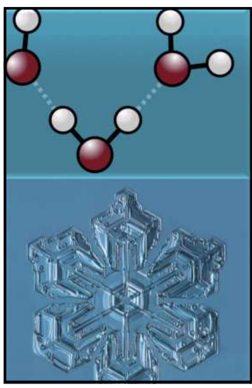
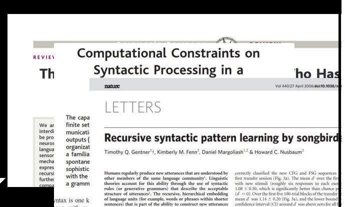
Neural representations for nested tree structures

Willem Zuidema
Institute for Logic, Language & Computation
University of Amsterdam

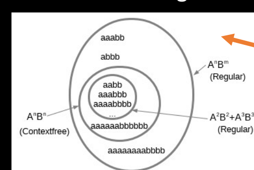
microscale



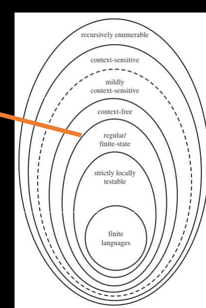


macroscale

Sets of strings



Sets of sets of strings



Common problem

- Many alternative hypotheses that we need to control for
- E.g., to distinguish A^nB^n sequences from $(AB)^n$ sequences it suffices to look for:
  1. the bigram *AA* (+)
  2. the bigram *BB* (+)
  3. the bigram *BA* (-)
  4. the start *AA* (+)
  5. the end *BB* (+)
  6. the start *AB* (-)
  7. the end *AB* (-)
  8. any sequence of *A*'s followed by *B*'s (A^nB^m)
  9. a mix of strategies 1-8
- Each of these alternative is plausible a priori, and none involves context-freeness (Zuidema, 2013, CogSci)

Zuidema (2013), reanalyzing Gentner et al. 2006

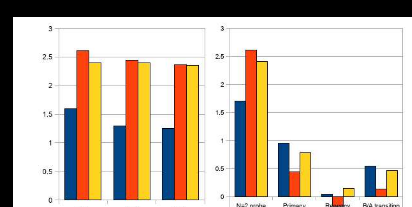
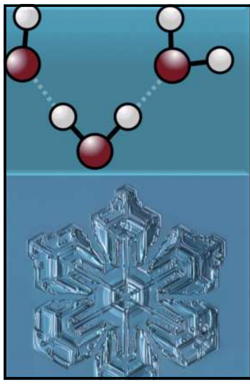
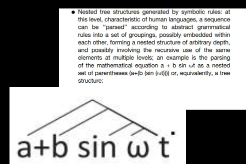


Figure 2: The d'-statistic calculated for the A^nB^n vs. $(AB)^n$ distinction (left) and for various controls (right). *Blue:* Gentner et al, *Red:* CFG, *Yellow:* MIX.



microscale

macroscale

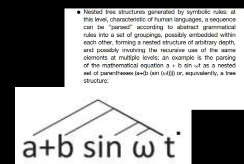


- Nested tree structures generated by symbolic rules: at this level, characteristic of human languages, a sequence can be "parsed" according to abstract grammatical rules into a set of groupings, possibly embedded within each other, forming a nested structure of arbitrary depth, and possibly involving the recursive use of the same elements at multiple levels; an example is the parsing of the mathematical equation a + b sin ωt as a nested set of parentheses (a+(b (sin (ωt)))) or, equivalently, a tree structure:

Dehaene et al. (2015)

This morning

- Increased sophistication of the hypotheses
  - UG -> merge -> dendrophilia -> Language of Thought w/ chunking;
  - Unique human ability to represent -> to learn -> to learn *quickly;*
  - Unique human ability for compositionality -> for symbols as reversible signs.

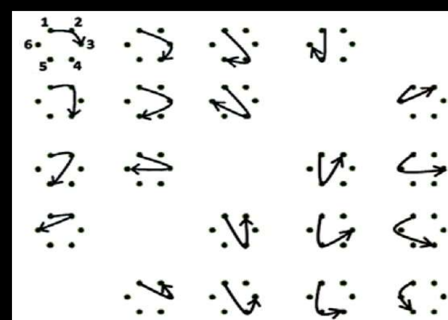


- Nested tree structures generated by symbolic rules: at this level, characteristic of human languages, a sequence can be "parsed" according to abstract grammatical rules into a set of groupings, possibly embedded within each other, forming a nested structure of arbitrary depth, and possibly involving the recursive use of the same elements at multiple levels; an example is the parsing of the mathematical equation a + b sin ωt as a nested set of parentheses (a+(b (sin (ωt)))) or, equivalently, a tree structure:

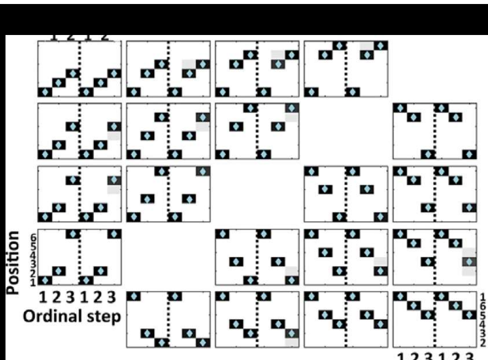
Dehaene et al. (2015)

This morning

- Great innovations in experimental paradigms:
  - 11 Tesla imaging
  - finger tracking
  - brain response
  - letting subjects produce rather than passively receive
  - touch screen interface (Jiang et al., 2018)



(Jiang et al., 2018)



(Jiang et al., 2018)

A different route: representation learning

- Neural language models have become amazingly good at learning subtleties of human language structure, including syntactic structure
- Internal states of the Neural language models give us the best available predictions of activation in the human brain
  - Although not as accurate as often claimed!

Sinclair et al (2022, TACL, *forthcoming*)

Structural Persistence in Language Models: Priming as a Window into Abstract Language Representations

Arabella Sinclair[1,2*] Jaap Jumelet[2*] Willem Zuidema[2] Raquel Fernández[2]
[1]School of Natural and Computing Sciences, University of Aberdeen [2] Institute for Logic, Language and Computation, University of Amsterdam
arabella.sinclair@abdn.ac.uk {j.w.d.jumelet|zuidema|raquel.fernandez}@uva.nl

Abstract

We investigate the extent to which modern, neural language models are susceptible to structural priming, the phenomenon whereby the structure of a sentence makes the same

an abstract notion of structure in their representations, and about the best ways to even assess the syntactic abilities of these models. A rich literature has emerged in the last few years addressing these questions, often taking inspiration from methodologies developed in theoretical linguistics

- Dative alternation:
  - gave the dog a bone vs. gave a bone to the dog
- Transitive alternation:
  - the actor followed the student vs. the student was followed by the actor

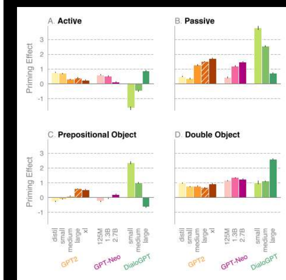
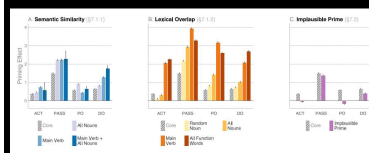


Figure 2: *Priming Effect* results of all models on the *core* corpus, across the four syntactic structures.



Conclusions

- There's a long history to determining the uniquely human ingredient that has given us language
- Proven to be a very difficult challenge
  - Theoretical and experimental innovations very welcome!
- Modern AI offers successful "representation learning" approaches that can be co-opted as hypothesis-generators on neural representations
- Modern LLMs are to big

And too data-hungry!

Taken together, our findings suggest that the neurological architecture of the monkey brain places no fundamental obstacle to the ultimate learning of supra-regular structures. Human uniqueness may lie in the speed with which such structures are learned, perhaps using a specific structure-sensitive algorithm [4, 38, 50], rather than in the mere capacity to acquire them. By introducing a task easily learnable by monkeys yet pre-

(Jiang et al., 2018)

(Zuidema et al, 2018, "Formal models of Structure Building in Music, Language and Animal Song")

- Chomsky Hierarchy
- Add semantics
- Add probabilities
- Make categories graded: vector grammars

freeness is uniquely human. Rodriguez's results demonstrate a continuum between finite-state and (at least some) context-free languages, which cast doubt on the validity of the focus on architectural constraints on structure building operations that dominates the CH. As such, while much theoretical work exploring their expressive power is still necessary, vector grammars provide another motivation to move on to probabilistic, non-symbolic models that go beyond the constraints of the CH.