Explaining semantic universals: Session 2

Milica Denić

UNIVERSITÉ DE GENÈVE

Question

What explains which meanings are lexicalized across languages?

Previous lecture: optimizing the simplicity/informativeness trade-off

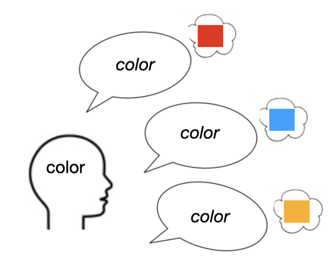
The simplicity/informativeness trade-off hypothesis

Languages are subject to two **competing pressures**:

The simplicity/informativeness trade-off hypothesis

Languages are subject to two **competing pressures**:

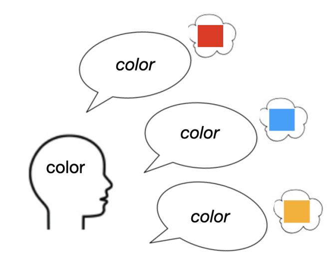
maximize **simplicity**
(minimize **complexity**)

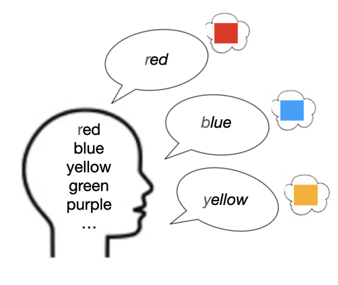


The simplicity/informativeness trade-off hypothesis

Languages are subject to two **competing pressures**:

maximize **simplicity** (minimize **complexity**)	maximize **informativeness** (minimize **communicative cost**)



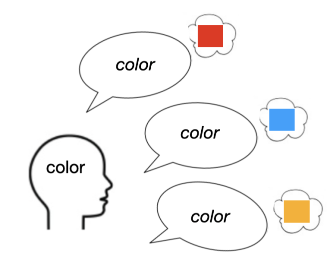


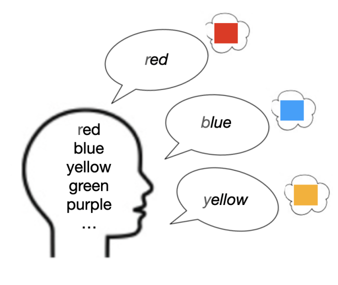
The simplicity/informativeness trade-off hypothesis

Languages are subject to two **competing pressures**:

maximize **simplicity**
(minimize **complexity**)

maximize **informativeness**
(minimize **communicative cost**)





Languages **optimize the trade-off** between these two pressures.

Simplicity/informativeness trade-off optimization: Successes

We discussed (*simplified versions of*) two case studies:

- Connectives: explaining item-level universals (Horn, 1972; Katzir and Singh, 2013; Uegaki, 2022)
- Kinship terms: explaining system-level universals (Kemp and Regier, 2012)

Complexity and informativeness of a language (within a semantic category)

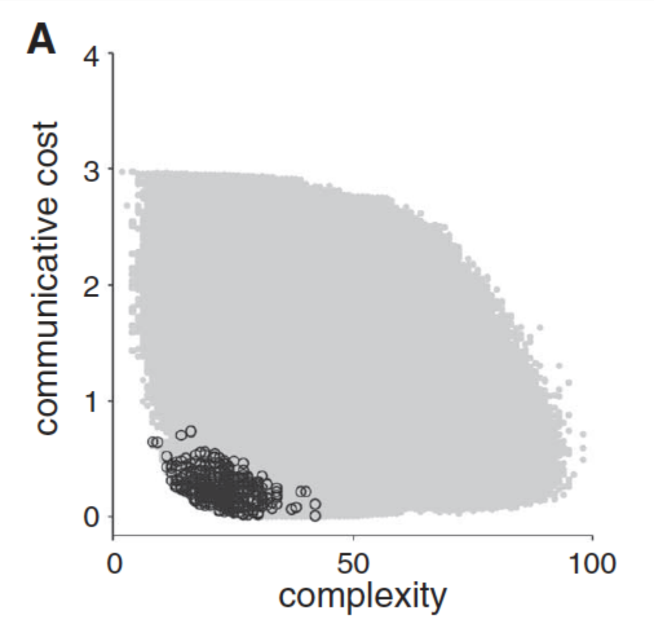
- **Complexity**: number of lexical items

Complexity and informativeness of a language (within a semantic category)

- **Complexity**: number of lexical items
- **Informativeness**: the probability that the speaker and listener successfully communicate

Example : Kinship terms (Kemp and Regier, 2012)

Example : Kinship terms (Kemp and Regier, 2012)



But what about morphosyntax?

But what about morphosyntax?

When measuring informativeness, a simplifying assumption was made: we communicate only using **single (monomorphemic) words**!

But what about morphosyntax?

When measuring informativeness, a simplifying assumption was made: we communicate only using **single (monomorphemic) words**!

But what about morphosyntax?

When measuring informativeness, a simplifying assumption was made: we communicate only using **single (monomorphemic) words**!

But the speaker could be more specific and say *paternal grandmother*!

Grandmother vs. paternal grandmother

Simplicity and informativeness not the (only) pressures being optimized?

Pressures shaping lexicons:

The role of morphosyntactic complexity of utterances

Denić and Szymanik (2024), *Cognitive Science*

Due to productive morphosyntax and compositional semantics, lexicon size doesn't always compete with informativeness.

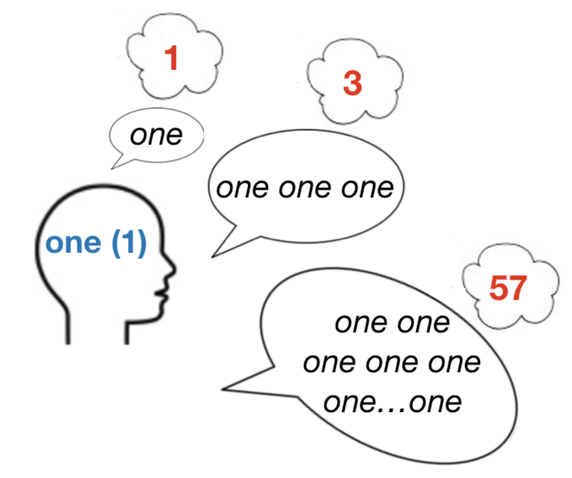
Due to productive morphosyntax and compositional semantics, lexicon size doesn't always compete with informativeness.

Number	English	Fulfulde
6	*six* (6)	*jowe-* (5) *-e-* (+) *-go* (1)
7	*seven* (7)	*jowe-* (5) *-e-* (+) *-didi* (2)
8	*eight* (8)	*jowe-* (5) *-e-* (+) *-tati* (3)
9	*nine* (9)	*jowe-* (5) *-e-* (+) *-nayi* (4)

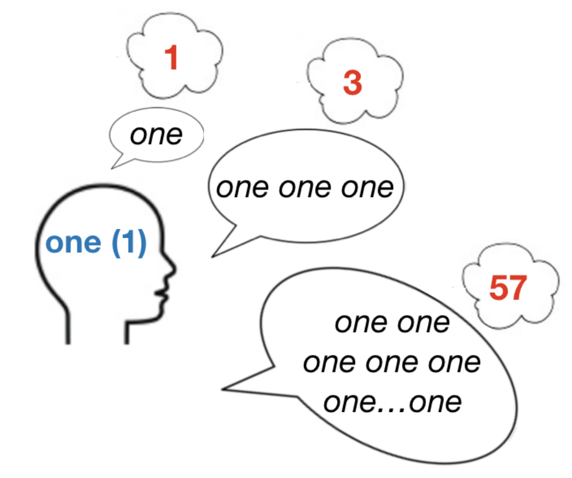
Recursive numeral systems are maximally informative.

Despite a small lexicon, any natural number can be referred to precisely.

Simplicity of lexicon/informativeness trade-off optimization makes wrong predictions for recursive numeral systems.



Simplicity of lexicon/informativeness trade-off optimization makes wrong predictions for recursive numeral systems.



Shortcoming: There is no **utterance complexity** in the model!

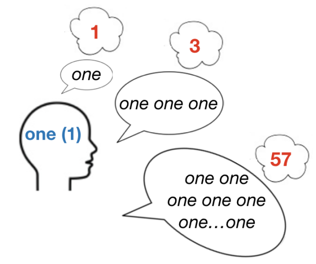
Simplicity of lexicon/simplicity of utterances trade-off

Denić and Szymanik (2024), *Cognitive Science*

Simplicity of lexicon/simplicity of utterances trade-off

Languages' number lexicons are under two **competing pressures**:

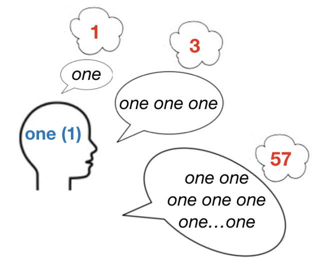
(a) minimize **lexicon size**



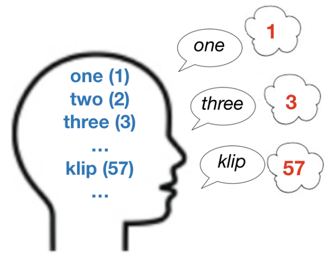
Simplicity of lexicon/simplicity of utterances trade-off

Languages' number lexicons are under two **competing pressures**:

(a) minimize **lexicon size**



(b) minimize **morphosyntactic complexity of utterances**

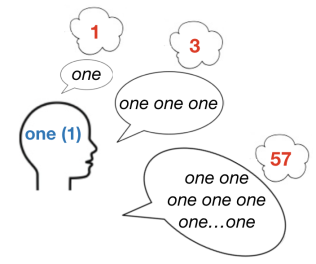


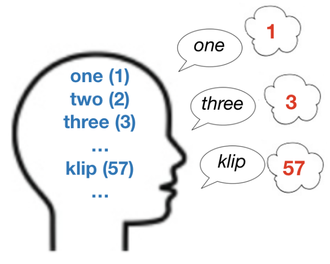
Simplicity of lexicon/simplicity of utterances trade-off

Languages' number lexicons are under two **competing pressures**:

(a) minimize **lexicon size**

(b) minimize **morphosyntactic complexity of utterances**





Languages **optimize the trade-off** between these two pressures.

Denić and Szymanik (2024), *Cognitive Science*

Natural language data

- 128 **recursive** numeral systems from *WALS* Numeral Bases chapter: genealogically and areally diverse language sample (Comrie, 2013)
- For each language, morphemes in each numeral for numbers 1-99 are identified.

128 languages differ along multiple dimensions.

- Which numbers are lexicalized:

128 languages differ along multiple dimensions.

- Which numbers are lexicalized:
  - Fulfulde: 1, 2, 3, 4, 5, 10
  - Khanty: 1, 2, 3, 4, 5, 6, 7, 8, 10, 20

128 languages differ along multiple dimensions.

- Which numbers are lexicalized:
  - Fulfulde: 1, 2, 3, 4, 5, 10
  - Khanty: 1, 2, 3, 4, 5, 6, 7, 8, 10, 20
- How morphosyntactically complex numerals are constructed:

128 languages differ along multiple dimensions.

- Which numbers are lexicalized:
  - Fulfulde: 1, 2, 3, 4, 5, 10
  - Khanty: 1, 2, 3, 4, 5, 6, 7, 8, 10, 20
- How morphosyntactically complex numerals are constructed:
  - Fulfulde: $9 = 5 + 4$
  - Khanty: $9 = 10 - 1$

There are **40 different types of numeral systems** among the 128 natural languages examined.

Measuring lexicon size

Lexicon size = **number of lexicalized numbers** in the range 1-99

Measuring lexicon size

Lexicon size = **number of lexicalized numbers** in the range 1-99

Example:

Fulfulde lexicalizes 1, 2, 3, 4, 5, 10 → lexicon size = 6

Measuring lexicon size

Lexicon size = **number of lexicalized numbers** in the range 1-99

Example:

Fulfulde lexicalizes 1, 2, 3, 4, 5, 10 → lexicon size = 6

Note: multiple morphemes for the same concept are not counted multiple times (e.g., *ten*, *-ty*, *-teen* in English). We'll come back to this!

Measuring morphosyntactic complexity of an utterance

Number	English	Fulfulde
6	*six* (6)	*jowe-* (5) *-e-* (+) *-go* (1)
7	*seven* (7)	*jowe-* (5) *-e-* (+) *-didi* (2)
8	*eight* (8)	*jowe-* (5) *-e-* (+) *-tati* (3)
9	*nine* (9)	*jowe-* (5) *-e-* (+) *-nayi* (4)

↓ **1 morpheme**

↓ **3 morphemes**

Measuring average morphosyntactic complexity

$$average\_morpho\_c(L) = \sum_{n=1}^{99} p(n) \cdot number\_of\_morphemes(L\text{'s numeral for } n)$$

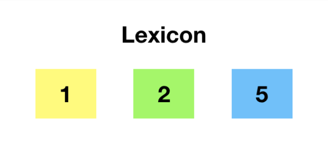
Example: English

$$average\_morpho\_c(L) = \sum_{n=1}^{99} p(n) \cdot number\_of\_morphemes(L\text{'s numeral for } n)$$

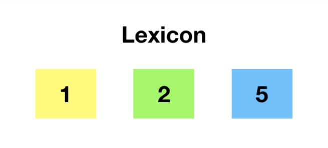
$$average\_morpho\_c(English) =$$
$$p(1) number\_of\_morphemes(one) + \cdots + p(99) number\_of\_morphemes(ninety\ nine)$$

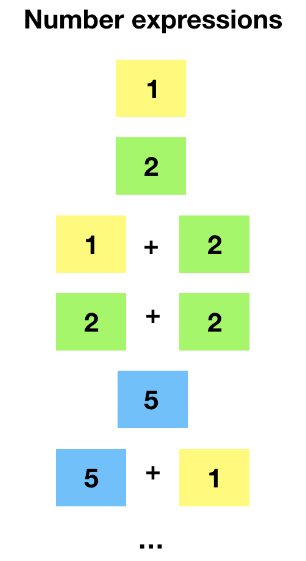
We need to determine which numeral systems are optimal among theoretically possible numeral systems.

We need to determine which numeral systems are optimal among theoretically possible numeral systems.



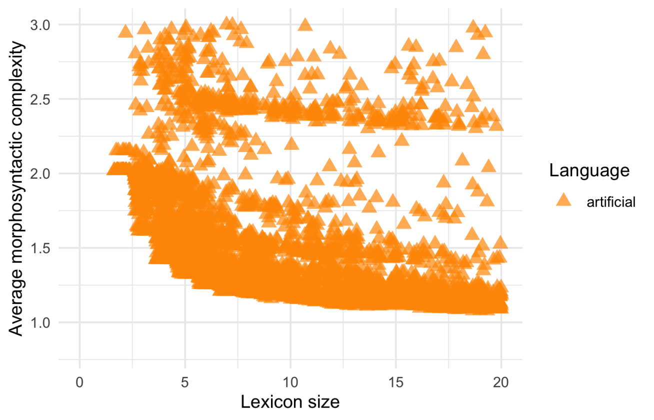
We need to determine which numeral systems are optimal among theoretically possible numeral systems.



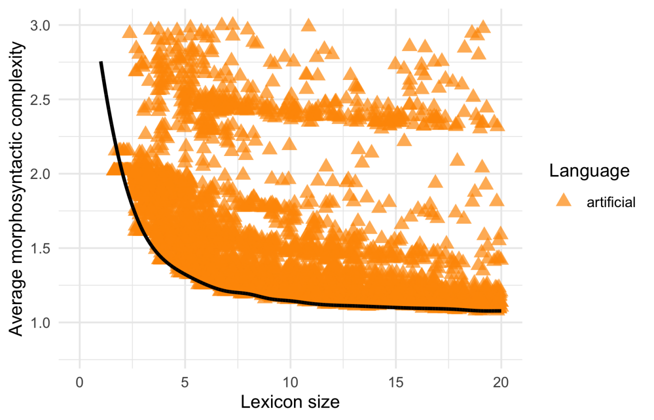


We need to determine which numeral systems are optimal among theoretically possible numeral systems.

We need to determine which numeral systems are optimal among theoretically possible numeral systems.

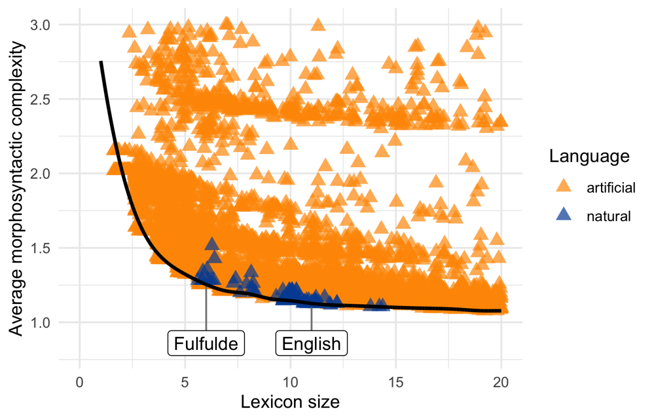


Pareto frontier is a set of Pareto-optimal numeral systems.



Pareto-optimal numeral systems cannot be improved on one dimension without becoming worse on the other.

Natural languages are (near-)optimal solutions to the pressures to minimize lexicon size and complexity of utterances.



Natural languages $N = 128$

Pressures shaping lexicons: Interim summary

Not two, but (at least) three pressures:

1. **minimize lexicon size**
2. **maximize informativeness**
3. **minimize morphosyntactic complexity of utterances**

Going back to other domains?

Going back to other domains?

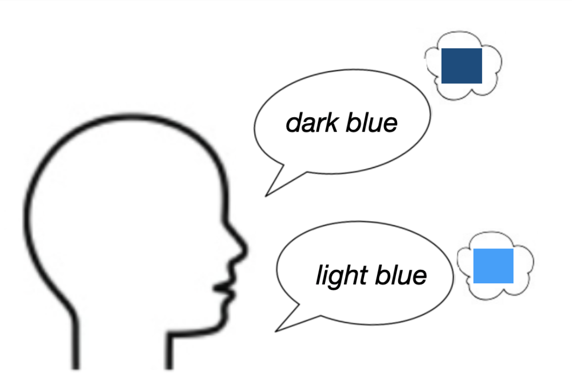
For related ideas for **connectives**, see Carcassi and Sbardolini (2023).

Pressures shaping lexicons: Interim summary

Recursive numeral systems are an extreme case re: informativeness (maximal), but the three pressures may interact elsewhere!

Pressures shaping lexicons: Interim summary

Recursive numeral systems are an extreme case re: informativeness (maximal), but the three pressures may interact elsewhere!



Future directions

What we have done here

What we have done here

- We started with the simplicity/informativeness trade-off hypothesis.

What we have done here

- We started with the simplicity/informativeness trade-off hypothesis.
- We found one of its limitations.

What we have done here

- We started with the simplicity/informativeness trade-off hypothesis.
- We found one of its limitations.
- This lead us to new discoveries about pressures shaping lexicons.

What we have done here

- We started with the simplicity/informativeness trade-off hypothesis.
- We found one of its limitations.
- This lead us to new discoveries about pressures shaping lexicons.

What else remains to be explained?

What else remains to be explained? Numerals

No morpheme denotes the division operator in any numeral.
(in our language sample)

What else remains to be explained? Numerals

No morpheme denotes the division operator in any numeral.
(in our language sample)

Example: 50 as $5 \cdot 10$ rather than $100/2$.

What else remains to be explained? Numerals

No morpheme denotes the division operator in any numeral.
(in our language sample)

Example: 50 as $5 \cdot 10$ rather than $100/2$.

Division is marked/more complex to represent (cognitive bias in language processing)?

What else remains to be explained? Numerals

No morpheme denotes the division operator in any numeral.
(in our language sample)

Example: 50 as $5 \cdot 10$ rather than $100/2$.

Division is marked/more complex to represent (cognitive bias in language processing)?

Curiosity: *half* is sometimes used! (Hurford, 1975)

What else remains to be explained? Numerals

Serbe

tri : 3, *deset* : 10, *trideset* : 30

Anglais

three : 3, *ten* : 10, *thirty* : 30

How can its existence be reconciled with the idea that languages aim to minimize the size of the lexicon?

What else remains to be explained? Numerals

There are optimal numeral systems with bases other than (multiples) of 5. Why are these not attested among human languages?

What we have done here

- We started with the simplicity/informativeness trade-off hypothesis.
- We found one of its limitations.
- This lead us to new discoveries about pressures shaping lexicons.

What else remains to be explained?

What we have done here

- We started with the simplicity/informativeness trade-off hypothesis.
- We found one of its limitations.
- This lead us to new discoveries about pressures shaping lexicons.

What else remains to be explained?

Next time:
We will do a parallel exercise for polarity items, and see where it leads us.

References

Bowler, M. (2014). Conjunction and disjunction in a language without'and'. In Semantics and linguistic theory, Volume 24, pp. 137–155.

Buccola, B., M. Križ, and E. Chemla (2022). Conceptual alternatives: Competition in language and beyond. Linguistics and Philosophy 45(2), 265–291.

Carcassi, F. and G. Sbardolini (2023). Assertion, denial, and the evolution of boolean operators. Mind & Language 38(5), 1187–1207.

Comrie, B. (2013). Numeral bases. In M. S. Dryer and M. Haspelmath (Eds.), The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Denić, M. (2023). Probabilities and logic in implicature computation: Two puzzles with embedded disjunction. Semantics and Pragmatics 16, 4–EA.

Denić, M. and E. Chemla (2020). Quantifier spreading in child language as distributive inferences. Linguistic Inquiry 51(1), 141–153.

Denić, M., E. Chemla, and L. Tieu (2018). Intervention effects in npi licensing: A quantitative assessment of the scalar implicature explanation. Glossa, 1–27.

Denić, M., V. Homer, D. Rothschild, and E. Chemla (2021). The influence of polarity items on inferential judgments. Cognition 215, 104791.

Denić, M. and Y. Sudo (2022). Donkey anaphora in non-monotonic environments. Journal of Semantics 39(3), 443–474.

Denić, M. and J. Szymanik (2022). Are most and more than half truth-conditionally equivalent? Journal of Semantics 39(2), 261–294.

Denić, M. and J. Szymanik (2024). Recursive numeral systems optimize the trade-off between lexicon size and average morphosyntactic complexity. Cognitive Science.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: most versus more than half. Natural Language Semantics 17(1), 63–98.

Horn, L. R. (1972). On the semantic properties of logical operators in English. Ph. D. thesis, University of California, Los Angeles.

Hurford, J. R. (1975). The linguistic theory of numerals, Volume 16. Cambridge University Press.

Hurford, J. R. (2007). A performed practice explains a linguistic universal: Counting gives the packing strategy. Lingua 117(5), 773–783.

Katzir, R. and R. Singh (2013). Constraints on the lexicalization of logical operators. Linguistics and Philosophy 36, 1–29.

Kemp, C. and T. Regier (2012). Kinship categories across languages reflect general communicative principles. Science 336(6084), 1049–1054.

Krifka, M. (2007). Approximate interpretation of number words. Humboldt-Universität zu Berlin, Philosophische Fakultät II.

Mascarenhas, S. (2014). Formal Semantics and the Psychology of Reasoning Building new bridges and investigating interactions. Ph. D. thesis, New York University.

Piantadosi, S. T., J. B. Tenenbaum, and N. D. Goodman (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. Cognition 123(2), 199–217.

References v

Uegaki, W. (2022). NAND and the communicative efficiency model. Unpublished manuscript, University of Edinburgh.

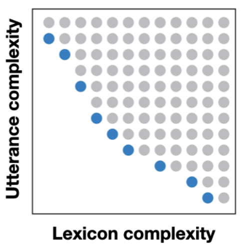
Werner, H. (1997). Die ketische Sprache, Volume 3. Otto Harrassowitz Verlag.

Xu, Y., E. Liu, and T. Regier (2020). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. Open Mind 4, 57–70.

Semantic representations: Evaluating competing hypotheses

Primitives:

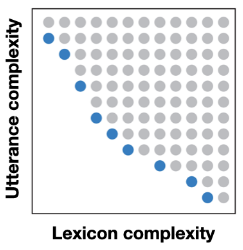
$+, \cdot,$ inverse



Semantic representations: Evaluating competing hypotheses

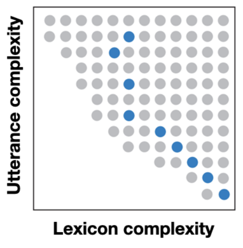
Primitives:

$+, \cdot,$ inverse



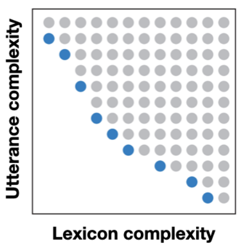
Primitives:

$+, \cdot, -,$ inverse

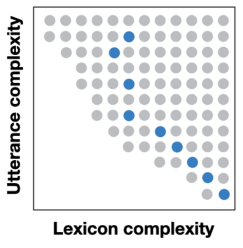


Semantic representations: Evaluating competing hypotheses

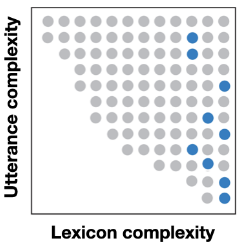
Primitives:
$+, \cdot,$ inverse



Primitives:
$+, \cdot, -,$ inverse



Primitives:
$+, \cdot, -, /,$ inverse



The availability of approximate interpretation is correlated with the morphosyntactic complexity of the numeral.

Example: The numeral for 50 more frequent in Norwegian than in Danish.

Krifka (2007)

Morphosyntax of numerals

Number-denoting morphemes can be digits (D) or bases (in my language sample, max. 5 bases).

NUMBER $\longrightarrow$ D
NUMBER $\longrightarrow$ PHRASE
NUMBER $\longrightarrow$ PHRASE + NUMBER | PHRASE - NUMBER
PHRASE $\longrightarrow$ BASE
PHRASE $\longrightarrow$ NUMBER $\cdot$ BASE

Hurford (2007, 1975)

Morphosyntax of numerals

Number-denoting morphemes can be digits (D) or bases (in my language sample, max. 5 bases).

NUMBER $\longrightarrow$ D
NUMBER $\longrightarrow$ PHRASE
NUMBER $\longrightarrow$ PHRASE + NUMBER | PHRASE - NUMBER
PHRASE $\longrightarrow$ BASE
PHRASE $\longrightarrow$ NUMBER $\cdot$ BASE

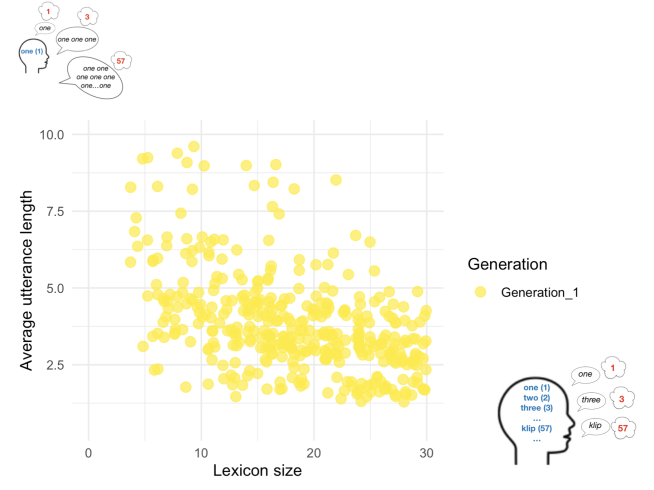
Example:
$D = \{1,2,3,4,5,6,7,8,9\}$, $M = \{10\}$
$40 = 4 \cdot 10$ and not $5 \cdot 8$

Hurford (2007, 1975)

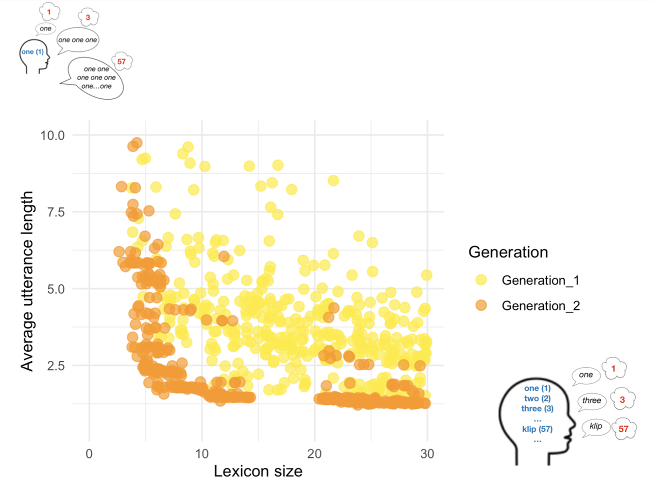
Why seek to reverse-engineer semantic representations? Intersection of semantics and cognitive science.

1. Lexical acquisition (e.g., Piantadosi et al., 2012)
2. Interface between language and reasoning:
    - Successes and failures (e.g., Mascarenhas, 2014)
    - Problem solving (e.g., verification, cf. Hackl, 2009)
3. Pragmatics (e.g., Buccola et al., 2022)

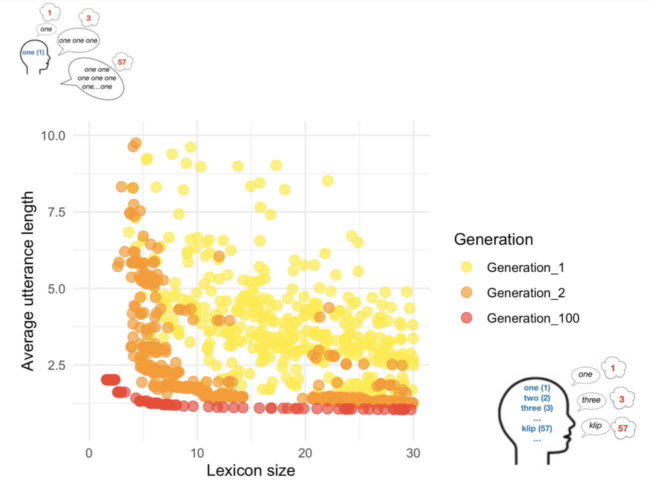
Simulating evolution: Generation 1 of artificial numeral systems



Simulating evolution: Generation 2 of artificial numeral systems



Simulating evolution: Generation 100 of artificial numeral systems



Measuring informativeness

Measure of informativeness of a language based on how likely S and L are to communicate successfully:

$$I(L) = \sum_{m \in M} \sum_{w \in L} P(m) P_S(w|m) P_L(m|w)$$

Complexity/informativeness approach to numeral systems

- Xu et al. (2020) investigate jointly recursive and restricted systems and argue that they optimize complexity/informativeness trade-off.

Complexity/informativeness approach to numeral systems

- Xu et al. (2020) investigate jointly recursive and restricted systems and argue that they optimize complexity/informativeness trade-off.
- Informativeness as before: average communication success

Complexity/informativeness approach to numeral systems

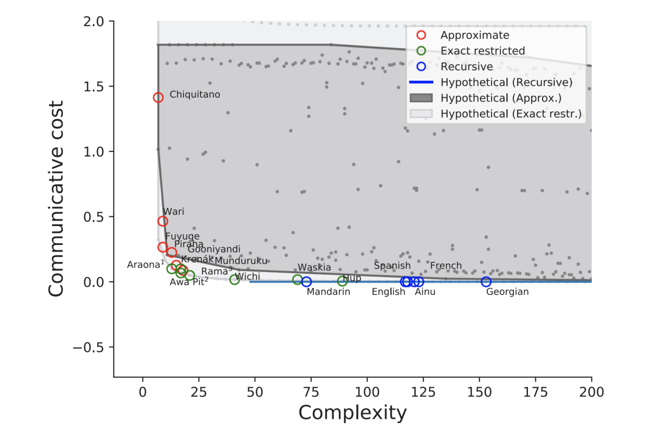
- Xu et al. (2020) investigate jointly recursive and restricted systems and argue that they optimize complexity/informativeness trade-off.
- Informativeness as before: average communication success
- Complexity = lexicon size + grammatical rules

Complexity/informativeness approach to numeral systems

- Xu et al. (2020) investigate jointly recursive and restricted systems and argue that they optimize complexity/informativeness trade-off.
- Informativeness as before: average communication success
- Complexity = lexicon size + grammatical rules

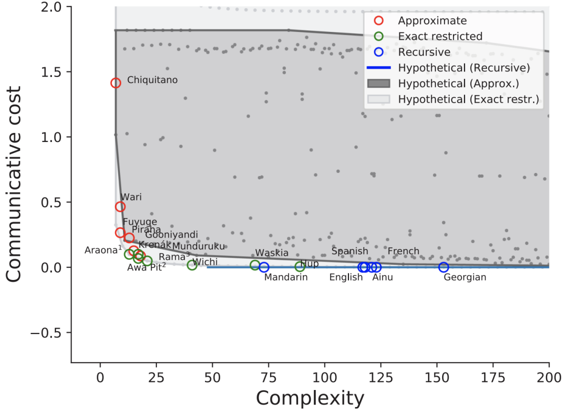
Complexity/informativeness approach to numeral systems

- Xu et al. (2020) investigate jointly recursive and restricted systems and argue that they optimize complexity/informativeness trade-off.
- Informativeness as before: average communication success
- Complexity = lexicon size + grammatical rules



Recursive vs. restricted

In fact, the Pareto frontier for recursive languages is very far!



Languages differ in which meanings they lexicalize.

A	B	A AND B
1	1	1
1	0	0
0	1	0
0	0	0

Bowler (2014)

Languages differ in which meanings they lexicalize.

A	B	A AND B
1	1	1
1	0	0
0	1	0
0	0	0

English

A and B

Bowler (2014)

Languages differ in which meanings they lexicalize.

A	B	A AND B
1	1	1
1	0	0
0	1	0
0	0	0

English

A and B

Warlpiri (Australia)

A manu B

analysis: exh (exh (A or B))

Bowler (2014)

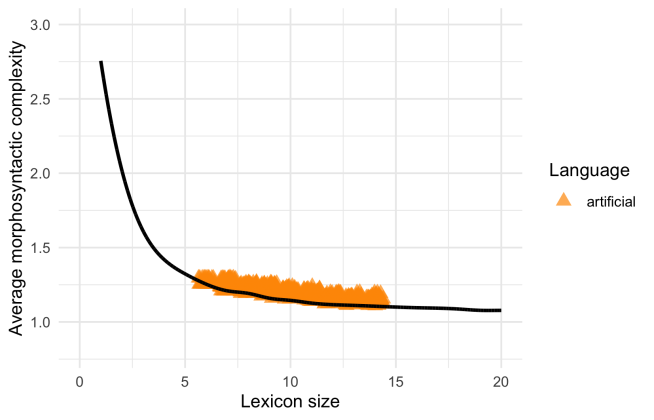
But there are also many similarities in number lexicons.

Number	**English**	**Fulfulde**
1	*one* (1)	*go'o* (1)
2	*two* (2)	*didi* (2)
3	*three* (3)	*tati* (3)
10	*ten* (10)	*sappo* (10)
100	*hundred* (100)	*teemerre* (100)

Explaining semantic universals: Example

All recursive numeral systems (in my sample) lexicalize 1-5.

Explaining semantic universals: Example



≈90% of theoretically possible numeral systems in relative proximity to the Pareto (and all of those on the Pareto) lexicalize 1-5.

Research theme 1: Language and reasoning

How do logical and probabilistic reasoning interact with language interpretation and production?

Research theme 1: Phenomena

- Polarity items (Denić et al., 2018, *Glossa*; Denić et al., 2021, *Cognition*)
- Scalar implicatures (Denić, 2023, *S&P*)
- Donkey anaphora (Denić and Sudo, 2022, *JoS*)
- Proportional quantifiers (Denić and Szymanik, 2022, *JoS*)
- Child language interpretation errors (Denić and Chemla, 2020, *LI*)

Research theme 1: Example result on scalar implicatures

(1) **All 20 of my friends are French or Spanish.**
$\rightsquigarrow$ At least one is French.
$\rightsquigarrow$ At least one is Spanish.

Denić (2023)

Research theme 1: Example result on scalar implicatures

(1) **All 20 of my friends are French or Spanish.**
$\rightsquigarrow$ At least one is French.
$\rightsquigarrow$ At least one is Spanish.

(2) **Both of my friends are French or Spanish.**
$\not\rightsquigarrow$ At least one is French.
$\not\rightsquigarrow$ At least one is Spanish.

Denić (2023)

Research theme 1: Example result on scalar implicatures

(1) **All 20 of my friends are French or Spanish.**
$\rightsquigarrow$ At least one is French.
$\rightsquigarrow$ At least one is Spanish.

(2) **Both of my friends are French or Spanish.**
$\not\rightsquigarrow$ At least one is French.
$\not\rightsquigarrow$ At least one is Spanish.

Empirical: Reducing domain size $\rightarrow$ Fewer scalar implicatures

Denić (2023)

Research theme 1: Example result on scalar implicatures

(1) **All 20 of my friends are French or Spanish.**
$\rightsquigarrow$ At least one is French.
$\rightsquigarrow$ At least one is Spanish.

(2) **Both of my friends are French or Spanish.**
$\not\rightsquigarrow$ At least one is French.
$\not\rightsquigarrow$ At least one is Spanish.

Empirical: Reducing domain size $\rightarrow$ Fewer scalar implicatures
Theoretical: New connection between probabilities and implicatures.

Denić (2023)

Finding what remains to be explained: Example

No morpheme denotes the division operator in any numeral.
(in my language sample)

Finding what remains to be explained: Example

No morpheme denotes the division operator in any numeral.
(in my language sample)

Ket (Werner, 1997)

(3) qol'ep ki
half hundred

Functional pressures (refined)?

Lexicon: Lexicalized numbers and lexicalized arithmetic operators?