Towards a Neurobiologically Inspired Model of Syntax Processing

Bernd J. Kröger (bernd.kroeger@rwth-aachen.de)
Department of Phoniatrics, Pedaudiology and Communication Disorders, RWTH Aachen University
Pauwelsstr. 30, 52074 Aachen, Germany

Abstract

A first version of a neurobiologically inspired neural network model for speech and language processing using a spiking neuron approach is introduced here. This model uses basic neural circuit elements for building up a large-scale brain model (i.e., elements for long-term and short-term memory, elements for activating and forwarding information (items) as neural states, elements for cognitive and sensorimotor action selection, elements for modeling binding of items, etc.). The resulting model architecture indicates three dense neural network modules, i.e., a module for lexical, for syntactic, and for semantic processing. Moreover, the model gives a detailed specification of the neural interaction interfaces between these modules. This large-scale model is capable of parsing syntactic simple but non-trivial sentences of Standard German and it clearly exemplifies the temporal-parallel as well as the hierarchical-sequential neural processes typically appearing in speech processing in the brain.

Keywords: neural network model; neurobiologically inspired model; large-scale model; speech processing; language processing; syntactic processing

Introduction

Theories and computer-implemented models of Natural Langue Processing (NLP) made progress over the last decades (e.g., Jurafsky & Martin, 2009; Jurafsky & Martin, 2023). Most natural language processing approaches comprise lexical, syntactic, and semantic processing as part of comprehension (Natural Language Understanding) as well as part of production (Natural Language Generation, see e.g., Khurana et al., 2023). Typical NLP approaches use text databases for lexical processing, i.e., for extracting grammatical word-type and semantic word information (Jurafsky & Martin, 2023, chap. 9 and chap. 24). NLP-based syntactic and semantic processing can be based on dependency or constituent grammatical concepts for constructing statistical or neural processing models (Zhang, 2020). Over the last decade, these NLP based deep neural models became more and more successful (e.g., BERT, Devlin et al., 2019; Koroteev, 2021; and GPT-2, Radford et al., 2019). These approaches already include attention modelling and the calculation of surprise (i.e., deviation from normal expectation) when new words or new sentences appear in a comprehension process (Vaswani et al., 2017; Verma, 2022). Even though some of these processes appear in human speech processing as well (Arana et al., 2023; Goldstein et al., 2022), the neural architecture, the learning mechanisms, and the core dynamics of human speech processing are not comparable with current NLP approaches (Pedrelli & Hinaut 2022).

Neurobiologically grounded and computer-implementable speech processing models take in consideration all behavioral and neurobiological data as they are condensed for example in function-specific box-and-arrow models (e.g., Friederici, 2011). The Friederici model aims for describing the neural processes of speech perception, speech comprehension, as well as of speech production including the left- and right-hemispheric differences. While left-hemispheric processing focusses on the identification of word and sentence meaning, i.e., is involved in lexical, syntactic, and semantic processing, the right hemisphere focusses on processing of prosody (i.e., intonation and accentuation) as well as on processing of emotional and non-linguistic aspects of speech. Four hierarchical processing stages are uncovered in this model, i.e., (i) acoustic-phonetic-phonological analysis (auditory cortex within the temporal lobe), (ii) initial syntactic processing (mainly left temporal), (iii) the computation of syntactic and semantic relations (mainly left temporal and frontal), and (iv) prosodic-segmental integration of the resulting syntactic-semantic activations (left hemispherical) combined with activations stemming from the right-hemispheric prosodic analysis. These processes take place partially in a hierarchical-serial and partially in a parallel processing fashion and the processing time window increases from less than 100 ms at the lowest processing level to above 500 ms at the highest processing level.

Another prominent model for speech processing comprising speech perception/comprehension and speech production is the dual stream model introduced by Hickok and Poeppel (2007). Here, a ventral and a dorsal processing stream are separated. In case of perception/comprehension the (initial) spectro-temporal (i.e., acoustic) and phonetic-phonological analysis is followed by lexical, syntactic, and semantic processing within a lexical, combinatorial, and conceptual network in the ventral stream, whereas the dorsal stream comprises the initial spectro-temporal and phonetic-phonological analysis which is directly connected with the articulatory network via a sensorimotor interface.

Neurobiologically inspired simulation models which are based on box-and-arrow models mainly exist for word processing. This is the spreading activation model of Dell (Schwartz et al., 2006), the Lichtheim approach of Ueno (2011), and the WEAVER model of Roelofs (2014). These models are designed for the simulation of speech production and speech perception/comprehension tasks like picture naming, picture identification, as well as for word and non-word repetition. Moreover, there exist first neurobiologically inspired models for sentence-domain syntactic-semantic processing (Hinaut & Dominey, 2013; Mitropolsky & Papadimitriou, 2023; Mitropolsky et al. 2021; Pedrelli & Hinaut 2022; Lindes 2018; Lindes 2022, based on Stocco et al. 2021;

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Venhuizen et al., 2019; Frank 2021; Krauska & Lau, 2022), but currently none of these simulation approaches is based on spiking neuron networks.

For developing such a neurobiologically inspired spiking neural network models (SNNs), several modeling environments including neural network simulation toolboxes are available, e.g., NEURON (Carnevale & Hines, 2006), NEST (Gewaltig & Diesmann, 2007), BRIAN (Goodmann & Brette, 2009), and NENGO (Bekolay et al., 2014). The NENGO software package which is based on the NEF-SPA theoretical framework (Neural Engineering Framework, NEF, and Semantic Pointer Architecture, SPA, see: Eliasmith, 2013; Eliasmith et al., 2012; Stewart & Eliasmith, 2014) as well introduces *compact basic neural network elements* which can be used for building up *neural processing modules* for simulating different sensorimotor and/or cognitive tasks (see Eliasmith et al., 2012; Crawford et al., 2015; Kröger, 2023). These are network elements for the realization of short-term and long-term memory, for the realization of item (i.e., vocabulary) networks including a modeling of word similarity or word distance, for the realization of feed-forward (co-)activation of linguistic items at different processing levels (i.e., associative memories), for the realization of action selection processes for sensorimotor and cognitive selection or decision processes, for the realization of binding of items (i.e., binding buffers) etc. In this approach, each set of phonological words, lemmata, or concepts is realized in form of distributed neural activation patterns within semantic pointer networks (e.g., Kröger et al. 2016).

Method

Based on our neurobiologically inspired model of word processing (Kröger et al., 2016; Kröger et al., 2020; Kröger et al., 2022) – which has been implemented by using the NEF-SPA approach and the Python based NENGO toolbox (Bekolay et al., 2014) – a trial-and-error process has been established for developing a syntactic processing module for parsing syntactically simple sentences of Standard German. For this process we made five assumptions: (i) The input to the syntax processing module is activated by the mental lexicon; this is a sequence of lemma (word-type) and semantic word information. (ii) The output of the syntactic component is a syntactic description of each a sentence like e.g., defined in constituent grammar as hierarchical set of constituents (Jurafsky & Martin, 2023, chap. 17) or in dependency grammar as a set of dependency arc relations between words (ibid., chap. 18). (iii) In order to be able to process word and phrase spans, some syntactic information needs to be stored in short-term memories (STMs). (iv) Because of time delay of syntactic information generated during syntactic processing, syntactic and semantic word information (i.e., the sequence of lemmata and word concepts) can be accessed both, firstly directly and temporally delayed. (v) The syntax module is interconnected with the action selection module (basal-ganglia and thalamus) for being capable of decision making, e.g., concerning choosing the correct syntactic elements, e.g., the correct dependency arcs or phrase constituents).

For our initial modeling trial a syntax module was developed which is capable of parsing grammatically simple sentences of Standard German like `subj-verb-obj` , like `subj-verb-obj-verbPart`, or like `subj-verb-prepPhra-verbPart` where `verb` represents single verbs like "trinkt" ("drinks", e.g., in "Benno trinkt Kaffee."; "Benno drinks coffee.") and where `verb-verbPart` represents verb constructions like "hat-getrunken" ("had drunk", e.g., in "Benno hat den Kaffee getrunken."), and where `prepPhra` represents prepositional phrases (e.g. "von Benno" like in "Der Kaffee wurde von Benno getrunken."; "The coffee was drunk by Benno."). In addition we allowed the realization of subject, object, and prepositional noun phrases as being composed of nouns, determiners (det) and adjectives (adj) (e.g., "der kalte Kaffee"; "the cold coffee", or "der große Junge"; "the big boy") leading to sentences like "Der kalte Kaffee wird von dem großen Jungen getrunken" ("The cold coffee is drunk by the big boy.").

The corpus which has been processed by our parsing module comprised 10 `subj-verb` phrases like "Benno trinkt", 20 `subj-verb-object` phrases like "Benno trinkt Kaffee", 20 `subj-verb-obj-verb` phrases like "Benno hat Kaffee getrunken (past tense)", and 20 `subj-verb-prepPhrase-verb` phrases like "Der Kaffee wurde von Benno getrunken (passive)". `Subj` as well as `obj` should be interpreted here as noun spans, i.e., as word spans comprising three different constituents, i.e., `noun` like "Kaffee", `det-noun` like "der Kaffee", or `det-adj-noun` like "der kalte Kaffee".

Results

A feasible neural network architecture for a correct parsing of simple sentences of Standard German is displayed in Fig. 1. The model was developed by a trial-and-error process using the sentence types outlined in the methods part of this paper. We used STMs for uncovering noun spans (see above) and verb spans (typical verb spans in Standard German comprise a so called "verb clamp" ("Satzklammer", e.g., Musan, 2022; and see the example given above: "hat-getrunken"). The module awaits up to two noun spans (incl. prepositional spans). These spans are detected from word type (lemma) information in combination with already online activated word span information by action selection processes and lead to the online generated activations of the arc type STMs as displayed in Fig. 2. Moreover, this information can be used in a further action selection process to activate a specific type of dependency arc for each lemma within the word sequence of the sentence under processing (for modeling action selection see e.g., Kröger et. al., 2016; Stewart & Eliasmith, 2014).

In addition, a preliminary semantic processing component has been implemented as part of our neural model for allowing semantic role assignment which is realized here by a further processing of the online generated syntactic relations (sequence of dependency arcs) and by associating subject and object with agent or patient and by associating the verb with the action and with the event tense. This semantic representation specifies the so-called semantic event described by the sentence (cf. Jurafsky & Martin, 2023, chap. 19-21).

The complete neural architecture of our model (Fig. 1) comprises three modules, i.e., mental lexicon, syntactic, and semantic processing module. The auditory-acoustic input activates phonological form states, and coactivates lemma states and word concept states using lexical associations via associative memories (AMs: p2w-AM and w2c-AM with p = phono, w = word, and c = concept). The generated sequence of lemma states (i.e., word types like noun, verb, adjective, determiner, etc.) is cleaned up from co-activated similar lemma states (via clean-up memory CAM) towards WTB (word type buffer, see Fig. 1) which is already a part of the syntactic processing module. In parallel the sequence of word meanings (concepts) is cleaned up as well and then forwarded towards the WCB (word concept buffer, Fig. 1) as part of the syntactic-semantic and semantic processing modules. The WTB states allow the activation of all word span states within the five arc type (or span) STMs (see Fig. 1 and Fig. 2) for generating the final syntactic parsing information (i.e., the sequence of dependency arc specifications activated in the dependency arc buffer, DAB, Fig. 1 and Fig. 3). The DAB states together with the WCB states allow the role assignment within the semantic processing component by activating appropriate concept states in the semantic target representation STMs (for agent, patient, action, event time, etc.). To associate subject and object with agent and patient correctly in an online approach (this model), it may be necessary to temporarily store some concepts in cache STMs (e.g., the subject noun for patient in case of a German passive verb construction).

The semantic target representation can be bound (binding buffer BiB, Fig. 1) towards a compound activation pattern representing the sentence meaning as one item (one S-pointer in terms of the SPA) in a further STM, i.e., the sentence meaning memory, SMM (Fig. 1 and see text below).

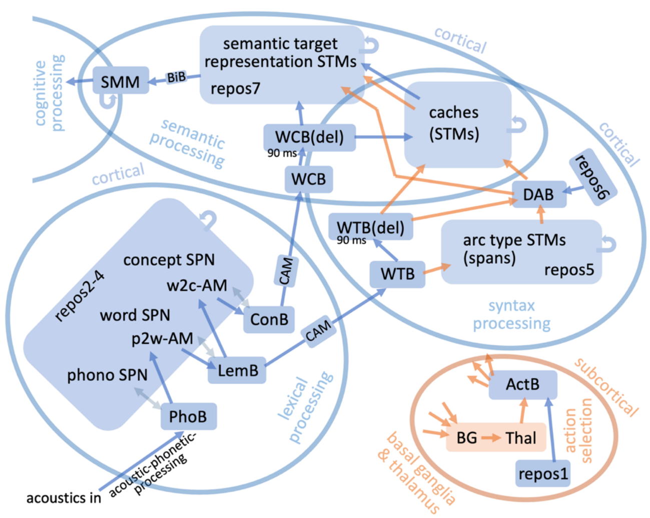

Figure 1: The neural network architecture for the perception part of speech processing within our neural model comprising acoustic-phonetic, lexical, syntactic, and semantic processing modules (for abbreviations see text).

The naming conventions used for specifying the neural network components in Fig. 1 are: *SPN – semantic pointer network* (Crawford et al., 2015; Kröger et al., 2016), *AM – associative memory*, *CAM – clean-up associative memory* (Kröger et al., 2022; Kröger, 2023), *XXB (XxxB) – buffer* for phonological states (Xxx=Pho), lemmata (Xxx=Lem), (word) concepts (Xxx=Con, XX=WC), word types (XX=WT), actions (Xxx=Act), and syntactic dependency arc information (DA), *XXB(del) – delay buffer* (neural activation is delayed here by 90 ms), *STM – short-term memory* for dependency arc types (also representing constituent span information), for syntactic-semantic cash information, and for sentence meaning SM following a binding operation done within a binding buffer (BiB). *Red arrows* indicate neural activation management between cortical neural network elements and the action selection component. *Repositories* are long-term storages for lexical information as well as for syntactic and semantic specifications. The associative memories within the mental lexicon use this repository information.

The *arc-type STM component* (Fig. 1, also called *word span component*) currently comprises five STMs, i.e., verb STM, aux STM, noun1 STM, noun2 STM and preposition STM. The action control component activates these sates within the STMs (see Fig. 2) leading to specific word spans and further on leading to a sequence of dependency arcs for the sentence (Fig. 3) within the DAB.

Particularly, it can be seen from Fig. 2 that the verb and aux STMs uncovers the verb-aux word span (see VerbSTM in Fig. 2). This span is labeled AUX in Fig. 4 and Fig. 5 and known as verb clamp in Standard German (see above and see also Foth, 2006). This verb clamp appears in case of past tense constructions like “hatte (das) getrunken” (“had drunk (that)”) where the object (“that”) is covered by the verb clamp. And the clamp appears in case of passive constructions like “wurde (von jemandem) getrunken” (“was drunk (by someone)”) where the preposition phrase (“by someone”) is covered by the verb clamp (see also Musan, 2022). The noun1 and noun2 STMs uncover the noun spans as subject or object (“der kalte Kaffee” (“the cold coffee”) and “der große Benno” (“(the) big Benno”)) ((see NounSTM in Fig. 2)). The preposition STM together with the following noun STM uncovers the preposition span (see PrepPhraSTM in Fig. 2).

The syntactic description of a sentence is activated in the dependency arc buffer (DAB, Fig. 1) and a typical DAB activation pattern is given in Fig. 3. Two different types of naming conventions can be seen in this figure, i.e., small types as used in the English dependency grammatical naming conventions, as used, e.g. by de Marneffe et al. (2021) and large types, following the German dependency grammatical naming conventions, as given, e.g., by Foth (2006).

The action selection procedure for activating the sequence of specific *dependency arc labels* (also called dependency arc names, arc types, or arc specifications, e.g., Root, Root2Subj, Root2Obj, Root2PrePhr, Subj2Det, Subj2Attr, etc.) is based on the word or phrase span information activated in the arc type (or word span) STMs (Fig. 1) together with the currently incoming (delayed) information from WTB(del), i.e., the de-

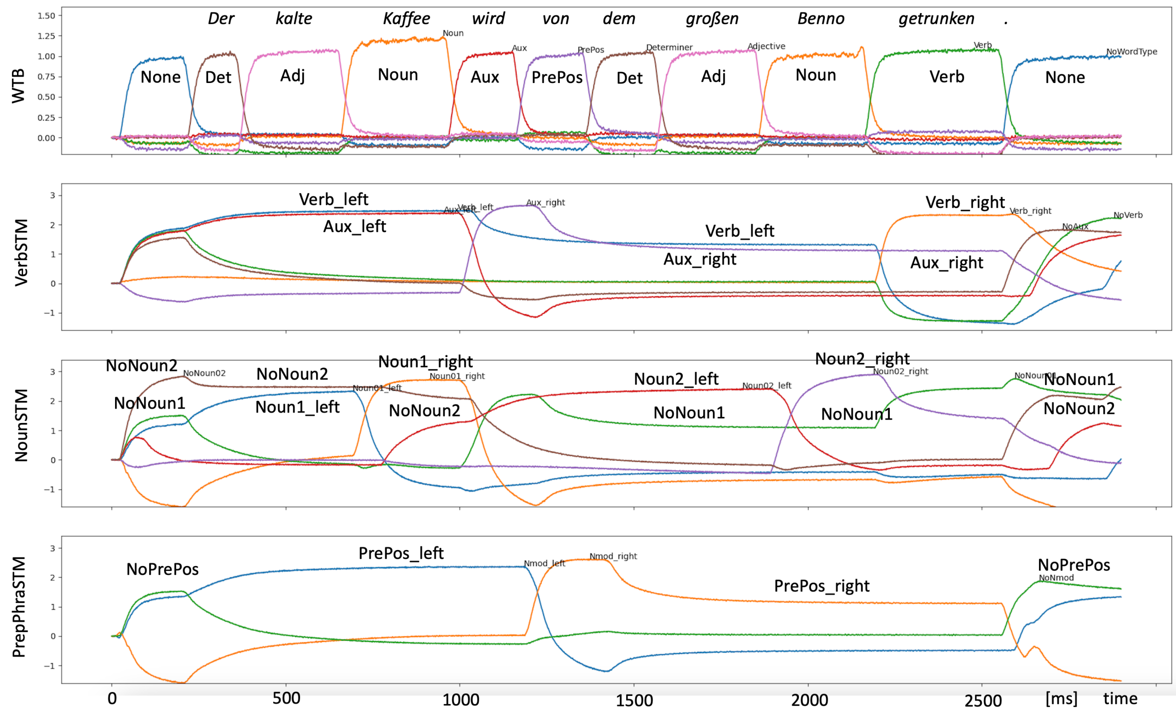

Figure 2: Activation signals in the word type buffer WTB (on top) and in the verb, aux, noun, and preposition STMs (below) in the arc type STMs component of the syntax processing module (Fig. 1). The activations auf verb and aux STM are overlayed and named verbSTM. The activations of noun1 STM and noun2 STM are overlayed in the row labeled as nounSTM. The processed sentence is: “Der kalte Kaffee wird von dem großen Benno getrunken” (“The cold coffee is drunk by (the) big Benno”).

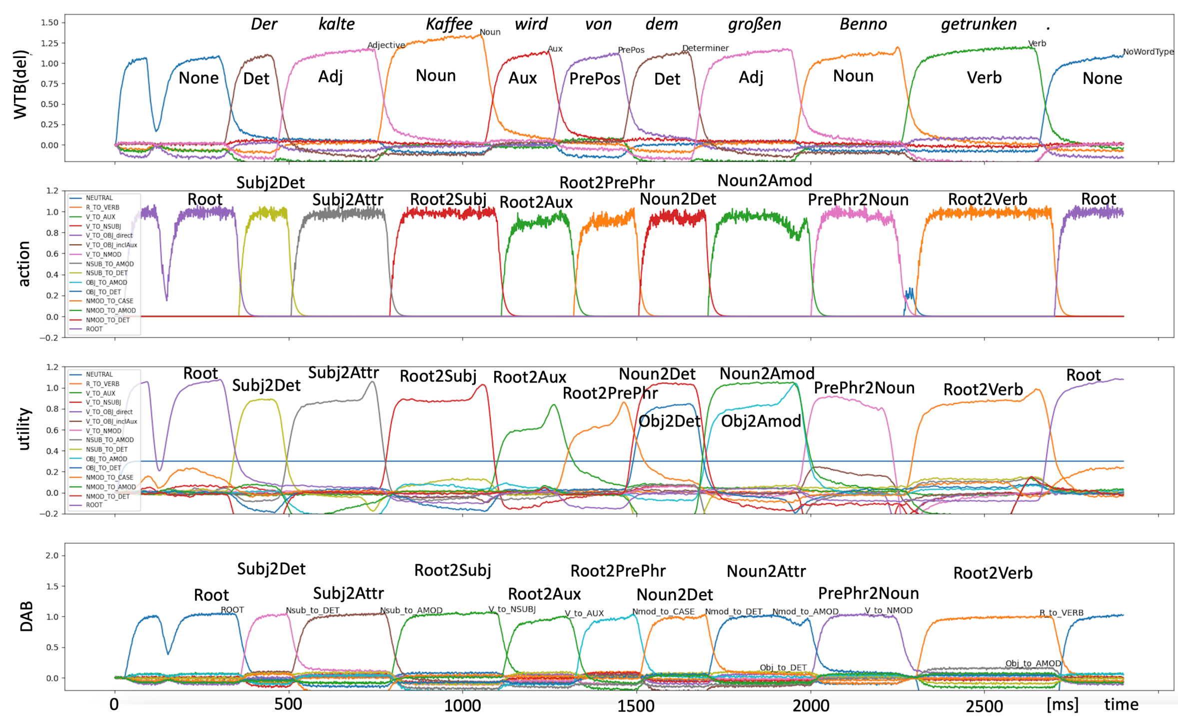

Figure 3: Activation signals in dependency arc buffer DAB (last row), delayed word type signal WTB(del) (top row), and utility values and action activations (second and third row) leading to a correct selection of dependency arc types. The processed sentence is identical in Fig. 2 and Fig. 3.

layed word type sequence. The resulting syntactic information can equivalently be written as dependency tree (Fig. 4) or as dependency arc structure (Fig. 5).

From the utility values appearing within the action selection process for arc types (activated in DAB), it can be seen, that, here the activation of adjective and determiner branches should probably not be separated for noun spans appearing as part of a prepositional word spans and for noun spans directly representing a subject or an object (see the comparable utility values in the action selection process for determiner and adjective branching of the prepositional phase ("Noun2Det" and "Noun2Attr") and for a (here not activated) object phrase ("Obj2Det", "Obj2Attr") in Fig. 3.

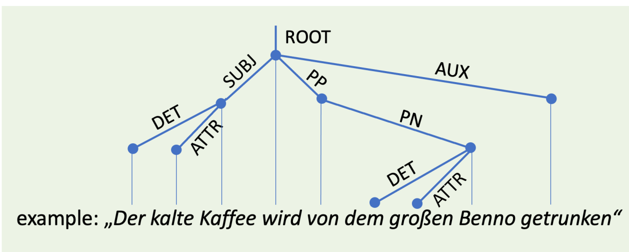

Figure 4: Dependency tree visualization for the syntactic structure of the processed sentence "Der kalte Kaffee wird von dem großen Benno getrunken" ("The cold coffee is drunk by (the) big Benno"). The dependency tree branch naming conventions are those used for Standard German (Foth, 2006).

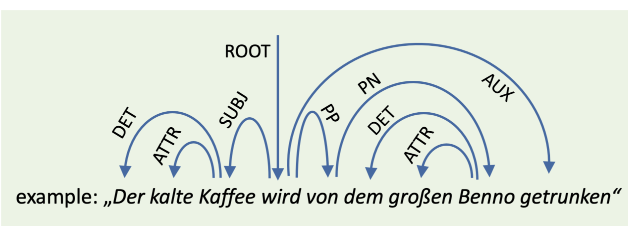

Figure 5: Equivalent dependency arcs visualization for the syntactic structure of the same sentence as given in Fig. 4.

If the syntactic analysis is competed the semantic target information STMs for event-time ("*present*: the event is happening right now" vs. "*past*: the event already happened") for action (e.g., "*trinken*"), for agent (e.g. "*Benno*") and for patient (e.g. "*Kaffee*") can be bound together (by using binding buffers) in the SMM (sentence meaning STM, see Fig. 1) resulting in an activation of one single neural pattern, e.g., "event_*time* * *present* + *action* * *trinken* + *agent* * *Benno* + *patient* * *Kaffee*". For the concept of adding neural activation patterns and for the concept of binding neural activation patterns in binding or convolution buffers see, e.g., Stewart & Eliasmith, 2014).

The model architecture displayed in Fig. 1 was incrementally developed by processing simple to complex sentences. The model was finally capable of processing all 70 test sentences in an error-free manner if starting with the already correct sequence of phonological words.

Discussion and Conclusions

A preliminary version of a neurobiologically inspired syntactic processing module has been implemented as part of a large-scale spiking neuron model for speech processing. This module takes the sequence of word types making up a sentence and generates a sequence of dependency arc specifications, which – together with the word concept information – can be used to generate and activate a neural state representation of the sentence meaning in form an event specification comprising event time, agent, patient, and action. The syntactic processing module within our model computer-implements the phrase level component postulated by Friederici (2011). Thus, the syntactic module can be assumed to be in the frontal operculum (lexical lemma activation) and in the anterior superior temporal gyrus (main syntactic processing), while further semantic processing (i.e., semantic role labeling) is done in the anterior as well as in the posterior temporal gyrus as well as in the inferior frontal gyrus (ibid., p. 1364f). Moreover, phrase level processing includes a syntactic rule system as part of the procedural memory and thus includes cortico-striatal connections and the basal ganglia-thalamus action selection system (Stocco et al., 2021).

It should be kept in mind that our neurobiologically inspired large-scale model cannot be interpreted in a sense that it is able to unfold the exact neurofunctional anatomy at a microscopic level. Moreover, neurobiologically inspired models like that under development here underline the functioning of (i.e., quantitatively computer-implement) a related (neurophysiological grounded) box-and-arrow models (e.g., Friederici, 2011) by instantiating and consolidating the qualitative processing steps outlined in that box-and-arrow models by using concrete and quantitative neurofunctional (and neurobiologically inspired) simulation elements (see the compact basic neural network elements as described in the introduction of this paper).

The developed neural network model gives an insight in the way how linguistic information could be processed in the brain: that is *hierarchical-sequentially* as well as *temporal-parallel*: (i) word type (lemma) information as well as word concept (semantic) information is sequentially forwarded from the mental lexicon to different higher-level modules (syntactic and semantic processing module); (ii) neural states (S-pointers in terms of the SPA) are aways activated in all state buffers and STMs of our model during the entire time interval of sentence processing. Thus, linguistic information is (i.e., meaningful items are) *permanently* generated in form of neural activation patterns within all buffers and short-term memories (STMs) and *permanently* forwarded towards other buffers and short-term memories within our model. Thus, syntactic, and semantic processing is done *online* and *in real-time* here (see also the model developed by Pedrelli & Hinaut 2022). It should be kept in mind that the storage of information in short-term memories is energy consuming and thus these time periods of (short-term) item storage need to be minimized. Thus, information is directly and always processed and forwarded to other processing modules in our model when it enters the syntactic-semantic processing system in form of an temporal continuous sequence of words (cf. the “now or never” bottleneck idea, Christiansen & Chater 2016; Pedrelli & Hinaut 2022). This is illustrated in our model by the fact that word span information, which is memorized for short time periods in the arc type (or word span) STMs, is permanently forwarded towards the action selection module and thus permanently used for decision making, i.e., for selecting dependency arc specifications which then are activated as S-pointers in the dependency arcs buffer (DAB).

Hierarchical processing is always associated with a hierarchy in state representations from acoustic via lexical towards syntactic and semantic representations. Pedrelli & Hinaut (2022) introduce four levels, i.e. phones, words, parts-of-speech, and semantic role labels. While words and parts-of-speech (as level-defining item classes) have a comparable temporal labeling frequency (parts-of-speech can be interpreted as syntactic word categories, see Jurafsky & Martin, 2023, chapter 8), the item frequency increases from phones to words to syntactic-semantic categories. The decrease in item frequency from words to semantic roles can be seen in our model by comparing the change of neural activation in word type buffer (WTB, Fig. 2) and in arc type STMs (all other STMs shown in Fig. 2). The change in neural states appearing in the arc type STMs is slower than in word buffer and the information appearing in the arc type STMs is the basis for semantic role labeling.

AI-based NLP models like BERT or GPT-2 as well include a kind of temporal processing because they use a step-by-step processing for incoming word sequences. Consequently, some researchers hypothesize a potential biological reality of current deep neural NLP models (Arana et al., 2023; Goldstein et al., 2022). For example, attention spotting on specific parts of sentences as well as the estimation of word prediction and estimation of surprise – i.e., deviation from expectation – are typical features of current deep network NLP models as well as of human speech processing (ibid.). Moreover, these deep neural NLP models as well include processing steps like lexical, syntactic, and semantic analysis, but it should be kept in mind that (i) the neural network architecture of these models is not comparable with the architecture of neurobiologically inspired neural network models (Pedrelli & Hinaut 2022), that (ii) the way of learning or training is different in comparison to natural speech acquisition procedures (ibid.) and that (iii) the way of temporal processing is different at least at the syntactic semantic level involved in romantic role labeling because these models need to process the whole sentence before meaning extraction can be done, while “our brain processes a sentence in an online and *anytime* fashion; we are able to partly understand the sentence (and even predict it) before it ends.” (ibid., p. 2655).

Beside the appearance of hierarchical-sequential and temporal-parallel processing in our modeling approach, is can be stated that – like in deep neural networks – *distributed vector representations* (also called *embeddings*, see Goldstein et al., 2022; McClelland et al., 2020) are used here at different levels of word information processing (S-pointer representations at phonological, lemma, and semantic level, see also Kröger et al. 2016). Distributed neural activity for item representation (one group of neurons represents different items by different activity patterns) is neurobiologically more realistic than a local representation (each item is represented by specific neurons) at cognitive levels of processing and distributed representations allows to quantify the degree of similarity between items.

Our approach so far models *bottom-up processing* while *top-down processing* is not included up to now. It is known that the mental lexicon influences word candidate activation in a top-down-manner even while the acoustic-phonetic and the phonetic-phonological analysis at segmental and at syllable level of a word is still ongoing (de Zubicaray et al., 2006). The same mechanism can be expected at the syntactic level. Word activation is influenced in a top-down-manner by early activated syntactic structure candidates for a currently processed sentence.

Our model has been developed in an incremental way by processing simple syntactic sentences first before we tried to process more complex sentences. Thus, it could be advantageous to model *the developmental process*, i.e. the development of the speech processing modules by considering onto- and phylogenetic aspects of speech and language development (Yang, 2013; Pearl, 2023). Concretely, it would be advantageous if we were able to motivate how the modules of the speech processing component grow during speech acquisition by integrating the basic neural processing elements (buffers, short-term memories, processing circuits for selection and activation of syntactic information) in a step-by-step manner into the model. This growing speech processing model should then be able to process utterances with increasing phonological, syntactic, and semantic complexity, like it appears in natural speech acquisition. The model should be able to replicate acquisition data for speech perception as well as for speech production (i.e., start to produce one-word than two- and more-word sentences etc., see Tracy, 2002). Therefore, it is important not to develop the perception side (the sentence comprehension side as it is introduced here) in isolation, but in addition to develop the sentence production side in parallel. This as well would allow to include all interactions appearing between production and perception at the semantic, at the syntactic, as well as at the lexical level of speech processing (Hickok & Poeppel, 2007).

Finally, it could be interesting not only to focus on one target language but to develop a speech processing model which is capable of processing more than one language. This allows to focus on the similarities and on the differences in semantic, syntactic, and lexical processing in different languages and may facilitate a later modeling of second and third language learning.

References

Arana, S., Lerousseau, J. P., & Hagoort, P. (2023). Deep learning models to study sentence comprehension in the human brain. *Language, Cognition and Neuroscience, Routledge Online Journal,* doi: 10.1080/23273798.2023.2198245

Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T., Rasmussen, D., Choo, X., Voelker, A., & Eliasmith, C. (2014). Nengo: a Python tool for building large-scale functional brain models. *Frontiers in Neuroinformatics, 7,* doi: 10.3389/fninf.2013.00048

Carnevale, N. T., & Hines, M. L. (2006). *The NEURON Book.* Cambridge, MA: Cambridge University Press.

Christiansen, M.H., & Chater, N. (2016) The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39,* e62. doi: 10.1017/S0140525X1500031X

Crawford, E., Gingerich, M., & Eliasmith, C. (2015). Biologically plausible, human-scale knowledge representation. *Cognitive Science, 40,* 782-821.

de Marneffe, M. C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal Dependencies. *Computational Linguistics, 47*, 255-308.

de Zubicaray, G., McMahon, K., Eastburn, M., & Pringle, A. (2006). Top-down influences on lexical selection during spoken word production: A 4T fMRI investigation of refractory effects in picture naming. *Human Brain Mapping, 27,* 864-873.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *ArXiv, abs/1810.04805,* doi: 10.48550/arXiv.1810.04805

Eliasmith, C. (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition.* New York, NY: Oxford University Press.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., et al., (2012). A large-scale model of the functioning brain. *Science, 338,* 1202-1205.

Foth, K. A. (2006). Eine umfassende Constraint-Dependenz-Grammatik des Deutschen. *Semantic Scholar,* https://api.semanticscholar.org/CorpusID:64322902

Frank, S.L. (2021). Toward Computational Models of Multilingual Sentence Processing. *Language Learning, 71,* 193-218.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Review, 91,* 1357-1392.

Gewaltig, M.O., & Diesmann, M. (2007). Nest (NEural Simulation Tool). *Scholarpedia 2,* 1430.

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., et.al (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience, 25,* 369-380.

Goodman D. F. M., & Brette R. (2009). The Brian simulator. *Frontiers in Neuroscience, 3,* 192-197.

Hickok G, & Poeppel D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393-402.

Hinaut, X., Dominey, P.F. (2013). Real-time parallel processing of grammatical structure in the fronto-striatal system: a recurrent network simulation study using reservoir computing. *PLoS One, 8(2),* e52946. doi: 10.1371/journal.pone.0052946

Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Upper Saddle River, NJ: Pearson Education, 2nd edition.

Jurafsky, D., & Martin, J.H. (2023). *Speech and Language Processing.* Draft version of third edition, https://web.stanford.edu/~jurafsky/slp3/ (accessed 27.12.2023)

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications, 82,* 3713-3744.

Koroteev, M. V. (2021). BERT: a review of applications in natural language processing and understanding. *ArXiv, abs/2103.11943,* doi: 10.48550/arXiv.2103.11943

Krauska, A., & Lau, E. (2022). Moving away from lexicalism in psycho- and neurolinguistics. *PsyArXiv*. https://doi.org/10.31234/osf.io/vyf94

Kröger, B. J., Crawford, E., Bekolay, T., & Eliasmith, C. (2016). Modeling interactions between speech production and perception: speech error detection at semantic and phonological levels and the inner speech loop. *Frontiers in Computational Neuroscience, 10, 51,* doi: 10.3389/fncom.2016.00051

Kröger, B. J., Stille, C., Blouw, P., Bekolay, T., & Stewart, T.C. (2020). Hierarchical sequencing and feedforward and feedback control mechanisms in speech production: A preliminary approach for modeling normal and disordered speech. *Frontiers in Computational Neuroscience, 14, 99,* doi: 10.3389/fncom.2020.573554

Kröger, B. J., Bekolay, T., & Cao, M. (2022). On the emergence of phonological knowledge and on motor planning and motor programming in a developmental model of speech production. *Frontiers in Human Neuroscience, 16, 844529,* doi: 10.3389/fnhum.2022. 844529

Kröger, B. J. (2023). The NEF-SPA Approach as a Framework for Developing a Neurobiologically Inspired Spiking Neural Network Model for Speech Production. *Journal of Integrative Neuroscience 22(5), 124,* doi: 10.31083/j.jin2205124

Lindes, P. (2018). The Common Model of Cognition and humanlike language comprehension. *Procedia Computer Science, 145,* 765-772.

Lindes, P. (2022). Constructing Meaning, Piece by Piece: A Computational Cognitive Model of Human Sentence Comprehension. Doctoral Thesis, University of Michigan. https://dx.doi.org/10.7302/4697

McClelland, J.L., Hill, F., Rudolph, M., Baldridge, J., & Schütze, H. (2020). Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models. *Proceedings of the National Academy of Sciences (PNAS), 17(42),* 25966-25974.

Mitropolsky, D., Collins, M. J., Papadimitriou, C. H. (2021). A Biologically Plausible Parser. *Transactions of the Association for Computational Linguistics 9*, 1374-1388.

Mitropolsky, D., & Papadimitriou, C. H. (2023). The Architecture of a Biologically Plausible Language Organ. *ArXiv, abs/2306.15364,* https://arxiv.org/abs/2306.15364

Musan, R. (2022). Syntax des Deutschen. In: Klabunde, R., Mihatsch, W., Dipper, S. (eds). *Linguistik im Sprachvergleich.* Berlin, Heidelberg: J.B. Metzler.

Pearl, L. (2023). Computational cognitive modeling for syntactic acquisition: Approaches that integrate information from multiple places. *Journal of Child Language, 50,* 1353-1373.

Pedrelli, L., Hinaut, X. (2022). Hierarchical-Task Reservoir for Online Semantic Analysis from Continuous Speech. *IEEE Transactions on Neural Networks and Learning Systems, 33,* 2654-2663.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *Semantic Scholar,* https://api.semanticscholar.org/CorpusID:160025533

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex, 59,* 33-48.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: evidence from picture naming. *Journal of Memory and Language, 54,* 228-264.

Stewart, T. C., & Eliasmith, C. (2014). Large-scale synthesis of functional spiking neural circuits. *Proceedings of IEEE, 102,* 881-898.

Stocco, A., Sibert, C., Steine-Hanson, Z., Koh, N., Laird, J. E., Lebiere, C. J., & Rosenbloom, P. (2021) Analysis of the human connectome data supports the notion of a "Common Model of Cognition" for human and human-like intelligence across domains. *NeuroImage, 235,* 118035. Doi: 10.1016/j.neuroimage.2021.118035

Tracy, R. (2002). Deutsch als Erstsprache: Was wissen wir über die wichtigsten Meilensteine des Erwerbs? *Informationsbroschüre 1/2002 der Forschungs- und Kontaktstelle Mehrsprachigkeit.* Mannheim, DE: Universität Mannheim und Institut für Deutsche Sprache. https://www2.ids-mannheim.de/prag/sprachvariation/fgvaria/Info-Spracherwerb-2002.PDF

Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron, 72,* 385-396.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *arXiv: 1706.03762* doi: 10.48550/arXiv. 1706.03762

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Expectation-based Comprehension: Modeling the Interaction of World Knowledge and Linguistic Experience. *Discourse Processes, 56(3),* 229-255.

Verma, P. (2022). Attention is all you need? Good embeddings with statistics are enough: large scale audio understanding without transformers / convolutions / BERTs / mixers / attention / RNNs or ... . *ArXiv: 2110.03183* doi: 10.48550/arXiv.2110.03183

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 6324-6327.

Zhang, M. (2020). A survey of syntactic-semantic parsing based on constituent and dependency structures. *Science China Technological Sciences, 63,* 1898-1920.