Brain & Language 150 (2015) 54–68

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Corticostriatal response selection in sentence production: Insights from neural network simulation with reservoir computing

Xavier Hinaut [a], Florian Lance [b], Colas Droin [b], Maxime Petit [b], Gregoire Pointeau [b], Peter Ford Dominey [b],*

[a] CNPS, UMR CNRS 8195, University Paris-Sud, Orsay, France
[b] INSERM Stem Cell and Brain Research Institute, Human and Robot Cognitive Systems, 18 Ave Lepine, 69675 Bron Cedex, France

ARTICLE INFO

Article history:
Received 16 February 2015
Revised 18 July 2015
Accepted 4 August 2015

Keywords:
Sentence production
Corticostriatal system
Language
Model
Neural network

ABSTRACT

Language production requires selection of the appropriate sentence structure to accommodate the communication goal of the speaker – the transmission of a particular meaning. Here we consider event meanings, in terms of predicates and thematic roles, and we address the problem that a given event can be described from multiple perspectives, which poses a problem of response selection. We present a model of response selection in sentence production that is inspired by the primate corticostriatal system. The model is implemented in the context of reservoir computing where the reservoir – a recurrent neural network with fixed connections – corresponds to cortex, and the readout corresponds to the striatum. We demonstrate robust learning, and generalization properties of the model, and demonstrate its cross linguistic capabilities in English and Japanese. The results contribute to the argument that the corticostriatal system plays a role in response selection in language production, and to the stance that reservoir computing is a valid potential model of corticostriatal processing.

1. Introduction

The goal of the current research is to present a model of sentence production based on the function of the primate corticostriatal system, extending our previous work on corticostriatal function in sentence comprehension. We situate this work in the context of related models, and background on the neuropsychology of corticostriatal function in sentence production, both of which are relevant to our proposed model. The transmission of meaning by language is one of the marvels of human cognition. Sentence production and comprehension are complementary, but asymmetric. In comprehension, it is possible to correctly extract only part of the message – for example only the thematic role assignment (who did what to whom). In production, the speaker must generate a specific linear string of words which communicates the intended meaning that in addition to thematic roles should include some notion of focus or importance, and other dimensions including time, mode and aspect (Klein, 2013). These dimensions can be considered in the larger context of phrasal semantics – meaning that can be communicated by the grammatical structure of the sentence (Dominey, 2005; Jackendoff, 2002). Here, we can consider a representation of the meaning of an event and its thematic roles in a predicate-argument structure, along with some indication of whether the focus is on the agent, object, recipient, etc. Our meaning representation is in a predicate-argument format, originally developed in the domain of describing object manipulation actions, e.g. "The ball was given to Jean by Marie" (Dominey & Boucher, 2005). There we adopted a representation with the predicate, corresponding to the action, and the arguments corresponding to the agent, the manipulated object, and the recipient. This resulted in our use of the PAOR – or predicate, agent, object, recipient-representation. Thus, our notion of object in the PAOR notation corresponds to the classic thematic role of patient. Both of these components (thematic roles, and focus) should be encoded in the phrasal semantics of the sentence. In comprehension, the reception of this sentence should allow the listener to reconstruct the intended meaning – the thematic roles and the focus structure constituting the speakers' construed meaning.[1] Part of the richness of language expressivity is the varying ability across languages to use word order as a mechanism for specifying the communicative focus and other aspects of phrasal semantics within the sentence, in addition to communicating "who did what to whom."

---

* Corresponding author.
E-mail address: peter.dominey@inserm.fr (P.F. Dominey).

[1] We note that in the more general case of discourse and dialog, once the processing has begun, there is significant context which specifies some of the intended meaning. This potentially reduces the difficulty of production and comprehension.

http://dx.doi.org/10.1016/j.bandl.2015.08.002
0093-934X/© 2015 Published by Elsevier Inc.

Our model can be considered in the larger context of models of language production, with those that focus on aspects of the word level processes of semantic retrieval, word repetition, and word production Roelofs (2014), or that may be more concerned with accounting for higher level behavior including alignment between speakers at multiple levels (e.g. alignment of grammatical structures, and situation models of the joint task that dialog participants are working on) that takes place during dialog (Pickering & Garrod, 2013). We are concerned with the production of sentences, multiple word utterances, that may have some degree of complexity including the use of embedded relative clauses. Takac, Benuskova, and Knott (2012) have modeled sentence production as a form of mapping from sensorimotor sequences to word sequences. They did not address issues of multiple non-canonical orders, relative clauses etc. Chang (2002) modeled sentence production using a dual path model that has one pathway for mapping message content to words and a separate pathway that enforces sequencing constraints, i.e. word order, based on Elman's simple recurrent network (SRN) (Elman, 1990, 1991, 1993). This model employs recurrent connections that are modified by back propagation of error. In order to simplify the difficult problem of assigning error to recurrent connections, the problem is simplified, by only taking one recurrent pass through the network into account for the learning, hence the term "simple". This model has been quite influential in cognitive science, including studies of language, e.g. (Christiansen & Chater, 1999; Elman, 1993) and sequence learning e.g. (Cleeremans & McClelland, 1991; Jiménez, Méndez, & Cleeremans, 1996; Servan-Schreiber, Cleeremans, & McClelland, 1991). Chang also set out to account for cross-linguistic differences, and thus demonstrated that the dual path model could account for word-order effects in English and in Japanese. Chang (2009) demonstrated that when the prominence of the thematic roles is expressed as part of the meaning, the model can appropriately learn different forms (e.g. active and passive) in English, and accommodate word scrambling in Japanese. The model was able to handle 50 different constructions with analogous structure in English and Japanese. This included 3 simple constructions, 9 sentential conjunctions, 6 phrasal conjunctions, 32 structures with relative clauses. In order to address relative clauses in more detail, Fitz, Chang, and Christiansen (2011) exploited the extended dual path model to accommodate multiple clauses. The meaning representation included three components: thematic roles (AGENT, PATIENT, RECIPIENT, etc.), concepts (lexical semantics), and event features to signal the number and relative prominence of event participants. Dell and Chang (2014) have recently applied their model of prediction and prediction error processing in sentence production to understanding aspects of aphasic production. Part of the goal of such modeling indeed should be not only to posit mechanisms of linguistic function, but also to establish links between linguistic function and the underlying neurophysiology.

The current research proposes a biologically inspired neural network model, in the reservoir computing framework, that learns to produce sentences. The link between reservoir computing and corticostriatal neurophysiology can provide useful insight into understanding aspects of higher cognitive function in human and non-human primates. Barone and Joseph (1989) observed PFC activity in macaque monkeys trained to perform a visuomotor sequencing task. For the first time, they observed PFC neurons that encoded a mixture of spatial and sequential rank selectivity. We modeled PFC as a network of leaky integrator neurons with fixed recurrent inhibitory and excitatory connections, and corticostriatal connections modifiable based on reward-related dopamine (Dominey, Arbib, & Joseph, 1995). This was the first instantiation of reservoir computing. The key notion is that the intrinsic dynamics of the fixed-connection reservoir provide an inherent capacity to represent arbitrary sequential structure. PFC neurons in the model displayed the same mixture of spatial location and sequence rank as observed by Barone and Joseph. We further demonstrated that in this configuration, PFC encodes task context, and striatum encodes action selection, again as observed in the primate (Dominey & Boussaoud, 1997), thus supporting the analogy between reservoir–readout and cortex–striatum. More recently the claim that cortex corresponds to a reservoir (based on dense local recurrent connections) has been supported by anatomy and physiology, and modeling (Nikolic, Hausler, Singer, & Maass, 2009; Rigotti, Rubin, Wang, & Fusi, 2010; Rigotti et al., 2013).

In this context, we attempt to determine if this approach to modeling the corticostriatal system can be applied to sentence production. We are particularly interested in the problem of how different word orders can be used to describe the same event, but with different focus. As will be described in more detail below, given a mental model with two events, and three arguments each, there is a small combinatorial explosion of the different ways that this meaning can be expressed in a sentence in English. The explosion is even greater in Japanese where there are fewer restrictions on word order.

When faced with this level of possible degrees of freedom, sentence production can take on an aspect of motor planning, in that the sequence of words to be produced is specific for a particular communicative goal, like a motor sequence trajectory may be specific for a particular action goal. The framework that we use to address this problem is based on the sequence processing capabilities of the corticostriatal system which plays a central role in the sequential organization of behavior, and action sequence selection (Hikosaka, Nakamura, Sakai, & Nakahara, 2002). In order to appreciate the functional significance of the corticostriatal system, one should recall that all of the primary and associative cortices including the language areas project to the striatum – the input nucleus of the basal ganglia (Alexander, DeLong, & Strick, 1986; Yeterian & Pandya, 1998). The integrity of the corticostriatal system is thus likely required both for language comprehension and production (Argyropoulos, Tremblay, & Small, 2013; Friederici & Kotz, 2003; Friederici, Kotz, Werheid, Hein, & von Cramon, 2003; Frisch, Kotz, von Cramon, & Friederici, 2003; Hochstadt, 2009; Hochstadt, Nakano, Lieberman, & Friedman, 2006; Kotz, Frisch, von Cramon, & Friederici, 2003).

We have previously examined how the corticostriatal system could implement aspects of the mechanism that learns to interpret sentences in language (Dominey, 2001, 2013; Dominey, Hoen, Blanc, & Lelekov-Boissard, 2003; Dominey & Inui, 2009; Dominey, Inui, & Hoen, 2009) where language is considered a structured inventory of grammatical constructions mapping sentence form to meaning (Goldberg, 1995, 2003). The model was based on the hypothesis that thematic role assignment (determining who did what to whom) can be determined by the order and position of closed class elements (grammatical function words and grammatical morphemes) (Dominey, 2001; Dominey & Inui, 2009; Dominey et al., 2003, 2009; Hinaut & Dominey, 2013). In this family of models, the input to the recurrent network was the sequence of activation of neurons coding the closed class words as they appeared in the sentence. This drove the recurrent network into a specific trajectory for each different sentence type. Learning in connections between the recurrent reservoir nodes and the output neurons allowed the output neurons to correctly decode the thematic roles for the open class words for input sentences. In the current research we invert this process, that is, we provide the input as activation of neurons coding the meaning of the desired sentence. Meaning is coded as the ordered set of open class elements, and their corresponding thematic roles, that we together refer to as the focus hierarchy. This drives the recurrent network through a specific trajectory of activation. We train the output connections to activate word-coding units in the appropriate order to generate the corresponding sentence to express the input meaning.