Computational Models of Language and Processing

Final report for the *Ethics of Research in Neuroscience* Course

Ludovica Pannitto
CIMeC - Center for Mind/Brain Sciences
University of Trento

September 1, 2019

We, humans, use language for a variety of purposes: we use it for communicating with others, as many other species do, but what's most striking is that we are able to use it creatively. In other words, our language is not a fixed set of symbol-meaning or symbol-intention pairings (i.e., a specific yell from a chimpanzee could mean "Danger" or "Food"), but rather a repository of complex conceptual objects that we use to convey our communicative intentions, combining these basic bits in new and unexpected ways, like children do with lego bricks.

This ability in manipulating conceptual units, despite seeming a very superficial, maybe even naive and intuitive aspect of human linguistic ability, is actually at the core of many properties that natural language exhibits and should be taken as both the starting point and the guiding light of any theory aimed at explaining how natural language, broadly speaking, develops.

Let's therefore define **linguistic creativity** as the **ability to reuse existing, small linguistic bits to build up new, unseen blocks**. Not all of the instances of this creative process happen to be *conscious* to the same extent. So, for example, if we were to hear the utterance *"Before one can say Mr. Darcy"* at a Jane Austen fans convention, we would immediately recognise it as a creative use of the *"Before one can say Jack Robinson"* idiom, while we probably wouldn't judge as creative a sentence like *"John sneezed the dust of the encyclopedia"*, or even less any sentence like *"John baked Mary a cake"*, that sounds completely normal to an average english speaker. From a linguistic perspective, however, we are likely to consider them instances of the same underlying process.

We, humans, are however not born with the ability to produce and comprehend such complex expressions: the progress to linguistic productivity is in fact shown gradually by children (Bannard et al., 2009), whose competence builds up on knowledge about specific items and on restricted abstractions before, if ever, getting to general categories and rules (Goldberg, 2006; Tomasello, 2003). The first utterances of a child are mainly echoing adult speech, and the ability to create totally new and creative linguistic instances is a matter of time and what we generally call *correct use of language*.

Some questions arise from this brief introduction:

- What are the **linguistic bricks** that we process and reuse?
- How does it happen that children **interiorize** such structures?

1 Language brick by brick

All theories of language development and use recognize that at the root of human linguistic ability is their capacity to handle symbolic structures: what theories do not agree on is the content of people's linguistic knowledge (is it rules or is it patterns?), on how this content is acquired and to what extent linguistic creativity is affected by this stored knowledge (Bannard et al., 2009).

A common assumption, often made by formal linguists and generativists[1], is that the child's linguistic knowledge is made up of abstract rules and categories (Hauser et al., 2002), which are often claimed to be innate in some way, hard-coded into the child's brain structure. These theories rely on two major assumptions, called respectively the *poverty of the stimulus assumption* (Chomksy, 1959; Chomsky, 1968) and the *continuity assumption* (Pinker, 1984). The former argues that children cannot possibly be exposed to enough input throughout the first stages of their childhood, to be able to generalize to their production: grammar is universal and genetically determined. The latter refers to the fact that, since grammar is an innate device, its structures are immutable and a person's competence stay the same throughout its life.

Both the postulates have been rejected by the so-called *usage-based models*: one central aspect that distinguishes the usage-based approaches and the generative ones is the emphasis that former pose on the linear and time-dependent nature of the linguistic signal (Elman, 1990). Probably due to our linguistic and ontological meta-analysis, we often think of language as a hierarchical structure on many levels: a sentence such as *John gives the book to Mary* is seen as the instantiation of a more abstract *[Subject verb direct-object indirect-object]* logical structure, which is in turn produced by a set of rules that generically make so that a sentence is made of a subject, a verb, an optional direct object and a

[1]As there are in all areas of research, the nuances that allow us to label a theory as part of one particular school of thought are many. For the sake of this brief presentation, let's just mention that there are four main families of theories: *functionalism*, that sees language as a tool we make use of in everyday life, and therefore emphasizes the importance of social context and communicative function, **structuralism**, which is focused on the inner organization of language in a system of symbols that are interconnected with each other, *generativism*, mainly based on the work of Noam Chomsky and the idea of *Universal Grammar* (UG), suggesting that language is made up of certain rules that apply to all humans and all languages, and **cognitivism**, born as a reaction to generativism and against the idea of a UG, stating that language is learned and emerges from human cognitive processes

number of non-compulsory indirect complements. Similarly, when we think of words we usually think of them as ontologically organized in hierarchies, linked by relations that define a *paw* as a part of a *dog*, connected to the class of *animals* and *living beings* in turn, for example.

The existence, and utter relevance of such hierarchies in language comprehension and production, is not denied by usage-based models. While generativists, and deductive models in general, see the linear nature of language just as the surface form of its hierarchical, recursive structure, usage-based theories, that are inductive in nature, advocate thet these hierarchical connection between items can emerge from the fact that language must be processed linearly and is subject to constrains posed by general-purpose memory and cognitive mechanisms (Christiansen and Chater, 2016). The existence and facilitatory role of higher-order structures remains unquestioned, and it's consistent with general observations about memory, such as the well known constraints on our ability to recall stimuli (Miller, 1956)[2], but the emergence of language-like structure from purely linear signal has for example been shown in recent experiments (Cornish et al., 2017), where the authors have demonstrated how important aspects of the sequential structure of language, as its characteristic reusable parts, may derive from adaptations to the cognitive limitations of human learners and users.

One of the major issues that these kind of inductive models have to face is the existence of non-adjacent structures with very variable aspect on the surface. These kind of long-distance dependencies are common in language: think about verb conjugation, that often involve an auxiliar verb and a fixed morphological ending (e.g., the *progressive is X-ing* form, where X can be filled with any suitable verb, or similarly *perfective construction has X-ed* in English), as well as higher order structures like the *correlative construction* (i.e., *the X-er, the Y-er*, as in *the more, the merrier*), but even more subtle things like agreement throughout the sentence or event-level dependencies: while it is intuitive that we, as speakers, are able to detect this kind of discontinuous patterns, evidence coming primarily from artificial grammar learning[3] is not so strong about it (Gomez, 2002; Newport and Aslin, 2004; Gómez and Maye, 2005), being influenced by a great number of factors such as internal variability and the nature of the elements in the pattern (e.g., patterns of homogeneous elements, such as patterns made up by solely consonants, are more easily recognised than eterogeneous patterns, such as patterns composed by both vowels and consonants).

[2]The number 7 has a particular role in our memory, as 7 ± 2 is the number of items that we are on average able to retain in short-term memory. The idea was originally formulated in Miller (1956), and finds also some empirical confirmation in common lists of items, such as the seven dwarfs, the seven seas or the seven deadly sins.

[3]In artificial grammar learning experiments, subjects are usually required, in the training phase, to memorize sets of strings generated by a specific grammar defined on purpose by researchers, and are later required to categorize new items as *ruleful* or *unruleful* with respect to the items seen during training. If speakers are able to tell apart grammatical and ungrammatical items, it could mean that a learning and generalization process has taken place during the training phase.

The discovery and treatment of non-adjacent dependencies have therefore a central role in the theories that subserve language comprehension and production. Be they rules or actual chunks, and be they managed by a dedicated mechanism or a general statistical process, they embody the building blocks that bridge the traditional lexical level to the sentence level, being therefore central to the issues of linguistic productivity and compositionality.

2 Through the Processing Glass

From a more strict linguistic standpoint, this kind of approaches have contributed to blur the traditional, manichaeistic distinction between *lexicon and grammar* (Elman, 2009), the former being the repository of meaning in a dictionary-like fashion and the latter being the grammatical device subserving the composition processes. A number of new architectures have been introduced in order to fill the gap left by the traditional dualistic model (MacDonald et al., 1994; Goldberg, 2003; Jackendoff, 2007; Christiansen and Chater, 2016).

Being also strongly supported by neural evidence (Kuperberg, 2007; Petten, 1993; Berkum et al., 1999; Hagoort et al., 2004; Nieuwland and Van Berkum, 2006), **processing, rather than abstract linguistic knowledge, with its physical and cognitive underpinnings, has gained centrality in linguistic research**.

This comes with some non negligible ethical implications, that we must be aware of, not only as researchers but also as citizens and natural language speakers. The first issue that emerges from the rejection of innatist theories is the apparent loss of democraticity that comes with them. **The innatist view equips every human being with the very same set of linguistic abilities**, no matter the differences in socio-economic conditions or environment that they grow up in. Language is then strictly tied to reasoning and knowledge sistematization, and so this extremely egalitarian view can be easily reflected in the idea that everyone has by nature equal possibility in accessing knowledge and, by extension, equal possibilities of self-realization. **Usage-based inductive models, by granting the input a much grander role in language development, appears much less democratic at a first glance.** The dangerous misunderstanding by which the environment a child is exposed to can determine his or her linguistic capacity (and, by extension, his or her reasoning skills) is behind the corner. As far as this is concerned, it must be clarified that **even inductive models pose some innate component, which is not unique to language but concerns learning in general**: what inductive models propose is that children are especially able to pick up patterns and learn how to reuse them creatively to convey their communicative intentions[4]. By using a single learning mechanism children are moreover able to eventually compensate lacks in linguistic input through knowledge acquired in other fields (e.g., from visual or sensory-motor input). The manipulation by which some languages can

[4]The question of how communicative intentions are acquired is a complicated matter that cannot be tackled from a strict linguistic perspective.

be *better* than others, some socio-cultural environment are *better* than others, is probably easier to make in this case with respect to fully innatist theories, and it is therefore extremely important, as researchers, to further investigate the role played by the input in early linguistic development, while advocating for the idea that democracy is something that is more correctly implemented at the level of society, namely the *population of speakers*, rather than the individual, and is nothing like *starting all with the same equipment*, but rather concerns the idea of giving everyone the possibility to fulfill their (in this case, communicative) needs.

The second point concerns species-specificity: while the linguistic ability is unique to human beings for universal-grammarians, learning is something that all species do. Even certainly being a human privilege, **usage-based theories are puting linguistic ability back on a continuum of skills**, in some sense back in the ecosystem as far as reasoning is concerned. This, while having clear ethical implications about our relationship with the animal world, also opens up a whole (partly) new array of possibilities for model learning and reasoning mechanisms, affecting decisions concerning the nature of linguistic conceptualization and abstractions and, most importantly, the processing mechanisms, which cannot be possibly faced from a pure linguistic standpoint, as it has been shown how general and unsupervised learning mechanisms are also active in language.

2.1 The word, voilà l'ennemi

Learning, that has now become our central topic of interest, also irrespective of the linguistic level, seem to entail two different aspects:

- **finding** (i.e., segmenting) the most relevant units to encode information
- **representing** (i.e., compressing) information so as to make is efficient to store and to reproduce

These two processes should be mutually informative and should be both considered when modeling or analyzing language: while the problem of *segmentation* has been largely taken for granted by semanticists, research on statistical learning and chunking has mainly focused on symbols, leaving aside issues concerning the function that the chunks have in communication. The two aspects are strictly tied to one another and the respective research outputs could be successfully integrated.

The question about *how do we attach meaning representations to linguistic symbols* has been central to usage-based models of language acquisition. In order to be better integrated with the statistical learning and cognitive-based community, we propose to pose the same question in a different formulation: **how do we identify the linguistic structures that are better suited, or more likely to cue the desired meaning?**

The proposal is therefore to provide a usage-based model of non-adjacent dependencies (i.e., construction), as they emerge from the linear linguistic stream through general purpose statistical mechanisms.

References

Colin Bannard, Elena Lieven, and Michael Tomasello. 2009. Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences*, 106(41):17284–17289.

Jos JA van Berkum, Peter Hagoort, and Colin M Brown. 1999. Semantic integration in sentences and discourse: Evidence from the n400. *Journal of cognitive neuroscience*, 11(6):657–671.

Noam Chomksy. 1959. Review of skinner's verbal behaviour. *Language*, 35:26–58.

Noam Chomsky. 1968. *Language and Mind*. New York: Harcourt Brace Jovanovich.

M. H. Christiansen. 2018. Implicit statistical learning: A tale of two literatures. *Topics in cognitive science*.

Morten H Christiansen and Nick Chater. 2016. The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39.

Hannah Cornish, Rick Dale, Simon Kirby, and Morten H Christiansen. 2017. Sequence memory constraints give rise to language-like structure through iterated learning. *PloS one*, 12(1):e0168532.

Jeffrey L Elman. 1990. Finding structure in time. *Cognitive science*, 14(2):179–211.

Jeffrey L Elman. 2009. On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive science*, 33(4):547–582.

Ram Frost, Blair C Armstrong, Noam Siegelman, and Morten H Christiansen. 2015. Domain generality versus modality specificity: the paradox of statistical learning. *Trends in cognitive sciences*, 19(3):117–125.

Adele E Goldberg. 2003. Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, 7(5):219–224.

Adele E Goldberg. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press on Demand.

Rebecca Gómez and Jessica Maye. 2005. The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7(2):183–206.

Rebecca L Gomez. 2002. Variability and detection of invariant structure. *Psychological Science*, 13(5):431–436.

Rebecca L Gomez, Louann Gerken, and Roger W Schvaneveldt. 2000. The basis of transfer in artificial grammar learning. *Memory & Cognition*, 28(2):253–263.

Peter Hagoort, Lea Hald, Marcel Bastiaansen, and Karl Magnus Petersson. 2004. Integration of word meaning and world knowledge in language comprehension. *science*, 304(5669):438–441.

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

Ray Jackendoff. 2007. A parallel architecture perspective on language processing. *Brain research*, 1146:2–22.

Tobias Johansson. 2009. Strengthening the case for stimulus-specificity in artificial grammar learning: no evidence for abstract representations with extended exposure. *Experimental Psychology*, 56(3):188–197.

Gina R Kuperberg. 2007. Neural mechanisms of language comprehension: Challenges to syntax. *Brain research*, 1146:23–49.

Maryellen C MacDonald, Neal J Pearlmutter, and Mark S Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological review*, 101(4):676.

George A Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81.

Elissa L Newport and Richard N Aslin. 2004. Learning at a distance i. statistical learning of non-adjacent dependencies. *Cognitive psychology*, 48(2):127–162.

Mante S Nieuwland and Jos JA Van Berkum. 2006. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience*, 18(7):1098–1111.

Cyma Van Petten. 1993. A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, 8(4):485–531.

Steven Pinker. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Martin Redington and Nick Chater. 1996. Transfer in artificial grammar learning: A reevaluation. *Journal of experimental psychology: general*, 125(2):123.

Michael Tomasello. 2003. *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.