

神経心理学モデルの基礎

浅川伸一

2018年12月27日

神経心理学的症状をニューラルネットワークをもじいて説明する試みは1980年代の第2次ニューロブームから存在した。2010年代から始まった第3次ニューロブームでは、その性能が向上し部分的には人間を上回る性能を示している。このような状況から第2次ブームでは未解決であったり実現が困難だと見なされた問題を解決できる緒があると考えられる。ここではこれらの概説を行う。

1 はじめに

認知神経心理学とその関連諸分野と機械学習や人工知能と呼ばれる分野との交流は現状では活発であるとは言い難い。しかし、これらは同一対象を別の視点から眺めていると見做しうる。計算機科学の黎明期から、人間の認知機能は、その障害から理解しようとしてきたようにも思われる (Turing, 1950)。2010年代以前は、人間の知的能力に比肩しうる程度の計算機上の実装が存在し得なかつたために神経心理学への貢献は存在してきた (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; McClelland & Rumelhart, 1986; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989b)。駄菓子菓子、人口に膾炙したとは言い難い。2012年以降の深層学習の発展に伴い (浅川, 2016b)¹、部分的には人間の認識能力を凌駕する機械学習モデル (He, Zhang, Ren, & Sun, 2015; Mnih et al., 2015) が発表されるにいたり、人間の認知能力と計算機モデルのそれとを直接比較する「ロボットは東大に入れるか」²などの試みがなされている。本稿では、上述のような視点から機械学習モデルによる神経心理学的症状の説明を試みる。

2 言語のモダリティ(様相)

言語能力はモダリティとして、「読む」「書く」「話す」「聞く」が考えられる。言語獲得、言語障害、言語能力の測定、母語以外の言語獲得、芸術分野への応用、機械学習系の自動翻訳を含む自然言語処理などを含めれば言語能力の包括範囲は多岐にわたる。言語は人間活動の基盤

とも言える基本能力でもあり人間の知的活動への貢献大きい。ここでは神経心理学症状に関連した言語の障害から言語機能を考える際の基本的な概念の概説を試みる。

神経心理学と前提として、自然科学の意味での実験を行うことができないという問題が挙げられる。臨床家が行う検査は実験であるが、自然科学の意味で全独立変数を実験者が統制可能ではないという意味で準実験である。倫理的制約から人間に対して人工的に脳を破壊することが困難なことから、脳損傷、脳挫傷、脳梗塞など脳に障害を持つ患者を観察することから言語機能を類推することが行われてきた。動物実験では脳の一部を破壊して観察することが可能であるが、人間と同じような言語機能を持つ種では破壊実験の実施は困難である。従って、脳の障害が発生した患者が基本資料となる。一般に言語の障害を総括して失語症 **aphasia** と呼ぶ

3 プローカ失語



図1 プローカ Wikipedia

¹ https://www.jstage.jst.go.jp/article/psychono/35/2/35_35_26/_pdf

² 東ロボプロジェクト

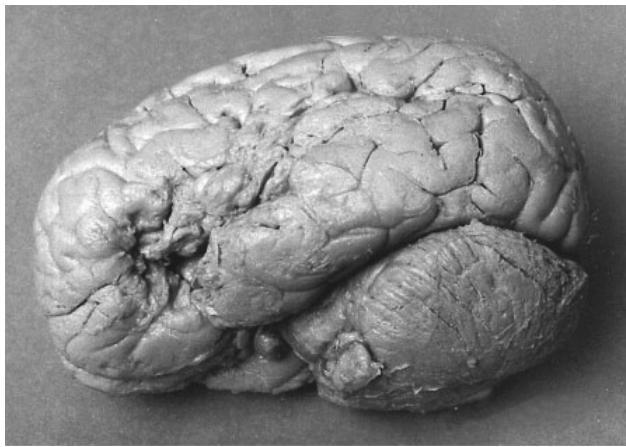


図2 tan の死後剖検写真 Dronkers et al. (2007)



図4 ウィルニッケ Wikipedia

言語の機能局在は Paul Broca による 1861 年(明治維新は 1868 年)に始まる。Broca の経験した症例タンは、ほぼ「タン」としか発話できなかつたため、皆からタンさんと(ムッシュ・タン)と呼ばれていたという。図2 に示すとおり、左半球前頭葉、ブロードマン 44 野、45 野領域が責任病巣であると一般には考えられて生きたようだ。

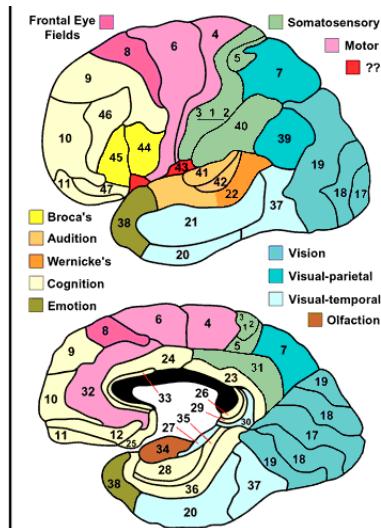


図3 ブロードマンの地図

5 ゲシュヴィンド失語

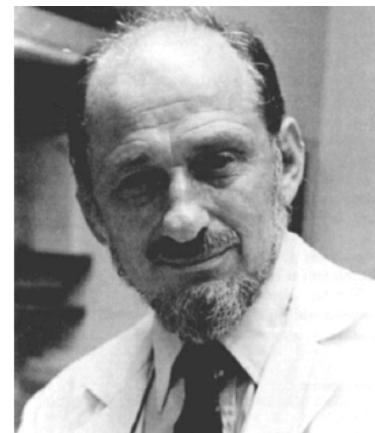


図5 ゲシュヴィンド Gread Thoughts Treasury

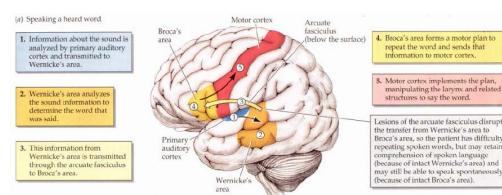


図6 ゲシュヴィンドの離断モデル

4 ウィルニッケ失語

カール・ウィルニッケによって提唱された主として聴覚入力情報からの言語情報処理のモデルである。

ドロンカーズ (2007)Dronkers et al. (2007) によれば、

- 症例「タン」は左前頭回を含む広範な領野が損傷されていた
- プローカ自身はこのことを知ることができなかつた(岩田先生談)

6 単語音読

6.1 入出力

言語情報処理を考える場合の入出力について考える。入力は、視覚(画像や動画などの場合と文字や記号などの2つに大別される)、聴覚(音声、音楽)、運動感覚がある。一方、出力は、音声、運動(書字、描画、指差し、視線の移動など)がある。これらの入出力を分類アルゴリズムを通して学習可能であると考えるのが一般的である。入出力のそれぞれをモダリティ(modality, 様相、様式)というが生理学における感覚入力の別に加えて、細分類がなされることもある。たとえば、視覚における、明るさ、色、線分の方位、聴覚における周波数帯域などは別モダリティと考えられる場合もある。各モダリティは独立した入力装置と見なさる場合が多い。

6.2 処理単位

言葉情報処理の単位には、その大小の別により、文字letter, 単語word, 文章sentence, 段落paragraph, それ以上の大きな単位がある。それぞれに認識、分類、生成、などの処理機構が考えられている。

6.3 単語音読課題に関する単語の分類

ここではまず最初に、英単語を視覚提示し、その単語を音読する場合を考える。

このとき英単語は綴りと読みとの関係が準規則的(quasi regular)であると言う。すなわち英単語には綴(正書法orthography)と発音(音韻(法)phonology)との間の対応関係が規則的な単語とそうでない単語が存在するからである。ここで、規則語、不規則語、という用語を用い、**一貫語**(consistent words), **非一貫語**(inconsistent words), **不規則語**(irregular words)の3種を区別する。一貫語と非一貫語は共に規則に従う読みであるが、同じ読みを共有する単語(隣人あるいは仲間と呼ぶ)の多少で区別される。例えばAVEという綴りは, GAVE, SAVE, CAVEなどの単語が存在するので一貫語に分類される。一方HAVEは、これらの単語とは読み方を共有していないため非一貫語に分類される。不規則語は独自の読み方をする単語、例えばyachtを指す。

以下で概説するように、上述のような準規則的な単語音読過程は2つの研究の潮流で調べられてきた。一方は、ニューラルネットワークや人工知能研究であり、他方は文字や単語の読みに障害を持つ児童や言語機能に損傷を持つ患者の神経心理学である。

ニューラルネットワークや人工知能研究においては、規則に従う単語の読みとそうでない単語の読みを区別するアルゴリズムの研究と関連し、ルールベースでの判

断やニューラルネットワークを用いたモデルが提案されてきた。同様の問題が、英単語の過去形、過去分詞形の判断、生成問題にも当てはまる Rumelhart & McClelland (1985)。

一方、読みに障害を持つ患者の研究からは、各種類の単語(一貫語、非一貫語、例外語、非単語)の読みの成績により言語の病理を診断する研究が行われてきている。

6.3.1 用語解説

正書法(論): orthography

音韻法(論): phonology

意味(論): semantics

それぞれの体系の中の要素をそれぞれ以下のように書くことがある

書記素: grapheme

音(韻)素: phoneme

意味(素): sememe

7 神経心理学モデル

7.1 ウィルニッケ-リヒトハイム図式 Wernicke-Lichtheim schema

上記は、単語音読の紹介であったが、ここでは歴史的な順序を考慮して、聴覚入力による言語理解とその病理モデルを紹介する。

図7に古典的な失語症のモデルである Wernicke-Lichtheim 図式を示した Lichtheim (1885)。図中のAは聴覚的イメージの座、Mは発話に用いられる運動表象の座。Bは言語の概念が保持されている部位、aは聴覚情報の入力系、mは運動機能の言語中枢である。このモデルによれば、言語理解は $a \rightarrow A \rightarrow B$ 、発話は $B \rightarrow M \rightarrow m$ 、復唱は $a \rightarrow A \rightarrow M \rightarrow m$ という経路を通ることになる。Mが損傷されると理解はできるが自発的な発話と復唱(言われた言葉をそのまま口に出して言うこと)ができなくなり(**プローカ失語**)、Aが損傷を受けると言語理解と復唱の障害が生じる(**ウェルニケ失語**)。実際この図式は患者の症状をよく説明し、モデルから予測される症状を呈する患者の発見もあって有効性が確認されている(岩田, 1996)。理論上、 $A \rightarrow M$ の結合が断たれると復唱のみができなくなり、言語理解も自発発話も保たれる**伝導失語** conduction aphasiaが(図中の3)、 $A \rightarrow B$ 間の結合が切れると復唱はできるが言語理解が不能な**超皮質性感覚失語** trans-cortical sensory aphasiaが(図中の6)、また、 $B \rightarrow M$ の結線が切断されると復唱はできるが自発発話が侵される**超皮質性運動失**

語 trans-cortical motor aphasia が生じることになる。さらに、 $M \rightarrow m$ では書字障害を伴わず発語に障害がみられる**純粹語嘔**が生じる。

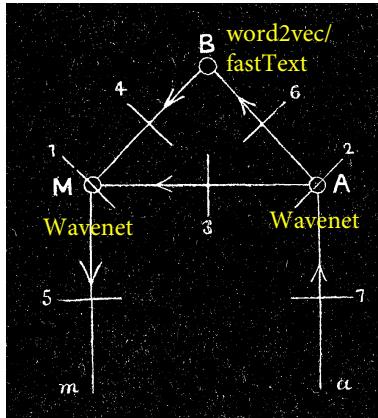


図 7 Lichtheim (1885)

7.2 3つの失読症:表層, 音韻, 深層

読字過程の障害には 3 種類ある。**音韻失読症** (phonological dyslexia), **表層失読症** (surface dyslexia), **深層失読症** (deep dyslexia) である。音韻性失読症の患者は実在する単語は読むことができるし、非単語を書き取ることもできるし、言われた単語を復唱することもできるが、非単語を読むことができない。たとえば ‘must’ は読むことができても実際には存在しないが発音可能な単語 nu st を発音することができない。二重経路モデル (DRCColtheart et al. (2001)) に照らして考えれば音韻失読は書記素から音韻へ変換（直接経路）の障害と考えることができる。音韻性失読に選択的障害をもつ患者は、通常の単語であれば問題なく読めるので、一見すると失読患者ではないように見える。表層失読の患者は、規則語や非単語を読むことができるが、低頻度の例外語（低頻度の例外語とは、語の出現頻度が低く、かつ、その単語の読みが、書記素と音韻とを結びつける簡単なルールには則っていない語のことをいう、たとえば ‘yacht’）を読むことができない。表層失読の患者は視覚性の誤り (dog を dot と言ったりする) もするが、意味の誤りはない。健常者であれば、たとえ自分の知らない単語であって、綴りと音の一般的対応関係から、それらしい発音をすることができる。ところが表層失読の患者はそれができない。表層失読はトライアングルモデルに沿って考えれば、意味への間接経路の障害と考えることができる。

深層失読の患者は音韻失読の患者と同じく非単語を読むことができない。このことに加えて、深層失読の患者は意味性の錯読を示す。たとえば ‘dog’ を ‘cat’ と言ったりする。また、視覚性の誤りを示すことがある。

ときには上記 2 つの混ぜ合わさった読みの誤り、たとえば ‘sympathy’ を ‘orchestra’ と言ったりする（おそらく ‘sympathy’ と綴りの似ている ‘symphony’ を介して）。この種の患者は、抽象語と具体語の成績にも差があることがある。具体語である「椅子」よりも抽象語である「真実」の読みの成績が悪い。二重経路モデルによる深層失読の説明では、書記素から音韻への直接経路の障害の他に意味経路の障害が加わったとされる。

7.3 二重経路モデル

印刷された文字を音読する場合を考える。Coltheart ら (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart & Rastle, 1994; Coltheart et al., 2001) によって開発された記号処理的な読みのモデルで、ある二重経路モデル (Dual Route Cascaded model) では、印制文字を音韻へ変換するための明示的な規則に基づく直接経路と、規則にあてはまらない単語を読むための単語ベースのルックアップテーブルをもつ間接経路（語彙経路）とから構成されている。ルックアップテーブルとは、それぞれの単語が登録された表のようなものだと考えればよい（書記素語彙目録という）。このルックアップテーブルにエントリのある単語項目は、この一対一対応関係のある間接経路を通じて読みられ、ルックアップテーブルに登録されていない単語は書記素音韻対応規則系を通じて発話に至る。

单一経路か二重経路かという議論ではなく、直接経路と間接経路の処理の違いに関する議論である。二重経路モデルとニューラルネットワークモデルであるトライアングルモデルとの違いは、二重経路モデルの特徴は明示的で記号処理的なルール（ルックアップテーブル）を用いるということにある。すなわち、二重経路モデルでは 2 つの経路間のどちらの経路を通じて印刷された文字が読みられるかを決めるための（離散的な）スイッチが仮定されていることである。

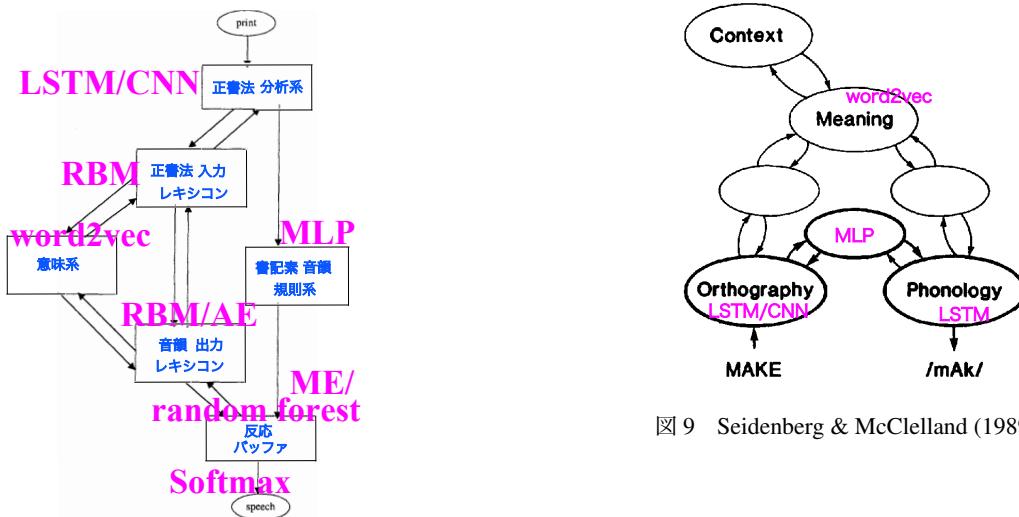


図8 二重経路モデル (Coltheart et al., 2001)への適用案

図9 Seidenberg & McClelland (1989a)

トライアングルモデルでは英語の単音節単語約3000語を読むことができる。初期のトライアングルモデル (Seidenberg & McClelland, 1989a)では、非単語を読むことができなかつたため、人間の読みのモデルとしては適当でないという批判を受けていた。

7.4.1 SM89

Wikergeln, 非単語の読み。Fig.10に示したのはSM89で実装されたモデルである。各モジュールはニューラルネットワークの1層であり、Fig.10の円内には用いられたユニット数が記されている。

7.4 トライアングルモデル

トライアングルモデル (Plaut et al., 1996; Seidenberg & McClelland, 1989a)では明示的で記号処理的なルール(ルックアップテーブル)あるいはスイッチの存在を仮定しない。代わりに、トライアングルモデルでは同時的、相互作用的処理が仮定される。書記素、音韻、意味の情報は各ユニット群内／群間に分散表現されており、類似した単語は、ユニット群内の類似した活性パターンとして表現されている。

トライアングルモデルにおける直接経路においては、多くの単語と発音規則が一致する規則語と高頻度の不規則語が学習される。一方低頻度の不規則語は意味系に依存すると仮定される。したがって直接経路は単語の頻度効果に、すなわち単語の統計情報(生起確率)に敏感である。規則語および高頻度例外語と低頻度例外語との処理の違いには労働の分割(division of labor)とよばれる作用が関与する。対照的に規則に基づく二重経路モデルでは高頻度不規則語を扱うことが期待できない。さらに、2つの経路間の離散的なスイッチをもつ代わりに、トライアングルモデルでは同時的、相互作用的処理による、軟らかい労働の分割が行われる。

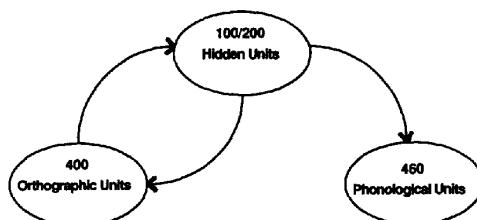


図10 Seidenberg & McClelland (1989a) Fig. 2より実装された部分と層内のユニット数

■SM89の入出力表現 Bag of Letters表現では、「TAB」と「BAT」とが同じ発音を与えることになってしまう。そこで(Wickelgren, 1969)の**3連図式(triples schema)**が用いられた。3連表現とはstopは $/_S_t, sT_o, tO_p, oP_{\#}/$ と表現される。この3連図式はfastTextでも用いられているJoulin et al. (2017)。MAKEは_MA, MAK, AKE, KE_と入力される。同様に出力表現/mAk/は_mA, mAk, aK_と表現された。

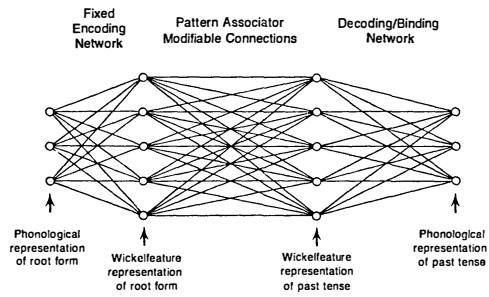


図 11 Rumelhart & McClelland (1985) Fig.1

■データ 2897 語の単音節語 monosyllabic words, うち 13 語は homogrpah (同型異音)

$$p = K \log (\text{frequency} + 2). \quad (1)$$

homograph は以下の通り: bass, bow, dave, house, lead, live, read, row, sow, tear, wind, wound

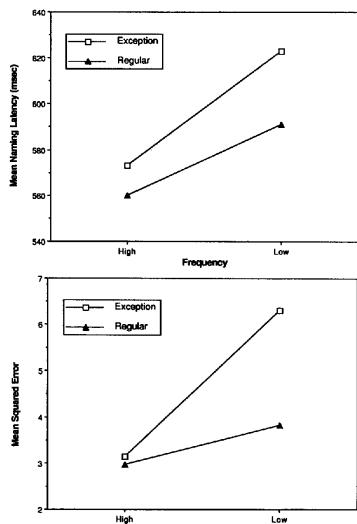


Figure 4. Results of the Taraban and McClelland (1987) study (upper graph) and the simulation data for 250 epochs (lower graph).

図 12 Seidenberg & McClelland (1989a) Fig. 4. Results of the Taraban and McClelland (1987) study (upper graph) and the simulation data for 250 epochs (lower graph).

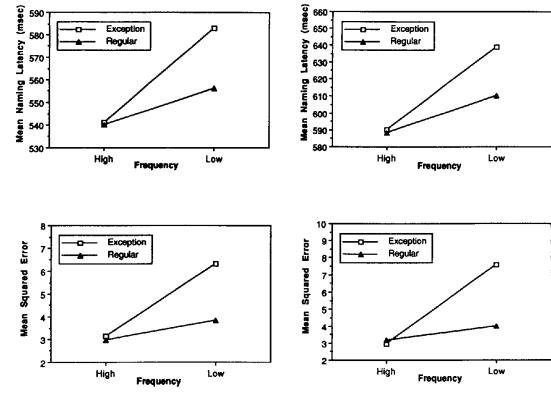


Figure 5. Results of the Seidenberg (1985c; left graphs) and Seidenberg, Waters, Barnes, & Tanenhaus (1984, Experiment 3; right graphs) studies: experiments (upper graphs) and simulations (lower graphs).

図 13 Seidenberg & McClelland (1989a) Fig. 5. Results of the Seidenberg (1985c; left graphs) and Seidenberg, Waters, Barnes, & Tanenhaus (1984, Experiment 3; right graphs) studies: experiments (upper graphs) and simulations (lower graphs).

■Regular inconsistent words In an important article, Glushko (1979) studied a class of words called regular inconsistent. These words, such as GAVE, PAID, and FOE, have two critical properties. Their pronunciations can be derived by rule; in fact, most of these words' neighbors rhyme (e.g., GAVE, PAVE, SAVE, BRAVE). However, each of these words has an exception word neighbor (e.g., HAVE, SAID, and SHOE, respectively).

Research in this area has been marked by the development and revision of several taxonomies based on different properties of words or perceptual units thought to be theoretically relevant. In part, this research was motivated by the fact that several models, incorporating very different representational and processing assumptions, all predict longer naming latencies for exception words compared with regular words. In the dual-route model (Coltheart, 1978), longer latencies result because readers attempt to pronounce exception words by applying grapheme-phoneme correspondence rules, resulting in a temporary misanalysis. In Glushko's (1979) model, a word is pronounced by analogy to similarly spelled neighboring words. The fact that the neighbors of an exception word are all regular was thought to interfere with generating its pronunciation. According to Brown (1987), the factor that determines naming latencies is the number of times a spelling pattern (word body) occurs with a particular pronunciation. A regular word such as DUST contains a word body, -UST, that is pronounced /ust/ in many words.

An exception word such as SWAMP contains a word body, -AMP, that is pronounced/omp/in only one word, the exception itself. Hence, the frequency of a spelling-sound correspondence could be the source of the exception effect.

We examined the model's processing of regular inconsistent words using stimuli from the Taraban and McClelland experiment described previously, which also included low/high freq, regular/exception/inconsistent, word/non-word

1. **high-frequency regular**
2. **low-frequency regular**
3. **high-frequency inconsistent**
4. **low-frequency inconsistent**
5. **matched regular word controls**

以下は Taraban & McClelland (1987) より

consistent: words have the same primary vowel sound and very similar pronunciations, like the rhyming words dish, wish, and fish the pronunciation will be easy to synthesize (Glushko, 1979, 1981).

These orthographic **neighbors** of the target-based on the identical spelling for the vowel and the consonant cluster ending- are phonological **friends** and tend to support the activation of the target pronunciation

An **exception** effect could result in the following way. When an exception word like have is the target, it supports its own activation, but also supports the activation of words that are visually similar to it, like save, wave, and gave -again, the similarity is based on the identical spelling of the vowel and consonant ending. Gave, save, and wave support the same vowel pronunciation, and they could **conspire** against the target word. These **conspirators** are **enemies** of have since they contain major phonological discrepancies with respect to the vowel and inhibit phonological features of this target word in favor of their own. Conspiracy relations of this sort could underlie longer processing times for exception words, since the exceptions must overcome the phonological interference from the conspirators.

High Freq exception: are, both, break, choose, come, do, does, done, foot, give, great, have, move, put, pull, said, says, shall, want, watch, were, what, word, work.

Low Freq. exception: bowl, broad, bush, deaf, doll, flood, gross, lose, pear, phast, pint, plow, rouse, sew, shoe, spook, swamp, swarm, touch, wad, wand, wash, wool, worm

High Freq regular-inconsistent: base, bone, but, catch, cool, days, dear, flew, flat, five, form, go, goes, grow, here, home, meat, paid, plant, roll, root, sand, small, speak

Low Freq regular-inconsistent: brood, cook, cord, cove, cramp, dare, fowl, gull, harm, hoe, lash, leaf, loss, mad, moose, moth, mouse, mush, pork, pose, pouch, rave, tint, toad

Plaut et al. (1996) では入力表現を工夫し、**オンセット** onset, **母音** vowel, **コード** code という 3 つの部分から構成される**母音中心表現**とよばれる表現を用いている。この入力表現の工夫により非単語の読みについても人間と同程度の能力があることが示された。トライアングルモデルの読むことができる単語は、単音節の単語である。すなわち、母音については 1 つだけのコーディングが必要である。加えて母音の前後に子音のクラスターが必要である。母音の前の子音をオンセット、母音の後の子音をコードという。したがって、単語 ‘make’ の入力表現は、オンセットが ‘M’、母音が ‘A’、コードが ‘K’ などと表現する。

8 意味障害のモデル

Hinton & Shallice (1991); Plaut & Shallice (1993); Warrington & Shallice (1979),

Mikolov, Chen, Corrado, & Dean (2013); Mikolov, Sutskever, Chen, Corrado, & Dean (2013); Mikolov, Yih, & Zweig (2013)

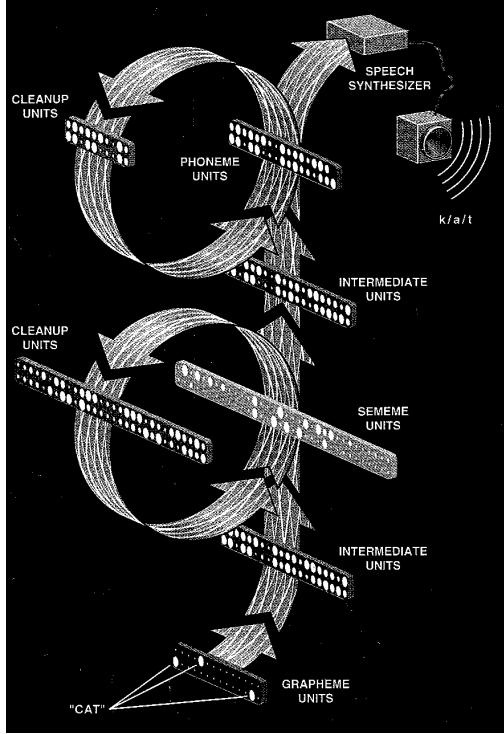
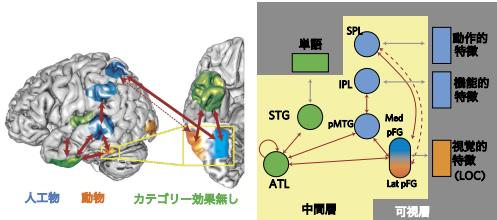


図 14 Chen et al. (2017) より



8.1 カテゴリー特異性

1. カテゴリー特異性 category specificity Warrington (1975); Warrington & Shallice (1979, 1984)
2. 基本概念優位性 basic level superiority
3. 上位概念保存 super-ordinate category preservation

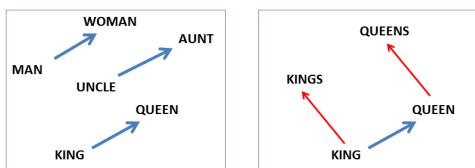


図 15 Mikolov et al. (2013) より

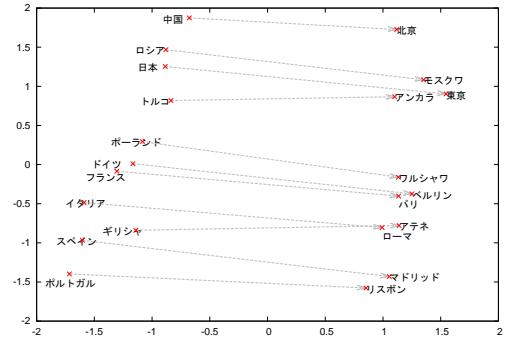


図 16 Mikolov et al. (2013) より

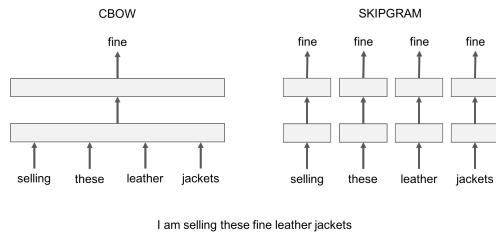


図 17 左:CBOW, 右:スキップグラム Joulin et al. (2017)

For a word w with N word vector sets $\{c(w)\}$ representing the words found in its contexts, and window size W , the empirical variance is:

$$\Sigma_w = \frac{1}{NW} \sum_i^N \sum_j^W (c(w)_{ij} - w)(c(w)_{ij} - w)^T \quad (2)$$

This is an estimator for the covariance of a distribution assuming that the mean is fixed at w . In practice, it is also necessary to add a small *ridge* term $\delta > 0$ to the diagonal of the matrix to regularize and avoid numerical problems when inverting.

9 失語症検査プロジェクト

深層学習の進展を背景に様々な課題を実用レベルで解決するモデルが提案されている浅川 (2016a)。このような成果を利用して、標準失語症検査を解かせるプロジェクトを立ち上げる。人工知能分野、機械学習分野への逆還流 (Hassabis, Kumaran, Summerfield, & Botvinick, 2017) を視野に入れる。

9.1 標準失語症検査への適用

- 聴く : 単語への理解, 短文の理解, 口頭命令に従う,
- 読む : 漢字単語の理解, 仮名单語の理解, 短文の理解, 書字命令に従う,
- 話す : 呼称, 動作説明, 語の列挙, まんがの説明,
- 書く : 漢字単語の書字, 仮名单語の書字, まんがの説明,
- 計算 : 加減算, 乗除算,

復唱：単語の復唱、文の復唱、
音読：漢字単語の音読、仮名单語の音読、短文の音読、
書取：漢字単語の書取、仮名单語の書取、短文の書取、
仮名一文字：仮名の理解、仮名一文字の音読、仮名一文字の書取

標準失語症検査 SLTA は言語能力の推量に用いられるが、SLTA 下位項目を解くために必要な機械学習モデルを考えれば、**CNN** (Krizhevsky, Sutskever, & Hinton, 2012), **SGD** (Bottou & Bousquet, 2007), **LSTM** (Greff, Srivastava, Koutník, Steunebrink, & Schmidhuber, 2015; Hochreiter, Bengio, Frasconi, & Schmidhuber, 2001; Hochreiter & Schmidhuber, 1997), **NIC** (Xu et al., 2015), **Seq2Seq** (Sutskever, Vinyals, & Le, 2014), **Wavenet** (Oord et al., 2016), **word2vec** (Mikolov et al., 2013, 2013), **強化学習** (Sutton & Barto, 1998, 2018)などが挙げられる。

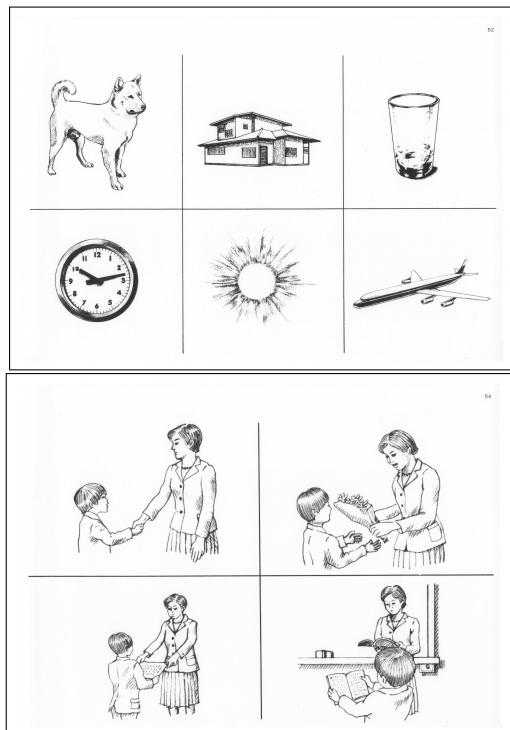


図 18 SLTA 図版

図 18 は SLTA の刺激図版である。これらの図版を用いた下位検査を解くためには、畳み込みニューラルネットワークによる画像認識、画像から言語情報を得るための言語モデル、教示を理解するための聴理解モデル、検査画像を選択するための注意モデル、言語応答のための発話モデルなどが必要となると考えられる。これらのモデルはすでに機械学習分野で提案されてきており、本稿の目的のために有益であると考えられる。

10 まとめ

昨今の深層学習、機械学習のモデルは人間に比肩する性能を示すようになってきた。本稿では、このような情勢に鑑み、神経心理学モデルを機械学習モデルを用いて構築する試みを考えた。

本稿の試みのように、神経心理学モデルに対して機械学習モデルを適用することにより、診断補助、治療計画立案、自立支援、評価手法の確立と提案、などこの分野に対する貢献が可能だろうと考えている。これにより、患者とその家族、脳神経外科医師、言語聴覚治療士、作業療法士、理学療法士、神経心理学者、高次脳機能障害に関心を持つ機械学習関係者との相互交流が促進され、これらの領域が活性化することが期待できる。本来、同じ目的を持つ領域を相互に結びつけることで、あらたな発展が関係するすべての領域にとって刺激的で良好な関係を構築できるのではないかと期待できる。進化上、他の種には存在しないという意味で人間にしか持ち得なかつた言語機能に光を当てることにより全脳アーキテクチャの解明と完成に向けた一里塚となると考える。

引用文献

- 浅川伸一. (2016a). 『Pythonで体験する深層学習』. 東京: コロナ社.
 浅川伸一. (2016b). 深層学習をめぐる最近の熱狂. 『基礎心理学研究』, 35, 149-162.
 浅川伸一. (2012). 脳損傷患者の症例から見た読字過程. 川崎恵里子(編)『認知心理学の新展開—言語と記憶』. 京都: ナカニシヤ出版.
 Bottou, L., & Bousquet, O. (2007). The tradeoffs of large scale learning. In *Advances in Neural Information Processing Systems* (Vol. 20). Cambridge, MA, USA: MIT Press.
 Chen, L., Lambon-Ralph, M. A., & Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*, 1, 1-10.
 Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589-608.
 Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1197-1211.
 Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204-256.
 Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., & Cabanis, E. A. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of leborgne and lelong. *Brain*, 130, 1432-1441.
 Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 674-691.
 Goldberg, Y., & Levy, O. (2014). word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.
 Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2015). LSTM: A search space odyssey. *arXiv:1503.04069*.
 Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95, 245-258.
 He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv:1512.03383*.
 Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74-95.
 Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (2001). Gradient flow in recurrent nets the difficulty of learning long-term dependencies. In S. C. Kremer & J. F. Kolen (Eds.), *A field guide to dynamical recurrent neural networks*. IEEE press.
 Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9, 1735-1780.

- 岩田誠. (1996). 『脳とことば—言語の神経機構』。東京: 共立出版。
- Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2017). FASTTEXT.ZIP: Compressing text classification models. In Y. Bengio & Y. LeCun (Eds.), *The proceedings of International Conference on Learning Representations (ICLR)*. Toulon, France.
- 近藤公久・浅川伸一. (2017). 日本語 wikipedia の word2vec 表現と語彙特性との関係. 『第 80 回日本心理学会発表論文集』久留米シティプラザ, 久留米。
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *in Advances in Neural Information Processing Systems 25*. Montréal, Canada.
- Levy, O., & Goldberg, Y. (2014a). Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)* (pp. 302–308). Baltimore, Maryland, USA.
- Levy, O., & Goldberg, Y. (2014b). Neural word embeddings as implicit matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 27, p. 2177-2185). Montréal CANADA: Curran Associates, Inc.
- Lichtheim, L. (1885). On aphasia. *Brain*, 433–484.
- McClelland, J. L., & Rumelhart, D. E. (1986). Amnesia and distributed memory. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 503–527). Cambridge, MA, USA: MIT Press.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Y. Bengio & Y. Lecun (Eds.), *Proceedings in the International Conference on Learning Representations (ICLR) Workshop*. Scottsdale, Arizona, USA.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Weinberger (Eds.), *Advances in neural information processing systems 26* (pp. 3111–3119). Curran Associates, Inc.
- Mikolov, T., Yih, W. tau, & Zweig, G. (2013). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL*. Atlanta, WA, USA.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstr, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518, 529–533.
- Oord, A. van den, Dieleman, S., Zeny, H., Simonyan, K., Vinyals, O., Kalchbrenner, A. G. N., Senior, A., & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint*.
- Plaut, D., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Plaut, D., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377-500.
- Rumelhart, D. E., & McClelland, J. L. (1985). On learning the past tenses of english verbs. In *Parallel distributed processing: Explorlation in tire microstructure of cognition*: (Vol. 2 Psychological and Biological Models). Cambridge, MA: Bradford Books/MIT Press.
- Seidenberg, M. S., & McClelland, J. L. (1989a). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Seidenberg, M. S., & McClelland, J. L. (1989b). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Weinberger (Eds.), *Advances in Neural Information Processing Systems (NIPS)* (Vol. 27, pp. 3104–3112). Montreal, BC, Canada.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). Cambridge, Massachusetts: The MIT Press.
- Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, 26, 608–631.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind A quarterly review of psychology and philosophy*, LIX 236, 433–460.
- Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, 27, 635–657.
- Warrington, E. K., & Shallice, T. (1979). Semantic access dyslexia. *Brain*, 102, 43–63.
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairment. *Brain*, 107, 829–854.
- Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychochological Review*, 76, 1-15.
- Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R. S., & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. *arXiv:1502.03044*.

付録 A ARPABET

表 1 ARPABET vowels

1-letter	2-letter	IPA	Example(s)
a	AA	ɑ	balm, bot
@	AE	æ	bat
A	AH	ʌ	butt
c	AO	ɔ	bought
W	AW	ɑʊ	bout
x	AX	ə	about
N/A	AXR	ə̄	letter
Y	AY	aɪ	bite
E	EH	ɛ	bet
R	ER	˥	bird
e	EY	eɪ	bait
I	IH	i	bit
X	IX	ɪ	roses, rabbit
i	IY	i	beat
o	OW	oʊ	boat
O	OY	ɔɪ	boy
U	UH	ʊ	book
u	UW	u	boot
N/A	UX	ʉ	dude

IPA		ARPAbet		IPA		ARPAbet	
Symbol	Symbol	Word	Transcription	Symbol	Symbol	Word	Transcription
[p]	[p]	parsley	['parsli]	[p aɪ r s l i y]			
[t]	[t]	tarragon	['tærəgən]	[t ae r ax g aa n]			
[k]	[k]	catnip	['kætnɪp]	[k ae t n i x p]			
[b]	[b]	bay	[beɪ]	[b eɪ]			
[d]	[d]	dill	[dɪl]	[d ih l]			
[g]	[g]	garlic	['garlik]	[g aa r l i x k]			
[m]	[m]	mint	[mɪnt]	[m ih n t]			
[n]	[n]	nutmeg	['nʌtmɪg]	[n ah t m eh g]			
[ŋ]	[ŋ]	ginseng	['dʒɪnsɪŋ]	[jh ih n s i x ng]			
[f]	[f]	fennel	['fēnl]	[f eh n el]			
[v]	[v]	clove	['klou̯]	[k l ow v]			
[θ]	[θ]	thistle	['θɪsl]	[th ih s el]			
[ð]	[ð]	heather	['hɛðər]	[h eh dh axr]			
[s]	[s]	sage	['seɪdʒ]	[s ey jh]			
[z]	[z]	hazelnut	['heɪzlənʌt]	[h ey z el n ah t]			
[ʃ]	[ʃ]	squash	['skwɑʃ]	[s k w a sh]			
[ʒ]	[ʒ]	ambrosia	['æmbrəʊʒɪə]	[a m b r ow zh ax]			
[tʃ]	[tʃ]	chicory	['tʃɪkəri]	[ch ih k axr iy]			
[dʒ]	[dʒ]	sage	['seɪdʒ]	[s ey jh]			
[l]	[l]	licorice	['lɪkɔrɪs]	[l ih k axr ix sh]			
[w]	[w]	kiwi	['kiwi]	[k iy w iy]			
[r]	[r]	parsley	['parsli]	[p aɪ r s l i y]			
[j]	[j]	yew	[ju]	[y uw]			
[h]	[h]	horseradish	['hɔ:sreɪdɪʃ]	[h a o r s r ae d ih sh]			
[ʔ]	[q]	uh-oh	['?əu̯]	[q ah q ow]			
[ɾ]	[dx]	butter	['bɑ:tə]	[b ah dx axr]			
[ɾ]	[nx]	wintergreen	['wɪntəgrɪn]	[w ih nx axr g r i n]			
[ɾ]	[el]	thistle	['θɪsl]	[th ih s el]			

Figure 4.1 IPA and ARPAbet symbols for transcription of English consonants.

IPA		ARPAbet		IPA		ARPAbet	
Symbol	Symbol	Word	Transcription	Symbol	Symbol	Word	Transcription
[i]	[iy]	lily	['lɪlɪ]	[i l i h l i y]			
[ɪ]	[ih]	lily	['lɪlɪ]	[i l ih l i y]			
[eɪ]	[ey]	daisy	['deɪzɪ]	[d ey z i]			
[ɛ]	[eh]	poinsettia	['poin'setiə]	[p oy n s eh dx iy ax]			
[æ]	[ae]	aster	['æstər]	[ae s t axr]			
[ɑ]	[aa]	pappy	['papɪ]	[p aa p i]			
[ɔ]	[ao]	orchid	['ɔrkɪd]	[ao r k ix d]			
[ʊ]	[uh]	woodruff	['wudrəf]	[w uh d r ah f]			
[oʊ]	[ow]	lotus	['lou̯rəs]	[l ow dx ax s]			
[u]	[uw]	tulip	['tulip]	[t uw l ix p]			
[ʌ]	[uh]	buttercup	['bʌtər,kʌp]	[b uh dx axr k uh p]			
[ə]	[er]	bird	['bɜ:d]	[b er d]			
[aɪ]	[ay]	iris	['arrɪs]	[ay r ix s]			
[əʊ]	[aw]	sunflower	['sʌnflaʊər]	[s ah n f l aw axr]			
[ɔɪ]	[oy]	poinsettia	['poin'setiə]	[p oy n s eh dx iy ax]			
[ju]	[y uw]	feverfew	['fi:vərfju]	[f iy v axr f y u]			
[ɔ]	[ax]	woodruff	['wudrəf]	[w uh d r ax f]			
[ə̄]	[axr]	heather	['heðər]	[h eh dh axr]			
[i]	[ix]	tulip	['tulip]	[t uw l ix p]			
[u]	[ux]		[]	[]			

Figure 4.2 IPA and ARPAbet symbols for transcription of English vowels

Word	jackal	raccoon	cougar	civet
IPA	[dʒækəl]	[ræk'kʌn]	[ku:gə]	[si:vɪt]
ARPAbet	[dʒ ae k el]	[r ae k uw n]	[k uw g axr]	[s ih v ix t]

Figure 4.3 The mapping between IPA symbols and letters in English orthography is complicated; both IPA [k] and English orthographic [c] have many alternative realizations

付録 B word2vec の目標関数

- skip gram:

$$J = \sum_{w \in D} \sum_{c \in C} \log P(c|w) \quad (3)$$

- CBOW:

$$J = \sum_{w \in D} \sum_{c \in C} \log P(w|c) \quad (4)$$

where,

D: 全語彙

C: 単語 w の隣接 $\pm h$ の範囲に出現する語

$P(c|w)$: 単語 w から隣接語群 C を予測する確率

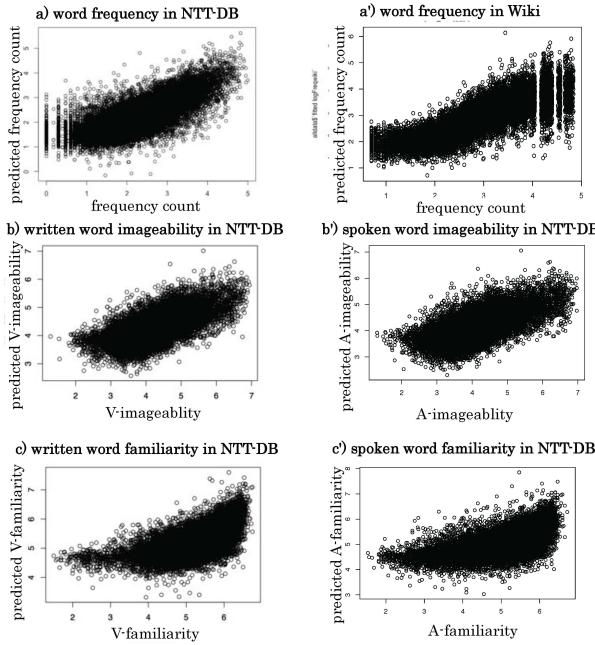


Fig. 1 各特性値と推定値の相関図

Fig. 1' 各特性値と推定値の相関図

図 19 Relations between NTT-DB and word2vec 近藤・浅川 (2017)

$P(\cdot)$ は以下のような softmax function:

$$P(c|w) = \frac{\exp(\mathbf{w}_w^\top \tilde{\mathbf{v}}_c)}{\sum_{w'} \exp(\mathbf{v}_w^\top \tilde{\mathbf{v}}_{w'})} \quad (5)$$

学習時には正事例だけを学習するのではなく、無関連語を排除するような負事例サンプリングが精度向上に寄与する Mikolov et al. (2013)。以下の式右辺第 2 項。

$$\log P(C|w) \propto \log \sigma(\mathbf{v}_w^\top \tilde{\mathbf{v}}_c) + \kappa \mathbb{E}_{r \sim P_n} [\log \sigma(-\mathbf{v}_w^\top \tilde{\mathbf{v}}_r)], \quad (6)$$

ここで右辺第 2 項の期待値は一様分布 P_n から単語 r を k 回サンプリングした平均、 σ はシグモイド関数 = $(1 + \exp(-x))^{-1}$ である。

Goldberg & Levy (2014); Levy & Goldberg (2014a) は word2vec は shifted PMI^{*3} であることを主張。

^{*3} $pmi(x, y) \equiv \log \frac{p(x, y)}{p(x)p(y)} = \log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)}$ https://en.wikipedia.org/wiki/Pointwise_mutual_information

B.0.1 Shifted PMI

$$M_{i,j} = PMI(w_i, c_j) - \log \kappa \approx \mathbf{w}_i^\top \tilde{\mathbf{w}}_j \quad (7)$$

単語と文脈の共起を PMI で計測して、分散表現（低次元ベクトル）を構成するのに近い Levy & Goldberg (2014b)。コーパス中の共起回数 $n(w, c)$ や出現頻度 $n(w)$ を用いて SGNS の目的関数を変形すると、

$$J = - \sum_{w \in D} \sum_{c \in C} \log \sigma(\mathbf{v}_w^\top \tilde{\mathbf{v}}_c) - \kappa \mathbb{E}_{r \sim p_n} [\log \sigma(-\mathbf{v}_w^\top \tilde{\mathbf{v}}_r)] \quad (8)$$

$$\begin{aligned} &= - \sum_{w \in D} \sum_{c \in C} n(w, c) \log \sigma(\mathbf{v}_w^\top \tilde{\mathbf{v}}_c) \\ &\quad - \sum_{w \in C} n(w) \kappa \mathbb{E}_{r \sim p_n} [\log \sigma(-\mathbf{v}_w^\top \tilde{\mathbf{v}}_r)] \end{aligned} \quad (9)$$

期待値の部分を明示的に計算すると、

$$\mathbb{E}_{r \sim p_n} [\log \sigma(-\mathbf{v}_w^\top \tilde{\mathbf{v}}_r)] = \sum_{r \in v_c} \frac{n(r)}{|D|} \log \sigma(-\mathbf{v}_w^\top \tilde{\mathbf{v}}_r) \quad (10)$$

$$\begin{aligned} &= \frac{n(c)}{|D|} \log \sigma(-\mathbf{v}_w^\top \tilde{\mathbf{v}}_c) \\ &\quad + \sum_{r \in v_c \setminus c} \log \sigma(-\mathbf{v}_w^\top \tilde{\mathbf{v}}_r) \end{aligned} \quad (11)$$

目的関数のうち、 w と c に関する部分だけを取り出すと、

$$\hat{l}(w, c) = -n(w, c) \log \sigma(\mathbf{v}_w^\top \tilde{\mathbf{v}}_c) - n(w, c) \kappa \frac{n(c)}{|D|} \log \sigma(-\mathbf{v}_w^\top \tilde{\mathbf{v}}_c) \quad (12)$$

$x = \mathbf{v}_w^\top \tilde{\mathbf{v}}_c$ として、 $l(w, c)$ を x で偏微分して 0 とおくと、

$$\frac{\partial l(w, c)}{\partial x} = -n(w, c) \sigma(-x) + \kappa n(w) \frac{n(c)}{|D|} \sigma(x) \quad (13)$$

$$= n(w, c) \{\sigma(x) - 1\} + \kappa n(w) \frac{n(c)}{|D|} \sigma x \quad (14)$$

$$= 0 \quad (15)$$

等式を整理すると、

$$\left\{ 1 + \frac{\kappa n(w) n(c)}{|D| n(w, c)} \right\} \sigma(x) = 1 \Leftrightarrow \exp(-x) = \frac{\kappa n(w) n(c)}{|D| n(w, c)} \quad (16)$$

したがって、

$$x = \mathbf{v}_w^\top \tilde{\mathbf{v}}_c \quad (17)$$

$$= \log \frac{|D| n(w, c)}{\kappa n(w) n(c)} \quad (18)$$

$$= \log \frac{|D| n(w, c)}{n(w) n(c)} - \log \kappa \quad (19)$$

$$= PMI(w, c) - \log(\kappa) \quad (20)$$