Learnability of Quantifiers

ML2: semantic universals, plus evolution and complexity

Shane Steinert-Threlkeld & Jakub Szymanik
University of Washington, Linguistics
ILLC
University of Amsterdam

Outline

1. Introduction
2. Quantifiers
   - RNNs + Encoding
   - Applications
3. Other Cases
   - Responsive Predicates
   - Color Terms
4. Evolution
5. Complexity
6. Conclusion

Recap

Yesterday:

- Formal learning theory: universals don't help
- Bayesian learning: restriction via conservativity doesn't help
- Intro to neural networks

Today:

- Walk through the tutorial notebook
- Apply neural learning to quantifiers (and responsive predicates and color terms)
- How learnability relates to evolution and complexity

Recap

Yesterday:

- Formal learning theory: universals don’t help
- Bayesian learning: restriction via conservativity doesn’t help
- Intro to neural networks

Today:

- Walk through the tutorial notebook
- Apply neural learning to quantifiers (and responsive predicates and color terms)
- How learnability relates to evolution and complexity

Explaining Universals

Natural Question

Why do the attested universals hold?

Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)

Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(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.

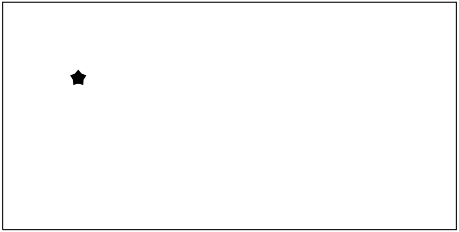
Compare: Poverty of the Stimulus argument for UG. (Chomsky 1980; Pullum and Scholz 2002)

Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)

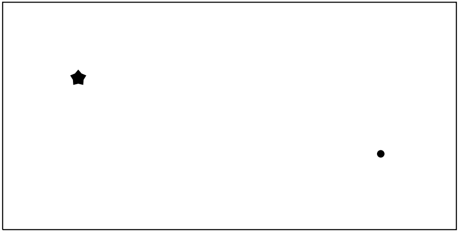


Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)

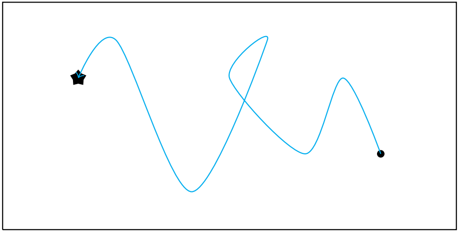


Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)

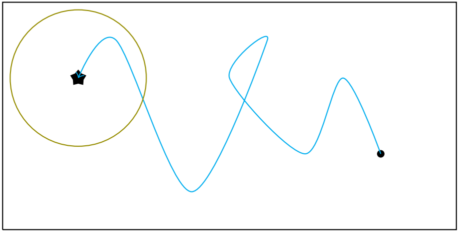


Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)

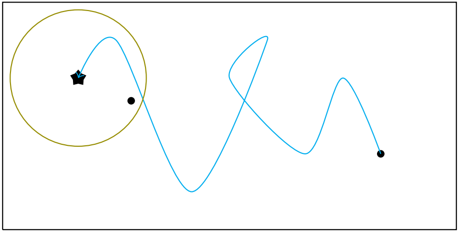


Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)

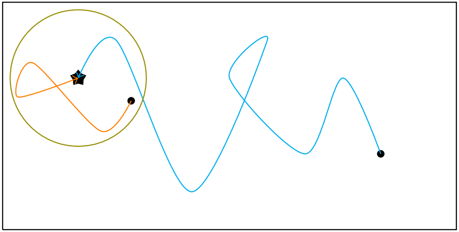


Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)

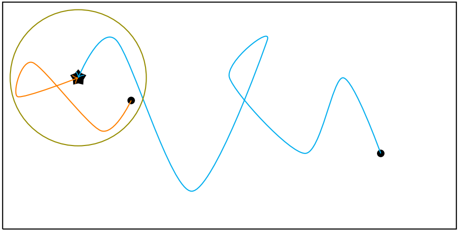


Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 1: *learnability* (as fencing-in; to be rejected).
(Barwise and Cooper 1981; Keenan and Stavi 1986; Szabolcsi 2010)



Answer must in a sense be true, but:

- Restriction may not help much. (Steven T Piantadosi, Tenenbaum, and Goodman 2013)
- Does not explain *which* universals are attested.

Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 2: *learnability* (as temperature).
(hints in van Benthem 1987; Peters and Westerståhl 2006)

Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 2: *learnability* (as temperature).
(hints in van Benthem 1987; Peters and Westerståhl 2006)

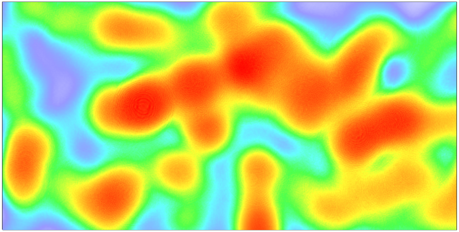
Universals aid learnability because expressions satisfying the universals are *easier* to learn than those that do not.

Explaining Universals

Natural Question

Why do the attested universals hold?

Answer 2: *learnability* (as temperature).
(hints in van Benthem 1987; Peters and Westerståhl 2006)



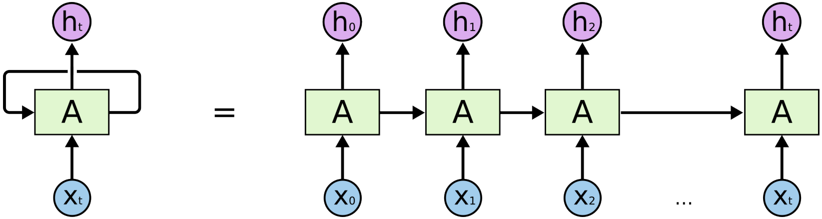
To the code!

https://github.com/shanest/nn-tutorial/blob/master/tutorial.ipynb

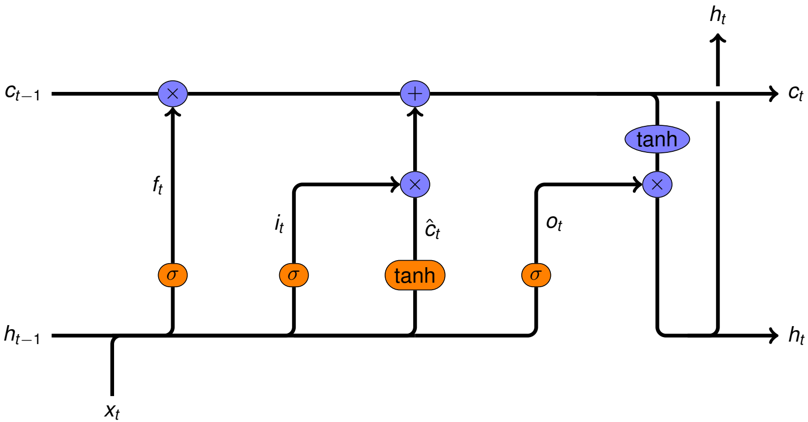
Outline

Outline

RNNs



Long Short-Term Memory Network



Hochreiter and Schmidhuber 1997

Quantifier Input

	$\in A$?	$\in B$?	x_i
o_1	✓	✓	$\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 1\end{bmatrix}$
o_2	✓	x	$\begin{bmatrix}0 & 1 & 0 & 0 & 0 & 1\end{bmatrix}$
o_3	x	✓	$\begin{bmatrix}0 & 0 & 1 & 0 & 0 & 1\end{bmatrix}$
o_4	✓	✓	$\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 1\end{bmatrix}$
o_5	x	x	$\begin{bmatrix}0 & 0 & 0 & 1 & 0 & 1\end{bmatrix}$

x_i: *i*th input to LSTM

- First four dimensions: where in the model is o_i
- Last two dimensions: label for quantifier.
  Quantifiers: ‘every’ and ‘some’ (two total)
  This example: Q = ‘some’

True label $y = \begin{bmatrix}1 & 0\end{bmatrix}$, because sentence is True.

Outline

Monotonicity

- Many Amsterdammers ride an omafiets to work.
  $\Rightarrow$ Many Amsterdammers ride a bike to work.

So: 'many' is *upward monotone*.

- Few Amsterdammers ride a bike to work.
  $\Rightarrow$ Few Amsterdammers ride an omafiets to work.

So: 'few' is *downward monotone*.

- At least 6 or at most 2 Amsterdammers ride an omafiets to work.
  $\not\Rightarrow$ (and $\not\Leftarrow$) At least 6 or at most 2 Amsterdammers ride a bike to work.

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

Monotonicity

- Many Amsterdammers ride an omafiets to work.
  $\Rightarrow$ Many Amsterdammers ride a bike to work.

So: ‘many’ is *upward monotone*.

- Few Amsterdammers ride a bike to work.
  $\Rightarrow$ Few Amsterdammers ride an omafiets to work.

So: ‘few’ is *downward monotone*.

- At least 6 or at most 2 Amsterdammers ride an omafiets to work.
  $\not\Rightarrow$ (and $\not\Leftarrow$) At least 6 or at most 2 Amsterdammers ride a bike to work.

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

Monotonicity

- Many Amsterdammers ride an omafiets to work.
  $\Rightarrow$ Many Amsterdammers ride a bike to work.

So: 'many' is *upward monotone*.

- Few Amsterdammers ride a bike to work.
  $\Rightarrow$ Few Amsterdammers ride an omafiets to work.

So: 'few' is *downward monotone*.

- At least 6 or at most 2 Amsterdammers ride an omafiets to work.
  $\not\Rightarrow$ (and $\not\Leftarrow$) At least 6 or at most 2 Amsterdammers ride a bike to work.

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

Monotonicity

- Many Amsterdammers ride an omafiets to work.
  $\Rightarrow$ Many Amsterdammers ride a bike to work.

So: ‘many’ is *upward monotone*.

- Few Amsterdammers ride a bike to work.
  $\Rightarrow$ Few Amsterdammers ride an omafiets to work.

So: ‘few’ is *downward monotone*.

- At least 6 or at most 2 Amsterdammers ride an omafiets to work.
  $\not\Rightarrow$ (and $\not\Leftarrow$) At least 6 or at most 2 Amsterdammers ride a bike to work.

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

Monotonicity

- Many Amsterdammers ride an omafiets to work.
  $\Rightarrow$ Many Amsterdammers ride a bike to work.

So: ‘many’ is *upward monotone*.

- Few Amsterdammers ride a bike to work.
  $\Rightarrow$ Few Amsterdammers ride an omafiets to work.

So: ‘few’ is *downward monotone*.

- At least 6 or at most 2 Amsterdammers ride an omafiets to work.
  $\not\Rightarrow$ (and $\not\Leftarrow$) At least 6 or at most 2 Amsterdammers ride a bike to work.

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

Monotonicity

- Many Amsterdammers ride an omafiets to work.
  $\Rightarrow$ Many Amsterdammers ride a bike to work.

So: ‘many’ is *upward monotone*.

- Few Amsterdammers ride a bike to work.
  $\Rightarrow$ Few Amsterdammers ride an omafiets to work.

So: ‘few’ is *downward monotone*.

- At least 6 or at most 2 Amsterdammers ride an omafiets to work.
  $\not\Rightarrow$ (and $\not\Leftarrow$) At least 6 or at most 2 Amsterdammers ride a bike to work.

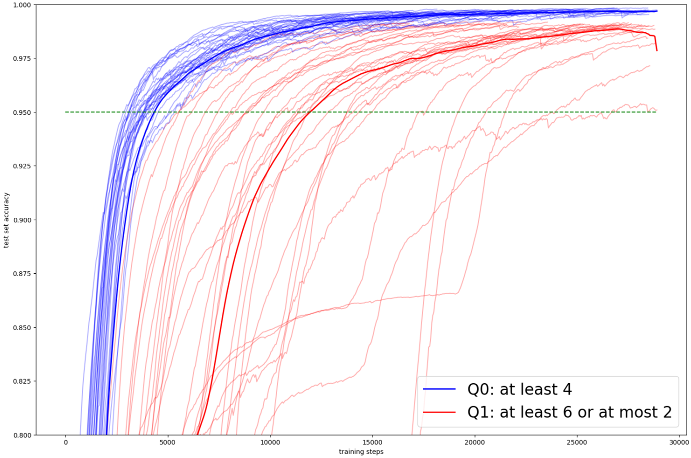
So: ‘at least 6 or at most 2’ is not monotone.

Monotonicity Universal

Monotonicity Universal

All simple determiners are monotone.
(Barwise and Cooper 1981)

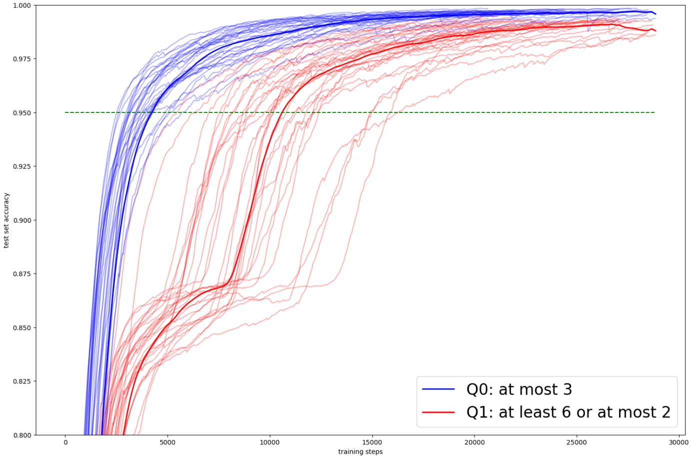
Monotonicity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, “Learnability and Semantic Universals”, in *Semantics & Pragmatics*.

Code and data: https://github.com/shanest/quantifier-rnn-learning.

Monotonicity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, "Learnability and Semantic Universals", in *Semantics & Pragmatics*.

Code and data: https://github.com/shanest/quantifier-rnn-learning.

Quantity

- At least three buildings at Science Park are blue.
  There are exactly as many blue and non-blue buildings on El Camino Real as at Science Park.
  $\Rightarrow$ At least three buildings on El Camino Real are blue.

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

- The first three buildings at Science Park are blue.
  There are exactly as many blue and non-blue buildings on El Camino Real as at Science Park.
  $\not\Rightarrow$ The first three buildings on El Camino Real are blue.

So: 'the first three' is not quantitative.

Quantity

- At least three buildings at Science Park are blue.
  There are exactly as many blue and non-blue buildings on El Camino Real as at Science Park.
  $\Rightarrow$ At least three buildings on El Camino Real are blue.

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

- The first three buildings at Science Park are blue.
  There are exactly as many blue and non-blue buildings on El Camino Real as at Science Park.
  $\not\Rightarrow$ The first three buildings on El Camino Real are blue.

So: ‘the first three’ is not quantitative.

Quantity

- At least three buildings at Science Park are blue.
  There are exactly as many blue and non-blue buildings on El Camino Real as at Science Park.
  $\Rightarrow$ At least three buildings on El Camino Real are blue.

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

- The first three buildings at Science Park are blue.
  There are exactly as many blue and non-blue buildings on El Camino Real as at Science Park.
  $\not\Rightarrow$ The first three buildings on El Camino Real are blue.

So: ‘the first three’ is not quantitative.

Quantity

- At least three buildings at Science Park are blue.
  There are exactly as many blue and non-blue buildings on El Camino Real as at Science Park.
  $\Rightarrow$ At least three buildings on El Camino Real are blue.

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

- The first three buildings at Science Park are blue.
  There are exactly as many blue and non-blue buildings on El Camino Real as at Science Park.
  $\not\Rightarrow$ The first three buildings 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

All simple determiners are quantitative.
(Keenan and Stavi 1986; Peters and Westerståhl 2006)

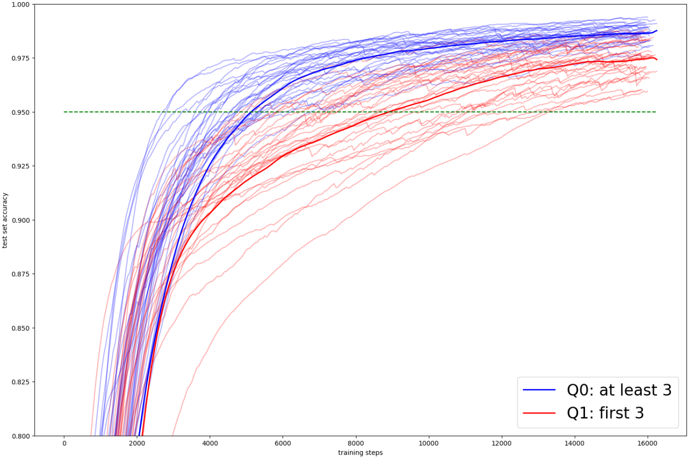
Quantity Universal

- 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

All simple determiners are quantitative.
(Keenan and Stavi 1986; Peters and Westerståhl 2006)

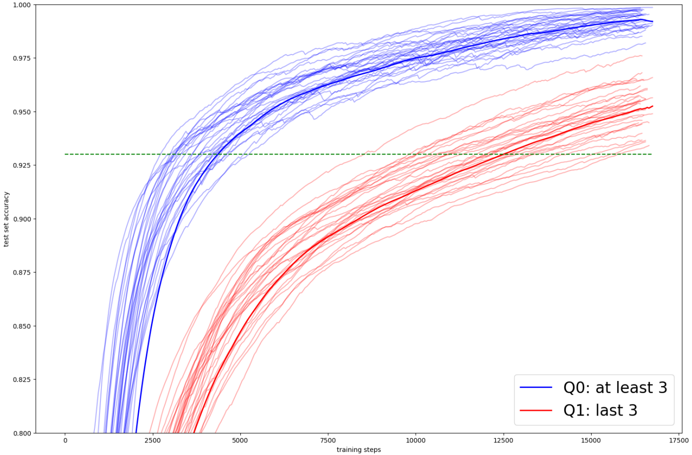
Quantity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, “Learnability and Semantic Universals”, in *Semantics & Pragmatics*.

Code and data: https://github.com/shanest/quantifier-rnn-learning.

Quantity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, "Learnability and Semantic Universals", in *Semantics & Pragmatics*.

Code and data: https://github.com/shanest/quantifier-rnn-learning.

Conservativity

- Many Amsterdammers ride an omafiets to work.
  $\equiv$ Many Amsterdammers are Amsterdammers who ride an omafiets to work.

So: ‘many’ is *conservative*.

- Only Amsterdammers ride an omafiets to work.
  $\not\equiv$ Only Amsterdammers are Amsterdammers who ride an omafiets to work.

So: ‘only’ is not conservative.

Conservativity

- Many Amsterdammers ride an omafiets to work.
  $\equiv$ Many Amsterdammers are Amsterdammers who ride an omafiets to work.

So: ‘many’ is *conservative*.

- Only Amsterdammers ride an omafiets to work.
  $\not\equiv$ Only Amsterdammers are Amsterdammers who ride an omafiets to work.

So: ‘only’ is not conservative.

Conservativity

- Many Amsterdammers ride an omafiets to work.
  $\equiv$ Many Amsterdammers are Amsterdammers who ride an omafiets to work.

So: 'many' is *conservative*.

- Only Amsterdammers ride an omafiets to work.
  $\not\equiv$ Only Amsterdammers are Amsterdammers who ride an omafiets to work.

So: 'only' is not conservative.

Conservativity

- Many Amsterdammers ride an omafiets to work.
  $\equiv$ Many Amsterdammers are Amsterdammers who ride an omafiets to work.

So: ‘many’ is *conservative*.

- Only Amsterdammers ride an omafiets to work.
  $\not\equiv$ Only Amsterdammers are Amsterdammers who ride an omafiets to work.

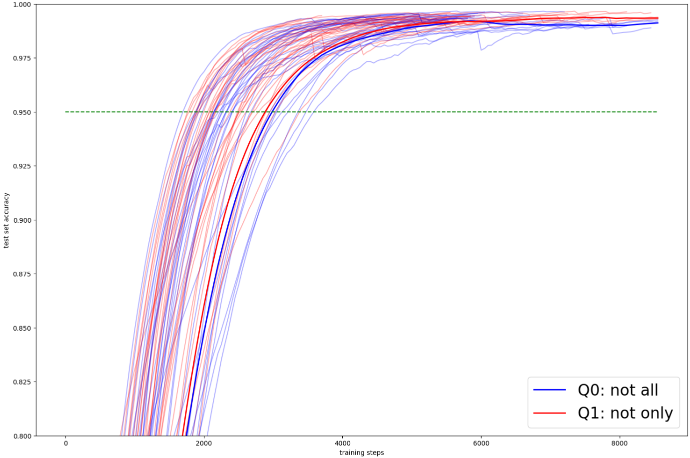
So: ‘only’ is not conservative.

Conservativity Universal

Conservativity Universal

All simple determiners are conservative.
(Barwise and Cooper 1981; Keenan and Stavi 1986)

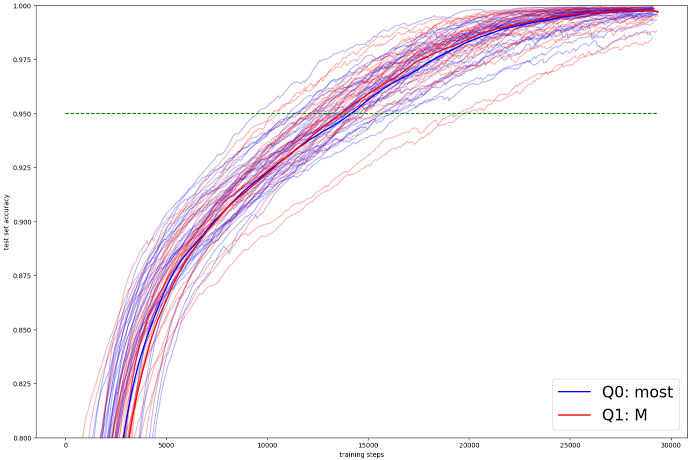
Conservativity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, “Learnability and Semantic Universals”, in *Semantics & Pragmatics*.

Code and data: https://github.com/shanest/quantifier-rnn-learning.

Conservativity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, "Learnability and Semantic Universals", in *Semantics & Pragmatics*.

Code and data: https://github.com/shanest/quantifier-rnn-learning.

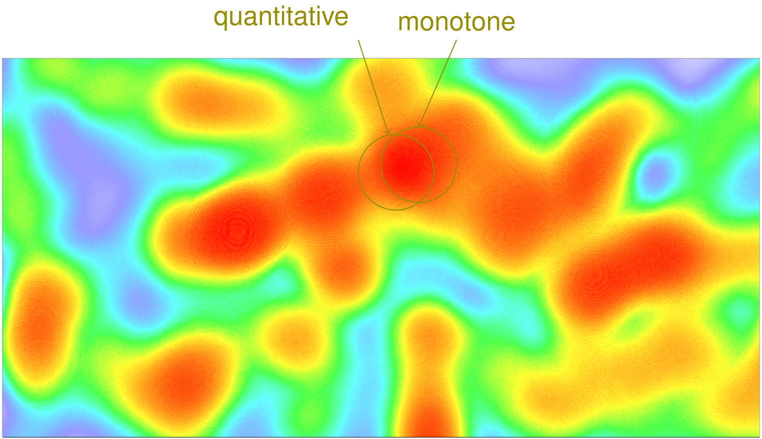
Conservativity: Discussion

- The data generation does not 'break the symmetry' between $A \setminus B$ and $B \setminus A$.
- Conservativity may be a syntactic/structural constraint, not a constraint on the lexicon.
  [See Fox 2002; Romoli 2015; Sportiche 2005, summarized Appendix to these slides]

Conservativity: Discussion

- The data generation does not 'break the symmetry' between $A \setminus B$ and $B \setminus A$.
- Conservativity may be a syntactic/structural constraint, not a constraint on the lexicon.
  [See Fox 2002; Romoli 2015; Sportiche 2005, summarized Appendix to these slides]

Quantifiers: Summary



$D_{\langle et, \langle et, t \rangle \rangle}$

Outline

Outline

Types of Clause-Embedding Predicates

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

Types of Clause-Embedding Predicates

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

Types of Clause-Embedding Predicates

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

Types of Predicates

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

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

Veridicality

- Maria knows that the canal has 7 bridges.
  $\rightsquigarrow$ The canal has 7 bridges.

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

- Maria knows how many bridges the canal has.
  The canal has 7 bridges.
  $\rightsquigarrow$ Maria knows that the canal has 7 bridges.

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

Veridicality

- Maria knows that the canal has 7 bridges.
  $\rightsquigarrow$ The canal has 7 bridges.

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

- Maria knows how many bridges the canal has.
  The canal has 7 bridges.
  $\rightsquigarrow$ Maria knows that the canal has 7 bridges.

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

Veridicality

- Maria knows that the canal has 7 bridges.
  $\rightsquigarrow$ The canal has 7 bridges.

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

- Maria knows how many bridges the canal has.
  The canal has 7 bridges.
  $\rightsquigarrow$ Maria knows that the canal has 7 bridges.

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

Veridicality

- Maria knows that the canal has 7 bridges.
  $\rightsquigarrow$ The canal has 7 bridges.

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

- Maria knows how many bridges the canal has.
  The canal has 7 bridges.
  $\rightsquigarrow$ Maria knows that the canal has 7 bridges.

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

Veridicality

- Maria knows that the canal has 7 bridges.
  $\rightsquigarrow$ The canal has 7 bridges.

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

- Maria knows how many bridges the canal has.
  The canal has 7 bridges.
  $\rightsquigarrow$ Maria knows that the canal has 7 bridges.

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

Veridicality

- Maria is certain that the canal has 7 bridges.
  $\not\to$ The canal has 7 bridges.

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

- Maria is certain about how many bridges the canal has.
  The canal has 7 bridges.
  $\not\to$ Maria is certain that the canal has 7 bridges.

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

Veridicality

- Maria is certain that the canal has 7 bridges.
  $\not\rightarrow$ The canal has 7 bridges.

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

- Maria is certain about how many bridges the canal has.
  The canal has 7 bridges.
  $\not\rightarrow$ Maria is certain that the canal has 7 bridges.

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

Veridicality

- Maria is certain that the canal has 7 bridges.
  $\not\rightarrow$ The canal has 7 bridges.

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

- Maria is certain about how many bridges the canal has.
  The canal has 7 bridges.
  $\not\rightarrow$ Maria is certain that the canal has 7 bridges.

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

Veridicality

- Maria is certain that the canal has 7 bridges.
  $\not\rightsquigarrow$ The canal has 7 bridges.

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

- Maria is certain about how many bridges the canal has.
  The canal has 7 bridges.
  $\not\rightsquigarrow$ Maria is certain that the canal has 7 bridges.

So: ‘be certain’ is *not* veridical with respect to interrogative complements.

So: ‘be certain’ is *veridically uniform*.

Veridicality

- Maria is certain that the canal has 7 bridges.
  $\not\rightarrow$ The canal has 7 bridges.

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

- Maria is certain about how many bridges the canal has.
  The canal has 7 bridges.
  $\not\rightarrow$ Maria is certain that the canal has 7 bridges.

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

The Veridical Uniformity Thesis

Veridical Uniformity Universal

All responsive predicates are veridically uniform.
(Spector and Egré 2015; Theiler, Roelofsen, and Aloni 2018)

Four Responsive Predicates

Predicate	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 predicates, exemplifying the possible profiles of veridicality.

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

Responsive Predicate Input

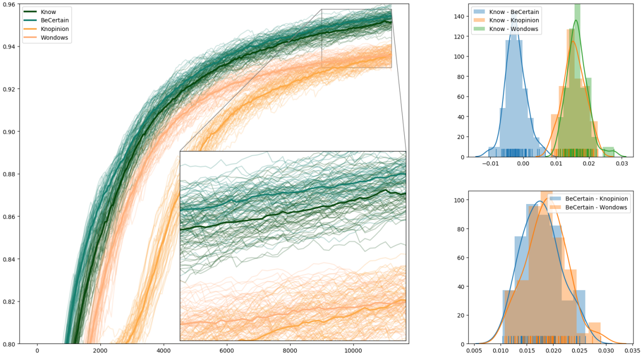
Suppose $W = \{w_1, w_2, w_3\}$, and we are considering an example with $Q = \{\{w_1\}, \{w_2, w_3\}\}$.

world	encoded
w_1	$\begin{bmatrix}1 & 0 & 0\end{bmatrix}$
w_2	$\begin{bmatrix}0 & 1 & 1\end{bmatrix}$
w_3	$\begin{bmatrix}0 & 1 & 1\end{bmatrix}$

We concatenate all of the following together:

- Encoding of each world
- A label for the predicate (e.g. $\begin{bmatrix}0 & 1 & 0 & 0\end{bmatrix}$)
- A label for the world of evaluation (e.g. $\begin{bmatrix}0 & 0 & 1\end{bmatrix}$)
- A vector (length $|W|$) for $\textsc{Dox}_w^a$ (e.g. $\begin{bmatrix}0 & 1 & 1\end{bmatrix}$)

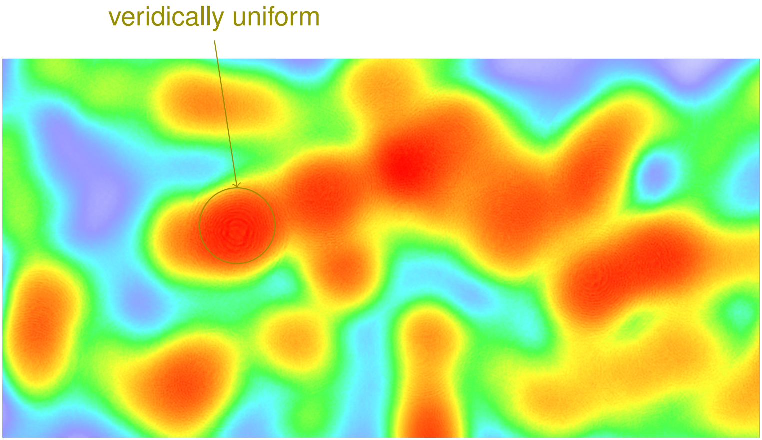
Veridical Uniformity: Results



Shane Steinert-Threlkeld, "An Explanation of the Veridical Uniformity Universal", in *Journal of Semantics*.

Code and data: https://github.com/shanest/responsive-verbs.

Responsive Predicates: Summary

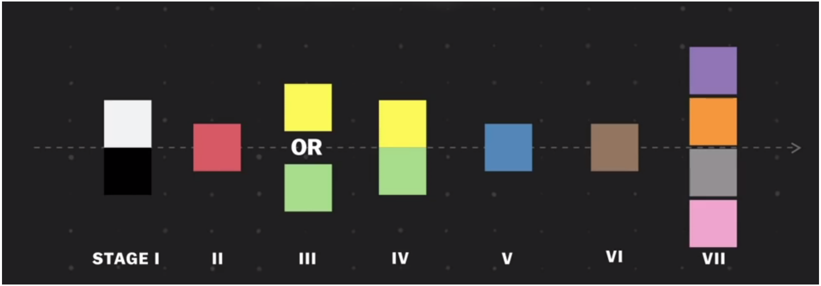


$D_{\text{responsive}}$

Outline

1. Introduction
2. Quantifiers
   - RNNs + Encoding
   - Applications
3. Other Cases
   - Responsive Predicates
   - Color Terms
4. Evolution
5. Complexity
6. Conclusion

The Order of Color Terms



Berlin and Kay 1969; E. Gibson, Futrell, Jara-Ettinger, Mahowald, Bergen, Ratnasingam, M. Gibson, Steven T. Piantadosi, and Conway 2017; Regier, Kay, and Khetarpal 2007

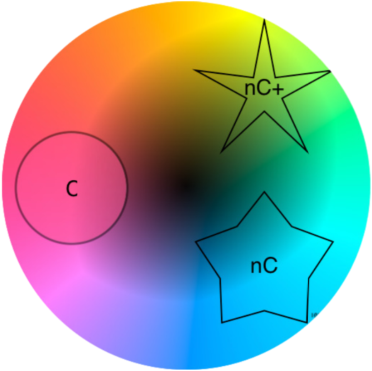
https://www.vox.com/videos/2017/5/16/15646500/color-pattern-language

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 $t \in (0, 1)$,

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

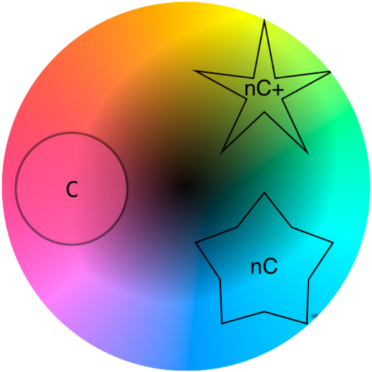


Convexity

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 $t \in (0, 1)$,

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



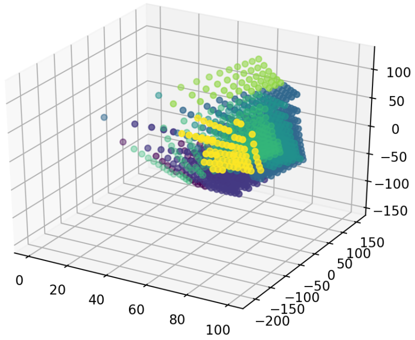
Convexity universal

Convexity Universal

All color terms denote convex regions of color space.
(Gärdenfors 2014; Jäger 2010)

Partitioning CIE-L*a*b* Space

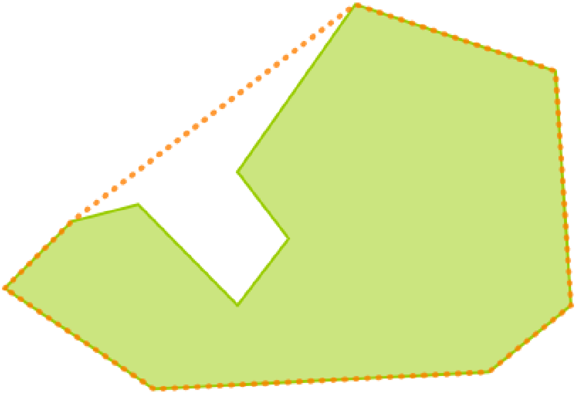
We generated 300 artificial color-naming systems by partitioning the CIELab color space into distinct categories. CIELab approximates human color vision. It is perceptually uniform, meaning that the distance in the space corresponds well with the visually perceived color change.



Example Partitions of 2D space

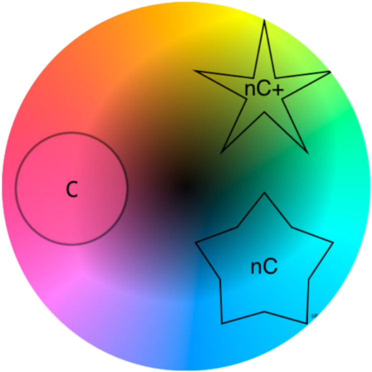
Degree of convexity

We measured the degree of convexity as the (weighted) average area of the convex hull of each color that is covered by that color.

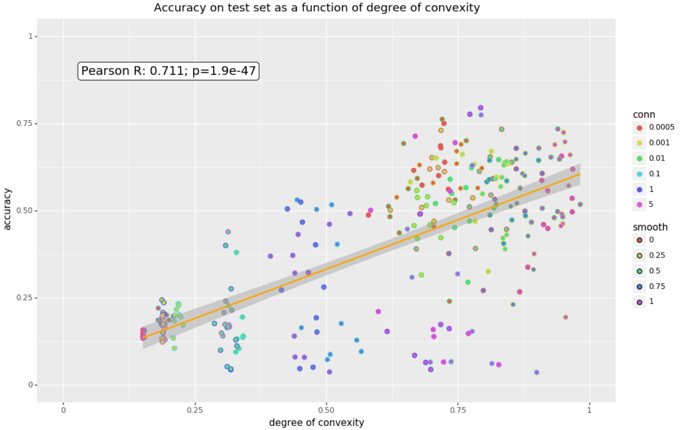


Degree of convexity

We measured the degree of convexity as the (weighted) average area of the convex hull of each color that is covered by that color.



Convexity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, “Ease of learning explains semantic universals”, *Cognition*.

Code and data: https://github.com/shanest/color-learning.

Convexity: Commonality Analysis

Variable	R^2	ΔR^2
conn	0.180	0.0003
smooth	0.008	0.0365
degree of convexity	**0.505**	**0.3726**
conn · smooth	0.054	0.0019
min size	0.014	0.0000
max size	0.001	0.0000
median size	0.000	0.0007
min / max	0.043	0.0014
max − min	0.000	0.0000

Shane Steinert-Threlkeld and Jakub Szymanik, “Ease of learning explains semantic universals”, *Cognition*.

Code and data: https://github.com/shanest/color-learning.

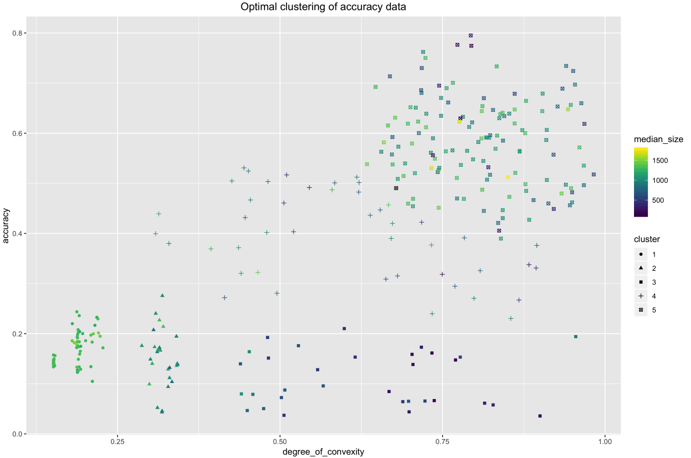
Controlling for Linear Separability

Variable	R^2	ΔR^2
degree of convexity	**0.505**	**0.1288**
linear separability	0.418	0.0005

Shane Steinert-Threlkeld and Jakub Szymanik, “Ease of learning explains semantic universals”, *Cognition*.

Code and data: https://github.com/shanest/color-learning.

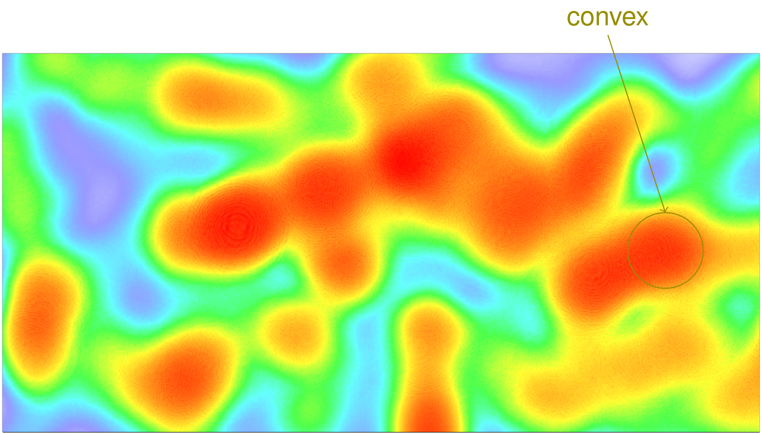
Cluster Analysis



Shane Steinert-Threlkeld and Jakub Szymanik, “Ease of learning explains semantic universals”, *Cognition*.

Code and data: https://github.com/shanest/color-learning.

Colors: Summary



D_{color}

Interim Summary

Ease of learning, measured as the speed of convergence of NNs, can explain the presence of linguistic universals in various semantic domains, including both function and content words.

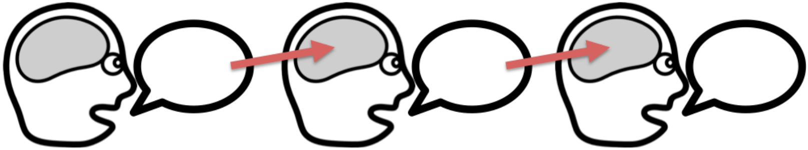
- Can the observed linguistic structure be explained by the learnability bias?
- Are there other / ‘better’ explanations?

Outline

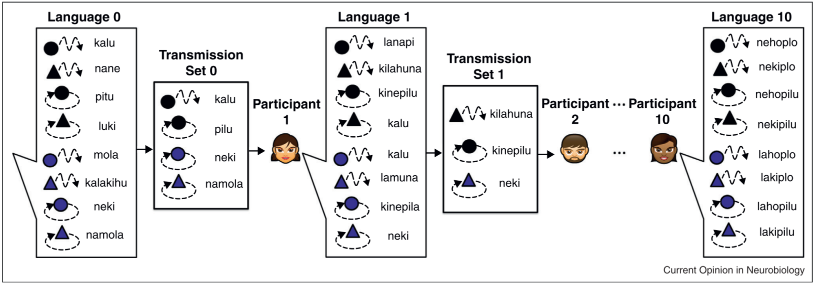
The Problem of Linkage (Kirby 1999)

Monotone quantifiers are easier to learn.	$\Longrightarrow$???	Natural language quantifiers are monotone.

Transmission



Iterated Learning



Kirby, Griffiths, and Smith 2014

Degree of Monotonicity: Intuition

Intuitively, quantifiers can be *more or less* monotone.

Information theoretically: how much information about the quantifier is provided by which models have true sub-models?

Degree of Monotonicity: Intuition

Intuitively, quantifiers can be *more or less* monotone.

-

\- +

\- + +

\- + + +

\- + + + +

... ...

at least *n*

Information theoretically: how much information about the quantifier is provided by which models have true sub-models?

Degree of Monotonicity: Intuition

Intuitively, quantifiers can be *more or less* monotone.

-

\- +

\- + +

\- + + -

\- + + - -

... ...

at least $n \prec$ between m and n

Information theoretically: how much information about the quantifier is provided by which models have true sub-models?

Degree of Monotonicity: Intuition

Intuitively, quantifiers can be *more or less* monotone.

-

\- +

\- + -

\- + - +

\- + - + -

... ...

at least *n* $\prec$ between *m* and *n* $\prec$ an odd number of

Information theoretically: how much information about the quantifier is provided by which models have true sub-models?

Degree of Monotonicity: Definition

Binary random variables:

- $\mathbb{1}_Q$: the truth-value of quantifier Q
- $\mathbb{1}_Q^{\prec}$: whether a model has a true sub-model

Mutual Information:

$$I(\mathbb{1}_Q; \mathbb{1}_Q^{\prec}) := H(\mathbb{1}_Q) - H(\mathbb{1}_Q | \mathbb{1}_Q^{\prec})$$

Degree of monotonicity:

$$\begin{aligned}\text{mon}(Q) &:= \frac{I(\mathbb{1}_Q; \mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)} \\ &= \frac{H(\mathbb{1}_Q) - H(\mathbb{1}_Q | \mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)} \\ &= 1 - \frac{H(\mathbb{1}_Q \mid \mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)}\end{aligned}$$

Degree of Monotonicity: Definition

Binary random variables:

- $\mathbb{1}_Q$: the truth-value of quantifier Q
- $\mathbb{1}_Q^{\prec}$: whether a model has a true sub-model

Mutual Information:

$$I(\mathbb{1}_Q;\mathbb{1}_Q^{\prec}) := H(\mathbb{1}_Q) - H(\mathbb{1}_Q|\mathbb{1}_Q^{\prec})$$

Degree of monotonicity:

$$\begin{aligned}\text{mon}(Q) &:= \frac{I(\mathbb{1}_Q;\mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)} \\ &= \frac{H(\mathbb{1}_Q) - H(\mathbb{1}_Q|\mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)} \\ &= 1 - \frac{H(\mathbb{1}_Q \mid \mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)}\end{aligned}$$

Degree of Monotonicity: Definition

Binary random variables:

- $\mathbb{1}_Q$: the truth-value of quantifier Q
- $\mathbb{1}_Q^{\prec}$: whether a model has a true sub-model

Mutual Information:

$$I(\mathbb{1}_Q; \mathbb{1}_Q^{\prec}) := H(\mathbb{1}_Q) - H(\mathbb{1}_Q | \mathbb{1}_Q^{\prec})$$

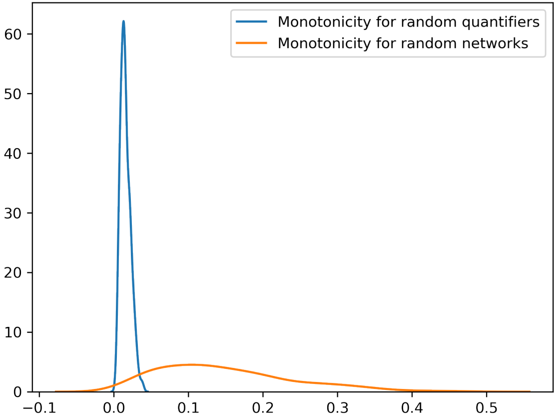
Degree of monotonicity:

$$\begin{aligned} \text{mon}(Q) &:= \frac{I(\mathbb{1}_Q; \mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)} \\ &= \frac{H(\mathbb{1}_Q) - H(\mathbb{1}_Q | \mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)} \\ &= 1 - \frac{H(\mathbb{1}_Q \mid \mathbb{1}_Q^{\prec})}{H(\mathbb{1}_Q)} \end{aligned}$$

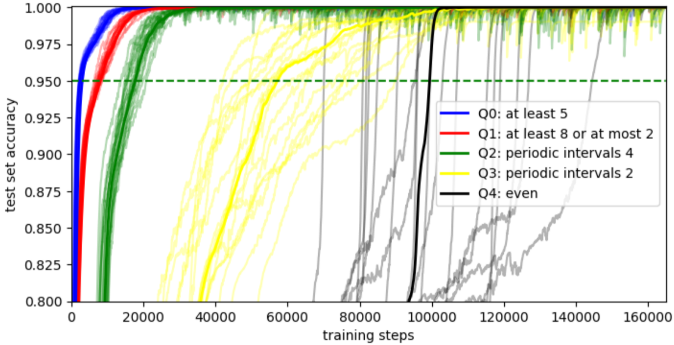
Degree of Monotonicity: Examples

- at least *n*: 1
- between 3 and 5: 0.752
- an even number of: 0.001

Degree of Monotonicity: Distribution

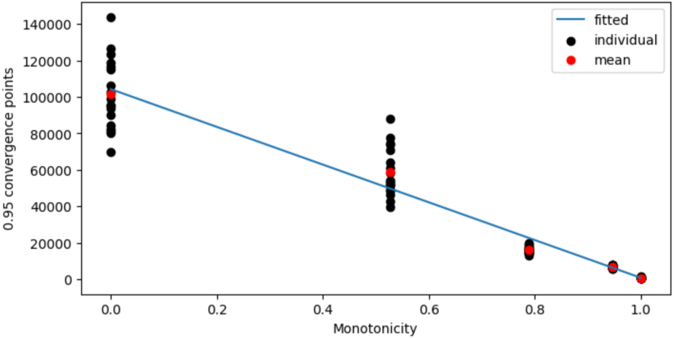


Degrees of Monotonicity and Learnability



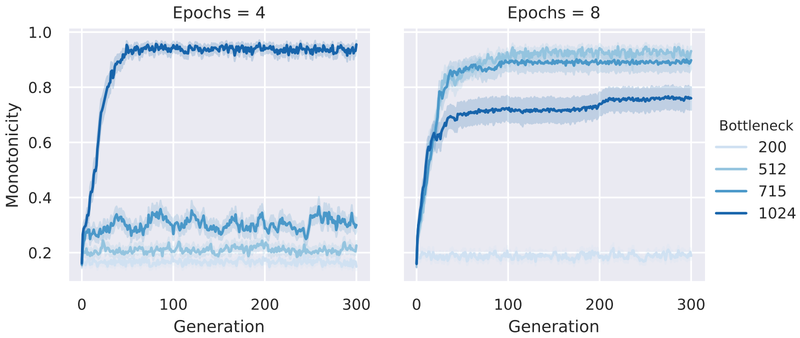
Zi Long Zhu, “Machine learning and semantic universals”, BSc Informatica thesis

Degrees of Monotonicity and Learnability



Zi Long Zhu, “Machine learning and semantic universals”, BSc Informatica thesis

Results



Fausto Carcassi, Shane Steinert-Threlkeld, and Jakub Szymanik, “Monotone Quatifiers Emerge via Iterated Learning”, *Cognitive Science*.

Code and data: https://github.com/thelogicalgrammar/NeuralNetIteratedQuantifiers

Outline

Learnability and Complexity

- Learnability *can* explain the presence of universals.
- But is it the only (or the best) such explanation?
- A natural idea: some notion of *complexity* explains both the universals and the learnability facts.

Learnability and Complexity

- Learnability *can* explain the presence of universals.
- But is it the only (or the best) such explanation?
- A natural idea: some notion of *complexity* explains both the universals and the learnability facts.

Learnability and Complexity

- Learnability *can* explain the presence of universals.
- But is it the only (or the best) such explanation?
- A natural idea: some notion of *complexity* explains both the universals and the learnability facts.

What's the right measure of complexity?

- Previous attempts fail to capture the distinctions:
  - Automata theory (van Benthem 1986)
  - Computational complexity (Szymanik 2016)
  - Formal learning theory (Gierasimczuk 2007, 2009; Tiede 1999)
- Let's try: information-theoretic perspective on GQs.

What's the right measure of complexity?

- Previous attempts fail to capture the distinctions:
  - Automata theory (van Benthem 1986)
  - Computational complexity (Szymanik 2016)
  - Formal learning theory (Gierasimczuk 2007, 2009; Tiede 1999)
- Let's try: information-theoretic perspective on GQs.

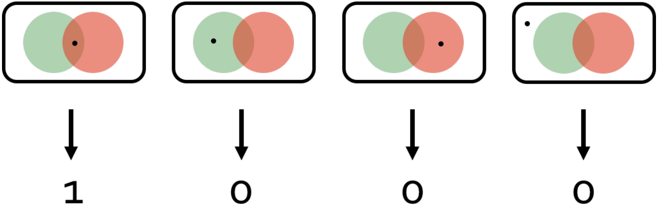
(Approximate) Kolmogorov Complexity

- $K(x)$—the length of the shortest program p that outputs x
- The drawback: K is uncomputable
- $LZ(x)$—Lempel-Ziv is a tractable approximation of K
- Recent applications: Dingle, Camargo, and Louis 2018; Feldman 2016; Valle-Pérez, Camargo, and Louis 2019

(Approximate) Kolmogorov Complexity

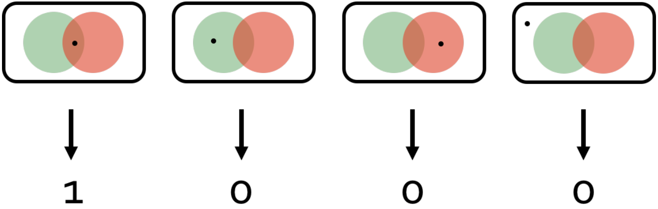
- $K(x)$—the length of the shortest program p that outputs x
- The drawback: K is uncomputable
- $LZ(x)$—Lempel-Ziv is a tractable approximation of K
- Recent applications: Dingle, Camargo, and Louis 2018; Feldman 2016; Valle-Pérez, Camargo, and Louis 2019

Method



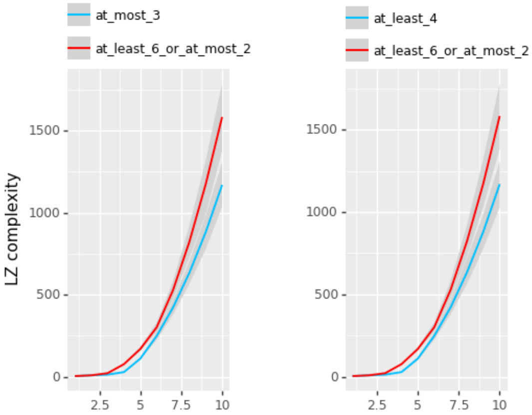
Idea: universals induce regularity/structure in the distribution of truth values across models, which aid compressibility.

Method



Idea: universals induce regularity/structure in the distribution of truth values across models, which aid compressibility.

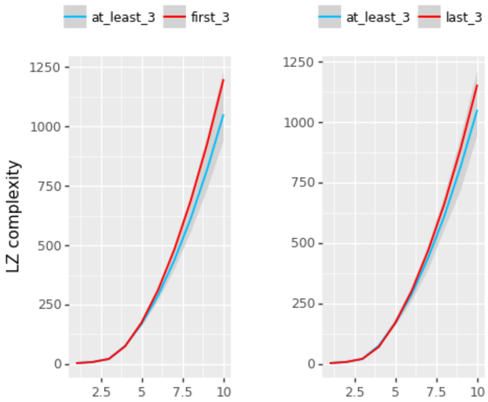
LZ Results: Monotonicity



Iris van de Pol, Paul Lodder, Leendert van Maanen, Shane Steinert-Threlkeld, Jakub Szymanik, “Quantifiers satisfying semantic universals have shorter minimal description length”, *Cognition*.

Code and data: https://tinyurl.com/quantifierLZ

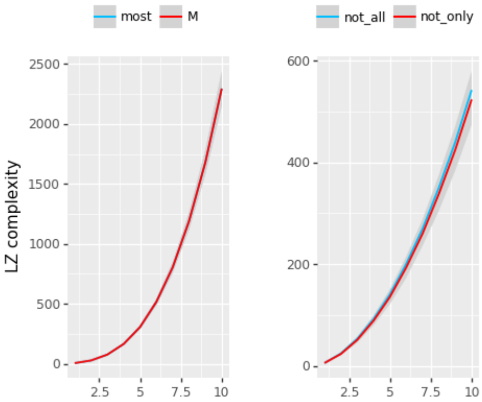
LZ Results: Quantity



Iris van de Pol, Paul Lodder, Leendert van Maanen, Shane Steinert-Threlkeld, Jakub Szymanik, "Quantifiers satisfying semantic universals have shorter minimal description length", *Cognition*.

Code and data: https://tinyurl.com/quantifierLZ

LZ Results: Conservativity



Iris van de Pol, Paul Lodder, Leendert van Maanen, Shane Steinert-Threlkeld, Jakub Szymanik, “Quantifiers satisfying semantic universals have shorter minimal description length”, *Cognition*.

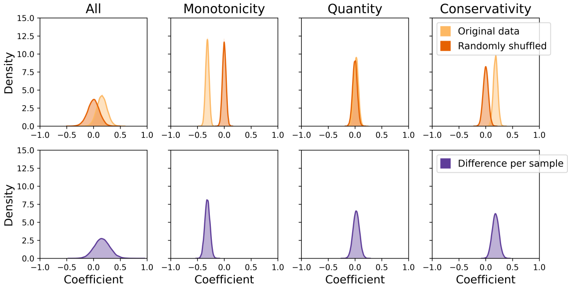
Code and data: https://tinyurl.com/quantifierLZ

Scaling Up Complexity Results

Moving beyond the minimal pair methodology:

- Generate a *large* body of quantifiers using a LoT grammar
- Measure: LZ, minimum expression length (as yesterday), all three universals
- Logistic regressions to predict presence/absence of each universal on the basis of complexity

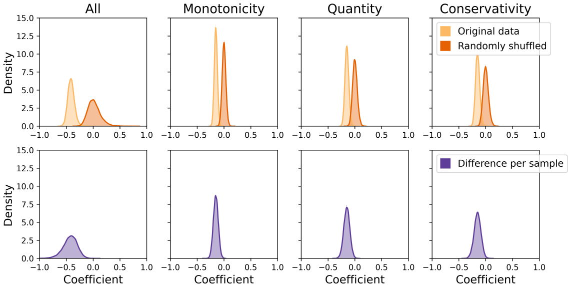
LZ Scaled Results



Iris van de Pol, Paul Lodder, Leendert van Maanen, Shane Steinert-Threlkeld, Jakub Szymanik, “Quantifiers satisfying semantic universals have shorter minimal description length”, *Cognition*.

Code and data: https://tinyurl.com/quantifierLZ

Minimum Expression Length Scaled Results



Iris van de Pol, Paul Lodder, Leendert van Maanen, Shane Steinert-Threlkeld, Jakub Szymanik, "Quantifiers satisfying semantic universals have shorter minimal description length", *Cognition*.

Code and data: https://tinyurl.com/quantifierLZ

Complexity: Conclusions

- Does complexity provide an independent explanation for the universals?
- LZ: monotonicity patterns the same as learnability, but Quantity and Conservativity are different.
- MEL: all three properties correlate well with complexity
- Much more work to be done: which properties are robust with respect to ‘most’ measures, learnability at scale as well

Outline

Explaining Universals

Why do semantic universals arise?

(I) Because expressions satisfying them are easier to learn.
(II) And languages tend to lexicalize easier-to-learn expressions.

We provided evidence for (I) by training neural networks to learn expressions from three very different linguistic domains, spanning function and content words: quantifiers, responsive predicates, and color terms.

For (II), combining iterated learning with neural network agents leads to the emergence of *monotone* quantifiers.

Explaining Universals

Why do semantic universals arise?

(I) Because expressions satisfying them are easier to learn.
(II) And languages tend to lexicalize easier-to-learn expressions.

We provided evidence for (I) by training neural networks to learn expressions from three very different linguistic domains, spanning function and content words: quantifiers, responsive predicates, and color terms.

For (II), combining iterated learning with neural network agents leads to the emergence of *monotone* quantifiers.

Explaining Universals

Why do semantic universals arise?

(I) Because expressions satisfying them are easier to learn.
(II) And languages tend to lexicalize easier-to-learn expressions.

We provided evidence for (I) by training neural networks to learn expressions from three very different linguistic domains, spanning function and content words: quantifiers, responsive predicates, and color terms.

For (II), combining iterated learning with neural network agents leads to the emergence of *monotone* quantifiers.

Explaining Universals

Why do semantic universals arise?

(I) Because expressions satisfying them are easier to learn.
(II) And languages tend to lexicalize easier-to-learn expressions.

We provided evidence for (I) by training neural networks to learn expressions from three very different linguistic domains, spanning function and content words: quantifiers, responsive predicates, and color terms.

For (II), combining iterated learning with neural network agents leads to the emergence of *monotone* quantifiers.

Future Directions

- More universals from more domains.
- ‘Scaling up’ the computational experiments, e.g.,
  Does CONS arise from a biased linguistic distribution?
  Mhasawade, Szabó, Tosik, and Wang 2018: NO
- IL: more realistic quantifiers, other case studies, full typological picture
- Simplicity-informativeness trade-off:
  - Quantifiers (Steinert-Threlkeld 2021)
  - Indefinites (Denić, Steinert-Threlkeld, Szymanik 2022)
  - Modals (Imel, Steinert-Threlkeld 2022)
- …

Future Directions

- More universals from more domains.
- ‘Scaling up’ the computational experiments, e.g.,
  Does CONS arise from a biased linguistic distribution?
  Mhasawade, Szabó, Tosik, and Wang 2018: NO
- IL: more realistic quantifiers, other case studies, full typological picture
- Simplicity-informativeness trade-off:
  - Quantifiers (Steinert-Threlkeld 2021)
  - Indefinites (Denić, Steinert-Threlkeld, Szymanik 2022)
  - Modals (Imel, Steinert-Threlkeld 2022)
- …

Future Directions

- More universals from more domains.
- 'Scaling up' the computational experiments, e.g.,
  Does CONS arise from a biased linguistic distribution?
  Mhasawade, Szabó, Tosik, and Wang 2018: NO
- IL: more realistic quantifiers, other case studies, full typological picture
- Simplicity-informativeness trade-off:
  - Quantifiers (Steinert-Threlkeld 2021)
  - Indefinites (Denić, Steinert-Threlkeld, Szymanik 2022)
  - Modals (Imel, Steinert-Threlkeld 2022)
- …

Future Directions

- More universals from more domains.
- ‘Scaling up’ the computational experiments, e.g.,
  Does CONS arise from a biased linguistic distribution?
  Mhasawade, Szabó, Tosik, and Wang 2018: NO
- IL: more realistic quantifiers, other case studies, full typological picture
- Simplicity-informativeness trade-off:
  - Quantifiers (Steinert-Threlkeld 2021)
  - Indefinites (Denić, Steinert-Threlkeld, Szymanik 2022)
  - Modals (Imel, Steinert-Threlkeld 2022)
- …

Tomorrow

(1) But what about *human* learning and development?
(2) Are the universals really cross-linguistically valid?
(3) Discussion

Outline

Confusion Matrices

	all		know		be-certain		knopinion		wondows	
label	1	0	1	0	1	0	1	0	1	0
1	15412.2	1176.4	3881.1	261.7	3878.5	240.8	3843.0	349.2	3809.6	324.7
0	587.8	14823.7	118.9	3738.3	121.6	3759.2	156.9	3650.9	190.4	3675.3

Table: Average confusion matrix across all 60 trials, in total and by verb. The rows are predicted truth-value, and the columns the actual truth value.

Distributions by Verb

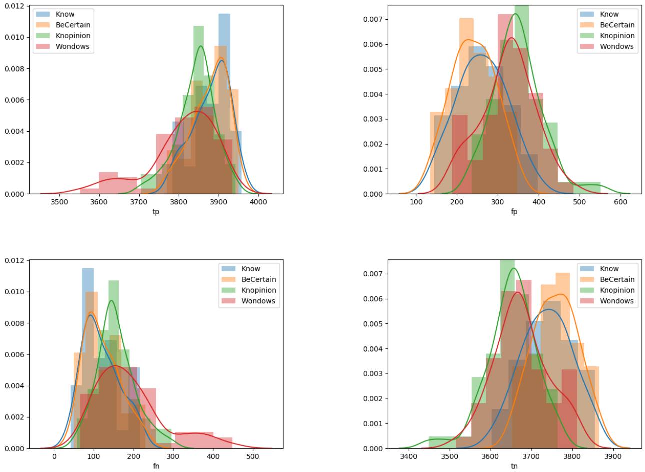


Figure: Distributions (Gaussian kernel density estimates) of the true/false positives/negatives by verb.

Accuracy by Semantic Properties of Input

factor	value	know	be-certain	knopinion	wondows
complement	declarative	0.983	0.986	0.954	0.983
	interrogative	0.923	0.924	0.921	0.841
$w \in \mathrm{DOX}_w^a$	1	0.964	0.957	0.954	0.947
	0	0.919	0.953	0.887	0.924
$\mathrm{DOX}_w^a \in P$	1	0.961	0.966	0.949	0.947
	0	0.945	0.943	0.929	0.922

Table: Accuracy by verb and various semantic features of the input, aggregated across all trials.

Outline

The Core Idea

Conservativity, neutrally stated: every sentence of the form “D NP VP” is truth-conditionally equivalent to “D NP is an NP that VP”.

Structural Conservativity: every sentence of the form “D NP VP” is truth-conditionally equivalent to $f(⟦NP⟧)(⟦VP⟧)$ for some conservative function f, *whether or not* D denotes a conservative quantifier.

The Core Idea

Conservativity, neutrally stated: every sentence of the form “D NP VP” is truth-conditionally equivalent to “D NP is an NP that VP”.

Structural Conservativity: every sentence of the form “D NP VP” is truth-conditionally equivalent to $f([\![\text{NP}]\!])([\![\text{VP}]\!])$ for some conservative function f, *whether or not* D denotes a conservative quantifier.

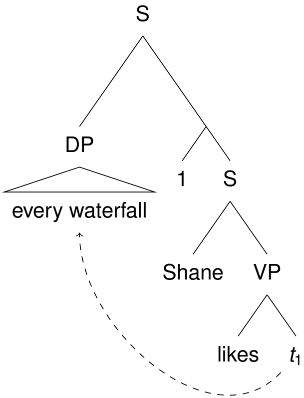
Movement à la Heim & Kratzer

Shane likes every waterfall.

Every waterfall is such that it is liked by Shane.

Movement à la Heim & Kratzer

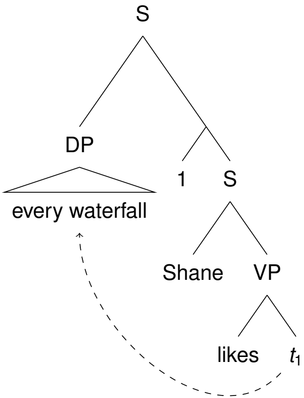
Shane likes every waterfall.



Every waterfall is such that it is liked by Shane.

Movement à la Heim & Kratzer

Shane likes every waterfall.



Every waterfall is such that it is liked by Shane.

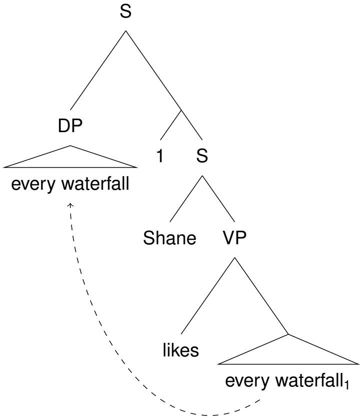
Movement as copying

Shane likes every waterfall.

Every waterfall is such that it is a waterfall liked by Shane.

Movement as copying

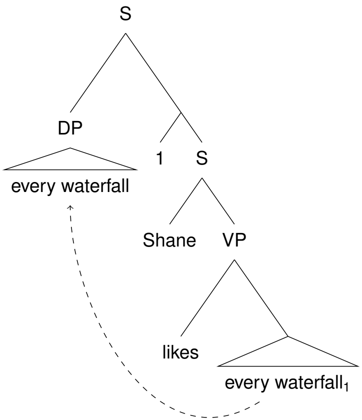
Shane likes every waterfall.



Every waterfall is such that it is a waterfall liked by Shane.

Movement as copying

Shane likes every waterfall.



Every waterfall is such that it is a waterfall liked by Shane.

Movement Without Type Mismatch

Every waterfall is tall.

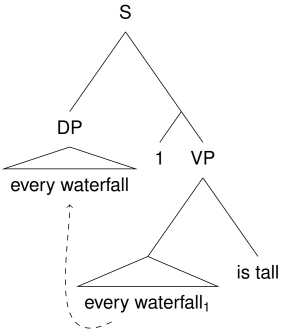
Key ingredient: VP internal subject hypothesis (e.g. Kratzer 1996).

Every waterfall is such that it is a waterfall that is tall.

Movement Without Type Mismatch

Every waterfall is tall.

Key ingredient: VP internal subject hypothesis (e.g. Kratzer 1996).

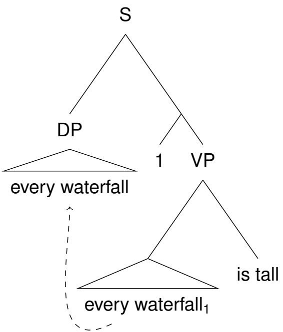


Every waterfall is such that it is a waterfall that is tall.

Movement Without Type Mismatch

Every waterfall is tall.

Key ingredient: VP internal subject hypothesis (e.g. Kratzer 1996).



Every waterfall is such that it is a waterfall that is tall.

Worked Example

Consider a hypothetical non-conservative determiner ‘equi’:

$$[\![\text{equi}]\!] = \{\langle M, A, B\rangle : A = B\}$$

With (i) copy theory of movement and (ii) VP-internal subjects:

‘Equi French people smoke cigarettes’ is true iff:

$$[\![\text{French people}]\!] = [\![\text{French people}]\!] \cap [\![\text{smoke cigarettes}]\!]$$

This is equivalent to: ‘All French people smoke cigarettes’!

Worked Example

Consider a hypothetical non-conservative determiner ‘equi’:

$$[\![\text{equi}]\!] = \{\langle M, A, B\rangle : A = B\}$$

With (i) copy theory of movement and (ii) VP-internal subjects:

‘Equi French people smoke cigarettes’ is true iff:

$$[\![\text{French people}]\!] = [\![\text{French people}]\!] \cap [\![\text{smoke cigarettes}]\!]$$

This is equivalent to: ‘All French people smoke cigarettes’!

Worked Example

Consider a hypothetical non-conservative determiner ‘equi’:

$$[\![\text{equi}]\!] = \{\langle M, A, B\rangle : A = B\}$$

With (i) copy theory of movement and (ii) VP-internal subjects:

‘Equi French people smoke cigarettes’ is true iff:

$$[\![\text{French people}]\!] = [\![\text{French people}]\!] \cap [\![\text{smoke cigarettes}]\!]$$

This is equivalent to: ‘All French people smoke cigarettes’!

Outline

Algorithm for Generating Color Systems

Algorithm 1 Generate an artificial color system

Parameters: temp (t), conn (c), initial ball size (b)
Inputs: a set X, distance measure d, number of categories N

```
UNLABELED ← X; LABELED_i ← ∅ (∀i ∈ {1, ..., N})
Choose x_1, ..., x_N uniformly at random from X
for i = 1, ..., N do
    LABELED_i += x_i; pop(x_i, UNLABELED)
    for all x ∈ NearestNeighbors(x_i, b) do
        LABELED_i += x; pop(x, UNLABELED)
    end for
end for
while UNLABELED ≠ ∅ do
    d_i ← 1/(min_{x' ∈ LABELED_i} d(x, x'))^{1/4}
    p_i ← e^{d_i/t} / Σ_j e^{d_j/t}
    Choose label i with probability p_i
    LABELED_i += x; pop(x, UNLABELED)
end while
for i = 1, ..., N, ordered by increasing size of LABELED_i do
    M_i ← ConvexHull(LABELED_i) \ LABELED_i
    R_i ← ClosestPoints(M_i, LABELED_i, c · |M_i|)
    for all x ∈ R_i do
        LABELED_i += x; pop(x, cell(x))
    end for
end for
```

Outline

References I

- Barwise, Jon and Robin Cooper (1981). “Generalized Quantifiers and Natural Language”. In: *Linguistics and Philosophy* 4.2, pp. 159–219.
- Benthem, Johan van (1986). *Essays in Logical Semantics*. Dordrecht: D. Reidel Publishing Company.
- – (1987). “Toward a Computational Semantics”. In: *Generalized Quantifiers: Linguistic and Logical Approaches*. Ed. by Peter Gardenfors. Kluwer Academic Publishers, pp. 31–71.
- 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.

References II

- Dingle, Kamaludin, Chico Q. Camargo, and Ard A. Louis (2018). "Input-output maps are strongly biased towards simple outputs". In: *Nature Communications* 9.1. ISSN: 20411723. DOI: 10.1038/s41467-018-03101-6.
- Feldman, Jacob (2016). "The simplicity principle in perception and cognition". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 7.5, pp. 330–340. ISSN: 19395086. DOI: 10.1002/wcs.1406.
- 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 III

- Gibson, Edward, Richard Futrell, Julian Jara-Ettinger, Kyle Mahowald, Leon Bergen, Sivalogeswaran Ratnasingam, Mitchell Gibson, Steven T. Piantadosi, and Bevil R. Conway (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.
- Gierasimczuk, Nina (2007). “The Problem of Lerning the Semantics of Quantifiers”. In: *6th International Tbilsi Symposium on Logic, Language, and Computation (TbiLLC)*. Ed. by Balder ten Cate and Henk Zeevat. Vol. 4363. Lecture Notes in Computer Science, pp. 117–126. DOI: 10.1007/978-3-540-75144-1_9.
- – (2009). “Identification through Inductive Verification Application to Monotone Quantifiers”. In: *7th International Tbilsi Symposium on Logic, Language, and Computation, TbiLLC 2007*. Ed. by P Bosch, D Gabelaia, and J Lang. Vol. 5422. Lecture Notes in Artificial Intelligence, pp. 193–205.

References IV

- 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.
- Jäger, Gerhard (2010). "Natural Color Categories Are Convex Sets". In: *Logic, Language, and Meaning: Amsterdam Colloquium 2009*. Ed. by Maria Aloni, Harald Bastiaanse, Tikitu de Jager, and Katrin Schulz, pp. 11–20. DOI: 10.1007/978-3-642-14287-1_2.
- 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.
- Kirby, Simon (1999). *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford Univeristy Press.

References V

- Kirby, Simon, Tom Griffiths, and Kenny Smith (2014). "Iterated learning and the evolution of language". In: *Current Opinion in Neurobiology* 28, pp. 108–114. DOI: 10.1016/j.conb.2014.07.014.
- Kratzer, Angelika (1996). "Severing the External Argument from its Verb". In: *Phrase Structure and the Lexicon*. Ed. by Johan Rooryck and Laurie Zaring. Vol. 33. Studies in Natural Language and Linguistic Theory. Springer Netherlands, pp. 109–137.
- Lahiri, Utpal (2002). *Questions and Answers in Embedded Contexts*. Oxford Univeristy Press.
- Mhasawade, Vishwali, Ildikó Emese Szabó, Melanie Tosik, and Sheng-Fu Wang (2018). "Neural Networks and Quantifier Conservativity: Does Data Distribution Affect Learnability?"
- Peters, Stanley and Dag Westerståhl (2006). *Quantifiers in Language and Logic*. Oxford: Clarendon Press.

References VI

- Piantadosi, Steven T, Joshua B Tenenbaum, and Noah D Goodman (2013). "Modeling the acquisition of quantifier semantics: a case study in function word learnability". In:
- 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.
- 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.

References VII

- Sportiche, Dominique (2005). “Division of labor between Merge and Move: Strict locality of selection and apparent reconstruction paradoxes”.
- Steinert-Threlkeld, Shane (2019). “An Explanation of the Veridical Uniformity Universal”. In: *Journal of Semantics*.
- Steinert-Threlkeld, Shane and Jakub Szymanik (2018). “Learnability and Semantic Universals”. In: *Semantics & Pragmatics*.
- – (2019). “Ease of Learning Explains Semantic Universals”.
- Szabolcsi, Anna (2010). *Quantification*. Research Surveys in Linguistics. Cambridge: Cambridge University Press.
- Szymanik, Jakub (2016). *Quantifiers and Cognition: Logical and Computational Perspectives*. Vol. 96. Studies in Linguistics and Philosophy. Springer. DOI: 10.1007/978-3-319-28749-2.

References VIII

- 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.
- Tiede, Hans-Joerg (1999). "Identifiability in the Limit of Context-Free Generalized Quantifiers". In: *Journal of Language and Computation*.
- Uegaki, Wataru (2018). "The semantics of question-embedding predicates". In: *Language and Linguistics Compass*.
- Valle-Pérez, Guillermo, Chico Q. Camargo, and Ard A. Louis (2019). "Deep learning generalizes because the parameter-function map is biased towards simple functions". In: *International Conference of Learning Representations (ICLR 2019)*. arXiv: 1805.08522.