Unnatural Language Semantics

Shane Steinert-Threlkeld

CLASP @ University of Gothenburg
30 March 2022

Overview

Unnatural Language Semantics

Natural Language Semantics

"...the analysis of the meanings of words and sentences in natural languages like Japanese and English."
— Elbourne 2011

Unnatural Language Semantics

Natural Language Semantics

"...the analysis of the meanings of words and sentences in natural languages like Japanese and English."
— Elbourne 2011

Unnatural Language Semantics

The analysis of the meanings of words and sentences in *un*natural languages *un*like Japanese and English.

Explaining Semantic Universals

Question

What is the range of *semantic* variation in human languages?

That is: (with respect to meaning) Which out of all of the logically possible languages that humans could speak, do they in fact speak?

Explaining Semantic Universals

Question

What is the range of *semantic* variation in human languages?

That is: (with respect to meaning) Which out of all of the logically possible languages that humans could speak, do they in fact speak?

Answer: cross-linguistically attested meanings are *easier to learn* than unattested ones.

Explaining Semantic Universals

Question

What is the range of *semantic* variation in human languages?

That is: (with respect to meaning) Which out of all of the logically possible languages that humans could speak, do they in fact speak?

Answer: cross-linguistically attested meanings are *easier to learn* than unattested ones.

Answer: cross-linguistically attested meanings optimize a trade-off between simplicity and informativeness.

Explaining Semantic Universals

Question

What is the range of *semantic* variation in human languages?

That is: (with respect to meaning) Which out of all of the logically possible languages that humans could speak, do they in fact speak?

Answer: cross-linguistically attested meanings are *easier to learn* than unattested ones.

Answer: cross-linguistically attested meanings optimize a trade-off between simplicity and informativeness.

Today: look at each of these answers. Discuss how to 'resolve' (or dissolve?) the tension.

Overview

Overview

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.

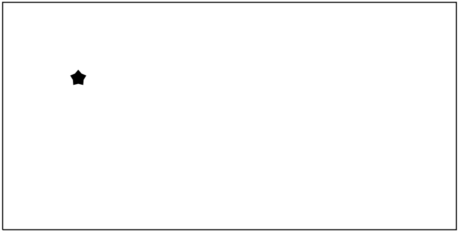
Compare: Poverty of the Stimulus argument for UG. (Chomsky 1980; Pullum and Scholz 2002)

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)

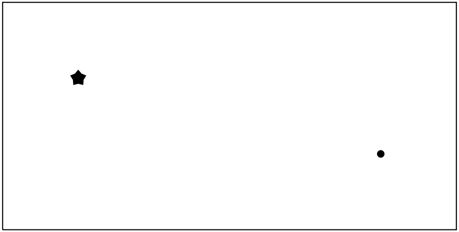


Explaining Universals

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)

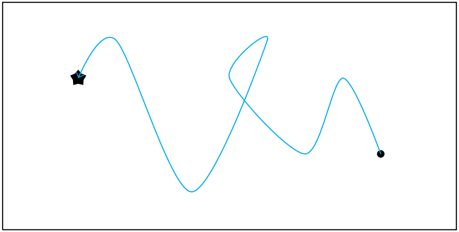


Explaining Universals

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)

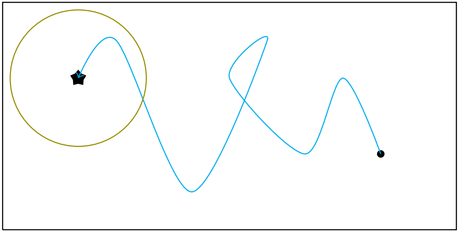


Explaining Universals

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)

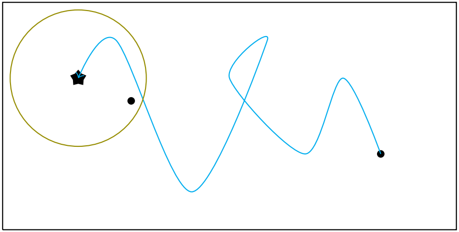


Explaining Universals

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)

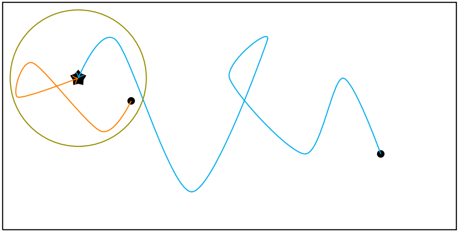


Explaining Universals

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)

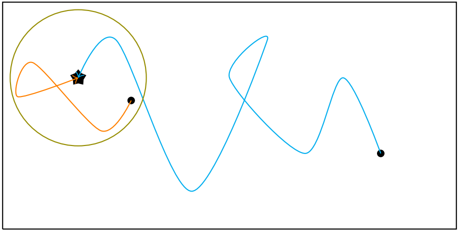


Explaining Universals

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. (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)

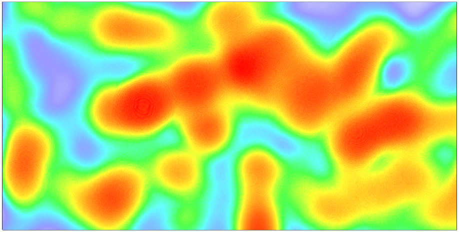
Universals aid learnability because expressions satisfying the universals are *easier* to learn than those that do not.

Explaining Universals

Natural Question

Why do the attested universals hold?

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



The Goal

General Hypothesis

Semantic universals arise because expressions satisfying them are easier to learn than those that do not.

The Goal

General Hypothesis

Semantic universals arise because expressions satisfying them are easier to learn than those that do not.

Auxiliary assumption: Languages tend to lexicalize easier-to-learn meanings.

The Goal

General Hypothesis

Semantic universals arise because expressions satisfying them are easier to learn than those that do not.

Auxiliary assumption: Languages tend to lexicalize easier-to-learn meanings.

Challenge: provide a model of learning according to which expressions satisfying semantic universals are in fact easier to learn.

Overview

Determiners

- Determiners:
  - Simple: *every*, *some*, *few*, *most*, *five*, ...
  - Complex: *all but five*, *fewer than three*, *at least eight or fewer than five*, ...

Determiners

- Determiners:
  - Simple: *every*, *some*, *few*, *most*, *five*, . . .
  - Complex: *all but five*, *fewer than three*, *at least eight or fewer than five*, . . .
- Semantics:
  - All languages have NPs, whose semantic function is to express generalized quantifiers. (Barwise and Cooper 1981)
  - 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

- Determiners:
  - Simple: *every*, *some*, *few*, *most*, *five*, . . .
  - Complex: *all but five*, *fewer than three*, *at least eight or fewer than five*, . . .
- Semantics:
  - All languages have NPs, whose semantic function is to express generalized quantifiers. (Barwise and Cooper 1981)
  - 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 Amsterdammers ride an omafiets to work.
  $\Rightarrow$ Many Amsterdammers ride a bike to work.

Monotonicity

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

So: 'many' is *upward 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.

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*.

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.

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

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

Monotonicity Universal

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

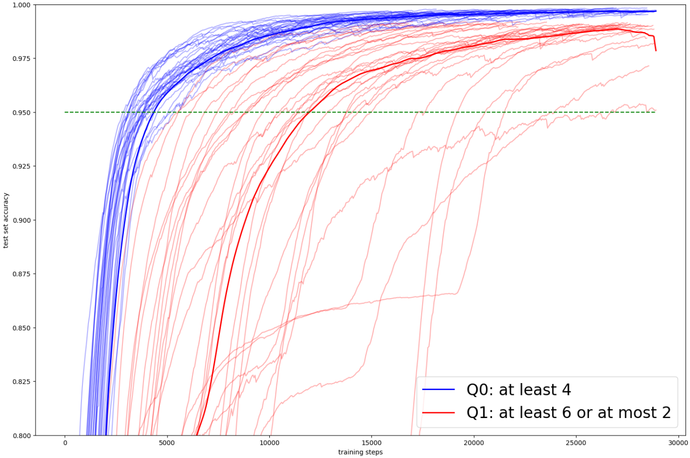
Monotonicity Universal

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

Monotonicity Universal

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

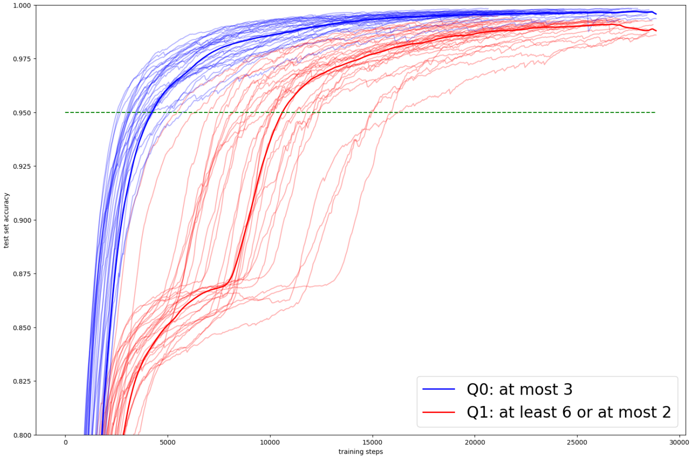
Monotonicity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, "Learnability and Semantic Universals", in *Semantics & Pragmatics*, `http://dx.doi.org/10.3765/sp.12.4`.

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*, `http://dx.doi.org/10.3765/sp.12.4`.

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.

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*.

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.

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 $\langle M, A, B, \ldots \rangle \in \mathrm{Q}$ 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 $\langle M', A', B', \ldots \rangle \in \mathrm{Q}$

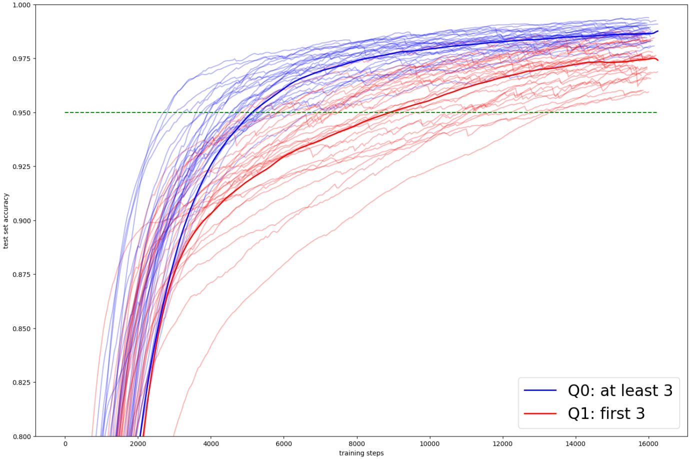
Quantity Universal

- 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)

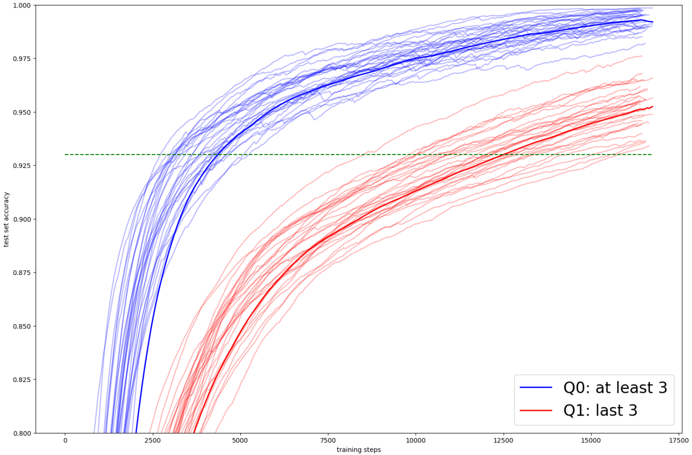
Quantity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, "Learnability and Semantic Universals", in *Semantics & Pragmatics*, `http://dx.doi.org/10.3765/sp.12.4`.

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*, `http://dx.doi.org/10.3765/sp.12.4`.

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.

Conservativity

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

So: 'many' is *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.

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

- Q is *conservative*:
  $\langle M, A, B\rangle \in \mathrm{Q}$ if and only if $\langle M, A, A \cap B\rangle \in \mathrm{Q}$

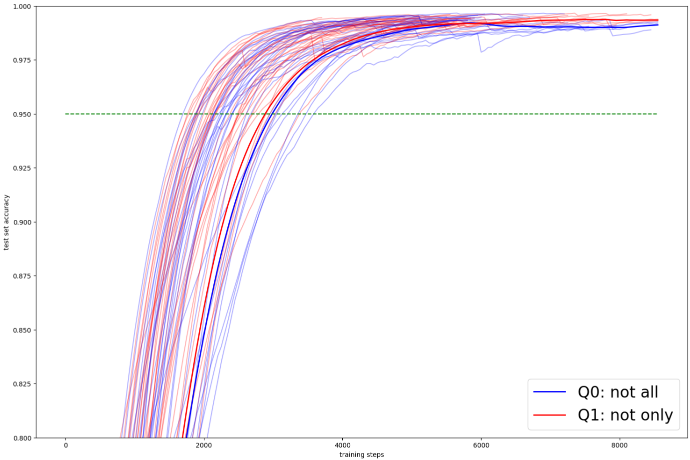
Conservativity Universal

- Q is *conservative*:
  $\langle M, A, B\rangle \in Q$ if and only if $\langle M, A, A \cap B\rangle \in Q$

Conservativity Universal

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

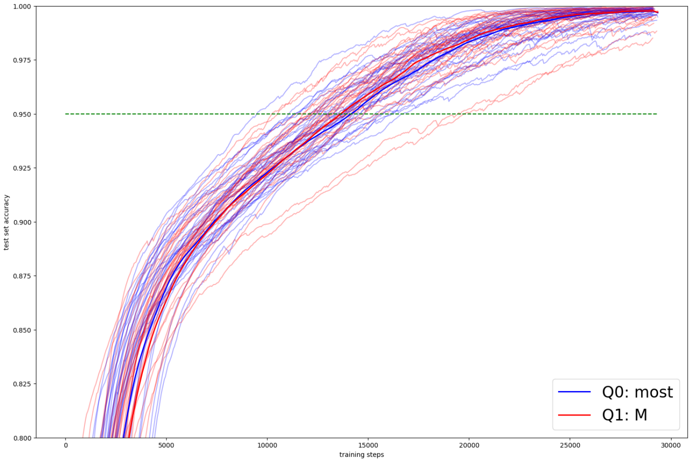
Conservativity: Results



Shane Steinert-Threlkeld and Jakub Szymanik, "Learnability and Semantic Universals", in *Semantics & Pragmatics*, `http://dx.doi.org/10.3765/sp.12.4`.

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*, `http://dx.doi.org/10.3765/sp.12.4`.

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

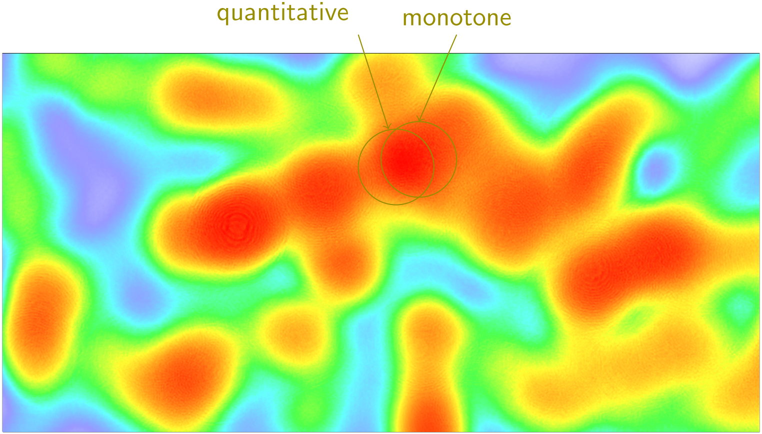
Conservativity: Discussion

- The data generation does not 'break the symmetry' between $A \setminus B$ and $B \setminus A$.

Conservativity: 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; Sportiche 2005; Romoli 2015]

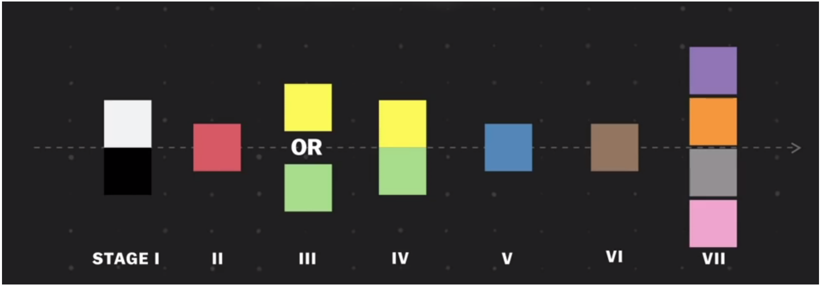
Quantifiers: Summary



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

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

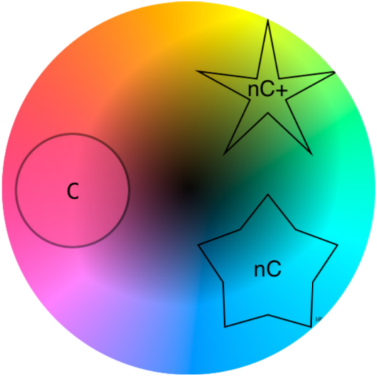
While natural languages vary in how many color terms they have and which specific colors are denoted, it seems that all color terms denote very ‘well-behaved’ regions of color space.

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 $t \in (0, 1)$,

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



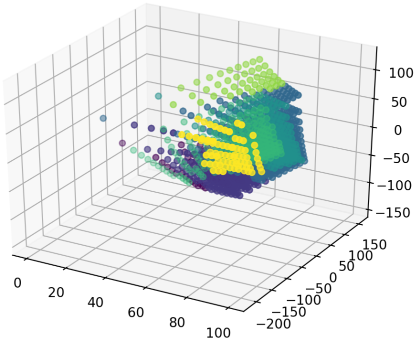
Convexity universal

Convexity Universal

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

Partitioning CIE-L*a*b* Space

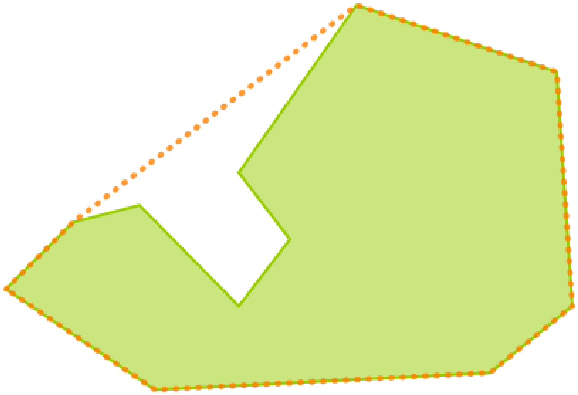
We generated 300 artificial color-naming systems by partitioning the CIELab color space into distinct categories. CIELab approximates human color vision. It is perceptually uniform, meaning that the distance in the space corresponds well with the visually perceived color change.



Example Partitions of 2D space

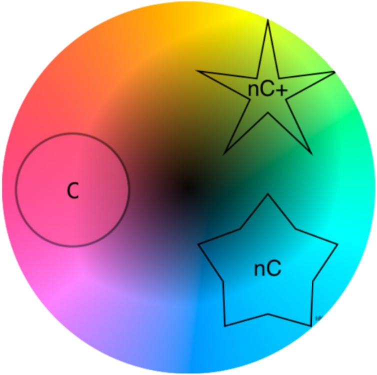
Degree of convexity

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

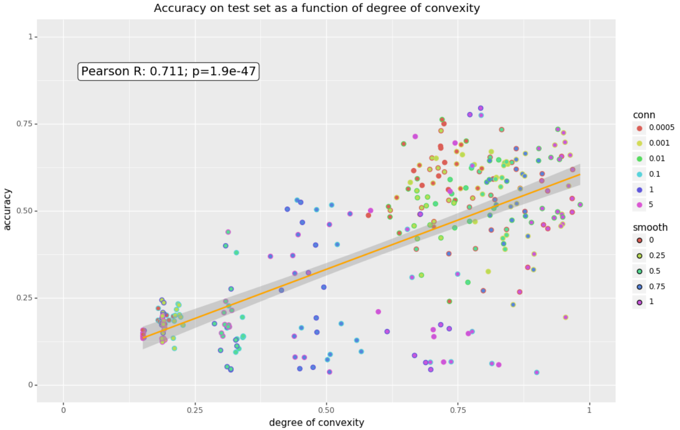


Degree of convexity

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", in *Cognition*, `https://doi.org/10.1016/j.cognition.2019.104076`.

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", in *Cognition*, `https://doi.org/10.1016/j.cognition.2019.104076`.
Code and data: https://github.com/shanest/color-learning.

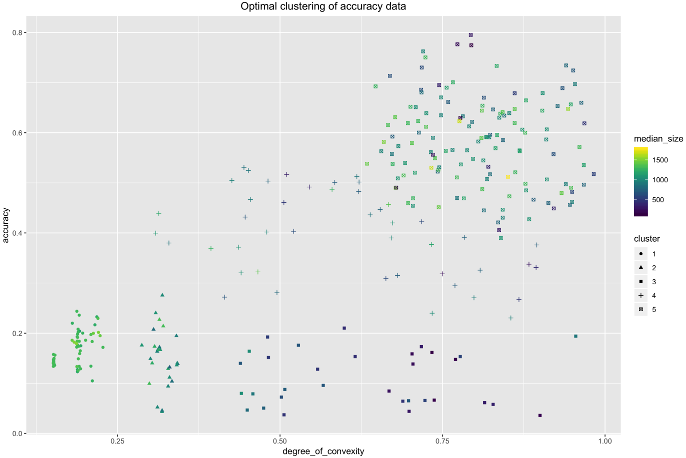
Controlling for Linear Separability

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", in *Cognition*, `https://doi.org/10.1016/j.cognition.2019.104076`.

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

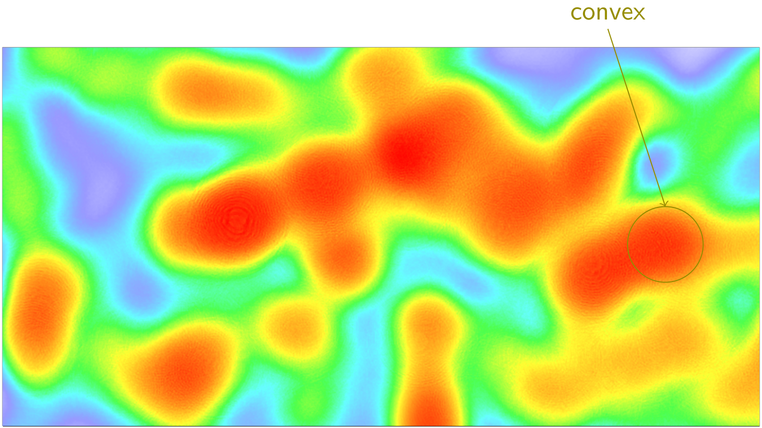
Cluster Analysis



Shane Steinert-Threlkeld and Jakub Szymanik, "Ease of learning explains semantic universals", in *Cognition*, `https://doi.org/10.1016/j.cognition.2019.104076`.

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

Colors: Summary



D_{color}

Overview

Types of Clause-Embedding Predicates

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

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.

Veridicality

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

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

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.

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*.

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.
  $\not\rightsquigarrow$ The canal has 7 bridges.

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.

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.

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.

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*.

The Veridical Uniformity Thesis

Veridical Uniformity Universal

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

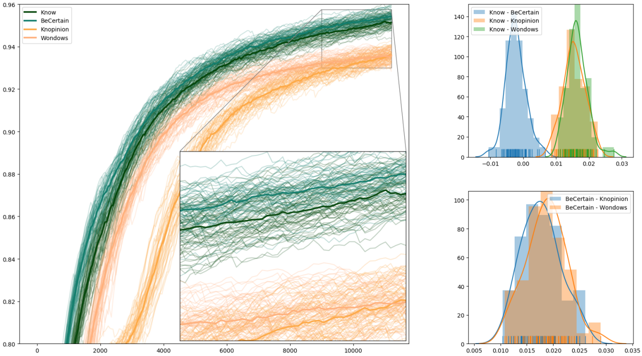
Four Responsive Predicates

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).

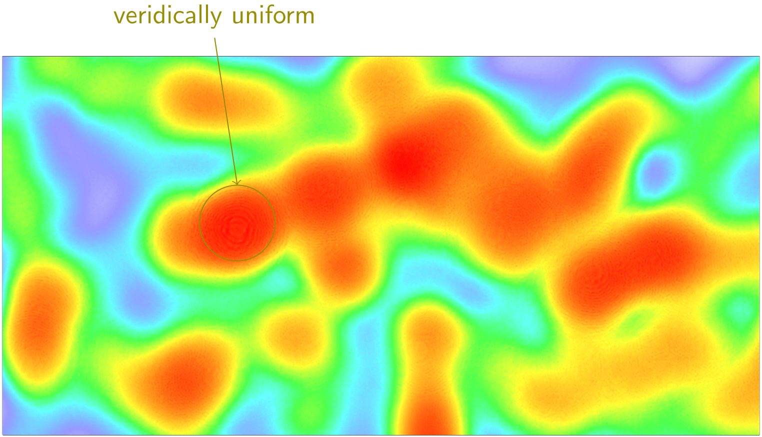
Veridical Uniformity: Results



Shane Steinert-Threlkeld, "An Explanation of the Veridical Uniformity Universal", in *Journal of Semantics*, `https://doi.org/10.1093/jos/ffz019`.

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

Responsive Predicates: Summary



$D_{\text{responsive}}$

Interim Summary

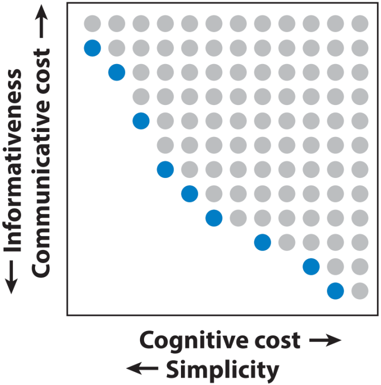
Ease of learning, measured as the speed of convergence of NNs, can explain the presence of linguistic universals in various semantic domains, including both function and content words.

Are there other explanations? If so, how to choose between them?

Overview

Overview

Simplicity/Informativeness Trade-off

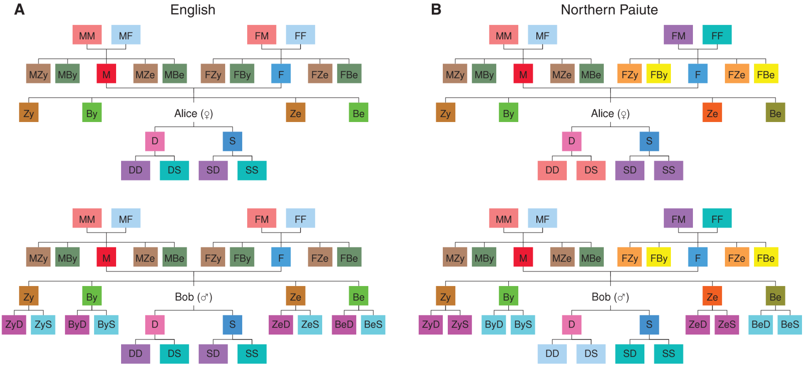


Kemp, Xu, and Regier 2018

Simplicity/Informativeness: Definitions

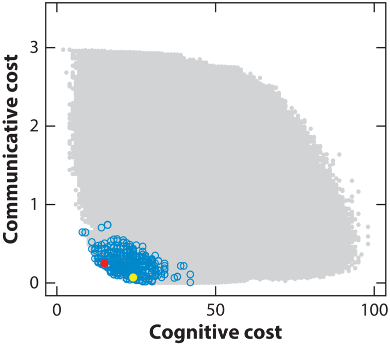
- Cognitive cost:
  Minimal description length in a 'language of thought' (or something similar)
- Communicative cost:
  Ease with which a Sender can convey an intended meaning to a Receiver using the language (Lewis 1969; Skyrms 2010)

Simplicity/Informativeness: Kinship



Kemp and Regier 2012

Simplicity/Informativeness: Kinship



Kemp and Regier 2012

Overview

Complexity of Quantifiers

Boolean	Set-Theoretic	Numeric
$\wedge, \vee, \neg$	$\cap, \cup, \subset, \lvert \cdot \rvert$	$/, +, -, >, =, \%$

Table: The operators in the grammar for generating quantifiers.

Informativeness of Quantifiers

$$I(L) := \sum_{\mathbb{M}} P(\mathbb{M}) \sum_{Q \in L} P(Q|\mathbb{M}) \sum_{\mathbb{M}' \in Q} P(\mathbb{M}'|Q) \cdot u(\mathbb{M}', \mathbb{M})$$

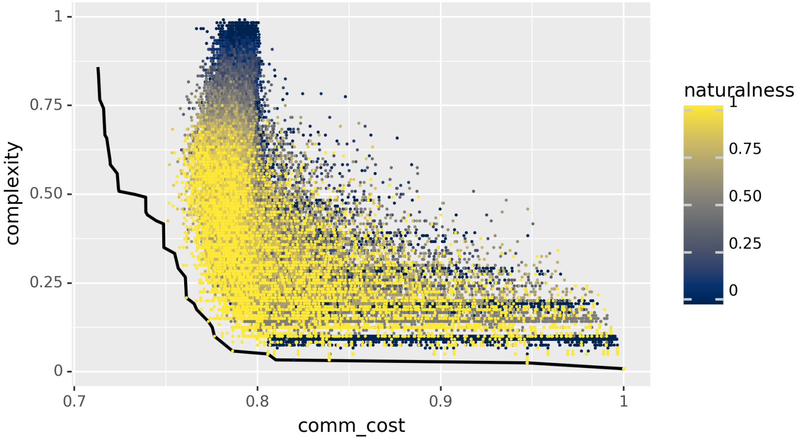
$$u(\mathbb{M}', \mathbb{M}) = \frac{1}{1 + d(\mathbb{M}', \mathbb{M})}$$

$$\text{where} \quad d(\mathbb{M}', \mathbb{M}) = \sum_{X \in A \setminus B, A \cap B, B \setminus A, M \setminus (A \cup B)} \max\{0, |X| - |X'|\}$$

Experiment Set-up

1. For each num of words $n \in \{1, \ldots, 10\}$, generate 8000 languages:
   - Choose $m \leq n$ uniformly at random
   - Sample m quantifiers from *quasi-natural* set (Keenan and Paperno 2012; Paperno and Keenan 2017):
     1. generalized existential
     2. generalized intersective
     3. proportional
   - All w/ minimal formulas w/ ≤ 12 operators
2. For each language, measure *distance to the Pareto frontier*
   - Estimated the true Pareto front using an evolutionary algorithm

Main Results



$$\beta_{naturalness} = -0.3$$

Shane Steinert-Threlkeld, "Quantifiers in natural language optimize the simplicity/informativeness tradeoff", in *Amsterdam Colloquium*

Experiment 2: Degrees of Semantic Universals

Key idea: use information theory to measure *degrees* of the various semantic universals.

Do artificial languages with higher degrees lie closer to the Pareto frontier?

Degree of Monotonicity

$$\mathbb{1}_{\mathsf{Q}}(\mathbb{M}) = 1 \text{ iff } \mathbb{M} \in \mathsf{Q}$$

$$\mathbb{1}^{\preceq}_{\mathsf{Q}}(\mathcal{M}) = 1 \text{ iff } \exists \mathbb{M}' \preceq \mathbb{M} \text{ s.t. } \mathbb{M}' \in \mathsf{Q}$$

Degree of Monotonicity

$$\mathbb{1}_{\mathsf{Q}}(\mathbb{M}) = 1 \text{ iff } \mathbb{M} \in \mathsf{Q}$$
$$\mathbb{1}_{\mathsf{Q}}^{\preceq}(\mathcal{M}) = 1 \text{ iff } \exists \mathbb{M}' \preceq \mathbb{M} \text{ s.t. } \mathbb{M}' \in \mathsf{Q}$$

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

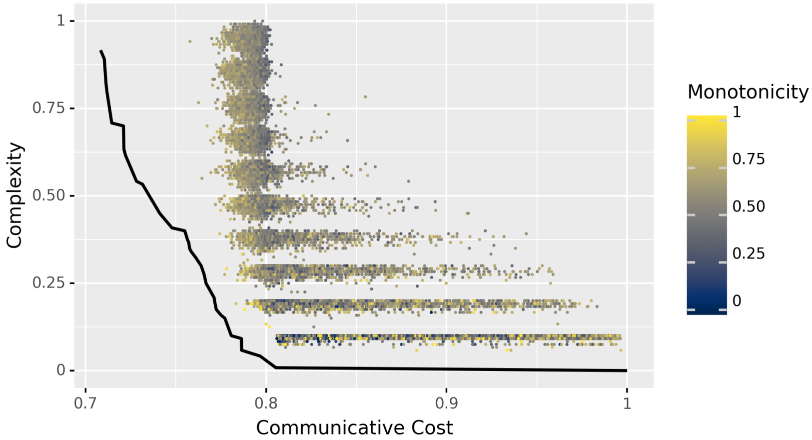
Degree of Monotonicity

$$\mathbb{1}_{\mathsf{Q}}(\mathbb{M}) = 1 \text{ iff } \mathbb{M} \in \mathsf{Q}$$
$$\mathbb{1}_{\mathsf{Q}}^{\preceq}(\mathcal{M}) = 1 \text{ iff } \exists \mathbb{M}' \preceq \mathbb{M} \text{ s.t. } \mathbb{M}' \in \mathsf{Q}$$

$$\begin{aligned}
\mathrm{mon}(\mathsf{Q}) &:= \frac{I(\mathbb{1}_{\mathsf{Q}}; \mathbb{1}_{\mathsf{Q}}^{\preceq})}{H(\mathbb{1}_{\mathsf{Q}})} \\
&= \frac{H(\mathbb{1}_{\mathsf{Q}}) - H(\mathbb{1}_{\mathsf{Q}} | \mathbb{1}_{\mathsf{Q}}^{\preceq})}{H(\mathbb{1}_{\mathsf{Q}})} \\
&= 1 - \frac{H(\mathbb{1}_{\mathsf{Q}} \mid \mathbb{1}_{\mathsf{Q}}^{\preceq})}{H(\mathbb{1}_{\mathsf{Q}})}
\end{aligned}$$

Note: different variables than $\mathbb{1}_{\mathsf{Q}}^{\preceq}$ for other universals. Final degree: average across A/B arguments of maximum of upward/downward degrees.

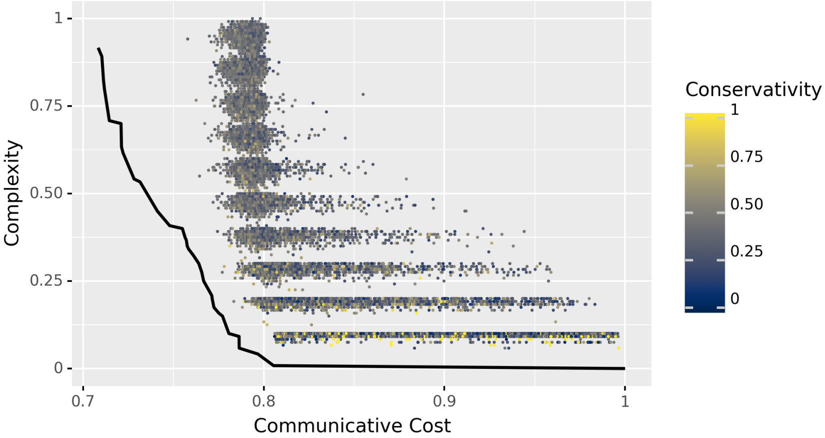
Results: Degree of Monotonicity



$\rho = -0.0590$ (boostrapped CI: $[-0.07460891, -0.04257208]$)

Shane Steinert-Threlkeld, "Quantifiers in Natural Language: Efficient Communication and Degrees of Semantic Universals", under review

Results: Degree of Conservativity



$\rho = 0.0725$ (bootstrapped CI: $[0.0565, 0.0883]$)

Shane Steinert-Threlkeld, "Quantifiers in Natural Language: Efficient Communication and Degrees of Semantic Universals", under review

Quantifiers: Summary

- More natural languages are more optimal
- Among a *random* sample of languages, degrees of monotonicity and conservativity are not strongly correlated with optimality
- Possibly: universals are an epiphenomenon of the more fundamental pressure for communication
- Todo: test with other sampling procedures

Overview

Indefinite Pronouns

Indefinite pronouns in English: *someone, anyone, no one, ...*

Haspelmath 2001

Indefinite Pronouns

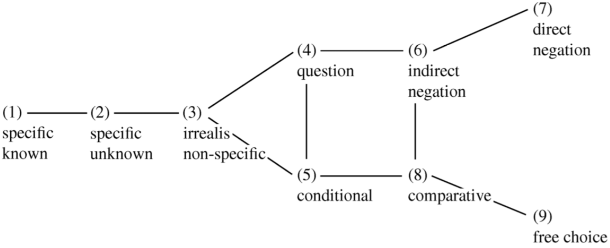
Indefinite pronouns in English: *someone, anyone, no one, ...*

Why indefinite pronouns?

1. Rare domain of *function words* for which rich cross-linguistic data-set is available.
2. There are numerous *semantic universals* in this domain that are to be explained.

Haspelmath 2001

Universals for Indefinites



Haspelmath 2001

Meaning Space

(1) **Specific known flavor** [the indefinite pronoun refers to a specific individual that the interlocutors can uniquely identify]:
Someone managed to mess this up — we all know who!

(2) **Specific unknown flavor** [the indefinite pronoun refers to a specific individual that the interlocutors cannot uniquely identify]:
I heard that *someone* failed, but I don't know who.

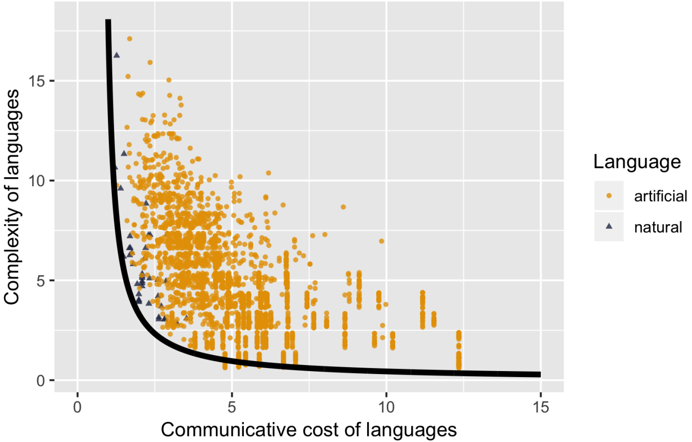
(3) **Non-specific flavor** [the indefinite pronoun is interpreted as an existential quantifier over some domain of possible referents, not referring to a specific individual]:
You should probably talk to *someone* else about this too.

(4) **Negative polarity flavor** [the indefinite pronoun is interpreted as an existential quantifier over a widened domain of possible referents]:
Less than three companies hired *anyone* this year.

(5) **Free choice flavor** [the indefinite pronoun is interpreted as a wide-scope universal quantifier over some domain of possible referents]:
You can hire almost *anyone* here: most of them great.

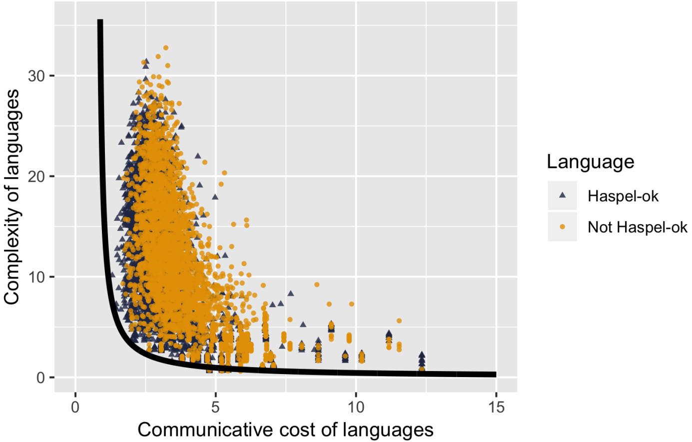
(6) **Negative indefinite flavor** [the indefinite pronoun is interpreted as a negated existential quantifier over some domain of possible referents]:
Who went to the party? *No one*.

Main Results



Milica Denić, Shane Steinert-Threlkeld, Jakub Szymanik "Complexity/informativeness tradeoff in the domain of indefinite pronouns", in *Proceedings of SALT*

Results: Experiment 2



Milica Denić, Shane Steinert-Threlkeld, Jakub Szymanik "Complexity/informativeness tradeoff in the domain of indefinite pronouns", in *Proceedings of SALT*

Overview

Modals: Universals

The target: expressions like many English auxiliaries (*may*, *must*, *can*, ...) used to express the relation of a clause to non-actual worlds.

Modals can be ambiguous in their *flavor*:

	epistemic	deontic	⋯
weak	✓	✓	
strong			

Table: English *may*

Modals: Universals

The target: expressions like many English auxiliaries (*may*, *must*, *can*, . . .) used to express the relation of a clause to non-actual worlds.

Modals can be ambiguous in their *flavor*:

	epistemic	deontic	· · ·
weak	✓	✓	
strong			

Table: English *may*

Modals can be ambiguous along their *force*:

	epistemic	deontic	· · ·
weak		✓	
strong		✓	

Table: St'át'imcets *-ka* (Rullmann, Matthewson, and Davis 2008)

Modals: Universals

On the basis of a detailed study of 6 typologically unrelated languages, Nauze (2008) proposes:

Single Axis of Variability

A modal may be ambiguous in either force or flavor, but *not both*.

Modals: Universals

On the basis of a detailed study of 6 typologically unrelated languages, Nauze (2008) proposes:

Single Axis of Variability

A modal may be ambiguous in either force or flavor, but *not both*.

This rules out hypothetical modals like *mought*:

	epistemic	deontic	···
weak	✓		
strong		✓	

Table: Hypothetical *mought*

Modals: Universals

On the basis of a detailed study of 6 typologically unrelated languages, Nauze (2008) proposes:

Single Axis of Variability

A modal may be ambiguous in either force or flavor, but *not both*.

This rules out hypothetical modals like *mought*:

	epistemic	deontic	$\cdots$
weak	✓		
strong		✓	

Table: Hypothetical *mought*

Vander Klok (2013), as reported by Matthewson (2016): within the root/epistemic domain, a modal *system* may have elements ambiguous along one or the other dimension, but not some modals ambiguous in one and some the other.

Complexity of Modals

Language of Thought: basic propositional language, with atoms for each force and for each flavor.

Shortest formula: write DNF for a modal meaning, apply algorithm generalized from Feldman (2001) to minimize.

Complexity of Modals

Language of Thought: basic propositional language, with atoms for each force and for each flavor.

Shortest formula: write DNF for a modal meaning, apply algorithm generalized from Feldman (2001) to minimize.

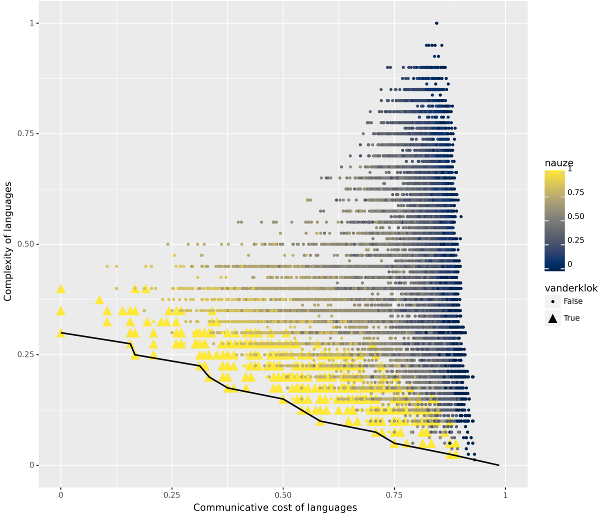
Modal	Meaning representation					Shortest Formula in LOT	Complexity (# of at
		e	d	c	t		
may	$\exists$	✓	✓			$\exists \wedge (e \vee d)$	3
	$\forall$						
		e	d	c	t		
mought	$\exists$	✓				$(\exists \wedge e) \vee (\forall \wedge d)$	4
	$\forall$		✓				
		e	d	c	t		
notcirc	$\exists$	✓	✓		✓	$\overline{c}$	1
	$\forall$	✓	✓		✓		

Informativeness of Modals

$$I(L) := \sum_{\mathbb{M}} P(\mathbb{M}) \sum_{m \in L} P(m|\mathbb{M}) \sum_{\mathbb{M}' \in m} P(\mathbb{M}'|m) \cdot u(\mathbb{M}', \mathbb{M})$$

$$u(\mathbb{M}', \mathbb{M}) = 0.5 \cdot \mathbb{1}_{\text{force}(\mathbb{M})=\text{force}(\mathbb{M}')} + 0.5 \cdot \mathbb{1}_{\text{flavor}(\mathbb{M})=\text{flavor}(\mathbb{M}')}$$

Modals: Main Results



Nathaniel Imel and Shane Steinert-Threlkeld, "Modals in natural langauge optimize the simplicity/informativeness trade-off", forthcoming in *Proceedings of SALT*

Overview

Explaining Universals

Why do semantic universals arise?

(I) Because expressions satisfying them are easier to learn.

Explaining Universals

Why do semantic universals arise?

(I) Because expressions satisfying them are easier to learn.
(II) Because languages optimize a trade-off between simplicity and informativeness.

Explaining Universals

Why do semantic universals arise?

(I) Because expressions satisfying them are easier to learn.
(II) Because languages optimize a trade-off between simplicity and informativeness.

General questions:

- Are these explanations in competition with each other?
- How can we adjudicate between them?
  (Especially in the presence of other pressures exerting their influence on linguistic structure)
- Tangentially related: can these inform model / dataset building for plausible biases in NLP systems?

One idea: using tools from language evolution: does one, but not the other, increase as languages evolve (in simulation, and in the lab)?

Extensions

- Does CONS arise from a biased linguistic distribution?
  Mhasawade et al. 2018: NO
- Generalizing the learnability experiments
- Iterated learning with neural agents produces *monotone* quantifiers
  Carcassi, Steinert-Threlkeld, and Szymanik 2021
- Do the degrees of universals aid learnability in a Bayesian setting?
  [D. Johnson CLMS thesis suggests not]
- More domains
  cf Uegaki; Enguehard and Spector on logical vocabulary
- Better cross-linguistic data (forthcoming modal database)
- Open source tool for doing efficiency analyses (ALTK)
- Information-bottleneck analyses in these domains / full comparison thereof

Other Things We're Working On

Emergent communication:

- Under what conditions do artificial agents learn to speak human-like languages (e.g. compositional, with functional vocabulary)?
- Using these tools for real NLP tasks: ongoing work on *unsupervised machine translation*.

Interpretability / analysis:

- LMs use monotonicity to assess NPI licensing (Jumelet et al 2021; Lapastora et al ongoing)
- Some parts-of-speech (but not others) are represented similarly cross-linguistically in multilingual models (Shapiro et al 2021)
- Representations of semantically similar tokens are more similar cross-linguistically [Shivin Thukral; ongoing]

Other Things We're Working On

Multilingual human/machine processing:

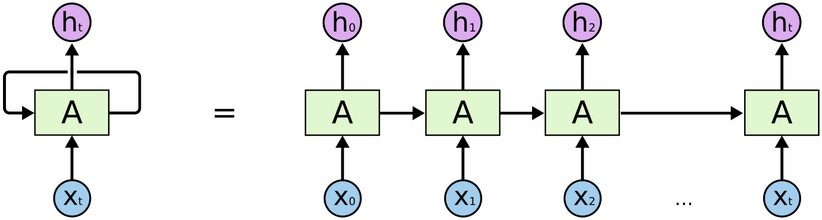
- Bilingual alignment transfers to multilingual alignment for unsupervised bitext mining [Tien and S-T 2022]
- Masked *segmental* language modeling [Downey et al 2021; 2022]
- Learning to translate by learning to communicate (i.e. EC fine-tuning of multilingual pretrained models) [S-T et al 2022; + active / ongoing]
- Artificial language learning at scale: which linguistic features are easiest to learn (as a function of native languages) [Shapiro, ongoing]

The End

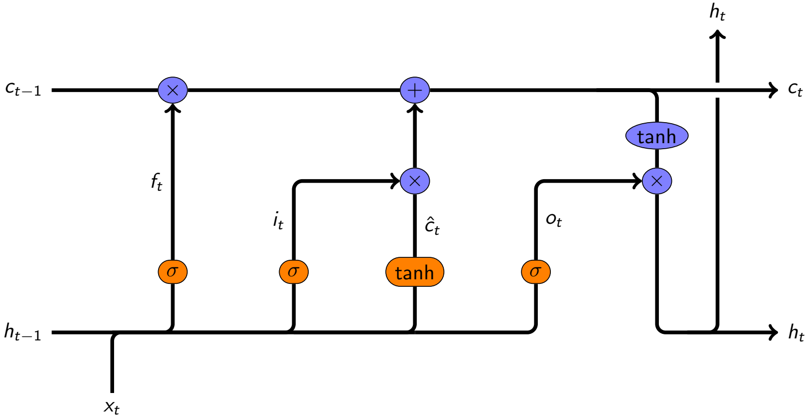
Thank you! Thoughts?

Overview

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: ith 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.

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. $\begin{bmatrix}0 & 0 & 1\end{bmatrix}$)

Overview

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

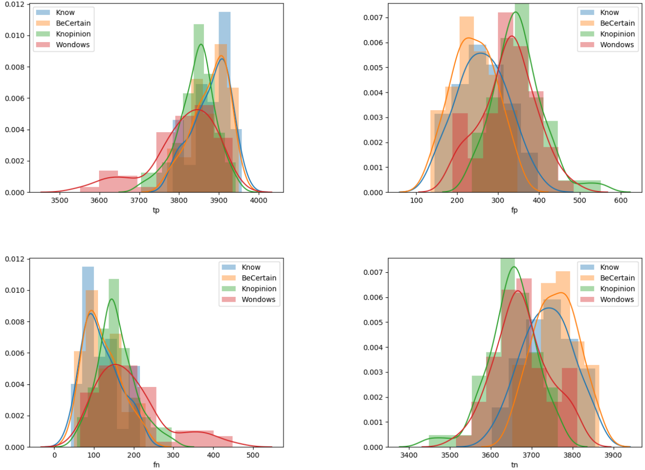


Figure: Distributions (Gaussian kernel density estimates) of the true/false positives/negatives by verb.

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
$w \in \mathrm{DOX}^a_w$	1	0.964	0.957	0.954	0.947
	0	0.919	0.953	0.887	0.924
$\mathrm{DOX}^a_w \in P$	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.

Overview

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} / ∑_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
```

Overview

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.
- Benthem, Johan van (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. 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.

References II

- Gärdenfors, Peter (2014). *The Geometry of Meaning*. The MIT Press.
- 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.
- 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 et al., pp. 11–20. DOI: 10.1007/978-3-642-14287-1_2.
- Keenan, Edward L and Denis Paperno, eds. (2012). *Handbook of Quantifiers in Natural Language*. Vol. 90. Studies in Linguistics and Philosophy. Dordrecht: Springer. DOI: 10.1007/978-94-007-2681-9.
- 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.

References III

- Kemp, Charles and Terry Regier (2012). "Kinship categories across languages reflect general communicative principles.". In: *Science* 336.6084, pp. 1049–1054. DOI: 10.1126/science.1218811.
- Kemp, Charles, Yang Xu, and Terry Regier (2018). "Semantic typology and efficient communication". In: *Annual Review of Linguistics*, pp. 1–23. DOI: 10.1146/annurev-linguistics-011817-045406.
- 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?" URL: https://arxiv.org/pdf/1809.05733.pdf.
- Paperno, Denis and Edward L Keenan, eds. (2017). *Handbook of Quantifiers in Natural Language: Volume II*. Vol. 97. Studies in Linguistics and Philosophy. Springer. DOI: 10.1007/978-3-319-44330-0.

References IV

- Peters, Stanley and Dag Westerståhl (2006). *Quantifiers in Language and Logic*. Oxford: Clarendon Press.
- 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: URL: http://colala.berkeley.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.

References V

- 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_.
- 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 (2019). "An Explanation of the Veridical Uniformity Universal". In: *Journal of Semantics*. URL: https://semanticsarchive.net/Archive/DI5ZTNmN/UniversalResponsiveVerbs.pdf.

References VI

- Steinert-Threlkeld, Shane and Jakub Szymanik (2019a). "Ease of Learning Explains Semantic Universals". URL: https://semanticsarchive.net/Archive/zM5ZGIxM/EaseLearning.pdf.
- – (2019b). "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.
- Theiler, Nadine, Floris Roelofsen, and Maria Aloni (2018). "A uniform semantics for declarative and interrogative complements". In: *Journal of Semantics* 35.3, pp. 409–466. 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.