A connectionist model of reading with error correction properties

Max Raphael Sobroza Marques, Xiaoran Jiang, Olivier Dufor, Claude Berrou, Deok-Hee Kim-Dufor

Télécom Bretagne
Télécom Bretagne, Technopôle Brest-Iroise, 29238 Brest, France
{max.sobrozamarques, xiaoran.jiang, olivier.dufor, claude.berrou, dh.kimdufor}@telecom-bretagne.eu

Abstract

Recent models of associative long term memory (LTM) have emerged in the field of neuro-inspired computing. These models have interesting properties of error correction, robustness, storage capacity and retrieval performance. In this context, we propose a connectionist model of written word recognition with correction properties, using associative memories based on neural cliques. Similarly to what occurs in human language, the model takes advantage of the combination of phonological and orthographic information to increase the retrieval performance in error cases. Therefore, the proposed architecture and principles of this work could be applied to other neuro-inspired problems that involve multimodal processing, in particular for language applications.

1. Typoglycemia: the error correction abilities of the brain when reading

Typoglycemia refers to the ability to read words wherein letters are transposed.

> Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are. [1]

The message says that according to a supposed study from Cambridge University, we can easily read sentences containing words with transposed letters. As stated by (Grainger and Whitney, 2004), this ability to correct words with scrambled letters indicates the existence of a special way of decoding input information that allows us to access the correct meaning of the word.

Based on an important literature in psycholinguistics on priming effects in reading, it has been shown that this ability involves for example the open bigram theory. The models such as the SERIOL model (Whitney, 2001) manipulating open bigrams provide a good alternative to solve this issue while only considering the orthographic information. However, many controversies about both theories and models subsist and experimental counterexample have been plentiful.

The aim of this work[2] is to implement a multimodal approach of a connectionist model to correct errors in reading process. For this purpose, we will use techniques based on neural cliques (See section 3.).

The dual-route cascaded (DRC) model (Coltheart et al., 2001) is a framework that includes the notion of multimodality in word recognition. Several studies found evidence to support this framework especially from dyslexia impairments. The proposed architecture is based on DRC framework. It is also composed of two routes: a lexical route and a sublexical route. The sublexical route is usually used for the recognition of non-words and it contains a grapheme-phoneme rule system that converts letters in phonemes representations. On the other hand, the lexical route includes orthographic and phonological representations of words.

The input errors are classified into five different categories, illustrated in the examples below.

Class	**Target word: *AIRPLANE***
Transposition	AIR**LP**ANE
Erasure	AIR_LANE
Substitution	AIR**K**LANE
Deletion	AIRPLNE
Insertion	AIR**F**PLANE

Table 1: Categories of typographical errors in words

Recent research links the behavior of correction reading with priming effect properties. Among them, we lay emphasis on *relative-position priming*, *transposition priming*, *superset priming* and *phonological priming*.

When a stimulus shares sub-sequences of letters with the target word, the reading process is facilitated. This effect is called relative-position priming (Grainger et al., 2006). The proportion of shared sub-sequences and the word length count to the presence of priming effect.

Transposition priming is an effect that facilitates the reading process when the stimulus has the same letters as the target word and when there are small variations in the order of the letters (Schoonbaert and Grainger, 2004). Moreover, (Perea et al., 2003) observed a stronger priming effect when transpositions are in adjacent positions. (Rayner et al., 2006) reported that sentences with transposed letters decrease the reading rate. However, these sentences are much easier to read than sentences with substitutions. They found evidence that when transpositions concern the ending or beginning of words, the sentences are more difficult to read. Furthermore, (Christianson et al., 2005) showed that the morpheme boundaries play an important role in visual recognition. For example, when the target word is *SUNSHINE*, the stimulus *SUNHSINE* is

[1] http://www.mrc-cbu.cam.ac.uk/people/matt.davis/Cmabrigde/

[2] This work was supported by the European Research Council under ERC Grant agreement number 290901 NEUCOD.

read more easily than *SUSNHINE*.

Superset priming is a phenomenon observed when unrelated letters are inserted in the stimulus and when all letters of the target word are preserved. It is demonstrated that each inserted letter linearly increases the processing cost of word recognition (Welvaert et al., 2008). Nevertheless, this gain (average of 11 ms per letter insertion) remains small compared to global processing of visual word recognition.

Finally, the experiments in (Van Orden, 1987) show that phonological sources of activation are used in word recognition. So, if primes have a phonological that overlap better those of target words, they contribute better to the recognition process. (Example: *TOATL* vs *TTAOL* for *TOTAL*).

All the mentioned effects above are related with word error correction capability in transposition, erasure, deletion and insertion cases.

2. The connectionist nature of cortical operation

The human neocortex is a complex circuit composed of tens of billions of neurons with a surface of 2600 cm^2. They are interconnected by a vast number of synapses (order of 10^{12}) (Mountcastle, 1997). The activation of one or more synapses can fire other neurons. In other words, a neuron is a processing unit which aggregates one or more inputs and combines them to obtain an output signal. The same idea is present in McCulloch-Pitts neuron model (McCulloch and Pitts, 1943).

Connectionist or neural network modeling is a specific computational method that simulates the behavior of interactions between neurons. These models have some advantages over other methods (Plaut, 2005). First, connectionist models are explicit about mechanisms and constraints in the brain. Second, they are a good tool to validate some hypotheses related to the representation of a cognitive or learning process. Third, neural networks have the ability to generalize input patterns. In our case, we will use this property to correct errors and provide invariability in the decoding process. For example, when reading the pattern *AIRLPANE*, our cortex provides invariability to read this word with the meaning *AIRPLANE*. Finally, some of these networks offer mechanisms to avoid loss of knowledge even if there are damaged connections or neurons. We call this property resilience.

3. Assembly coding and neural cliques, coding and decoding principles

3.1. Clique-based neural networks (CBNN)

(Gripon and Berrou, 2011) proposed a new model of a neuro-inspired associative memory. This model combines the Willshaw-type model (Willshaw et al., 1969), a clustered structure and distributed codes to encode and decode mental information. The advantages to use this type of network are the ability to store patterns with a good performance when retrieving partially damaged messages, robustness and biological plausibility as explained in (Berrou et al., 2014).

Binary tournament-based neural networks, as an extension to deal with sequences efficiently, were introduced in (Jiang et al., 2015a). To do so, non-oriented connections of a clique-based model are replaced with oriented connections (chain of tournaments). Therefore, when a clique is activated, sequences related to it may also be triggered.

As in any associative memory, there are two different procedures. The first one is storing. In this procedure, an input pattern is given to the network and then connections are drawn. The second one is message retrieval in which for a given input the network will activate the pattern with maximum correspondence.

3.1.1. Model representation and storing procedure

In CBNN, a message is represented by a fully interconnected sub-graph (namely clique) with binary connections. There are five important concepts to describe a clique-based neural network:

Fanal: A node in the network.

Message: A vector of c fanals.

Connection: A non-oriented edge in a sub-graph.

Cluster: A group of l fanals. In a clique, there is at most one fanal per cluster.

Clique: A group of c fanals that encodes a message in the network. This is a fully interconnected sub-graph, as depicted in Fig. 2 b).

3.2. Coding neural cliques with twin neurons

In clique-based associative memories, if the stored messages materialize correlated data, the retrieval quality decreases. (Boguslawski et al., 2014) proposed a method to avoid this problem, the principle of *twin neurons*. After each step of message storing, if the number of outgoing connections of a fanal exceeds a given threshold value, then a new fanal (*twin neuron*) is created. From this moment, new connections will use this cloned fanal, thus limiting the number of connections per node. An example is depicted in Fig. 2 a).

3.3. Message retrieval

The decoding procedure or message retrieval consists in an iterative process with two steps: *message passing* and *selection of winners*. In the first step, input fanals are activated in the network using an aggregation rule. There are two main aggregation rules in CBNN: the Sum-of-Sum (SoS) and Sum-of-Max (SoM) (Aboudib et al., 2014). In the second step, an selection rule is applied. For example, one fanal with the highest activity is elected in each cluster, named Local Winner Take-All rule (LWTA), or fanals with less activity are eliminated using a threshold filter, named Losers Kicked-Out rule (LsKO) (Jiang et al., 2015b).

3.4. Decoding with boosting

The performance, using a mentioned selection rule (LWTA or LsKO), drops dramatically in scenarios where input pattern distortions and/or severe interference (when cliques have overlapped fanals). Recently, (Jiang et al., 2015b) proposed an approach based on heuristic retrieval in order to increase the performance in these situations. It consists of several iterations composed of three steps.

First, a fanal is chosen out of input graphs of activation. Second, this fanal receives a strong impulse activity. This activity will propagate towards its adjacent fanals. In the last step, a decoding rule is applied. The iterations stop after reaching a specific condition. This procedure is detailed in (Jiang et al., 2015b).

3.5. Decoding blurred messages

Blurred inputs are stimuli that some input patterns are not present in the stored message (Example: *AIR**R**PLANE*). (Gripon and Jiang, 2013) proposed solutions in order to adapt the model of clique-based network to allow decoding these inputs in the case of stored uniform data. The algorithm uses the SoM rule and the Local Winner-Take-All (LWTA) propagation rule. Based on biological plausibility, the principle of divergence of neural connections (Thivierge and Marcus, 2007) is strong evidence of the existence of blur decoding in the neocortex.

4. The proposed architecture

Reading is a complex cognitive process that involves interactions between several functional modules in the brain: visual, phonological and semantics modules.

As we mentioned in the introduction, we propose an architecture based on DRC model. For that reason, a Text-To-Phoneme converter software[3] for French language implements the Grapheme-Phoneme Rule System, depicted in Fig. 1.

(Leaver et al., 2009) found evidence of an anticipation effect of sound sequences in the brain. This effect is modeled by chain of tournaments in the phonological network.

As such, the proposed architecture is an interface network of phonological and orthographic information combining. The model ends up in a hub and other networks, which are not considered in this work, can be connected to the hub to create any kind of output system (speech or writing system). (Van den Heuvel and Sporns, 2013) indicate the existence of nodes, named *hubs*, in the cortical network that have an important role in linking several modules.

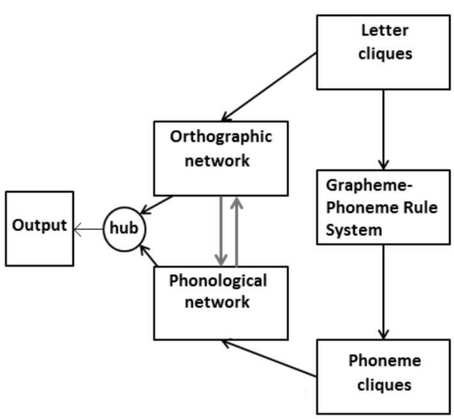

Figure 1: Illustration of the network architecture based on the DRC framework.

[3]LIA_PHON v1.2, under GPL license, available in `http://lia.univ-avignon.fr/chercheurs/bechet/download_fred.html`

4.1. Orthographic network

In this network, letters of words are encoded in positional clusters. This model is an implementation of a clique-based network where fanals are letters, clusters are positions of letters and words are cliques. Nevertheless, according to (Boguslawski et al., 2014), clique-based neural networks are adapted only for data with uniform distribution. When the messages are correlated, the performance of this type of network decreases. In our application, the words are extremely correlated data. For this reason, we chose to use the technique of twin neurons. An example of the proposed orthographic network is illustrated in Fig. 2.

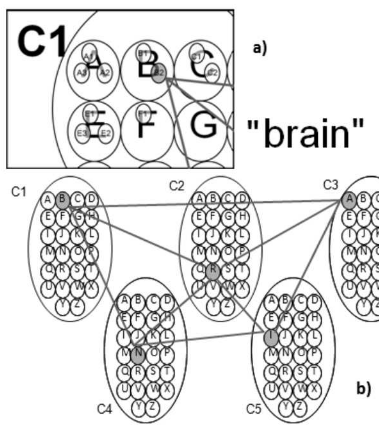

Figure 2: Illustration of the storing procedure for the word BRAIN in the orthographic network. In b), the network has $\chi = c = 5$ clusters and each one contains $l = 26$ groups of fanals representing letters. The most frequent letters are represented in a) with more fanals according to the twin neurons principle.

We use a decoding strategy based on blurred messages and the boosting approach described in 3.5. and 3.4. respectively. The purpose of this strategy is to achieve error correction abilities in the orthographic network. So, we considered a blur decoding capable of activating fanals of letters in different positions in the word. Then, this decoding system eliminates fanals that do not match those of the target word, thanks to the LWTA rule applied in the third step of the boosting decoding. The parameter b is the length of the activation window[4].

During the boosting decoding procedure, the parameter b is adjusted[5] to obtain an activated clique or several cliques. The number of clusters is adapted to the word length before the activation process. For the cases of insertion, the inserted letters are merged either in the previous or in the next cluster randomly. For the cases of deletion, we consider that the network does not know which cluster correspond to each letter. To model this phenomenon,

[4]Example: if $b = 3$ the same letter pattern is activated in 3 adjacent clusters. Using the example of Fig. 2 activated fanals are {(B,C5);(B,C1);(B,C2);(R,C1);(R,C2);(R,C3);(A,C2);(A,C3);(A,C4);(I,C3);(I,C4);(I,C5);(N,C4);(N,C5);(N,C6)}

[5]$b_0 = 1$ and then $b_{t+1} = 2 * b_t + 1$ (for $t = 0, 1, 2, 3, ...$) until the stopping condition is reached or $b > c$

empty clusters are randomly marked among all the available clusters.

Figure 3 shows examples of activation of input letters with different values of b. However, in certain cases, other words could also be activated, for example, in Fig. 3 c) the target word *BRING* instead of *BRAIN* is also a possibility. Thus, we need to look into an additional mechanism to disambiguate these cases. Our strategy is to combine orthographic and phonological information to increase the degree of certainty in the correction mechanism of words.

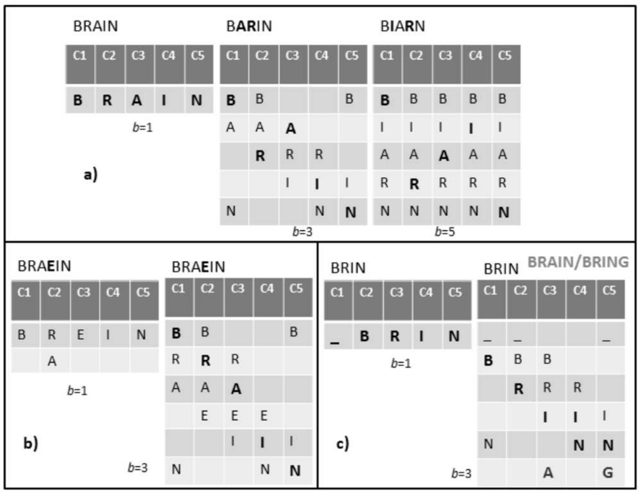

Figure 3: Illustration of activation of input fanals using a blur decoding strategy. In the table: a) transposition case; b) insertion case; c) deletion case.

4.2. Phonological network

The results presented in (Perea and Carreiras, 2006) show that the brain encodes the order of letters at the orthographic level rather than the phonological level and transposition priming effect is less present at phonological level. Therefore, we consider that the phonological network is less flexible in transposition error cases. In that way, the chain of tournaments has an important role in fixing the order of sub-sequences of phonemes of a word. (Example: sub-sequences of the word $/breIn/$ are $/b/$, $/r/$, $/e/$, $/I/$, $/n/$, $/br/$, $/re/$, $/eI/$, $/In/$, $/bre/$, $/reI/$, $/eIn/$, $/breI/$ and $/reIn/$)

The architecture of the phonological network is a hierarchical multilayer network in which nodes represent cliques. The hierarchical structure is formed by triangular patterns connecting two consecutive layers h and $h+1$ (See Fig. 4).

Bidirectional arrows represent connections between two levels in the network. One-way arrows represent a chain of tournaments (sequences). Each clique encodes a sub-sequence of phonemes of a word.

An example of the proposed phonological network is illustrated in Fig. 4.

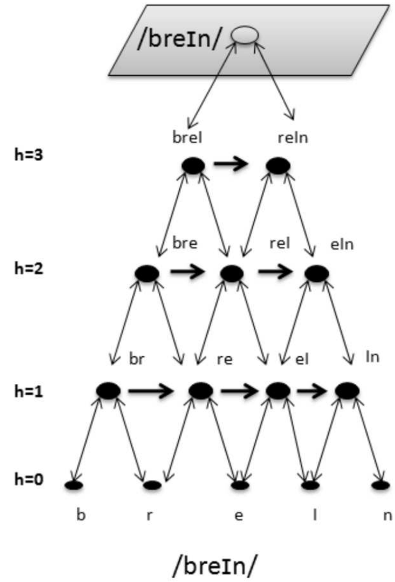

Figure 4: Storing procedure of the word $/breIn/$ in the phonological network. h is the level number in the architecture. The nodes are cliques. In each triangular pattern, two cliques at the bottom level (h level) aggregate to form a new clique at the top level ($h+1$ level). All cliques of sub-sequences of phonemes (e.g. $/br/$ or $/reI/$) are fully connected with the main clique of word $/breIn/$. For the sake of clarity, these connections are not represented in the figure.

In the learning procedure, fanals are randomly chosen to compose cliques of single phonemes. Those cliques can share fanals together. At the next level ($h+1$ level), the procedure is repeated to form cliques which represent the aggregated sequence of phonemes. In addition, all fanals of cliques of sub-sequences are fully connected with the main clique representing the word.

Sequences are used in order to anticipate the activation of cliques in the phonological network.

A bottom-up decoding approach is implemented for decoding. For each triangular pattern, as depicted in Fig. 4, the clique at the top layer is obtained using the propagation of *feedforward* activities of two cliques at the bottom level. Then, the activity is propagated to the sequence on the right. Therewith, a decoding procedure is applied with the SoM and WTA rules.

4.3. Combining orthography and phonology in a hub neural network

The clique-based hub network is capable of integrating decisions from two different modules (phonological and orthographic). A bipartite clique-based network architecture is used in order to integrate decisions from the orthographic network to the activated cliques at the upper level of the phonological network. All fanals of the activated cliques in the phonological and orthographic networks are connected to a clique in the hub. The decoding procedure happens in parallel within the two networks. The last step is to propagate the activity of these cliques towards the hub.

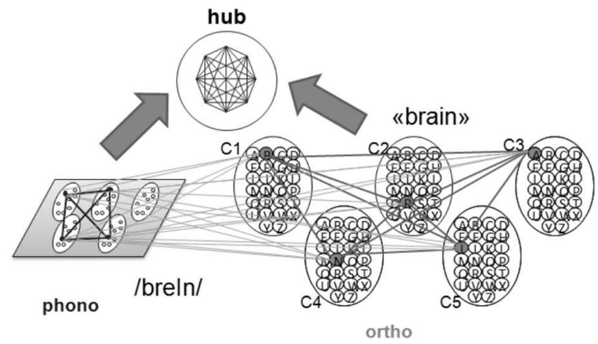

Figure 5: Illustration of the decoding process in a bipartite clique-based network. The propagation is performed from the right to left side and from the phonological network. All fanals of both cliques are fully connected (*feedforward* connections) with a clique in the hub network

5. Results

For the tests, we used a lexical database[6] of French language to select the stimuli. All networks were created with fixed parameters. (For each level of the phonological network: $c = 8$, $\chi = 200$, $l = 200$ and for the orthographic network: $ConnectionLimitTwin = 25$, $\chi = c = 9$ and $numberOfFrenchLetters = 54$[7]. For the hub: $c = 8$, $\chi = 300$, $l = 300$).

In the first procedure, due to the random picking of fanals in the network, we built 20 different networks. Each network learns 6,163 French 7-letter lemmas (in the orthographic network) and its corresponding phonemes (in the phonological network). Then, 200 samples of each type of error were tested in these 20 networks. The results of these tests are in Table 2.

Two criteria are used to evaluate the performance of word correction in each network: *matching*[8] rate and *error*[9] rate. The matching rate corresponds to the number of expected fanals (correct fanals) among the activated fanals in the last layer of the phonological network, the orthographic network or the hub. The error rate is calculated by computing the number of failures to find the exact clique within 4,000 tests in each situation. Even if the correct clique is found with additional fanals, the error rate equals one because it is a strict criterion. The final recognition rate is calculated based on the ratio of correctly recognized words (or phonemes) to the number of total words ($recognitionRate = 1 - errorRate$). The error rate is not calculated for the phonological network because several cliques are elected each time ($errorRate = 1$).

In the last procedure, the network learns 79 unique French words and phonemes of a translated typoglycemia text test bench proposed in (Starzyk et al., 2009). Then, the translated text was modified according to transposing rules used in the paper. As showed below:

> Je n'ariravis pas à criroe que je psuse effeiceetvmnt cmprnodere ce que j'éaits en trian de lrie: le puovior phnémnoéal de l'episrt haumin. Sleon une éqpuie de recehrche de l'Uvinertisé de Cmabrigde, ce n'est pas l'odrre des ltteers qui cmopte dnas un mot, la suele coshe ipmrotnate est que la pmeirére et la drenèire sioent à la bnnoe pclae. Le rsete puet êrte dnas un dsérorde ttoal et vuos puoevz tujoruos lrie snas polbrème. C'est prace que le creaveu hmauin ne lit pas chuaqe ltetre une par une, mias le mot cmome un tuot. Cttee cnotidion s'alepple la Typoglycémie. Inocrblyae, non ? Ouias, et vuos aevz tojruous psneé que l'oroathgrphe éatit impoartnte.

This network is then able to store words with variable length. For this purpose: $\chi = c = 16$ and each word contains $c - lengthWord$ padding characters at the end (Example: importante, importante######).

Testing set	Rate	Network		
		Phono	Ortho	Hub
Transposition (adjacent)	Match	70.0	100.0	98.2
	Error	-	5.0	2.0
Transposition (1 between)	Match	57.0	99.0	90.7
	Error	-	26.6	9.8
Erasure (1 letter)	Match	100.0	100.0	94.5
	Error	-	11.5	5.5
Deletion (1 letter)	Match	85.0	99.7	88.2
	Error	-	43.2	11.62
Insertion (1 letter)	Match	99.0	100.0	99.0
	Error	-	5.8	1.1
Insertion (2 letters)	Match	93.0	99.8	98.0
	Error	-	12.2	2.2
Benchmark with 79 words	Match	76.9	100.0	100.0
	Error	-	0.0	0.0

Table 2: Recognition rates (percentage) of the network.

Here are some examples of recognition ambiguities for the first procedure: (*COLONLE, colonel, colonne*); (*_EINDRE, ceindre, geindre, feindre, teindre, peindre*); (*NCRTASHER, cascher, crasher, castrer, cracher, catcher*).

The present model has a retrieval accuracy rate of 100% of 111 tested words in the proposed typoglycemia benchmark for French language. We can compare this result with the state-of-the-art for the English benchmark[10]. For instance, we have an accuracy of 94.67% for the hidden Markov models (HMM) and 84.36% for the Levenshtein distance methods. 100% accuracy was obtained for the LTM model based on the spatio-temporal memory proposed in (Starzyk et al., 2009) and the episodic neural memory model based on the fusion adaptive resonance theory proposed in (Wang et al., 2012). None of the two last models consider phonological information of words, the number of learned messages using the test bench is limited to 73 words and there is no insertion, deletion or erasure in the testing set.

[6] Lexique.org is a French lexical database of lexical information of 135,000 words and 55,000 lemmas.

[7] Accentuated and special characters are included.

[8] $matching = numberCorrectFanals/c$

[9] It provides zero if $numberActivatedFanals = c$ and $matching = 1$ else it provides one

[10] The English benchmark has 107 words, among which there are 73 unique words.

The result shows that the error retrieval rate globally decreased in our hub network using a multimodal approach. Transpositions in adjacent letters are more easily recognized than non-adjacent cases. Finally, insertions of one or two letters have a strong superset priming effect (with 1.1% and 2.2% recognition error rates respectively).

Those results indicate the model seems to mimic quite adequately some of the behavioral performances described in the literature of priming effects.

6. Future work

Our future work will be to adapt the orthographic model to allow for words with variable length and consider the word frequencies by reinforcement[11] of connections. Then, we will study an extension of this model for sentences. To overcome this issue, we will consider the context and semantic information.

7. References

Aboudib, A., V. Gripon, and X. Jiang, 2014. A study of retrieval algorithms of sparse messages in networks of neural cliques. In *COGNITIVE 2014 : the 6th International Conference on Advanced Cognitive Technologies and Applications*.

Berrou, C., O. Dufor, V. Gripon, and J. Xiaoran, 2014. Information, noise, coding, modulation: What about the brain? In *Turbo Codes and Iterative Information Processing (ISTC), 2014 8th International Symposium on*.

Boguslawski, B., V. Gripon, F. Seguin, and F. Heitzmann, 2014. Huffman coding for storing non-uniformly distributed messages in networks of neural cliques. In *AAAI 2014: the 28th Conference on Artificial Intelligence*, volume 1.

Christianson, K., R. Johnson, and K. Rayner, 2005. Letter transpositions within and across morphemes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6):1327.

Coltheart, M., K. Rastle, C. Perry, R. Langdon, and J. Ziegler, 2001. Drc: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1):204.

Grainger, J., J. Granier, F. Farioli, E. Van Assche, and W. Van Heuven, 2006. Letter position information and printed word perception: the relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4):865.

Grainger, J. and C. Whitney, 2004. Does the huamn mnid raed wrods as a wlohe? *Trends in cognitive sciences*, 8(2):58–59.

Gripon, V. and C. Berrou, 2011. Sparse neural networks with large learning diversity. *Neural Networks, IEEE Transactions on*, 22(7):1087–1096.

Gripon, V. and X. Jiang, 2013. Mémoires associatives pour observations floues. In *Proceedings of XXIV-th Gretsi seminar*.

Jiang, X., V. Gripon, C. Berrou, and M. Rabbat, 2015a. Storing sequences in binary tournament-based neural networks. *Neural Networks and Learning Systems, IEEE Transactions on*, PP(99):1–1.

Jiang, X., M. Marques, P. Kirsch, and C. Berrou, 2015b. Improved retrieval for challenging scenarios in clique-based neural networks. In *Advances in Computational Intelligence*. Springer.

Leaver, A., J. Van Lare, B. Zielinski, A. Halpern, and J. Rauschecker, 2009. Brain activation during anticipation of sound sequences. *The Journal of Neuroscience*, 29(8):2477–2485.

McCulloch, W. and W. Pitts, 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

Mountcastle, V., 1997. The columnar organization of the neocortex. *Brain*, 120(4):701–722.

Perea, M. and M. Carreiras, 2006. Do transposed-letter similarity effects occur at a prelexical phonological level? *The Quarterly journal of experimental psychology*, 59(9):1600–1613.

Perea, M., S. Lupker, and S. Kinoshita, 2003. Transposed-letter confusability effects in masked form priming. *Masked priming: State of the art*:97–120.

Plaut, D., 2005. Connectionist approaches to reading. *The science of reading: A handbook*:24–38.

Rayner, K., S. White, R. Johnson, and S. Liversedge, 2006. Raeding wrods with jubmled lettres there is a cost. *Psychological science*, 17(3):192–193.

Schoonbaert, S. and J. Grainger, 2004. Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes*, 19(3):333–367.

Starzyk, J., H. He, et al., 2009. Spatio–temporal memories for machine learning: A long-term memory organization. *Neural Networks, IEEE Transactions on*, 20(5):768–780.

Thivierge, J. and G. Marcus, 2007. The topographic brain: from neural connectivity to cognition. *Trends in neurosciences*, 30(6):251–259.

Van den Heuvel, M. and O. Sporns, 2013. Network hubs in the human brain. *Trends in cognitive sciences*, 17(12):683–696.

Van Orden, G., 1987. A rows is a rose: Spelling, sound, and reading. *Memory & Cognition*, 15(3):181–198.

Wang, W., B. Subagdja, A. Tan, J. Starzyk, et al., 2012. Neural modeling of episodic memory: Encoding, retrieval, and forgetting. *Neural Networks and Learning Systems, IEEE Transactions on*, 23(10):1574–1586.

Welvaert, M., F. Farioli, and J. Grainger, 2008. Graded effects of number of inserted letters in superset priming. *Experimental Psychology*, 55(1):54–63.

Whitney, C., 2001. How the brain encodes the order of letters in a printed word: The seriol model and selective literature review. *Psychonomic Bulletin & Review*, 8(2):221–243.

Willshaw, D., O. Buneman, and H. Longuet-Higgins, 1969. Non-holographic associative memory. *Nature*.

[11]The reinforcement of connections in a multilayer clique-based neural network is an unpublished problem.