Natural language processing in artificial intelligence (NLP AI) and natural language processing algorithms relating to grammar as a foreign language

Article · October 2018

CITATION
1

READS
11,513

1 author:

Javier Julian Enriquez
Universitat Politècnica de València
84 PUBLICATIONS **29** CITATIONS

SEE PROFILE

Natural language processing in artificial intelligence (NLP AI) and natural language processing algorithms relating to grammar as a foreign language

Abstract

Universal Artificial Intelligence (UAI) within an algorithm design framework might play a major role in pointing up important technological contribution and innovative development that could be useful for content analysis (semantic) of large databases, for example. In view of that, this paper aims to provide a background for further research based on other authors' analysis such as James Le, Luonge, Maldonado Fonken, Sutskever, or Vinyals, whose results may be the epistemological framework in Applied Linguistics and trends of theoretical language research in comparative form between native and foreign languages.

Key words: Universal Artificial Intelligence (UAI), algorithm design framework, content analysis (semantic), theoretical framework, Applied Linguistics, comparative form.

INTRODUCTION

Hutter (2004) undertakes highly focused research on Universal Artificial Intelligence (UAI), thus pointing out that we need to take closer examination of this approach looking for theories that can guide our search for intelligent algorithms, which measure an agent's ability to achieve goals in a wide range of environments, whose relevant research fields are:

- Computer science (artificial intelligence, machine learning);
- Engineering (information theory, adaptive control);
- Economics (rational agents, game theory);
- Mathematics (statistics, probability);
- Psychology (behaviorism, motivation, incentives),
- Philosophy (reasoning, induction, knowledge).

Progressing in the exposed sense, according to Everitt and Hutter (2018), Universal Artificial Intelligence (UAI) is an increasingly well-studied foundational theory for Artificial Intelligence (AI), based on ancient principles in the philosophy of science and modern developments in information and probability theory, which also offers a deeper appreciation of fundamental problems such as the induction problem and the exploration-exploitation dilemma. UAI, then, as indicated by these authors, is a formal, foundational theory for AI that gives a precise answer to the question of what is the optimal thing to do for essentially any agent acting in essentially any environment. In doing so, the UAI theory is composed of the following four components underpinned by foundational theories that have contributed greatly to scientific progress in many fields, as Everitt and Hutter (2018: 18) point out:

- Framework. Defines agents and environments, and their interaction.
- Learning. The learning part of UAI is based on Solomonoff induction. The general learning ability this affords is the most distinctive feature of UAI.
- Goal. In the simplest formulation, the goal of the agent will be to maximize reward.
- Planning. (Near) perfect planning is achieved with a simple expectimax search.

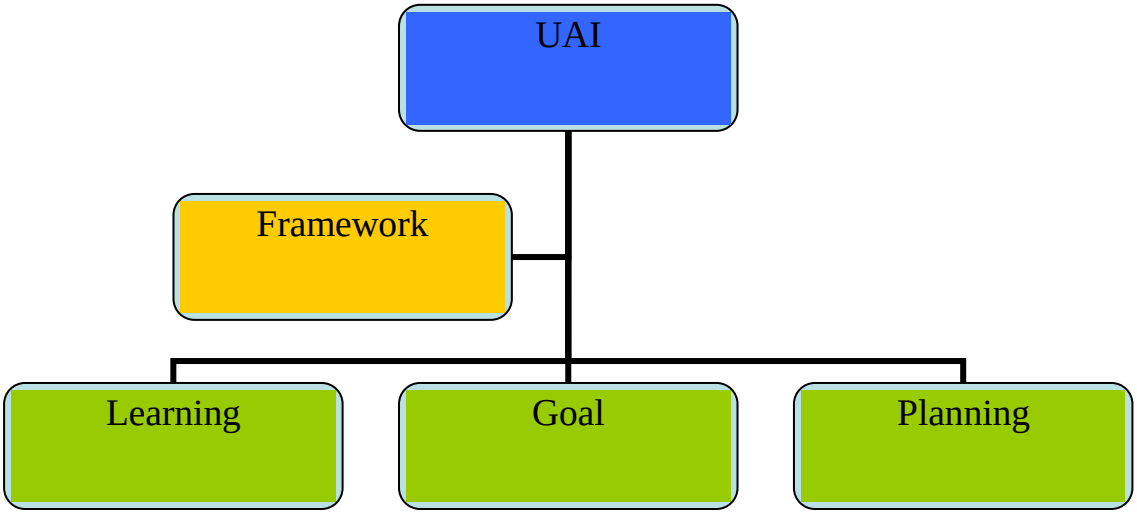


Figure 1. The UAI theory (Adaptation based on Everitt and Hutter's study, 2018)

As a result, this epistemological model may provide us with a framework for further research; particularly, its study might include, for example, the use of the *Master Algorithm*, conceived of as *machine learning* and *new switchboard for Higher Education*. In this respect, we should underline the importance that language, as a system, is acquired from linguistic data presented in natural settings, concerned as well with the path followed by linguistics proper to the configuration of a pragmatic or communicative paradigm from a Blending and Online Learning perspective (Tom Vander Ark, 2015).

Similarly, we believe that we could endeavor to include such epistemological principles within an algorithm design framework from the knowledge, procedure, skill, value, and critical analysis perspective underpinned by the aforementioned foundational theories that would also contribute greatly to scientific progress in many fields by utilizing the *application of holistic strategic management PTT Know How* for future education, for example. Appropriately, several interesting implications may arise from these approaches from our point of view:

1. A breakthrough and groundbreaking piece of research for development of solutions, thus focusing on the study of cultural deficiencies and its impact on leadership, strategic management and governance;
2. Challenges by implementing creative and pioneering solutions for future technology;
3. Rigorous methods and procedures, thus offering a sophisticated insight into instructional as well as educational approaches;
4. Application of technology in instruction, computer-assisted language learning, and learner-centered instruction;
5. A relationship between education, literacy, and identity within a sociocultural context at the cutting edge of discussions about what matters as humankind and civilization learn.

Fittingly, Vinyals et al. (2015) undertake highly focused research on Grammar as a Foreign Language by using a Recurrent Neural Network with attention

mechanism to generate sentence parse trees. Hence, according to Vinyals et al. (2015), syntactic constituency parsing is a fundamental problem in linguistics and natural language processing that has a wide range of applications. As stated, then, by Vinyals et al. (2015), this problem has been the subject of intense research for decades, and as a result, there exist highly accurate domain-specific parsers. These authors, thus, highlight that the computational requirements of traditional parsers are cubic in sentence length, and while linear-time shift-reduce constituency parsers improved in accuracy in recent years, they never matched state-of-the-art. Furthermore, Vinyals et al. (2015) emphasize that standard parsers have been designed with parsing in mind; they underline that the concept of a parse tree is deeply ingrained into these systems, which makes these methods inapplicable to other problems.

In doing so, they show that generic sequence-to-sequence approaches can achieve excellent results on syntactic constituency parsing with relatively little effort or tuning. In addition, while they find the model of Sutskever et al. (2014) not to be particularly data efficient, they discover that the attention model of Bahdanau et al. (2014) is highly data efficient, as it matches the performance of the Berkeley Parser when trained on a small human-annotated parsing dataset. Finally, they show that synthetic datasets with imperfect labels could be highly useful, as their models have substantially outperformed the models that have been used to create their training data. Accordingly, they suspect it is the case due to the different natures of the teacher model and the student model: the student model is likely viewed the teacher's errors as noise, which it is able to ignore. As a result, Vinyals et al. (2015) point out that such an approach is so successful that they obtain a new state-of-the-art result in syntactic constituency parsing with a single attention model, which also means that the model is exceedingly fast. Consequently, these authors conclude that this work showing domain independent models with excellent learning algorithms can match and even outperform domain specific models.

THEORETICAL BACKGROUND

In line with the above, the trends of theoretical language research in comparative form between native and foreign languages indicate a trend toward new approaches, as is the case of *neural machine translation*, by achieving a translation performance comparable to the existing state-of-the-art phrase-based system on the task of English-to-French translation, for example, as suggested by Bahdanau et al. (2014). This model proposed by Bahdanau et al. (2014: 1-2), then, predicts a target word based on the context vectors associated with certain source positions and all the previous generated target words, which show that such a method of jointly learning to align and translate attains significantly improved translation performance over the basic encoder–decoder approach. Like this manner, Bahdanau et al. (2014: 2) point out:

> In the Encoder–Decoder framework, an encoder reads the input sentence, a sequence of vectors x = (x1, · · · , xTx), into a vector c. 2 The most common approach is to use an RNN such that
>
> $$ht = f (xt, ht−1) (1)$$
>
> and
>
> $$c = q (\{h1, · · · , hTx \}),$$
>
> where ht ∈ R n is a hidden state at time t, and c is a vector generated from the sequence of the hidden states. f and q are some nonlinear functions. Sutskever et al. (2014) used an LSTM as f and q ({h1, · · · , hT }) = hT , for instance.

Along these lines, Bahdanau et al. (2014: 9) test the proposed model, called RNNsearch, on the task of English-to-French translation. Their experiment reveals that the proposed RNNsearch outperforms the conventional encoder–decoder model (RNNencdec) significantly, regardless of the sentence length and that it is much more robust to the length of a source sentence:

> Let us consider another sentence from the test set:
>
> This kind of experience is part of Disney's efforts to"extend the lifetime of its series and build new relationships with audiences via digital platforms that are becoming ever more important," he added.
>
> The translation by the RNNencdec-50 is

> Ce type d'experience fait partie des initiatives du Disney pour "prolonger la dur ´ ee´ de vie de ses nouvelles et de developper des liens avec les ´ lecteurs numeriques ´ qui deviennent plus complexes

Progressing in the exposed sense, Bird et al. (2009: 27) explore language bottom-up with the help of texts and the Python programming language in addition to exploiting our knowledge of language and computation by building useful language technologies. Equally, they take the opportunity to step back from the practical details of code in order to paint a bigger picture of natural language processing. As so, Bird et al. (2009: 33) undertake a detailed research on the difficult task of building technologies that understand language by using superficial yet powerful techniques instead of unrestricted knowledge and reasoning capabilities, thereby equipping us with the knowledge and skills to build useful Natural Language Processing systems, and to contribute to the long-term aspiration of building intelligent machines. For this to be achieved, these authors design the Python programming language together with an open source library called the *Natural Language Toolkit* (NLTK), which includes extensive software, data, and documentation, all freely downloadable from http://www.nltk.org/. Example:

> For our purposes, we will think of a text as nothing more than a sequence of words and punctuation. Here is how we represent text in Python, in this case the opening sentence of Moby Dick: >>> sent1 = ['Call', 'me', 'Ishmael', '.']

Anyway, the process of improving or developing, then, or of getting nearer to achieving a practical Natural Language Processing (NLP) approach may become somewhat of a tricky issue. However, some authors, such as Collobert et al. (2011: 2527-2528), believe that their contribution to this field of study represents an important milestone to tackle training algorithms scaling linearly that are most able to benefit from such tremendous progress in computer hardware. In that case, using NLP tagger as a strong basis, these authors rely on large unlabeled data sets and let the training algorithm discover internal representations that prove useful for all the tasks of interest. Example:

> Also called shallow parsing, chunking aims at labeling segments of a sentence with syntactic constituents such as noun or verb phrases…
>
> Semantic Role Labelling (SRL) aims at giving a semantic role to a syntactic constituent of a sentence, State-of-the-art SRL systems consist of several stages: producing a parse tree, identifying which parse tree nodes represent the arguments of a given verb, and finally classifying these nodes to compute the corresponding SRL tags (Collobert et al., 2011: 2496-2497).

Following with the review of approaches in the topic that concerns us, Jean et al. (2014: 1-2) propose a method based on the importance sampling that allows us to use a very large target vocabulary without increasing training complexity about Neural Machine Translation (NMT) model to solving machine translation. Then, to deal with Neural Machine Translation and Limited Vocabulary Problem, they suggest an approximate training algorithm based on (biased) importance sampling that allows us to train an NMT model with a much larger target vocabulary, which uses a single neural network trained jointly to maximize the translation performance. Like this, Jean et al. (2014: 5) evaluate the proposed approach in English→ French and English→ German translation tasks. They train the neural machine translation models using only the bilingual, parallel corpora made available as a part of WMT'14. For each pair, the datasets they use are:

- English→French2 :

Europarl v7, Common Crawl, UN,

News Commentary, Gigaword

- English→German:

Europarl v7, Common Crawl, News Commentary

On English →French and English →German translation tasks, Jean et al. (2014: 8) note that the neural machine translation models trained using the proposed method performs as well as, or better than, those using only limited sets of target words, even when replacing unknown words:

As performance of the RNNsearch-LV models increased when only a selected subset of the target vocabulary was used during decoding, this makes the proposed learning algorithm more practical.

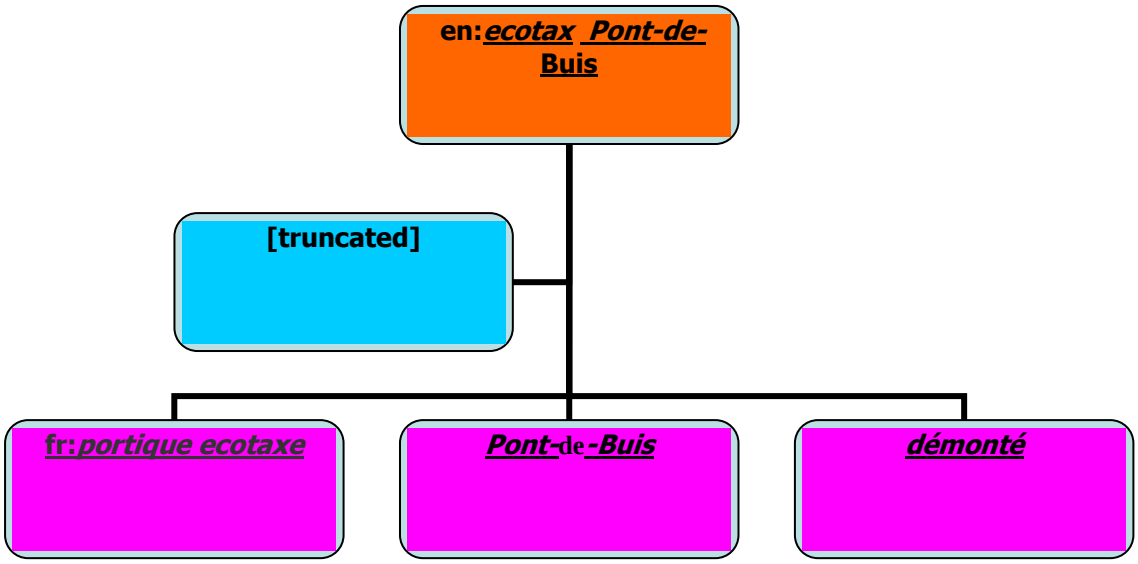
In spite of these assumptions, as indicated by Luonge et al. (2014: 1-2), conventional Neural Machine Translation (NMT) systems are unable to translate rare words because they have a fixed modest-sized vocabulary, which forces them to use the *unk* symbol to represent the large number of out-of vocabulary (OOV) words, as illustrated in Figure 3:

Example of the rare word problem – An English source sentence (*English*), a human translation to French (*French*), and a translation produced by one of our neural network systems *(nn*) before handling OOV words. We highlight words that are unknown to our model. The token *unk* indicates an OOV word. We also show a few important alignments between the pair of sentences.

en: The *ecotax* portico in *Pont-de*-Buis , . . . [truncated] . . . , was taken down on Thursday morning

fr: Le *portique ecotaxe* ´ de *Pont-de-Buis* , . . . [truncated] . . . , a été *démonté* jeudi matin

nn: Le *unk* de *unk* `a *unk* , . . . [truncated] . . . , a été pris le jeudi matin



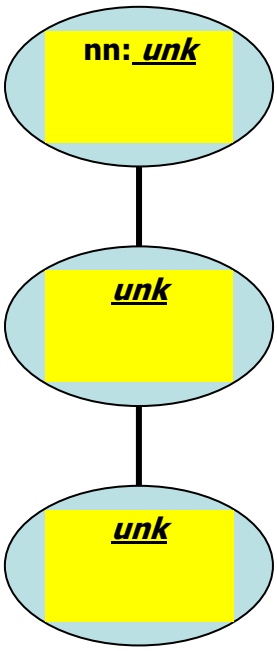


Figure 3. Example of the rare word problem (Adaptation based on Luonge et al study, 2014)

To try hard to find out more information about NMT, Luonge et al. (2014: 2) use, then, the model of Sutskever et al. (2014), which uses a deep LSTM to encode the input sequence and a separate deep LSTM to output the translation. It follows that, consistent with Luonge et al. (2014: 2), the encoder reads the source sentence, one word at a time, and produces a large vector that represents the entire source sentence in which the decoder is initialized with this vector and generates a translation, one word at a time, until it emits the end-of-sentence symbol < eos >. A key advantage of these authors' technique is the fact that it is applicable to any NMT system and not only to the deep LSTM model of Sutskever et al. (2014). This way, Luonge et al. (2014: 9) rightly and cogently state, "A technique like ours is likely necessary if an NMT system is to achieve state-of-the-art performance on machine translation."

As follows, we must, next, delve back into Deep Neural Networks (DNNs) model, put forward by Sutskever et al. (2014: 1-2), which shows that a straightforward application of the Long Short-Term Memory (LSTM) architecture can solve general sequence-to-sequence problems. This method,

thus, is based on a multilayered Long Short-Term Memory (LSTM) to map the input sequence to a vector of a fixed dimensionality, and then another deep LSTM to decode the target sequence from the vector, as illustrated in figure 4:

> Our model reads an input sentence "ABC" and produces "WXYZ" as the output sentence. The model stops making predictions after outputting the end-of-sentence token. Note that the LSTM reads the input sentence in reverse, because doing so introduces many short-term dependencies in the data that make the optimization problem much easier.

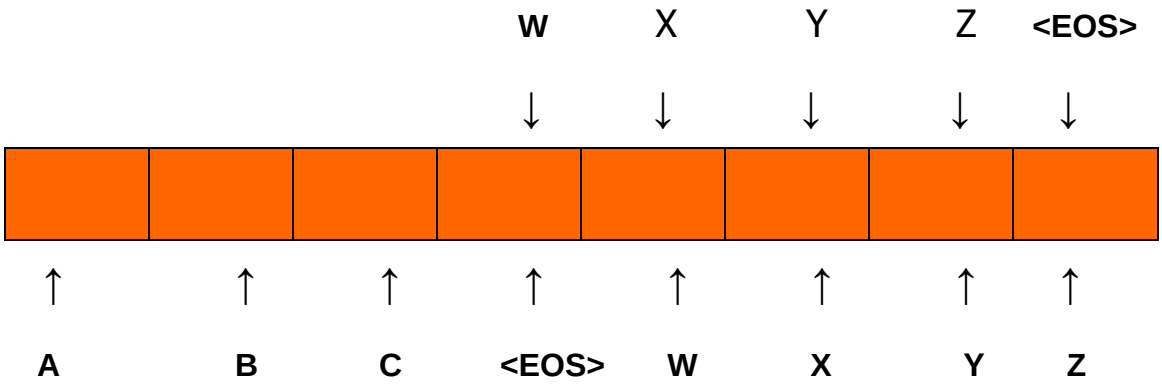


Figure 4. Multilayered Long Short-Term Memory (LSTM) model based on an English to French translation task (Adaptation based on Sutskever et al study, 2014)

As can be observed, as maintained by Sutskever et al. (2014: 6), one of the attractive features of their model is its ability to turn a sequence of words into a vector of fixed dimensionality.

Like this manner, Sutskever et al. (2014: 8) show that a large deep LSTM with a limited vocabulary can outperform a standard SMT-based system whose vocabulary is unlimited on a large-scale MT task. They conclude that the success of their simple LSTM-based approach on MT suggests that it should do well on many other sequence learning problems, provided they have enough training data.

METHODOLOGY AND DISCUSSION

We have reviewed some theoretical background carried out by several authors, who have provided us with important research by showing domain independent models with excellent learning algorithms that can match and even outperform domain specific models in search of a theoretical framework in Applied Linguistics and trends of theoretical language research in comparative form between native and foreign languages. In this sense, we believe that this algorithmic paradigm for determining such domain specific models might be well carried out by a groundbreaking and valuable reference tool, *the Matrix-Q translator as an epistemological model to put into practice* (Maldonado Fonken, 2018). The Matrix-Q Knowledge Approach (MKA) entails a very important and innovative epistemological framework both in theoretical and practical terms for other users to put its tools into practice. In this way, it does focus mostly on PSL and PSPL (Primordial Symbolic Language and Primordial Symbolic Programming Language); the first one can be found in any culture-related language, while the second one is an advanced version of the first one, developed by utilizing Matrix-Q Knowledge, tools and skills, whose purpose is Artificial Intelligence (AI) programming. Accordingly, Maldonado Fonken (2018) states his own views quite openly as he reveals several interesting implications, which may arise from this cutting edge Project for future entrepreneurs and users:

- It may really contribute to the acquisition of users' target skills and competences as in the technical knowledge necessary as in the values that if observed would bring humanity to a more advanced stage of development, in culture and civilization globally, sustainable development, in addition to personally, individual growth and collective ability to communicate and collaborate.
- It may actually contribute to the acquisition of competences as regards matrix of cultural knowledge (ancient cultures and civilizations, the awareness as well as modern scientific knowledge, like for example neuroscience, education, language, and others).
- Its proposal may clearly show that The PTT Brain Mapping System - PTT *Neuroscience* is an incredible advance to develop a brain Mapping System based on PTT Primordial Mathematics, PTT Primordial Algorithms, and PTT Know How. In this sense, this approach may entail, for example,

a wider framework of Second Language Acquisition by putting forward a systematic classification from a neuroscience perspective according to this new model and new PTT Tool being developed.

- It may undoubtedly show that Matrix-Q Civilization: 9 Types of Choices and the Matrix-Q Alternative for Human Species Conscious Evolution is an incredible advance in engineering, innovation, and methodology, which entails the necessary knowledge, experience, and resources for the organizations, societies, companies, practitioners, consultants, and costumers in their target field. In this sense, its analysis makes clear the fundamental and theoretical values that underline this empirical and scientific research, which are evident throughout all its great value-laden pragmatic and groundbreaking pieces of research. That is, an education system in which students do promote their critical spirit through scientific methods, emotional intelligence, global culture development, while influencing their leadership development, communicative and collaborative skills in favor of society.

The next figure shows the aforementioned algorithmic paradigm for determining such domain specific models:

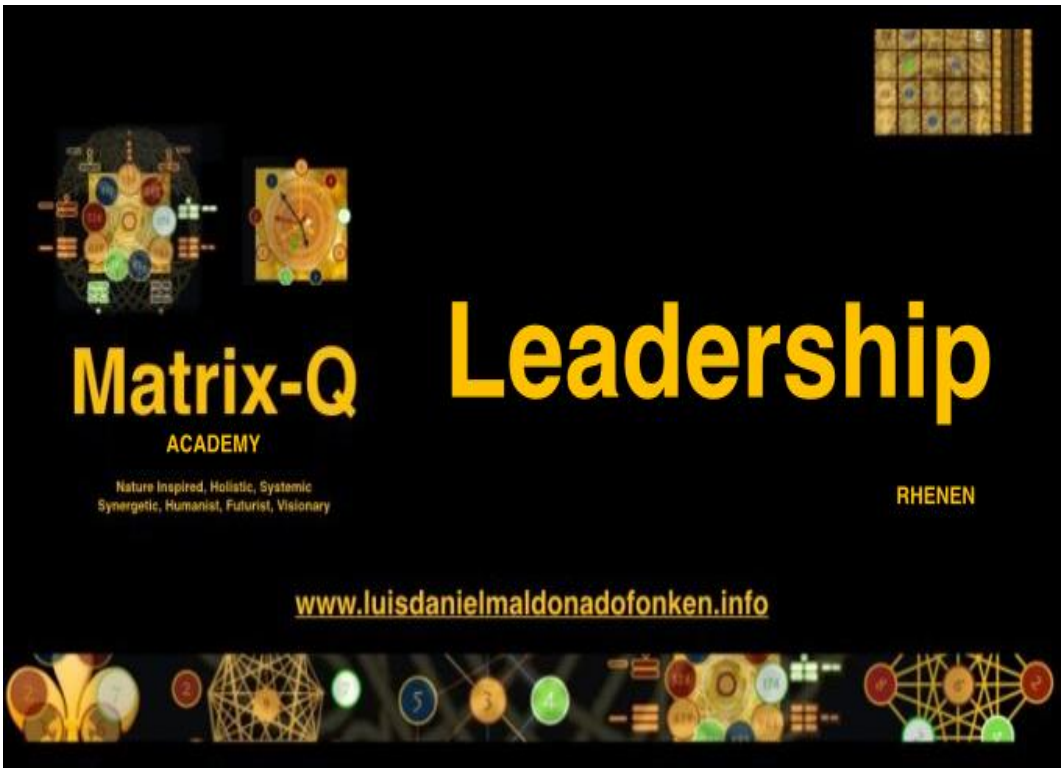


Figure 5. The Matrix-Q Translator Research Project. (Adaptation based on Maldonado Fonken's study, 2018)

Along these lines, Maldonado Fonken (2018: 3) emphasizes that the Matrix-Q Translator Project goal is to develop a Matrix-Q Tool which is able to read and organize, identify and recognize, nature inspired language and translate it into ciphers, algorithms, and Matrix-Q PSL & PSPL Languages (Primordial Symbolic Language, and Primordial Symbolic Programming Language):

> A translator of this kind will communicate fields of study, principles, methods, tools and knowledge utilized by ancient cultures and civilizations through geometrical values (structures) as well as through tones (frequencies) and algorithms (sequence of ciphers), which are correlated to specific cognitive skills, brain activity, body regions, emotions, behavior, archetypes, knowledge and methods and in general to 9+ classes or natural, artificial and human phenomena.

This tool, subsequently, as stated by Maldonado Fonken (2018: 4), will be useful for multidisciplinary research as a single language platform that serves for purpose of communication and collaboration between professional disciplines, from natural science to social science, from technology to leadership, from human potential studies to quantum physics.

Now then, could these key concepts bring a certain model, which might idealize the setting within a given conceptual framework?

As regards these key concepts, in keeping with Maldonado Fonken (2018: 3), we have seen earlier that a translator of this kind, for example, will communicate fields of study, principles, methods, tools and knowledge utilized by algorithms (sequence of ciphers), which are correlated to specific cognitive skills. Correspondingly, in our view, a model capable of idealizing the target setting within this given framework would be that proposed by Vinyals et al. (2015: 5), in generating well-formed trees, for instance:

> The LSTM+A model trained on WSJ dataset only produced malformed trees for 25 of the 1700 sentences in our development set (1.5% of all cases), and the model trained on full high-confidence dataset did this for 14 sentences (0.8%). In these few cases where LSTM+A outputs a malformed tree, we simply add brackets to either the beginning or the end of the tree in order to make it balanced.

As we can see, these authors underline their model by making assumptions to remove the error included by natural variations in the concerned system[1].

The figure 6 illustrates this model:

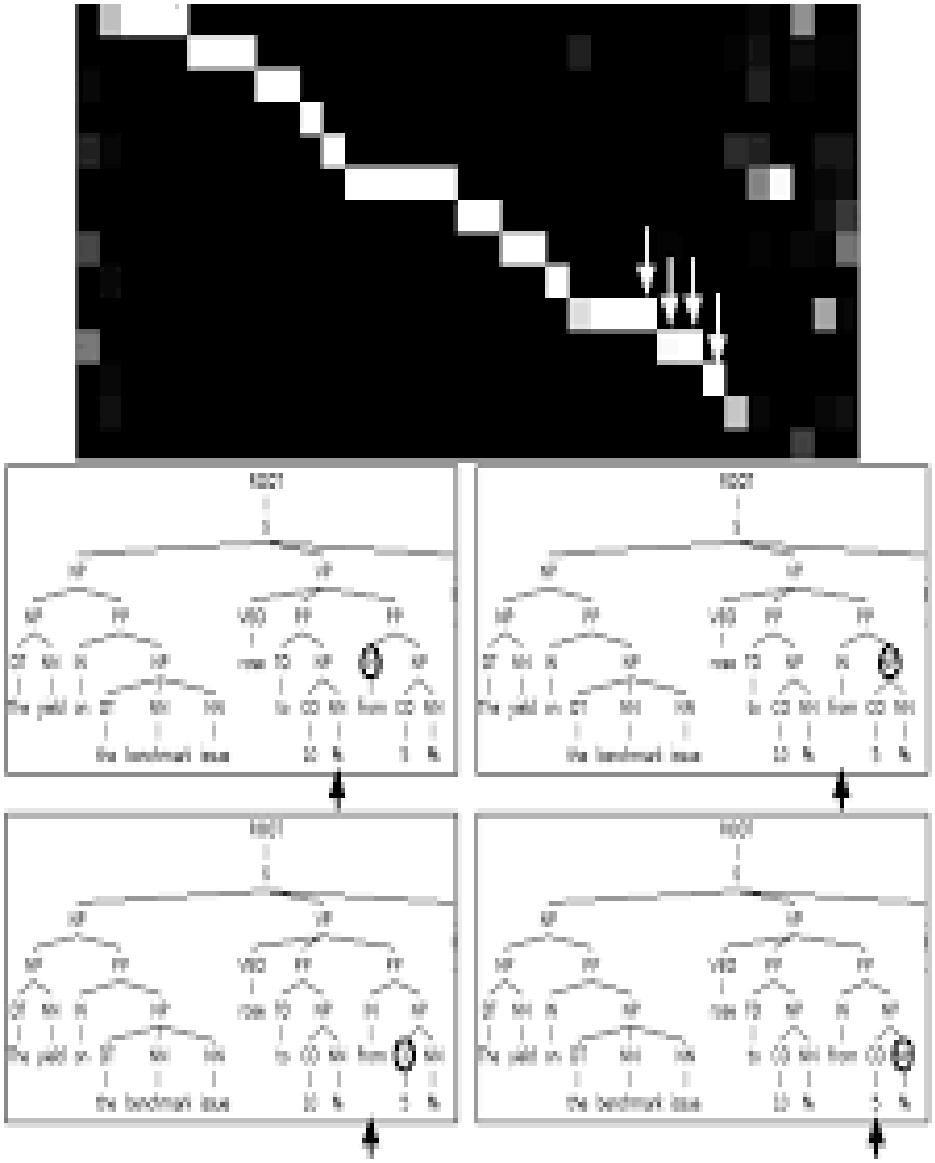


Figure 6. Attention matrix. (Adaptation based on Vinyals et al study, 2015)

[1] Adaptation based on Admin's study, April 3, 2013. Source : (https://www.differencebetween.com/difference-between-model-and-vs-framework/)

In this manner, consistent with Vinyals et al. (2015: 7), figure 6 shows an example of the attention model trained only on the WSJ dataset. In relation to their statements, it is clear that the model focuses quite sharply on one word as it produces the parse tree from the attention matrix, where each column is the attention vector over the inputs. Equally, they indicate that it is also clear that the focus moves from the first word to the last monotonically, and steps to the right deterministically when a word is consumed.

Followed by these implicit assumptions in that case, could any models fit into a specific methodology within a given conceptual framework, with very good criteria-referenced tools and knowledge utilized by algorithms correlated to specific cognitive skills?

Our reasoning being based on the premise that an algorithm is a sequence of steps designed in order to solve a challenge as pointed out by Maldonado Fonken (2018), Le (2016) states in the same way that "decision tree" as a method, allows us to approach the problem in a structured and systematic way to arrive at a logical conclusion. Let us look at the figure 7 to get the impression of how it looks like:

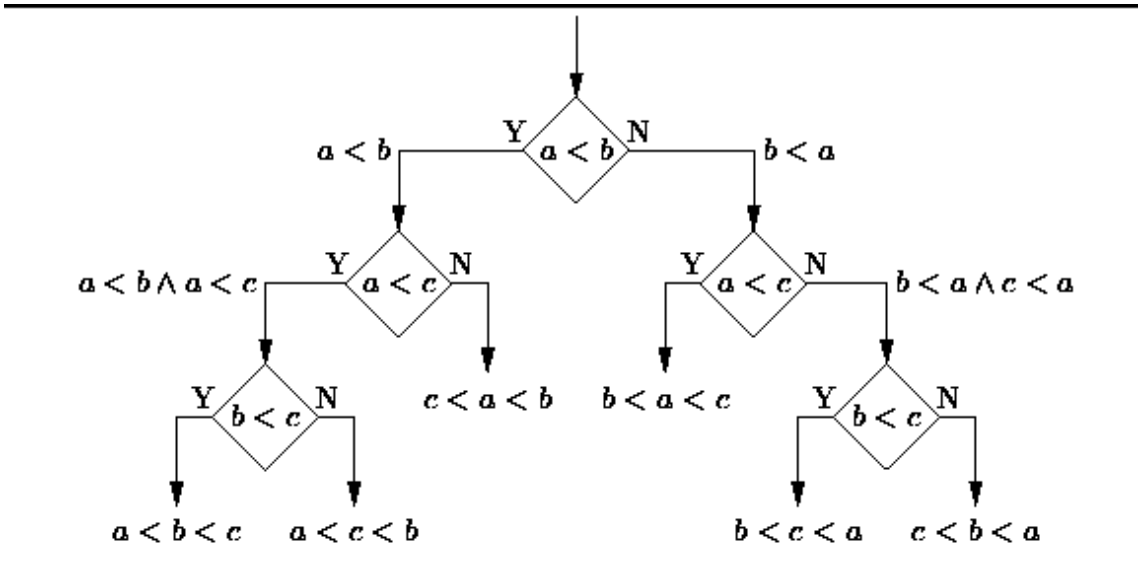


Figure 7. Decision Tree. (Adaptation based on James Le's study, 2016)

Suitably, based on the rationale behind these hypotheses, so therefore, we would be able to verify the validity of the model as a specific methodology put forward by Le (2016) within the given conceptual framework algorithms correlated to specific cognitive skills suggested by Maldonado Fonken (2018). Such results may bring about one of the epistemological framework in Applied Linguistics and trends of theoretical language research in comparative form between native and foreign languages that we have been trying to seek. However, these approaches might be considered only as a theoretical construct that other users may not get it as a verification of hypotheses to validate.

Maybe, still some cognitive strategies and skills should be taken into account in the Applied Linguistics setting in relation to theoretical language research in comparative form between native and foreign languages to validate such theories. To this end, O'Malley and Chamot (1990: 119-120) carry out an ESL descriptive study, which might shed light on this tricky issue regarding some learning cognitive strategies in Second Language Acquisition:

- Auditory representation. Planning back in one's mind the sound of a word, phrase, or longer language sequence.
- Elaboration. Relating new information to prior knowledge, relating different parts of new information to each other, or making meaningful personal associations with the new information.
- Recombination. Constructing a meaningful sentence or larger language sequence by combining known elements in a new way.
- Grouping. Classifying words, terminology, or concepts according to their attributes or meaning.

This latest hypothesis, then, might solve to some extent ambiguity, one of the major problems of natural language, which is usually faced in syntactic level that has subtask as lexical and morphology, which are concerned with the study of words and word formation, as maintained by Khurana et al. (2017: 3).

After we have outlined our proposals of the concepts and practices involved in this paper, let us go on to the conclusions.

CONCLUSION

Throughout this paper, we undertook a critically review on Natural Language Processing in Artificial Intelligence (NLP AI) and Natural Language Processing algorithms relating to grammar as a foreign language. Particularly, based on other authors' background, we were intent on providing important technological contribution and innovative development that could be useful for both content and semantic analysis of large databases, whose results may be the theoretical framework in Applied Linguistics and trends of theoretical language research in comparative form between native and foreign languages. On the basis, then, of this rationale, other experts and researchers would be able to add further prospective models fitting into specific methodologies within any given conceptual frameworks by evaluating first-rate criteria-referenced tools of very definite views on algorithms correlated to specific cognitive skills for both content and semantic analysis of large databases leading to the theoretical framework just suggested.

During the whole review of this article, we consequently analyzed and explored previous theoretical and hypothetical background. Like this, we looked at Grammar as a Foreign Language by using a Recurrent Neural Network with attention mechanism to generate sentence parse trees, which show that generic sequence-to-sequence approaches, can achieve excellent results on syntactic constituency parsing with relatively little effort or tuning, in accordance with some authors such as Vinyals. In this respect, we believe that we could complement this theory with some learning cognitive strategies in relation to Second Language Acquisition as that related to constructing a meaningful sentence or larger language sequence by combining known elements in a new way, as indicated by O'Malley and Chamot in an attempt to solve one of the major problems of natural language, ambiguity. Similarly, we would be able to complement this proposed model with a target word based on the context vectors associated with certain source positions and all the previous generated target words, since they show that such a method of jointly learning to align and translate attains significantly improved translation performance over the basic encoder–decoder approach, as maintained by Bahdanau.

Considering our overview of the key concepts and practices as well as of the theoretical constructs involved in this methodology analysis, it is also worthwhile to have another look at other authors. In this manner, in this paper we focused on the Python programming language, which equips us with the knowledge and skills to build useful Natural Language Processing systems, and to contribute to the long-term aspiration of building intelligent machines, in keeping with some authors such as Bird. As follows, we also delved into Deep Neural Networks (DNNs) model, put forward by Sutskever, which shows that a straightforward application of the Long Short-Term Memory (LSTM) architecture can solve general sequence-to-sequence problems. On the other hand, on the importance sampling that allows us to use a very large target vocabulary without increasing training complexity about Neural Machine Translation (NMT) model to solving machine translation, consistent with Jean, for instance.

In view of that, we suggested a conceptual framework, according to Maldonado Fonken, as the connectivity amongst all models of research undertaken. We supported this view because Matrix-Q Tool is able to read, organize, identify, and recognize, nature inspired language and translate it into ciphers, algorithms, and Matrix-Q PSL & PSPL Languages (Primordial Symbolic Language, and Primordial Symbolic Programming Language). These key concepts focusing on a translator of this kind, subsequently, will communicate fields of study, principles, methods, tools and knowledge utilized by algorithms (sequence of ciphers), which are correlated to specific cognitive skills. Correspondingly, in our view, some models capable of idealizing the target setting within this given framework would be those proposed by Vinyals and James Le, for example. As far as Vinyals' model is concerned, we looked, thus, at the LSTM+A model trained on WSJ dataset, which focuses quite sharply on one word as it produces the parse tree from the attention matrix, where each column is the attention vector over the inputs. In addition, it moves from the first word to the last monotonically, and steps to the right deterministically when a word is consumed. For the same reason, just as we suggested that an algorithm is a sequence of steps designed in order to solve a challenge as pointed out by Maldonado Fonken, so we noticed that James' model "decision

tree" is a method that allows us to approach the problem in a structured and systematic way to arrive at a logical conclusion. Suitably, based on the rationale behind these hypotheses, so therefore, we implied that we would be able to verify the validity of the model as a specific methodology put forward by James Le within the given conceptual framework algorithms correlated to specific cognitive skills proposed by Maldonado Fonken.

Anyway, such premises may bring about one of the epistemological framework in Applied Linguistics and trends of theoretical language research in comparative form between native and foreign languages that we have been trying to seek. However, new empirical research may be needed as these approaches might be considered only as a theoretical construct that other users may not get it as a verification of hypotheses to validate. Even so, we believe that this investigation may provide us with a useful starting point for further research whereby using either a survey or focus groups could bring in key data to support the validity and reliability of these theories to develop this target conceptual framework.

BIBLIOGRAPHICAL REFERENCES

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Source of work cited:

(https://arxiv.org/pdf/1409.0473.pdf)

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".

Source of work cited:

(http://117.3.71.125:8080/dspace/bitstream/DHKTDN/6460/1/Natural%20Language%20Processing%20with%20Python.4149.pdf)

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*(Aug), 2493-2537.

Source of work cited:

(http://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf)

Everitt, T., & Hutter, M. (2018). Universal Artificial Intelligence. In *Foundations of Trusted Autonomy* (pp. 15-46). Springer, Cham.

Source of work cited:

(https://link.springer.com/chapter/10.1007/978-3-319-64816-3_2).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Source of work cited:

(http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.676.4320&rep=rep1&type=pdf)

Hutter, M. (2004). *Universal artificial intelligence: Sequential decisions based on algorithmic probability*. Springer Science & Business Media.

Source of work cited:

(http://www.hutter1.net/ai/suaibook.pdf).

Jean, S., Cho, K., Memisevic, R., & Bengio, Y. (2014). On using very large target vocabulary for neural machine translation. *arXiv preprint arXiv:1412.2007*.

Source of work cited:

(https://arxiv.org/pdf/1412.2007.pdf)

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2017). Natural Language Processing: State of The Art, Current Trends and Challenges. *arXiv preprint arXiv:1708.05148*.

Source of work cited:

(https://arxiv.org/ftp/arxiv/papers/1708/1708.05148.pdf)

Le, J. (2016). *The 10 Algorithms Machine Learning Engineers Need to Know*. KDnuggets.

Source of work cited:

(https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html)

Luong, M. T., Sutskever, I., Le, Q. V., Vinyals, O., & Zaremba, W. (2014). Addressing the rare word problem in neural machine translation. *arXiv preprint arXiv:1410.8206*.

Source of work cited:

(https://arxiv.org/pdf/1410.8206.pdf)

Maldonado Fonken, L. D. (2018). The Matrix-Q Translator Research Project. 10.13140/RG.2.2.12490.26561.

Source of work cited:

(https://www.researchgate.net/publication/326989207_The_Matrix-Q_Translator_Research_Project)

Manning, C. D., Manning, C. D., & Schütze, H. (1999).*Foundations of statistical natural language processing*. MIT press.

Source of work cited:

(http://thuvien.thanglong.edu.vn:8081/dspace/bitstream/DHTL_123456789/4027/1/cs511-1.pdf)

O'Malley, M. J., & Chamot, A. U. (1990).*Learning strategies in second language acquisition*. Cambridge University Press.

Schmidhuber, J., & Hochreiter, S. (1997). Long short-term memory. *Neural Comput*, *9*(8), 1735-1780.

Source of work cited:

(http://people.idsia.ch/~juergen/lstm2003tutorial.pdf)

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104-3112).

Source of work cited:

(http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf)

Tom Vander Ark, (2015). 8 Ways Machine Learning Will Improve Education. **Source of work cited**:

(http://www.gettingsmart.com/2015/11/8-ways-machine-learning-will-improve-education/).

Vinyals, O., Kaiser, Ł., Koo, T., Petrov, S., Sutskever, I., & Hinton, G. (2015). Grammar as a foreign language. In *Advances in Neural Information Processing Systems* (pp. 2773-2781).

Source of work cited:

(https://papers.nips.cc/paper/5635-grammar-as-a-foreign-language.pdf)

RESOURCES IN INTERNET

1. **Difference Between Model and Framework**- April 3, 2013 posted by **Admin:**

   (https://www.differencebetween.com/difference-between-model-and-vs-framework/)

2. **The 10 Algorithms Machine Learning Engineers Need to Know** - 18 Aug 2016 posted by **James Le**:

   (https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html)