

## Uniqueness of Language Vectorization by the mathematical structure.

Word has evident mathematical structure for the almost unique language vectorization, In semantics, a word immediately becomes a **metric space measure** <close meaning = parallelism, non-overlapping meaning = orthogonality in vector space>. By the word vectorization, document one is also accomplished because it is series vector of word one. <http://777true.net/Evidence-of-telepathy-And-about-on-the-Serious-and-Wonderful-AI.pdf>

### [ 1 ] : Rationale for Vectorization=Mathematical Structure in Language.

#### ① Vector Metric Space (Function one). 2023/4/29,5/2,3

People say “*the meaning is near for this, or not.*” This is decisive evidence that words has a metric.

example\_1): *bad*  $\perp$  *good* <orthogonality>

this is no nearness=no overlapping in the meaning to none identifying.

example\_2): *bad* vs *worse*<parallelism>

this is nearness=overlapping in the meaning to identifying, **but not complete one.**

#### ② Set Hierarchy by Including Property.

People say *this meaning is included that, or not.* This is decisive evidence that words has a Set Hierarchy. Those are combined with **a common abstract property.**

example\_3): **culture**  $\supset$  **religion**  $\supset$  christianism  $\supset$  protestant  $\supset$  Baptist church.

**culture**  $\supset$  **religion**  $\supset$  islamism  $\supset$  Sunni  $\supset$  Salafism

This shows that those vector has **common component** in some dimension.

\*culture=spiritual way of life,, VS civilization=technical, material way of life.

culture  $\perp$  civilization.....however some words may be overlapping

example\_4): **Art** is spiritual, while it could be also technical such as architecture.

**language** is spiritual, while it could be also technical such as business dealing.

☞; Assuming vector component of culture(1,0)  $\perp$  civilization(0,1), then those may be

(0.5,0.5). Thus it is rather complicated even for us mankind, while AI may be troublesome.

Purpose of language is to **identify** an information in many learned data,

Detects the target case that is proposed as an internal information=(cognitive AI),

By **synthesizing** many input data as trial-and-error combination in multiple learned data, detecting the target project is creative act=(generative AI)

After all, the supreme essence of language is function of **identifying or non-one.**

#### 3 keywords in Words Categorization for the manual vectorization:

*orthogonal*=no overlapping in the meaning between two words...good  $\perp$  bad

*parallel*=overlapping in the meaning between two words,.....tiny//small

*inclusion hierarchy*: culture  $\supset$  religion  $\supset$  christianism  $\supset$  protestant  $\supset$  Baptist church(**proper noun**)

☞; hierarchy height corresponds to degree of **conceptual, generality, abstractness.**

thereby those words become important in whole view statement. While hierarchy lowest is proper nouns the most concrete precise statement,

## [ 2 ] : Language Vectorization may need Transfer Learning ? ! ! 。

① Also author's aim is high accuracy of language vectorization for AI tasks of **higher generalized problems** such as abstract moral, ideology, and religion. Now author could not have comprehension on AI principle by internet learning. However he considered incompleteness of the vectorization methods employed by researchers at that front line.

Their method seem not agree with ideology of [ 1 ]. But following might be an exception ??.

<https://www.tryeting.jp/column/6839/>

**Retrofitting** (1985) is a method that uses a **semantic dictionary called WordNet** to adjust groups of words with similar meanings so that they form similar vectors.

<https://wordnet.princeton.edu/>

## ② The Difficulty in Language Vectorization.

(1) Now language vectorization in AI development, **the most difficulty** is due to the no math measure such as "**image or sound**" (function in space and time) especially in language, they employ **self learning** method, however authors opinion is **transfer learning** <that is, messy brain transplant> may be absolutely necessary, the reason are,

(2) **Measure similarity between sentences with "Sentence Bert"**.

<https://zenn.dev/yoshikawat64m/articles/c242b11d21be68>

AI conclusion is rounded off by threshold setting, so the result will be to some extent, but as far as the language processing is concerned, it feels like making a shack? ?

(3) Similarity Strength = learning growth in neural network.

Infants are lingual geniuses, acquire linguistic semantic identities from the frequency of commonality experiences of **actual situation patterns** and **phonetic patterns** in **different-dimensional productive environments**. In US development, they are trying to measure the semantic extraction from the statistical correlation frequency among the same dimensional environment of the document pattern and the document pattern. In trying to measure semantic extraction, attacking with CPU brute force seems little lazy? ? ,

(4) Can you it possible that in closed silent library of books with characters only, but without picture or etc, infants can really get language??.

**cipher**:: those who try to read cipher <closed library> already had a language.

\* **cipher** may be a pseudo randomization of words with the reversibility.

(5) Once again language recognition in brain.

A tag of a word is **accurately** connected with a collective of **not word events in brain**.

Can a tag of a word **accurately** connect with a collective same words in documents ?.

From the beginning with nothing word tags, can above method success??

Now life can be generated only from life, but none life from not life at all.

There must previously be least words which can define new comer words.

**[ 3 ] : Hand Made Vectorization an Example of Noun.**

☞:less learned author can not tell [ 3 ] accurate,but it may be many misunderstand.

**I :Whole View on the Strategy.**

- (1)List up all the most frequently used words= $\{U_k\}$  by fixed amount.  $1 \leq k \leq 4000?$ .
  - (a)List up All the pieces of speech,except nouns,those are fewer,but important.
  - (b)List up All the **collective(abstract)nouns?**,but proper ones.
- Then words out of the list $\{v_i\}$  may be almost rare proper nouns.  $4000? \leq i \leq 100,000??$ .

(2)Vectorization by categorizing with **orthogonality** in **a inclusion hierarchy(parallelism)**→ **II**

				.....	$U_{H-2,p}$	$U_{H-1,p,q}$ proper noun	$U_{H,p,q,r}$ blank components for v proper noun
--	--	--	--	-------	-------------	---------------------------	--

This image is total length of vectors with hierarchy of H boxes,with pure component  $\{0, 1\}$ ,. The upper parts are parallel components,while the last one ="1" is orthogonal.

(3)The words out of the list $\{v_i\}$ .  $4000? \leq i \leq 100,000??$ .

The words are expected to be **almost proper nouns.**→(b)

(a)Word overlapped with orthogonal words can be linear sum of the orthogonal ones< **II** (5)>.

Some  $\{v_i\}$  could be linear sum of words in  $\{U_k\}$ ,those are same dimension with  $\{U_k\}$ .

example)**welfare**=**goodness** in **life** condition for **health,happiness,prosperity**.

**LGBT**=Lesbian, Gay, Bisexual,Transgender/Transsexual

**extinction**=death, no-revival

**security**=safety, keeping, preservation, sureness, guarantee, peace preservation.

By emerging new word,the vector dimension format should be fixed still,

→**blank space for components of v proper noun**

**II :Vectorization of "noun" as an example.**

**Orthogonal categorization with down processing of inclusion hierarchy order.**

\*The other piece of speech seem easier to establish orthogonality than collective nouns?.

**(1)Orthogonal Categorization**

=**semantic dissimilarity<vector orthogonality ⊥ classification>**

It is self-evident that the purpose of language = semantic logic judgment is accomplished because of this,

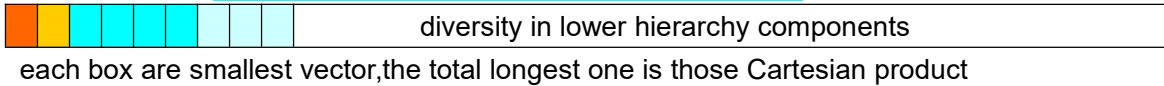
an example)),**matter(hard)⊥none matter(soft)**=culture⊥civilization

good ⊥bad;yes⊥no,long⊥short,

(2) **Categorization by Inclusion Hierarchy Order in Ensemble Nouns.**

- (a)  $B \subset A$  = Elements of set B are simultaneously elements of A (set theory), and if premise is B, the conclusion is A (logic).  
 (b) an example)),  
 civilization ⊥,,, culture (the biggest collective noun) ⊃ religion ⊃ christianism ⊃ protestant  
 ⊃ Baptist church (proper noun = the lowest hierarchy)

(c) Diversity with keeping Semantic Similarity in higher hierarchy .



(d) Note each hierarchy  $H_k$  has vector component dimension =  $d_k$

$H_1$	2	matter(hard)=0				non matter(soft)=1				vector(1,1)					
$H_2$	4	life=0		non-life=1		culture=0		0.5, 0.5		civilization=1					
$H_3$	$d_3$	artificial		natural											
...		Certainly this categorization task is hard enough, however, once accomplished,													
...		the strong contribution is forever !!!													
		*the dimension of $H_3 = d_3$ .													
$H_H$		(1.1, .....00100)													

(e) **dimensional compressed representation of orthogonal vectors**

The words vector with  $\{0, 1\}$  components are very long component vector, so those need equivalent dimension compressed representation, which is easily synthesized, because in each hierarchy  $H_k$ , component = 1 emerges one time only, so its position number in  $H_k$  become equivalent representation in the hierarchy  $H_k$ ,

(f) If there are 10 kinds of parts of speech, the direct product is added to the head of the word vector with a 10-dimensional classification unit vector.

(3) **categorizing by complete and incomplete orthogonality in a hierarchy:**

an example)),  
**ideology** ⊃ {Christianism ⊥ Islamism ⊥ Buddhism ⊥ Communism ⊥ Existentialism ⊥ ...  
 ⊥ democracy ⊥ authoritarianism ⊥ fascism ⊥ racial spremacism ⊥ NAZISM .....}??

Those gray portion has common property behind distinct words. In the later, it is to show overlapping similarity becomes linear some of vectors.

$$\text{NAZISM} = \alpha \text{ fascism} + \beta \text{ authoritarianism} + \gamma \text{ racial spremacism} + \delta \text{ atheism} + \epsilon \text{ anti-Judaism} + \zeta \text{ belligerence} + \eta \text{ monarchism} + \dots$$

A synonyms nominate the single same vector.

tiny? = small, little, pocket-size, petite, diminutive, bantam, cute, pretty, lovely, dear, charming, wee., measly, puny

(4) Listing up all the possible nouns in a Hierarchy rank=r.

orthogonality checking matrix  $\# = n_r(n_r - 1) / 2$

	n <sub>1</sub>	n <sub>2</sub>	n <sub>3</sub>		n <sub>r</sub>
n <sub>1</sub>		X	O	X	X
n <sub>2</sub>	X		X	O	O
n <sub>3</sub>	O	X		O	O
	X	O	O		X
n <sub>d</sub>	X	O	O	X	

collecting all possible nouns in a hierarchy=r by any how ,and then checking each orthogonality between n<sub>j</sub> and n<sub>k</sub>.  
dissimilarity yes>50 → orthogonal=X  
 by many test observers.

In a hierarchy H<sub>k</sub>, extracting possible many complete orthogonal nouns is to determine

d<sub>jk</sub>=dimension of vector in jth element of higher hierarchy noun in H<sub>k</sub>, [1 ≤ j ≤ d<sub>k-1</sub>]

$$d_k = \sum_{j=1}^{d_{k-1}} d_{jk}$$

example)), life ⊃ animal ⊥ plant ⊥ bacteria ⊥ fungi ⊥ protista ⊥ monera.....5dim  
 animal ⊃ mammal ⊥ birds ⊥ reptiles ⊥ .... ⊥ .... ... 9dim  
 plant ⊃ .....  
 bacteria ⊃ .....  
 .....

(5) **incomplete orthogonality** (overlapping by multiple order “m”).

1	2	3	4	5	6	7	8	9	10
		√0.1		√0.7	√0.2				

an example)), √ is due to V<sup>2</sup>=1.

above is 10 dimension orthogonal categorization, while some noun concern multi-nearness m=3 with (3,5,6) components, then the vector component is not such pure (0000100000), but (0,0,√0.1,0,√0.7,√0.2,0,0,0,0), these partitioning is by nearness estimation with each component, (0,0,√0.1,0,√0.7,√0.2,0,0,0,0) = √0.1V<sub>3</sub> + √0.7V<sub>5</sub> + √0.2V<sub>6</sub>,

That is, **incomplete orthogonal noun** is linear sum of the orthogonal ones

Ordinal Language Translation by adjoint “or.”

(0,0,√0.1,0,√0.7,√0.2,0,0,0,0) is {V<sub>3</sub>, or V<sub>5</sub>, or V<sub>6</sub>,} by each tendency {√0.1, √0.7, √0.2}

(6) **The Final Lowest Hierarchy in Nouns.**

Counting up all proper nouns in a Hierarchy final=f. Those are complete orthogonal.

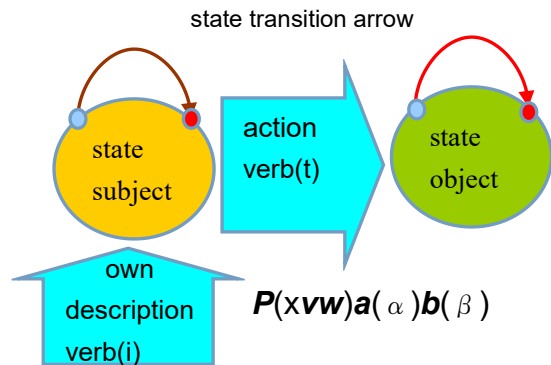
**APPENDIX\_1:A Model of Language in Predicative Logic.**

After all,a language has strong aspect of idiom out of simple grammar rule. It is told French is rule mode,while English is not ?!.

Thereby,language interpretation by deterministic logic could not success!!

Strong learning strategy will advance AI toward the completeness ?

Deepen precise categorization will not success ?



adjoint	$P(xvw)a(\alpha)b(\beta)$ $P=action(vt),$ $statement(vi) \rightarrow b=0$			$a=subject, b=object,$ $,c=object$	int erj ec tio n
	xvw		pre posi tion	$\alpha . \beta =noun$ collective,proper pronoun	
	v=verb(i) verb(t)	x=auxilliary verb			
	w=adverb			$\alpha . \beta =adjective$	

**Conditional Proposition the Logic  $A \subset B, if cause is A, the result is B.$**

General our task is judging true or false on Conditional Proposition for deriving true B.

(1) **Proof:** proposition A is true,  $A \subset B$  is tautology, then conclusion B is true.

True" is a realization in LOGIC by observation logic<Birkov & Neumann>.

General our task is judging true or false on Conditional Proposition for deriving true B.

(2) **tautology..**

$(A \supset B)$  must be tautology...  $(A \supset B) \equiv n(A \wedge nB) \equiv nA \vee B \rightarrow (B=A)$  make (1) tautology.

In predicative logic,A cause had already stated possible result=B in an ensemble=A.

Thereby result B is one element in cause ensemble=A.

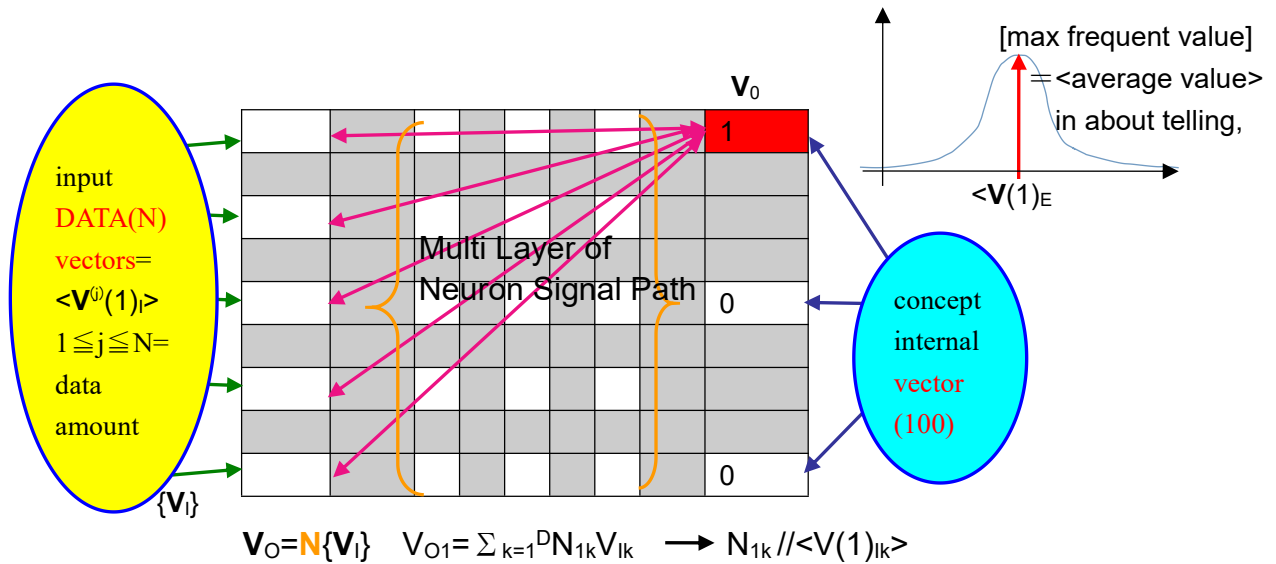
Example)A={rain falls for a day by 19cm/hour at xcxx}  $\supset$  B={river flooding,cliff decaying,necessity of umbrella,collecting water},

After all, cause is everything toward generating result as an element in the cause.

This is valid both for caustic and probabilistic.

**APPENDIX\_2:Extracting a Meaning(Conception)from Many Data  
(Learning with teacher).**

**=The Linear Combination Neuron Model as Growing Parallel Vector  
by Max Frequent Input as Average Value .**



Adjusting neuron path conductivity= $N_{1k}$ (weight coefficient)=[ $N_{11}, N_{12}, \dots, N_{1D}$ ]  
by internal recognition signal vector(100)is to grow a component  
 $N_{1k}$  parallel to input data average= $\langle V(1)_{ik} \rangle$ .

(1)Each input data(~apple)vector(0001010001000.....00011) is very higher dimension= $N$ ,  
and all the data vectors is normalized as  $1=|V_i|^2$  by the normalization coefficient.

(2)Output vector= $V_0(3 \times 1)$  is generated by multiplying neuron matrix= $N(3 \times D)$   
with input vector= $V_i(D \times 1)$  due to linear combination model of neuron layer..  
 $V_0 = N\{V_i\}$ .

Neuron layers are multiple, however after all, the result is single.

(3)Teacher signal demands (100)vector, that is **max value** in  $V_{01}$  component.

$$V_{01} = (1/N) \sum_{j=1}^N \sum_{k=1}^D N_{1k} V^{(j)}_{ik} \rightarrow N_{1k} // \langle V(1)_{ik} \rangle ; ; \langle V(1)_{ik} \rangle = (1/N) \sum_{j=1}^N N V^{(j)}_{ik}$$

That is, max value is realized by  $N_{1k}$  parallel with  $\langle V(1)_{ik} \rangle$

(4) $\langle V(1)_{ik} \rangle$  is **most frequent value** in the data, or the average value.

A concept is generated by the most frequent value in the data,

(5)In (Quantum)LOGIC, true is **realization at probability =1**, in this sense,  
the most frequent value is very reasonable.

(6)Most remarkable fact for neuron model is both input and recognition one cooperate neuron path growing  
to establish a conception as an output, **however as for AI technology**, result we want is unique and only.

$N_1 // \langle V(1)_i \rangle$ ; those vector inner product is max value (near parallel).

or **maximum frequent event** is memorized as a concept.

**$N_1$  vector** is the memory itself obtained by the learning <frequent data input>.

### **APPENDIX\_3:Big Error Correction on chapter [ 2 ]. 2023//5/3**

(7) {Meaningful events for us $\sim$ Higher realization probability}  
 $\Leftrightarrow$  {Higher realization probability $\sim$ Meaningful events for us}

From view of (5),above statement must be true !!.

(8)In accomplishing language vectorization,method of detecting higher correlation between {words,sentence}becomes true.  
then authors opinion [language vectorization by transfer learning \[ 2 \]](#) : become not always true.