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 Rosch, Mervis, Gray, Johnson, & Boyes-Braem (1976)
3. **上位概念保存** super-ordinate category preservation Warrington (1975)

analogy by vector space

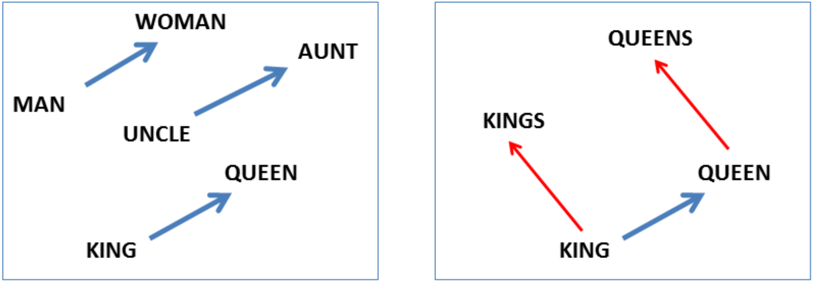

Figure 1: Mikolov, Yih, & Zweig (2013) より

Sample example of word2vec

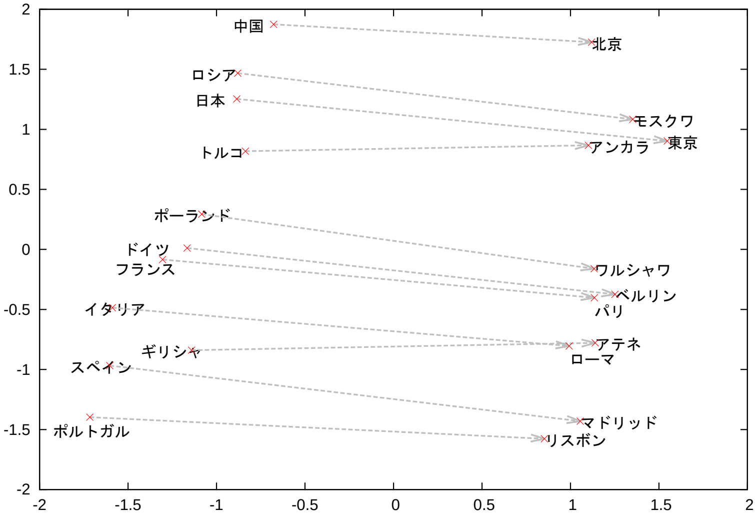

Figure 2: Mikolov, Sutskever, Chen, Corrado, & Dean (2013) より

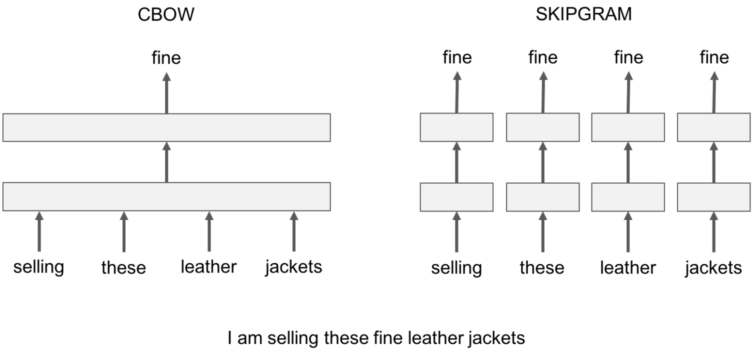

Figure 3: 左:CBOW，右:スキップグラム Joulin et al. (2017)

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:

$$\Sigma_w = \frac{1}{NW} \sum_{i}^{N} \sum_{j}^{W} \left(c(w)_{ij} - w\right)\left(c(w)_{ij} - w\right)^{\top} \tag{1}$$

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.

Objective function of word2vec

- skip gram:

$$J = \sum_{w \in D} \sum_{c \in C} \log P(c|w) \tag{2}$$

- CBOW:

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

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**

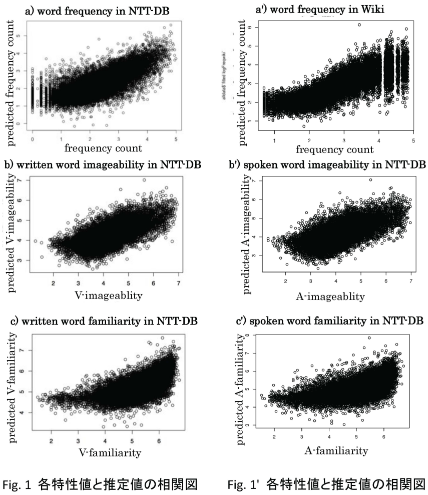

Figure 4: Relations between NTT-DB and word2vec 近藤・浅川 (2017)

Negative Sampling

$P(\cdot)$ は以下のような softmax function:

$$P(c\,|w) = \frac{\exp(\boldsymbol{w}_w^\top \tilde{\boldsymbol{v}}_c)}{\sum_{w'} \exp(\boldsymbol{v}_w^\top \tilde{\boldsymbol{v}}_{w'})} \tag{4}$$

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。

$$\log P(C\,|w) \propto \log\sigma\left(\boldsymbol{v}_w^\top \tilde{\boldsymbol{v}}_c\right) + \kappa\mathbb{E}_{r\sim P_n}\left[\log\sigma\left(-\boldsymbol{v}_w^\top \tilde{\boldsymbol{v}}_r\right)\right], \tag{5}$$

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[1] であることを主張。

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

Shifted PMI

$$M_{i,j} = PMI\left(w_i, c_j\right) - \log\kappa \approx \boldsymbol{w}_i^\top \widetilde{\boldsymbol{w}}_j \tag{6}$$

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(Skip-gram with Negative Sampling) **の目的関数を変形すると，**

$$\begin{aligned} J &= -\sum_{w\in D}\sum_{c\in C}\log\sigma\left(\boldsymbol{v}_w^\top\widetilde{\boldsymbol{v}}_c\right) - \kappa\mathbb{E}_{r\sim P_n}\left[\log\sigma\left(-\boldsymbol{v}_w^\top\widetilde{\boldsymbol{v}}_r\right)\right] && (7) \\ &= -\sum_{w\in D}\sum_{c\in C} n(w,c)\log\sigma\left(\boldsymbol{v}_w^\top\widetilde{\boldsymbol{v}}_c\right) \\ &\quad -\sum_{w\in C} n(w)\,\kappa\mathbb{E}_{r\sim p_n}\left[\log\sigma\left(-\boldsymbol{v}_w^\top\widetilde{\boldsymbol{v}}_r\right)\right] && (8) \end{aligned}$$

期待値の部分を明示的に計算すると，

$$
\begin{aligned}
E_{r\sim p_n}\left[\log\sigma\left(-\boldsymbol{v}_w^\top \widetilde{\boldsymbol{v}}_r\right)\right] &= \sum_{r\in v_c}\frac{n(r)}{|D|}\log\sigma\left(-\boldsymbol{v}_w^\top \widetilde{\boldsymbol{v}}_r\right) &(9)\\
&= \frac{n(c)}{|D|}\log\sigma\left(-\boldsymbol{v}_w^\top \widetilde{\boldsymbol{v}}_c\right)\\
&\quad + \sum_{r\in v_c\backslash c}\log\sigma\left(-\boldsymbol{v}_w^\top \widetilde{\boldsymbol{v}}_r\right) &(10)
\end{aligned}
$$

目的関数のうち，w と c に関する部分だけを取り出すと，

$$\updownarrow(w,c) = -n(w,c)\log\sigma\left(\boldsymbol{v}_w^{\top}\widetilde{\boldsymbol{v}}_c\right) - n(w,c)\,\kappa\frac{n(c)}{|D|}\log\sigma\left(-\boldsymbol{v}_w^{\top}\widetilde{\boldsymbol{v}}_c\right) \tag{11}$$

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

$$\begin{aligned} \frac{\partial l(w,c)}{\partial x} &= -n(w,c)\,\sigma(-x) + \kappa n(w)\,\frac{n(c)}{|D|}\sigma(x) && (12)\\ &= n(w,c)\{\sigma(x)-1\} + \kappa n(w)\,\frac{n(c)}{|D|}\sigma x && (13)\\ &= 0 && (14) \end{aligned}$$

等式を整理すると，

$$\left\{1 + \frac{\kappa n(w)\, n(c)}{|D| n(w,c)}\right\} \sigma(x) = 1 \Leftrightarrow \exp(-x) = \frac{\kappa n(w)\, n(c)}{|D| n(w,c)} \tag{15}$$

したがって，

$$\begin{aligned} x &= \boldsymbol{v}_w^\top \widetilde{\boldsymbol{v}}_c && (16) \\ &= \log \frac{|D| n(w,c)}{\kappa n(w)\, n(c)} && (17) \\ &= \log \frac{|D| n(w,c)}{n(w)\, n(c)} - \log \kappa && (18) \\ &= PMI(w,c) - \log(\kappa) && (19) \end{aligned}$$

文献

Goldberg, Y., & Levy, O. (2014). *word2vec* explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jĕgou, 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**.

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 embeddingas 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). Montrèal CANADA: Curran Associates, Inc.

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.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

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.