

DEEP LEARNING AND COMPETITION IN PSYCHOLINGUISTIC RESEARCH

Roman Taraban

roman.taraban@ttu.edu

Philip H. Marshall

philip.marshall@ttu.edu

Texas Tech University, USA

Received November 12, 2017; Revised December 3, 2017; Accepted December 5, 2017

Abstract. MacWhinney, Bates, and colleagues developed the Competition Model in the 1980s as an alternate to Chomskyan models that encapsulate syntax as a special-purpose module. The Competition Model adopted the functional perspective that language serves communicative goals and functions. In contrast to the premise that knowledge of language is innate, the Competition model asserts that language is learned and processed through general cognitive mechanisms that identify and weight phonological, morphological, syntactic, and semantic cues in the language experiences of the learner. These weighted cues guide the language user in the comprehension and production of language forms. The present article provides background on the Competition Model, describes machine simulations of linguistic competition, and extends the principles of the Competition Model to new machine models and applications through deep learning networks.

Keywords: *deep learning, Competition Model, cues, machine learning*

Тарабань Роман, Маршал Філіп Х. Поглиблене навчання та конкуренція в психолінгвістичних дослідженнях.

Анотація. Б. МакВінні, Е. Бейтс разом із колегами розробили модель конкуренції в 1980-х роках як альтернативу моделям Н. Чамського, які містять синтаксис як спеціальний модуль. Модель конкуренції започаткувала функціональну перспективу, відповідно до якої мова слугує комунікативній меті та функціям. На відміну від положення про те, що знання мови – вроджене, модель конкуренції постулює, що мова вивчається й обробляється за допомогою загальних когнітивних механізмів, які ідентифікують та «взважають» фонологічні, морфологічні, синтаксичні та семантичні сигнали в мовному досвіді суб'єкта навчання. Ці сигнали скеровують користувача мови до розуміння та породження мовних форм. Цю статтю присвячено опису основ моделі конкуренції, машинного моделювання мовної конкуренції. Автори спробували поширити принципи конкурентної моделі на нові машинні моделі та додатки за допомогою мереж поглибленого навчання.

Ключові слова: *поглиблене навчання, модель конкуренції, сигнали, машинне навчання*

1. Introduction

While I (RT) was on sabbatical leave in Ukraine in spring 2017 and using Google Translate on the internet, I was struck by how well it seemed to translate from English to Ukrainian. Checking into this a bit further, I found similar impressions from others who commented on the noticeable improvement. From the New York Times Magazine:

Late one Friday night in early November, Jun Rekimoto, a distinguished professor of human-computer interaction at the University of Tokyo, was online preparing

*for a lecture when he began to notice some peculiar posts rolling in on social media. Apparently Google Translate, the company's popular machine-translation service, had suddenly and almost immeasurably improved. Rekimoto visited Translate himself and began to experiment with it. He was astonished.*¹

Google first launched a translation utility in 2006. As it turns out, they changed the processing paradigm from one in which the machine tried to match phrases from one language to another, to a system that used deep learning networks. This change took place quietly and incrementally beginning in September 2016, and it was not long before users began noticing a definite difference. Statistics showed a leap in accuracy, with errors in Google translation declining 58 %-87%² with the new system, depending on the specific language being translated. The goal, of course, is for 100 % accuracy, and with deep learning networks, Google may very well approach that goal soon across a broad range of languages.

The new machine power to translate exponentially more effectively was attributed largely to deep learning networks, so I was determined to learn more about how deep learning networks worked. Expert language translation, whether carried out by humans or machines, requires native-like knowledge of both languages in order to capture a textbook level of accuracy regarding factors like case, gender, and number formation, involving morphology, inflections, and vowel changes, but also a level of colloquial knowledge that allows a person to sound as natural as a native speaker. Therefore, deep learning networks provide a valuable opportunity to look into the workings of these machines, to ask how they have achieved this lion's leap in language processing and, further, whether the operations of these artificial systems at all resemble theories of how humans process language.

2. Methods

In the remainder of this paper, we provide a brief overview of the nature of deep-learning networks. This includes a description of the networks' composition and operation. Using this background knowledge we describe how deep learning networks actually reflect well-established, but still quite controversial, theories of how natural language is acquired and processed in humans. Finally, deep learning networks have created new opportunities for advancing linguistic and psycholinguistic theories, as well as unanticipated opportunities for applied linguistic research, which we consider.

3. The Study

3.1. Deep Learning Networks

A key process in theories across diverse domains is computational modeling. Computational models are mathematical models that account for the behavior of complex systems through simulations on a computer – e.g., weather forecasting.³ In

¹ https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html?_r=0

² <http://www.sciencemag.org/news/2016/09/google-s-new-translation-software-powered-brainlike-artificial-intelligence>

³ https://en.wikipedia.org/wiki/Computational_model

academic disciplines, like cognitive psychology, artificial intelligence (AI), and linguistics, computation refers to structured approaches to representing and solving problems and, in some applications, testing theories of thinking. At a general level, we can think of computation as a set of rules or algorithms governing the mapping between inputs, intermediate states, and outputs. This includes a range of applications, from paper-and-pencil depictions of syntactic transformations to simulations of language acquisition in a machine.

A governing element in computational models involving intelligent behavior, like language acquisition and use, is knowledge representation. The typical modeling approach in the cognitive sciences and artificial intelligence research up to the 1980s was some form of rule-based representation, one example being production rules, which are IF (condition) THEN (action) rules for considering the current situation (condition) and taking some action based on current conditions. Production rules have been used to model reading and comprehending text, as one example (e.g., Just & Carpenter, 1980). A major shortcoming of production rules, and other rule-based models, was that the rules had to be hand coded. The models did not have a capacity for learning and self-modification.

The situation changed dramatically in the 1980s when researchers introduced neural models as a theoretical alternative to rule-based models. Neural models were referred to variously as connectionist models, parallel-distributed processing models, and neural networks (McClelland & Rumelhart 1986). Neural networks were considered brain-style models because they imitated properties of processing in the human brain, including the interconnection of a massive number of simple neuron-like processors, parallel processing across these processing units, and processes of activation and inhibition. Importantly, neural networks incorporated simple, but powerful, learning rules. They were able to learn on their own. When exposed to a sample of input materials, these networks extracted and applied the regular patterns in the input in ways comparable to human learning and processing.

Neural network research lulled in the late 1990s, due to slow computer processor speed and limited memory capacity. Recent technical breakthroughs allowed processing speeds to increase exponentially, and large-scale distributed memory storage increased the memory capacity of machine systems. Neural networks provided the foundation for present-day deep learning networks. The technical breakthroughs, as well as a near-obsession worldwide with deep learning, has revitalized neural network models and artificial intelligence research. Indeed, the machine achievements are so compelling that well-known figures from Elon Musk¹ to Steven Hawking² have speculated on the potential collapse of civilization due to AI.

So what are deep learning networks? Deep learning networks are an extension of neural networks and represent a class of computational models, which includes Recurrent Neural Networks, Convolution Neural Networks, Artificial Neural Networks, and Neural Machine Translation, among others.³ Deep learning networks apply machine learning approaches, typically implementing Bayesian-based learning

¹ <https://www.cnn.com/2017/09/04/elon-musk-says-global-race-for-ai-will-be-most-likely-cause-of-ww3.html>

² <http://www.bbc.com/news/technology-30290540>

³ https://en.wikipedia.org/wiki/Artificial_neural_network

algorithms. A deep learning network consists of input and output units and several layers of internal processing units. Input is passed through the system to an output layer, and the system strengthens the influence of some inputs and minimizes the influence of others, in order to achieve the best match between inputs and desired outputs. Deep learning networks are used quite extensively for a wide variety of applied problems, like vision based classification (e.g. distinguishing different breeds of dogs), marketing analysis, and social trends. These networks are also used for natural language processing.

How do deep learning networks inform theories of human processing of language, and how do deep learning networks relate to linguistic and psycholinguistic theories? According to Noam Chomsky, perhaps the most influential language theorist of the 20th century, neural networks are not very good models of human language processing.¹ For Chomsky, deep learning networks have applied value, but do not contribute to our understanding of the nature of language representation and processing in humans. The advent of the technology that has propelled Google Translate to significantly lower translation errors allows for fast search of massive databases of information and returns plausible outputs. However, for Chomsky this is inadequate. It may allow one to predict the most likely translation for a sentence from one language to another, but it does not reveal anything fundamental about the nature of language, or its essence. Chomsky acknowledges that deep learning networks may have practical value, but because they do not add to our understanding of language and the fundamental computations carried out in language processing, they do not represent good basic science. Gary Marcus made similar points and extended the case against statistical learning like that implemented by deep learning networks.² Intelligent machines, Marcus notes, need to be able to compute causal relationships, make logical inferences, and extract and integrate abstract knowledge about objects. These are things, Marcus asserts, deep learning networks cannot do.

3.2. The Competition Model

One goal that is achieved by deep learning networks, as well as other machine-based tools for language analysis, is discovering the statistical regularities in the input data. Deep learning networks have been successful at language translation because they choose the most probable pattern-match from one language to another. Basically, these networks tease apart signal from noise. Can this property of deep learning networks be aligned with the search for the intrinsic properties of language structure and processing? A psycholinguistic theory called the Competition Model developed by MacWhinney and colleagues (Bates & MacWhinney, 1989; MacWhinney, 1987; 2001), and further expanded in the Unified Competition Model (Li & MacWhinney, 2013; MacWhinney, 2008; 2012), suggests that this is possible. “The Competition Model is a framework for the crosslinguistic study of language use. It is designed to capture facts about the comprehension, production, and acquisition of language by real human beings, across a variety of qualitatively and quantitatively distinct language types” (Bates & MacWhinney, 1989: 3). The Competition Model applies to first (L1) and second (L2) language learning (MacWhinney, 2012), and the model also accounts for language disorders (Presson & MacWhinney, 2011).

¹ <https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/>

² <http://bmia.bmt.tue.nl/people/BRomaneny/Courses/BMIE-DeepLearning2017/Wikipedia-DeepLearning2017.pdf>

The Competition Model is rooted in the premise that “the forms of natural languages are created, governed, constrained, acquired and used in the service of communicative functions” (Bates & MacWhinney, 1982). The model provides significant resources to language processing through its interactive nature. Specifically, different linguistic data types (phonological, morphological, syntactic, and semantic) are brought to bear immediately and in parallel in the ongoing activation and selection of linguistic operators in language production and comprehension. The interactive, parallel-processing, and integrative characteristics of the Competition Model are in sharp contrast to models in generative linguistics. Hauser, Chomsky, and Fitch (2002), for example, make a distinction between the faculty-of-language in the narrow sense (FLN) and a faculty-of-language in the broad sense (FLB). The FLN is a narrow syntactic computational system that generates internal representations and maps these to the FLB, which maps formal semantic specifications of the FLN to conventional symbolic and conceptual representations. The Competition Model rejects this type of modular approach to performance and the related encapsulation of function as characteristic of the language processing mechanism. Instead, the model postulates that the functions of the FLN and FLB function in parallel and interactively.

In sum, the Competition Model approaches the different aspects of language, including syntax, phonology, semantics, and pragmatics as being interconnected and interactive. Mechanistically, the Competition Model works within a probabilistic framework, wherein the operative elements are weighted cues representing elements of an utterance. Cue strength is determined by the empirical properties of the language instances in the learner’s experience. The Model has been validated in English and cross-linguistically in studies on Hungarian, German, Russian, Hebrew, and French, among others (e.g., Devescovi, D’Amico, Smith, Mimica, & Bates, 1998; MacWhinney, 2001; MacWhinney, Bates, & Kliegl, 1984; MacWhinney, Leinbach, Taraban, & McDonald, 1989; Sokolov, 1988).

3.3. Deep Learning and Competition

A fundamental property of deep learning parallels a central principle of the Competition Model, which is computing the probabilistic weighting of cues in the input according to the predictive strength (reliability) of the cues. The basic equation for estimating cue strength in the Competition Model is the conditional probability:

$$P(X | y) = P(X \cap y) / P(y)$$

where X is a linguistic category or operation, given some cue y . An example from German is $P(\text{Subject} | \textit{der}) = P(\text{Subject} \cap \textit{der}) / P(\textit{der})$, in other words, the probability that a noun phrase (NP) is the sentence Subject is estimated by the probability that Subject and *der* occur together divided by the probability of the occurrence of the definite article *der*. Alternatively, one could express this relationship in terms of the frequency with which the two variables occur in the experience of the learner, i.e., the frequency with which Subject and *der* occur together divided by the frequency of the definite article *der*. The parallel processing of these cues during language use allows the language system to achieve fast and accurate comprehension and production of linguistic forms.

Cue weights dynamically change as language skill develops in an individual learner, and cue weights vary depending on whether the language is L1 or L2. Further, cue-based processing describes the nature of language learning and processing across different languages. As stated by Bates and MacWhinney (1989: 12), “We describe linguistic representations in terms of a complex set of weighted form–function mappings, a dynamic knowledge base that is constantly subject to change. The emphasis in the Competition Model on probabilistic rules does not mean that we ignore the powerful laws that separate one language from another.” The predictive validity of this principle was demonstrated in the applied developmental research of MacWhinney and Bates, and colleagues, well before neural networks and deep learning were developed. It has been substantiated through a number of empirical tests and computer simulations in a number of languages, including German (MacWhinney et al., 1989; Taraban, McDonald, & MacWhinney, 1989), French (Taraban & Roark, 1996), Russian (Taraban & Kempe, 1999), and artificial languages (Taraban, 2004).

3.4. Deep Learning and Discourse

The psycholinguistic principle of competition involves the discovery by the language learner of the probabilistic weighting of a range of phonological, morphological, syntactic, and semantic markers in a language. This competition and probabilistic perspective provides an alternative to traditional Chomskyan approaches that encapsulate syntax and separate it from meaning and function. Bates and MacWhinney (1989:5) regard this as a limitation: “Chomsky (1975) argues for a kind of autonomy of syntax that would cut it off from the pressures of communicative functions. In the Chomskyan vision, language is pure and autonomous, unconstrained and unshaped by purpose or function.” Research has shown that competition and deep learning networks provide a plausible account of learning and processing at the sentence level. Language use, however, goes beyond the level of individual utterances, and occurs in the context of a larger discourse (Langacker 1989).

Deep learning networks provide a means for the development of the Competition Model beyond the sentence, at the level of discourse. Discourse applies to spoken and written communications. Individuals and groups have linguistic registers. These consist of the specific vocabulary and ways of communicating that characterize individual thinking, as well as the shared conversations and practices of a group (Taraban 2017). Linguistic registers allow individuals to activate specific mindsets and engage in shared behaviors with others of similar mind. Discourse analysis allows for a discovery of those mindsets through the application of deep learning, Bayesian algorithms, and other analytic methods capable of extracting the patterns of communication in a corpus of written or spoken samples.

3.5. Deep Learning in the Classroom

Returning from my sabbatical leave in Ukraine, I (RT) was energized by the technological breakthroughs in psycholinguistics taking place through machine models and applications. I decided to restructure my graduate seminar in

psycholinguistics to incorporate many of the freely available machine applications. My goals were threefold: 1) to use machine applications as a means of introducing students to the perspective embodied in the Competition Model of language arising from the acquisition of probabilistic cues for the comprehension and production of linguistic forms; 2) to provide a clear contrast to more traditional Chomskyan grammars that encapsulate syntax from the use of language to communicate meaning; and 3) to provide students with hands-on experiences using machine-based language resources. As the fall semester comes to an end, my students are building chatbots for their final class presentations. Chatbots are computer programs that interact with a human user via a computer or some other device. Chatbots can be designed to carry on conversations on particular topics or to complete routine tasks, like registering a person for an event, taking a sandwich order, or advising on the best statistical method to apply to a set of data. Earlier in the semester, students completed assignments using IBM Watson Natural Language Classifier,¹ Linguistic Inquiry and Word Count (LIWC),² Latent Semantic Analysis (LSA),³ and Coh-Metrix.⁴ All of these computer applications are based on the probabilistic nature of language. In these programs, in deep learning applications, and in the Competition Model, “probabilistic” does not mean “by chance.” Rather, these applications identify the most regular patterns – i.e., the high probability patterns – in the input. This allows these applications to analyze ordinary texts for meaning and coherence and to output assessments of analytic thinking, self-confidence, emotionality, rhetorical themes, and coherence of the composition, among others.

4. Conclusion

In this paper we have tried to communicate a sense of the excitement surrounding new possibilities associated with deep learning. Although the media tends to focus on end-of-the-world scenarios associated with this technology, there are more-immediate possibilities for expanding a class of psycholinguistic theories that take a probabilistic cue-based approach to language learning and processing. It is high time to go beyond an analysis of sentences to incorporate discourse analyses as well. Deep learning networks afford the resources to explore ways of integrating language comprehension starting with the morphological constituents at the sentence level up to individual and group registers at the discourse level. Students need to be exposed to these ideas and related tools in order to prepare them for the next wave of psycholinguistic research and instruction (MacWhinney, 2015).

References

- Bates, E. & MacWhinney, B. (1982). A functionalist approach to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: the state of the art*. New York: Cambridge University Press.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In: *The Crosslinguistic Study of Sentence Processing*, (pp 3-76). B. MacWhinney and E. Bates (Eds.), New York: Cambridge University Press.

¹ <https://www.ibm.com/watson/services/natural-language-classifier/>

² <https://liwc.wpengine.com/>

³ <http://lsa.colorado.edu/>

⁴ <http://cohmetrix.com/>

- Devescovi, A., D'Amico, S., Smith, S., Mimica, I., & Bates, E. (1998). The development of sentence comprehension in Italian and Serbo-Croatian: Local versus distributed cues. In: *Syntax and Semantics: Vol. 31. Sentence Processing: A Cross-Linguistic Perspective*, (pp. 345-377). D. Hillert (Ed.), San Diego: Academic Press.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What it is, who has it, and how did it evolve? *Science*, 298, 1569-1579.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Langacker, R. (1989). *Foundations of cognitive grammar. Vol. 2: Applications*. Stanford: Stanford University Press.
- Li, P., & MacWhinney, B. (2013). Competition model. In: *The Encyclopedia of Applied Linguistics*. C. A. Chapelle (Ed.), Malden, MA: Wiley.
- MacWhinney, B. (1987). The competition model. In: *Mechanisms of Language Acquisition*, (pp.249-308). B. MacWhinney (Ed.). Hillsdale, NJ: Lawrence Erlbaum.
- MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In: *Cognition and Second Language Instruction*, (pp. 69-90). P. Robinson (Ed.), New York: Cambridge University Press.
- MacWhinney, B. (2008). A Unified Model. In: *Handbook of Cognitive Linguistics and Second Language Acquisition*, (pp. 341-371). P. Robinson & N. Ellis (Eds.). Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney B. (2012). The logic of the Unified Model. In: *The Routledge Handbook of Second Language Acquisition*, (pp. 211-227). S. Gass and A. Mackey (Eds.). New York: Routledge.
- MacWhinney, B. (2015). Multidimensional SLA. In: *Usage-Based Perspectives on Second Language Learning*, (pp. 22-45). S. Eskilde and T. Cadierno (Eds.). New York: Oxford University Press.
- MacWhinney, B., Bates, E. & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127-150.
- MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language*, 28, 255-277.
- McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*. Cambridge, MA: MIT Press.
- Presson, N. & MacWhinney, B. (2011). The Competition Model and language disorders. In: *Handbook of Psycholinguistic and Cognitive Processes*, (pp. 31-48). J. Guendozi, F. Loncke, and M. Williams (Eds.). New York: Psychology Press.
- Sokolov, J. L. (1988). Cue validity in Hebrew sentence comprehension. *Journal of Child Language*, 15, 129-156.
- Taraban, R. (2004). Drawing learners' attention to syntactic context aids gender-like category induction. *Journal of Memory and Language*, 51(2), 202-216.
- Taraban, R. (2017). Hate, white supremacy, PTSD, and metacognition. In: *Improve With Metacognition* [online]. L. Scharff, A. Richmond, & J. Draeger (Eds.). Retrieved from: www.improvewithmetacognition.com.
- Taraban, R., & Kempe, V. (1999). Gender processing in native and non-native Russian speakers. *Applied Psycholinguistics*, 20, 119-148.
- Taraban, R., McDonald, J., & MacWhinney, B. (1989). Category learning in a connectionist model: Learning to decline the German definite article. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic categorization* (pp. 163-193). Philadelphia: Benjamins.
- Taraban, R., & Roark, B. (1996). Competition in learning language-based categories. *Applied Psycholinguistics*, 17, 125-148.