos won the race.
  $\rightsquigarrow$ Meica knows that Carlos won the race.

So: ‘know‘ is *veridical with respect to interrogative complements*.
So: ‘know‘ is *veridically uniform*.

Veridicality

- Meica is certain that Carlos won the race.
  $\not\rightarrow$ Carlos won the race.

Veridicality

- Meica is certain that Carlos won the race.
  $\not\to$ Carlos won the race.

So: 'be certain' is not veridical with respect to declarative complements.

Veridicality

- Meica is certain that Carlos won the race.
  $\not\rightarrow$ Carlos won the race.

So: ‘be certain‘ is not veridical with respect to declarative complements.

- Meica is certain about who won the race.
  Carlos won the race.
  $\not\rightarrow$ Meica is certain that Carlos won the race.

Veridicality

- Meica is certain that Carlos won the race.
  $\not\rightarrow$ Carlos won the race.

So: 'be certain' is not veridical with respect to declarative complements.

- Meica is certain about who won the race.
  Carlos won the race.
  $\not\rightarrow$ Meica is certain that Carlos won the race.

So: 'be certain' is not veridical with respect to interrogative complements.

Veridicality

- Meica is certain that Carlos won the race.
  $\not\rightarrow$ Carlos won the race.

So: 'be certain' is not veridical with respect to declarative complements.

- Meica is certain about who won the race.
  Carlos won the race.
  $\not\rightarrow$ Meica is certain that Carlos won the race.

So: 'be certain' is not veridical with respect to interrogative complements.
So: 'be certain' is *veridically uniform*.

The Veridical Uniformity Thesis

Veridical Uniformity Universal (Spector and Egré 2015; Theiler, Roelofsen, and Aloni 2018)

All responsive verbs are veridically uniform.

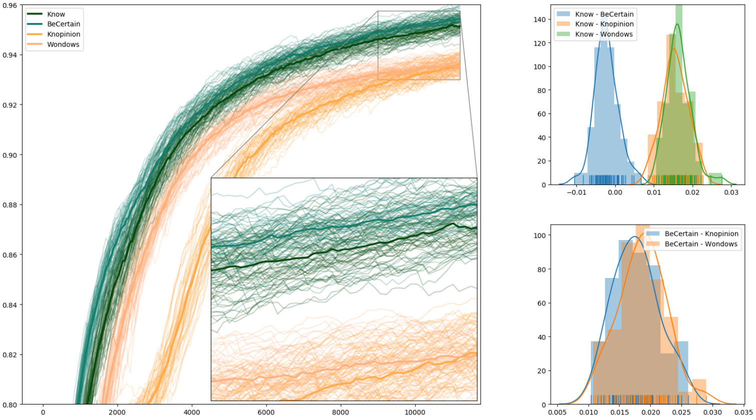
Four Responsive Verbs

Verb	Lexical Entry: $\lambda P_T.\lambda p_{\langle s,t\rangle}.\lambda a_e.\forall w \in p : \ldots$	Veridical Declarative	Veridical Interrogative
know	$w \in \text{DOX}^a_w \in P$	✓	✓
wondows	$w \in \text{DOX}^a_w \subseteq \text{info}(P)$ and $\text{DOX}^a_w \cap q \neq \emptyset\, \forall q \in \text{alt}(P)$	✓	x
knopinion	$w \in \text{DOX}^a_w$ and $(\text{DOX}^a_w \in P$ or $\text{DOX}^a_w \in \neg P)$	x	✓
be-certain	$\text{DOX}^a_w \in P$	x	x

Table : Four verbs, exemplifying the possible profiles of veridicality.

The semantics are given in terms of *inquisitive semantics* Ciardelli, Groenendijk, and Roelofsen 2018

Veridical Uniformity: Results



Steinert-Threlkeld 2018

Overview

Explaining Universals

Why do semantic universals arise?

Because expressions satisfying them are easier to learn.

Explaining Universals

Why do semantic universals arise?

Because expressions satisfying them are easier to learn.

We looked at three very different domains:

- Function words: quantifiers
- Content words: attitude verbs, color terms

Explaining Universals

Why do semantic universals arise?

Because expressions satisfying them are easier to learn.

We looked at three very different domains:

- Function words: quantifiers
- Content words: attitude verbs, color terms

In each, a general purpose and biologically-inspired model of learning made good on this answer. We take this as strong evidence that learnability does indeed explain semantic universals.

Future Directions

- Relation between learnability and (descriptive) complexity, e.g.,
  Does Kolmogorov complexity of Qs predict learnability (with Iris van de Pol)?
  What are the corresponding minimal programs over a LoT (with Steven Piantadosi)?

Future Directions

- Relation between learnability and (descriptive) complexity, e.g.,
  Does Kolmogorov complexity of Qs predict learnability (with Iris van de Pol)?
  What are the corresponding minimal programs over a LoT (with Steven Piantadosi)?
- 'Scaling up' the computational experiments, e.g.,
  Does CONS arise from the biased linguistic distribution?
  Mhasawade et al. 2018: NO

Future Directions

- Relation between learnability and (descriptive) complexity, e.g.,
  Does Kolmogorov complexity of Qs predict learnability (with Iris van de Pol)?
  What are the corresponding minimal programs over a LoT (with Steven Piantadosi)?
- 'Scaling up' the computational experiments, e.g.,
  Does CONS arise from the biased linguistic distribution?
  Mhasawade et al. 2018: NO
- More universals from more domains, e.g., thematic roles

Future Directions

- Relation between learnability and (descriptive) complexity, e.g.,
  Does Kolmogorov complexity of Qs predict learnability (with Iris van de Pol)?
  What are the corresponding minimal programs over a LoT (with Steven Piantadosi)?
- 'Scaling up' the computational experiments, e.g.,
  Does CONS arise from the biased linguistic distribution?
  Mhasawade et al. 2018: NO
- More universals from more domains, e.g., thematic roles
- Integration with models of the evolution of language
  Recent news: iterated learning with NNs yields monotonicity (with Fausto Carcassi).

Future Directions

- Relation between learnability and (descriptive) complexity, e.g.,
  Does Kolmogorov complexity of Qs predict learnability (with Iris van de Pol)?
  What are the corresponding minimal programs over a LoT (with Steven Piantadosi)?
- 'Scaling up' the computational experiments, e.g.,
  Does CONS arise from the biased linguistic distribution?
  Mhasawade et al. 2018: NO
- More universals from more domains, e.g., thematic roles
- Integration with models of the evolution of language
  Recent news: iterated learning with NNs yields monotonicity (with Fausto Carcassi).
- Studies with humans and animals, e.g.,
  Chemla et al. have recently shown that humans and baboons are biased towards convexity. This should be extended to other universals.

Future Directions

- Relation between learnability and (descriptive) complexity, e.g.,
  Does Kolmogorov complexity of Qs predict learnability (with Iris van de Pol)?
  What are the corresponding minimal programs over a LoT (with Steven Piantadosi)?
- ‘Scaling up’ the computational experiments, e.g.,
  Does CONS arise from the biased linguistic distribution?
  Mhasawade et al. 2018: NO
- More universals from more domains, e.g., thematic roles
- Integration with models of the evolution of language
  Recent news: iterated learning with NNs yields monotonicity (with Fausto Carcassi).
- Studies with humans and animals, e.g.,
  Chemla et al. have recently shown that humans and baboons are biased towards convexity. This should be extended to other universals.
- . . . and more!

References I

- Barwise, Jon and Robin Cooper (1981). "Generalized Quantifiers and Natural Language". In: *Linguistics and Philosophy* 4.2, pp. 159–219. DOI: 10.1007/BF00350139.
- Berlin, Brent and Paul Kay (1969). *Basic Color Terms: Their Universality and Evolution*. University of California Press.
- Chomsky, Noam (1980). *Rules and Representations*. Oxford: Basil Blackwell.
- Ciardelli, Ivano, Jeroen Groenendijk, and Floris Roelofsen (2018). *Inquisitive Semantics*. Oxford Univeristy Press. URL: http://semanticsarchive.net/Archive/WFjYTUwN/book.pdf.
- Fox, Danny (2002). "Antecedent-Contained Deletion and the Copy Theory of Movement". In: *Linguistic Inquiry* 33.1, pp. 63–96. DOI: 10.1162/002438902317382189.
- Gärdenfors, Peter (2014). *The Geometry of Meaning*. The MIT Press.

References II

- Gibson, Edward et al. (2017). "Color naming across languages reflects color use". In: *Proceedings of the National Academy of Sciences* 114.40, pp. 10785–10790. DOI: 10.1073/pnas.1619666114.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. The MIT Press. URL: https://www.deeplearningbook.org/.
- Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Computation* 9.8, pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
- Hunter, Tim and Jeffrey Lidz (2013). "Conservativity and learnability of determiners". In: *Journal of Semantics* 30.3, pp. 315–334. DOI: 10.1093/jos/ffs014.
- Jäger, Gerhard (2010). "Natural Color Categories Are Convex Sets". In: *Logic, Language, and Meaning: Amsterdam Colloquium 2009*. Ed. by Maria Aloni et al., pp. 11–20. DOI: 10.1007/978-3-642-14287-1_2.

References III

Keenan, Edward L and Jonathan Stavi (1986). "A Semantic Characterization of Natural Language Determiners". In: *Linguistics and Philosophy* 9.3, pp. 253–326. DOI: 10.1007/BF00630273.

Lahiri, Utpal (2002). *Questions and Answers in Embedded Contexts*. Oxford Univeristy Press.

Mhasawade, Vishwali et al. (2018). "Neural Networks and Quantifier Conservativity: Does Data Distribution Affect Learnability?" In: *arXiv preprint arXiv:1809.05733*.

Nielsen, Michael A (2015). *Neural Networks and Deep Learning*. Determination Press. URL: http://neuralnetworksanddeeplearning.com/.

Peters, Stanley and Dag Westerståhl (2006). *Quantifiers in Language and Logic*. Oxford: Clarendon Press.

References IV

- Piantadosi, Steven T (2013). "Modeling the acquisition of quantifier semantics: a case study in function word learnability". In: URL: http://colala.bcs.rochester.edu/papers/piantadosi2012modeling.pdf.
- Pullum, Geoffrey K. and Barbara C. Scholz (2002). "Empirical assessment of stimulus poverty arguments". In: *The Linguistic Review* 18.1-2, pp. 9–50. DOI: 10.1515/tlir.19.1-2.9.
- Regier, Terry, Paul Kay, and Naveen Khetarpal (2007). "Color naming reflects optimal partitions of color space". In: *Proceedings of the National Academy of Sciences* 104.4, pp. 1436–1441. DOI: 10.1073/pnas.0610341104.
- Romoli, Jacopo (2015). "A Structural Account of Conservativity". In: *Semantics-Syntax Interface* 2.1, pp. 28–57. URL: https://www.academia.edu/8736879/A_structural_account_of_conservativity_final_version_.

References V

- Spector, Benjamin and Paul Egré (2015). "A uniform semantics for embedded interrogatives: an answer, not necessarily the answer". In: *Synthese* 192.6, pp. 1729–1784. DOI: 10.1007/s11229-015-0722-4.
- Sportiche, Dominique (2005). "Division of labor between Merge and Move: Strict locality of selection and apparent reconstruction paradoxes". URL: http://ling.auf.net/lingbuzz/000163.
- Steinert-Threlkeld, Shane (2018). "An Explanation of the Veridical Uniformity Universal". URL: https://semanticsarchive.net/Archive/DI5ZTNmN/UniversalResponsiveVerbs.pdf.
- Steinert-Threlkeld, Shane and Jakub Szymanik (2018a). "Ease of Learning Explains Semantic Universals".
- – (2018b). "Learnability and Semantic Universals". In: *Semantics & Pragmatics*. URL: http://semanticsarchive.net/Archive/mQ2Y2Y2Z/LearnabilitySemanticUniversals.pdf.
- Szabolcsi, Anna (2010). *Quantification*. Research Surveys in Linguistics. Cambridge: Cambridge University Press.

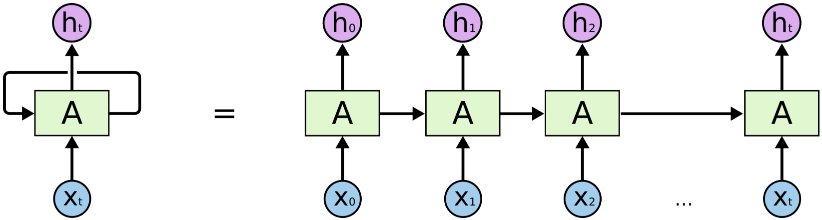
References VI

Theiler, Nadine, Floris Roelofsen, and Maria Aloni (2018). "A uniform semantics for declarative and interrogative complements". In: *Journal of Semantics*. DOI: 10.1093/jos/ffy003.

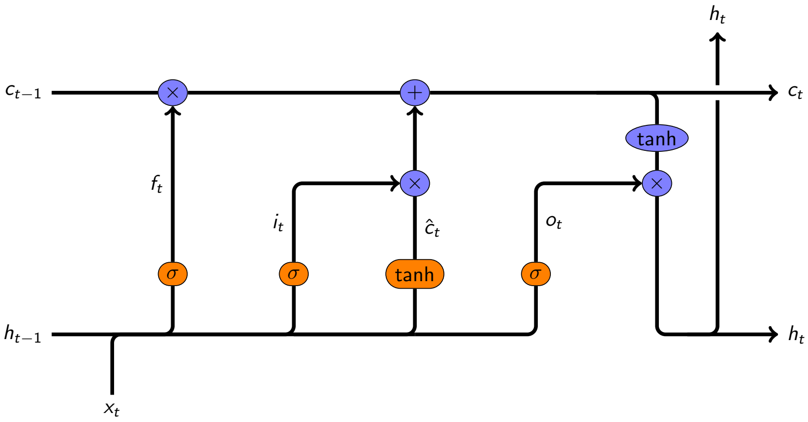
Uegaki, Wataru (2018). "The semantics of question-embedding predicates". In: *Language and Linguistics Compass*. URL: https://semanticsarchive.net/Archive/DQ3MDgwN/paper.pdf.

Overview

RNNs



Long Short-Term Memory Network



Hochreiter and Schmidhuber 1997