Hypothesis

Shiro Takagi

2021/4/12

1 question

Q: How to make an intelligent system that learns to manipulate symbols?

2 survey

2.1 cognitive science

Chomsky says that the power to taming syntax is innate property of human brain (universal grammar) [1]. Ibbotson says that "the complexity of language emerges not as a result of a language-specific instinct but through the interaction of cognition and use" (usage-based theory) [2]. According to the usage-based theory, linguistic structure develops by 1. categorization, 2. chunking, 3. rich memory, 4. analogy, and 5. cross-modal association [3, 2]. Children use a limited number of reliable short frames.

" *Overall it seems there is good evidence to support the usage-based prediction that language structure emerges in ontogeny out of experience (viz. use) and when a child uses core usage-based cognitive processes – categorization, analogy, form-meaning mapping, chunking, exemplar/item-based representations – to find and use communicatively meaningful units.* [2] "

The meaning of symbols is established by convention [4, 5, 6].

2.2 transformers

Transformer learns syntactic information [7, 8, 9, 10].

2.3 comparative study

Watson et al. claims that "*nonadjacent dependency processing, a crucial cognitive facilitator of language, is an ancestral trait that evolved at least 40 million years before language itself*" [11]. Wilson et al. explains that sufficient cues play crucial role for human nonadjacent dependency learning [12]. Okanoya and Merker propose the hypothesis that human language is established through string-context mutual segmentation: "*song strings and behavioral contexts are mutually segmented during social interactions*" [13].

2.4 Neo-Vygotskian Theory

The key novelties in human evolution were all, in one way or another, adaptations for an especially cooperative, indeed hypercooperative, way of life [14]. Tomasello states that there are two human-unique capabilities: joint intentionality and collective intentionality. He says that Human developed preparation for these abilities and children first acquire joint intention and then collective intention in their developmental stages.

3 Discussion

3.1 transformers

If transformers really capture syntax, **how it develops the syntactic representation during the pre-training?** Previous studies seem to find that pre-trained transformer have syntactic representation but how to do that remains to be answered. If I can single out the cause of the syntax emergence, I may be able to model a guiding principle for an intelligent agent to learn syntax.

3.2 syntax

It might be plausible that infants first identify phrases in sentences. To identify phrases, it might be necessary that the phrase is used in multiple sentences. By observing the subset of sentences in multiple sentences repeatedly, infants could identify which subset is the the phrase. However, a phrase less likely to appear many times in multiple contexts. Thus, I could hypothesize that infants first understand too common and too often appearing phrase.

Then, they understand that theres is the concept "phrase" in their society. Finally, they could start to generalize their knowledge and to do try and error to manipulate phrase order. In sum, **I hypothesize that artificial intelligence should follow the following path for language acquisition: phrase identification - phrase order arrangement - phrase manipulation.** If this is the case, I should create an environment where identifying key phrase will give reword to the agent. This view is similar to that in [13] which I explained above.

3.3 keyword detection

Dual-coding memory may be a key because visual information is pseudo label there [15]. Even if no instructor exists, the agent can learns the concept him/herself.

3.4 collective intention

A distinct feature of human intelligence is its mastery of languages. On the other hand, Tomasello argues that humans differ from other animals in terms of collective intentionality. If this is the case, **a natural hypothesis from this is that human develop language skills thanks to their collective intentionality.**

3.5 Action and Symbolic Manipulation

Human understand a notion of "research" and know that it includes "science", for example. Or, I can use a more abstract notion of "object" and apply an "operation" on it. Human beings seem to excel at these symbolic manipulation. I believe that memory is "used" for these operations.

I support the view that the semantics of neural representation is designed through the culture the agent is in [4]. Repeated activations of neurons constructs an abstract concept and the symbols are attached to these concepts and segments the neural representation space. Thus, the abstract concept is formed through energy minimization and semantics is grounded to these abstract through social interactions. I hypothesize that human attach the symbol because it improves the predictability of internal and external states. In other words, I do so because it is useful. I think that the relation is also formed through this process.

I think that the symbolic operation is a generalization of the physical action in the neural representation space. When I move a cup, the cup will change its position after the action. In the similar vein, I act on a symbol and produce another symbol. Planning, making a hypothesis, proving a proposition, all of these mental actions can be regarded as like this. I find a book that presents a similar idea a bit [16]. Also, the idea of *motor control origin of Merge* presented in [17, 18] proposed a similar view. There are some counter arguments on this view as well [19]. This kind of idea seems to have its long history [20, 21]. I also think that symbolic operation occurs just because it is "useful".

Therefore, I believe that the temporal characteristics of the environment humans live in matters for symbolic manipulation.

3.6 Language and Localization

A research proposes an interesting view on localization of human language function in our brain [22]. They say that "*our hypothesis holds that the left hemisphere of the vertebrate brain was originally specialized for the control of well-established patterns of behavior under ordinary and familiar circumstances. In contrast, the right hemisphere, the primary seat of emotional arousal, was at first specialized for detecting and responding to unexpected stimuli in the environment ... In other words, the left hemisphere became the seat of self-motivated behavior, sometimes called top-down control. (We stress that self-motivated behavior need not be innate; in fact, it is often learned.) The right hemisphere became the seat of environmentally motivated behavior, or bottom-up control.*" [22]. This view is consistent with [13], where authors emphasize the importance of "song" and repeated phrases for the emergence of human language. An important implication from this literature is the connections among repeated action, top-down control, and language. **An intelligent agent may come to manipulate symbols after gaining top-down control through repeated actions.**

They present another interesting observation. They note that "*In humans the right hemisphere "takes in the whole scene," attending to the global aspects of its environment rather than focusing on a limited number of features. That capacity gives it substantial advantages in analyzing spatial relations. Memories stored by the right hemisphere tend to be organized and recalled as overall patterns rather than as a series of single items. In contrast, the left hemisphere tends to focus on local aspects of its environment.*" **If this is**

the case, I may assume that locality preference bias may be a good starting point for nurturing symbolic manipulations.

They also hypothesize that "*to assess an incoming stimulus, an organism must carry out two kinds of analyses simultaneously. It must estimate the overall novelty of the stimulus and take decisive emergency action if needed (right hemisphere). And it must determine whether the stimulus fits some familiar category, so as to make whatever well-established response, if any, is called for (left hemisphere).*" **If I adopt this view, I may say that I can assume two kind of intrinsic motivations. One is the preference for the novelty, which is studied for a long time [23]. The other is the preference for categorization, which is not explored but can be important for higher cognitive functions.** They also say that "*Perhaps, then, those hemispheric specializations initially evolved because collectively they do a more efficient job of processing both kinds of information at the same time than a brain without such specialized systems.*"

3.7 Language and Brain

I find [24] really good read. The author describe the ontogeny of the linguistic skills of human children.

Under age 3, the author state that "*The available data from the neurocognitive studies reviewed in this section provide consistent evidence that very early on an infant is able to extract language-relevant information from the acoustic input. Infants appear to be equipped with the ability to identify those language sounds relevant for the phonetics of the target language, to perceive prosodic cues that allow them to chunk the input into phrases, and to recognize positional regularities and dependencies that are crucial for the syntax of the target language. Moreover, associative learning allows infants to rapidly acquire names of objects and actions, and the relation between them. All the processes mainly involve the temporal cortices of both hemispheres with a shift toward the left hemisphere with increasing age.*"

She also says that "The findings provide suggestive evidence that the full language capacity can only be reached once the brain has fully matured." For syntactic ability, she find that "*the dorsal fiber tract connecting BA 44 in the inferior frontal gyrus to the posterior temporal cortex is crucial for the full achievement of syntactic abilities.*"

She points out the relations of semantics and syntax in human ontogeny. She says that : "*3- to 4- and 6- to 7-year-old children, in contrast to adults, do* *not process syntax independently from semantics as indicated by their activation pattern. Semantics and syntax interacted with each other in the superior temporal cortex. Interestingly, it is not until the end of the 10th year of life that children show a neural selectivity for syntax, segregated and independent from semantics, in the left inferior frontal gyrus similar to activation patterns seen in the adult brain. Thus, for processing the more complex object-first sentences compared to the less complex subject-first sentences, it is not until early adolescence that a domain-specific selectivity of syntax can be observed at the brain level.*"

She summarizes these findings as follows: "*during the first months of life, language processing is largely input-driven and supported by the temporal cortex and the ventral part of the language network. Beyond the age of 3 years, when top-down processes come into play, the left inferior frontal cortex and the dorsal part of the language network are recruited to a larger extent.*" An important point of this is linguistic capability on semantics develops faster than that on syntax. This implies that **I should have agents to learn semantics first to develop their symbol manipulation skills.**

The author also discuss language evolution. She introduces a classical view on language development that language development can be related to imitation because Broca's area is related with both mirror system and language [25, 26]. Though the validity of this theory seem to be under debase and the author of the book takes a negative position on this view, I find the hypothesis interesting and promising.

By comparing human and non-human, the author clarifies an important distinction: "*During the evolution of language two crucial abilities had to evolve: these are first, sensorymotor learning, and second, the ability to process hierarchical structures. Sensory-motor learning of simple rule-based sequences is an ability that is present songbirds. However, the ability to process hierarchical structures is not present in songbirds. Therefore, it is conceivable that the ability to process structural hierarchies is what should be considered as a crucial step toward the language faculty.*" If this were the case, **I may have to consider how to install the ability to process structural hierarchies for the agent.**

She finds the structural differences between human brain and non-human brain indicate the human faculty on syntax processing: "*First, cytoarchitectonic analyses demonstrate a leftward asymmetry of Broca's area in the inferior frontal gyrus in humans, but not in non-human primates. Second, the dorsal connection between BA 44 in Broca's area and the superior tem-*

poral cortex is stronger in the human brain than in the non-human primate brain. These structures may have evolved to subserve the human capacity to process syntax, which is at the core of the human language faculty."

I put the summary of this book below: "*Humans differ from non-human primates in their ability to build syntactic structures. This statement has set the scene for our assessment of language in this book. Many sections of Language in Our Brain have focused on syntax, and yet our feeling—when considering language—may rather tend to favor semantics and meaning as the most important aspects of language. We do, after all, want to communicate meaning. In order to do so, however, we need syntax to combine words into phrases and sentences. ... Language is what enables us to plan the future together with others, and to learn from the past through narratives. Language forms the basis of our social and cultural interaction not only in the "here and now," but most importantly also in the "there and then." This evolutionary step is realized in a brain that—as far as we know—underwent only minor, although important changes compared to our closest relatives. I have argued that the syntactic specificity of Broca's area, in particular its posterior part, namely BA 44 and its connection to the temporal cortex granting the interplay between these two regions, is what accounts for one of these changes. From an evolutionary perspective it remains an open question whether humans develop language to cover those needs and whether the emergence of language determined the homo loquens as a cultural being.*"

3.8 Recursive Merge and Displacement

Hauser et al. propose a hypothesis that the only unique properties of human language are *recursive merge* and *displacement* [27].

3.9 Environment

It matters to consider what situation is good for developing language ability. For example, if you an agent is in a simple environment and it has only one action, it might not have to learn complex signal like language.

4 Hypotheses

In sum, I hypothesize that following things are important to make agents learn to process symbols:

- Mastery of compositional/hierarchical physical action
- Preference on repetition, locality, and categorization
- Rich capacity to understand semantics
- Collective intention
- Preference on complex chunk arrangement
- Statistical learning of syntax from texts/speech

I will evaluate these hypotheses one by one.

5 A Probable path toward human faculty of language

The basic function of the brain is prediction. I hypothesize that every function of human brain can be explained as prediction. When an agent is exposed to an environment, it first learns the intuitive physics of the world by prediction. Because finding the pattern is the best for prediction, the agent naturally learn o extract patterns from the world. As a result, the agent forms an abstract concept of the objects in the world. This is a precursor of the symbol. This is supported by the preference on repetition and categorization, which is installed in human brain, especially in the left side of it.

This conceptualization is applicable to actions as well. Human can acquire an abstract notion of action by repetition and categorization. Then, they can combine several actions and form another action. They can use this constructed action to compose another action. In this way, human learn to construct hierarchical action. This is a physical manifestation of recursive merge, an important characteristics of language. At this point, human get the precursors for symbolization and its operation.

Because humans are social animal, they communicate with each other. Communication arises because it is useful for them to live in the world. Without communication, abstract concept can be in any form as long as it makes sense to him/herself. However, once you have to communicate with others, the abstract concept should be grounded to something common to the members, symbol. Thus, they define symbols and semantics attached to them through cultural interaction. The first time to name a semantics as a symbol is important since one the agents do so, it can learn the action of naming symbol itself. The agents might to this by string-context mutual segmentation.

This development is necessary in evolution process. However, in the human ontogeny, infants are exposed to language spoken by other humans. Thus, it is easier to learn semantics and syntax. Infants can associate linguistic semantics with abstract concept. Syntax is also learned from tons of demonstration of human language. In addition, since phonological segmentation of a sentence is tightly connected with syntactic information, it makes easier children to learn syntax.

In this whole process, what I think particularly important are abilities/properties to learn hierarchical pattern and compose hierarchical actions.

6 Hypothesis

I hypothesize that I can have an agent to learn to manipulate symbol by installing the capabilities to manipulate action sequences.

References

[1] Noam Chomsky. *Syntactic structures.* Walter de Gruyter, 2002.

[2] Paul Ibbotson. The scope of usage-based theory. *Frontiers in psychology*, 4:255, 2013.

[3] Maria Angélica Furtado da Cunha. Language, usage and cognition. cambridge: Cambridge university press. 252 págs. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, 29(1), 2010.

[4] Adam Santoro, Andrew Lampinen, Kory Mathewson, Timothy Lillicrap, and David Raposo. Symbolic Behaviour in Artificial Intelligence. *arXiv preprint arXiv:2102.03406*, 2021.

[5] Tadahiro Taniguchi, Emre Ugur, Matej Hoffmann, Lorenzo Jamone, Takayuki Nagai, Benjamin Rosman, Toshihiko Matsuka, Naoto Iwahashi, Erhan Oztop, Justus Piater, and Florentin Wörgötter 1. Symbol Emergence in Cognitive Developmental Systems: a Survey. *IEEE transactions on Cognitive and Developmental Systems*, 11(4):494 – 516, 2018.

[6] James L McClelland, Felix Hill, Maja Rudolph, Jason Baldridge, and Hinrich Schütze. Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models. *Proceedings of the National Academy of Sciences*, 117(42):25966–25974, 2020.

[7] Emily Reif, Ann Yuan, Martin Wattenberg, Fernanda B Viegas, Andy Coenen, Adam Pearce, and Been Kim. Visualizing and measuring the geometry of bert. *Advances in Neural Information Processing Systems*, 32:8594–8603, 2019.

[8] John Hewitt and Christopher D Manning. A structural probe for finding syntax in word representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4129–4138, 2019.

[9] Yoav Goldberg. Assessing bert's syntactic abilities. *arXiv preprint arXiv:1901.05287*, 2019.

[10] Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R Bowman, Dipanjan Das, et al. What do you learn from context? probing for sentence structure in contextualized word representations. *arXiv preprint arXiv:1905.06316*, 2019.

[11] Stuart K Watson, Judith M Burkart, Steven J Schapiro, Susan P Lambeth, Jutta L Mueller, and Simon W Townsend. Nonadjacent dependency processing in monkeys, apes, and humans. *Science advances*, 6(43):eabb0725, 2020.

[12] Benjamin Wilson, Michelle Spierings, Andrea Ravignani, Jutta L Mueller, Toben H Mintz, Frank Wijnen, Anne Van der Kant, Kenny Smith, and Arnaud Rey. Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*, 12(3):843–858, 2020.

[13] Kazuo Okanoya and Bjorn Merker. Neural substrates for string-context mutual segmentation: A path to human language. In *Emergence of communication and language*, pages 421–434. Springer, 2007.

[14] Michael Tomasello. *Becoming human: A theory of ontogeny*. Belknap Press, 2019.

[15] Felix Hill, Olivier Tieleman, Tamara von Glehn, Nathaniel Wong, Hamza Merzic, and Stephen Clark. Grounded language learning fast and slow. In *International Conference on Learning Representations*, 2021.

[16] Jeff Hawkins. *A Thousand Brains: A New Theory of Intelligence*. Basic Books, 2021.

[17] Koji Fujita. Recursive merge and human language evolution. In *Recursion: Complexity in cognition*, pages 243–264. Springer, 2014.

[18] Koji Fujita. On the parallel evolution of syntax and lexicon: A merge-only view. *Journal of Neurolinguistics*, 43:178–192, 2017.

[19] Emiliano Zaccarella, Giorgio Papitto, and Angela D Friederici. Language and action in broca's area: Computational differentiation and cortical segregation. *Brain and Cognition*, 147:105651, 2021.

[20] Karl Spencer Lashley. *The problem of serial order in behavior*, volume 21. Bobbs-Merrill Oxford, United Kingdom, 1951.

[21] Patricia M Greenfield et al. Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and brain sciences*, 14(4):531–595, 1991.

[22] Peter F MacNeilage, Lesley J Rogers, and Giorgio Vallortigara. Origins of the left & right brain. *Scientific American*, 301(1):60–67, 2009.

[23] Jürgen Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, 2010.

[24] Angela D Friederici. *Language in our brain: The origins of a uniquely human capacity*. MIT Press, 2017.

[25] Marco Iacoboni, Roger P Woods, Marcel Brass, Harold Bekkering, John C Mazziotta, and Giacomo Rizzolatti. Cortical mechanisms of human imitation. *science*, 286(5449):2526–2528, 1999.

[26] Marco Iacoboni, Istvan Molnar-Szakacs, Vittorio Gallese, Giovanni Buccino, John C Mazziotta, and Giacomo Rizzolatti. Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol*, 3(3):e79, 2005.

[27] Marc D Hauser, Noam Chomsky, and W Tecumseh Fitch. The faculty of language: what is it, who has it, and how did it evolve? *science*, 298(5598):1569–1579, 2002.