

Transformer におけるフィードフォワードネットワークの作用

小林悟郎¹ 栗林樹生^{1,2} 横井祥^{1,3} 乾健太郎^{1,3}

¹ 東北大学 ² Langsmith 株式会社 ³ 理化学研究所

goro.koba@dc.tohoku.ac.jp {kuribayashi, yokoi, inui}@tohoku.ac.jp

概要

近年 Transformer ネットワーク [1] が大きな成功を収め、Transformer を用いたモデルの分析が盛んに行われている。しかし多くの研究は注意機構に焦点を絞り、他のモジュールを無視して分析を行っている。特に、多くの学習パラメータを持つフィードフォワードネットワークを分析に組み込む試みはこれまでほとんどない。本研究ではフィードフォワードネットワークにも光を当て、層を構成するすべてのモジュールを考慮した分析方法を提案する。BERT[2] を対象とした実験では、フィードフォワードネットワークが単語埋め込み間の混ぜ合わせに興味深い影響をもたらすことを明らかにする。

1 はじめに

近年、Transformer ネットワークが自然言語処理の幅広いタスクに性能向上をもたらし、基盤技術となりつつある。その成功理由の解明および更なる性能改善を目指して Transformer を用いたモデル (BERT など) の分析が盛んに行われている [3]。Transformer は入力系列の情報を混ぜ合わせながら単語埋め込みを更新していく。そこで既存分析の主な関心は、モデルが単語埋め込みを更新する際に文脈情報をどのように混ぜ合わせているかにある。多くの既存研究では、直接的に単語埋め込み同士の混ぜ合わせを実現する注意機構の振る舞いに焦点を絞った分析が行われてきた [4, 5, 6, 7, 8, 9, 10, 11, 12]。

しかし、Transformer を構成する要素は注意機構だけではなく、その他のモジュールも間接的に混ぜ合わせに影響を及ぼしうる。そこで我々はこれまで、Transformer をより精緻に分析するために分析範囲を拡張することを試みてきた [12, 13]。その結果、分析範囲を拡張することでモデル内部の観察が大きく変わることが分かっており、対象モジュールを絞った限定的な分析を続けることには不完全な知見をコミュニティ内に積み上げてしまう危険がある。

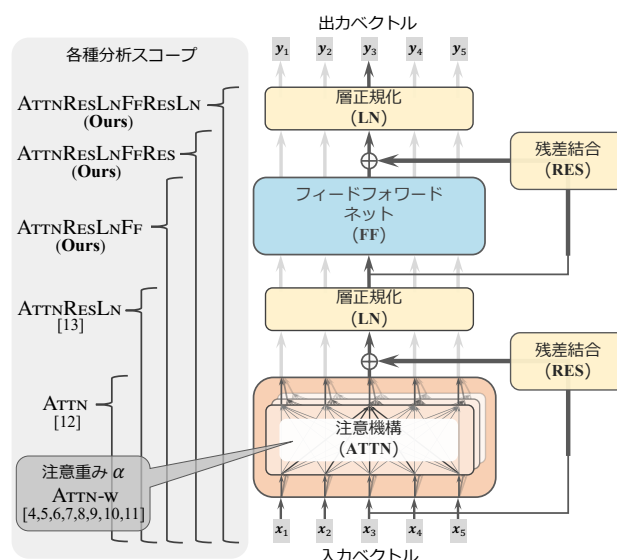


図1 Transformer 層の概要と各種分析方法のスコープ。

本研究では新たにフィードフォワードネットワークを分析に組み込み、分析範囲を層全体まで拡張する。代表的なモデルである BERT を対象とした実験の結果、フィードフォワードネットワークによって混ぜ合わせの振る舞いに変化すること、さらに言語的に関係深い単語ペア (e.g., 意味的つながりが強い単語ペア) の相互作用が増幅されることなどがわかった。

2 準備：Transformer

Transformer は入力されたベクトル系列 (トークン列) を混ぜ合わせながら更新する層を積み上げて構成される。各層は (1) 注意機構, (2) フィードフォワードネットワーク, (3) 残差結合, (4) 層正規化の 4 種類の要素からなる (図 1)。Transformer を用いたモデルを分析する多くの既存研究は、注意機構だけに焦点を当ててきた [4, 5, 6, 7, 8, 9, 10, 11, 12]。最近、我々が注意機構に加えて残差結合と層正規化を考慮することを提案した [13] が、フィードフォワードネットワークの影響ははまだ考慮されていない。本研究ではフィードフォワードネットワークを分析に組み込み、層内の混ぜ合わせにもたらされる影響を調査する。

注意機構 (ATTN) は単語埋め込み列 $\{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$ を混ぜ合わせることで各埋め込み x_i を更新する機構である。具体的には、各単語ペア (x_i, x_j) の関係性に相当する注意重み $\alpha_{i,j} \in [0, 1]$ を計算しその値に基づいて周囲の埋め込みを混ぜ合わせることで、関係深い単語の埋め込みから情報を多く集めながら埋め込みを更新することを実現する。

フィードフォワードネット (FF) は活性化関数を間に挟んだ 2 層の線形ニューラルネットワークである。各単語埋め込みに対して独立に同じ変換を行うため、この機構単体では埋め込み同士に混ぜ合わせのような相互作用は生まれない。以降ではフィードフォワードネットを FF と省略して表記する。

残差結合 (RES) は注意機構および FF の前後を短絡する形で、それら機構への入力ベクトルをその出力ベクトルに足し込み、元の情報を残す。

層正規化 (LN) は各ベクトルに要素平均を 0、要素分散を 1 にする正規化を行った上で、各要素をスケールするアフィン変換を行う。

3 関連研究

本節では Transformer を用いたモデル (BERT など) の内部挙動を分析する既存手法を概観し、本研究の位置付けについて述べる。

注意機構の分析 モデル内部での単語間の混ぜ合わせを調べるために多くの研究で注意機構に焦点が当てられてきた。特にアテンション重み α の値を観察するのが最も典型的な分析方法である [4, 5, 6, 7, 8, 9, 10, 11] (図 1 の ATTN-w)。この方法は機構内の α 以外の影響を無視してしまうが、我々の以前の研究ではその点を改良した注意機構の包括的な分析方法を提案した [12] (図 1 の ATTN)。

分析範囲を拡張する試み これらの分析では Transformer を構成する 4 要素 (セクション 2) のうち注意機構だけしか考慮されない。そこで注意機構以外の構成要素も分析に組み込むことも過去の研究で試みた。具体的には、ベクトルノルムを用いることで層の前半部分 (注意機構・残差結合・層正規化) を包括的に分析する方法を提案した [13] (図 1 の ATTNRESLN)。そして実際に、この分析範囲の拡張によってモデル内での単語間の混ぜ合わせの見え方が大きく変化することを報告した。

既存の問題点と本研究の立ち位置 しかしながら Transformer の学習可能パラメータの半数以上 (埋め込み層を除く) を占める FF を分析に組み込む試み

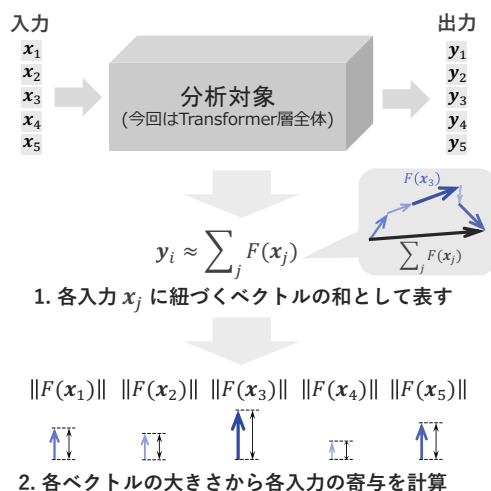


図 2 ベクトルノルムに基づく分析の概要。

は現状ほとんどない。数少ない試みは FF を独立に調べるもので [14, 15], 注意機構など直前までのモジュールによる混ぜ合わせに与える影響については明らかになっていない。本研究は FF を注意機構などの周囲のモジュールと合わせて分析に組み込むことで、FF が単語同士の混ぜ合わせにどのような影響を与えるのかを調べる。

4 提案手法

Transformer を構成する 4 種類の要素全て (2 節) を考慮した、1 層の包括的な分析法を提案する。

4.1 ベクトルノルムに基づく分析

本研究の提案手法では我々が以前提案したベクトルノルムを用いた方法 [12] を採用する。この方法はアテンション重み α を観察するような方法と違い、分析範囲を自然に拡張できる。また、勾配を用いて分析対象の入出力間の変化を分析する方法と違い、入力ベクトルによる影響も考慮できる。ベクトルノルムを用いた分析の概要を図 2 に示す。分析は次の 2 段階で行われる：

1. 分析したい処理を各入力 x_j に紐づくベクトルの和に書き直す (加法分解する)。
2. 分解された各ベクトルのノルムを測ることで出力 y_i に対する各入力 x_j の貢献を測定する。

この方法はベクトルの総和において長いベクトルほど出来上がるベクトルに強く影響を与えるという性質を基に考案されている。3 節で述べた注意機構の包括的な分析 (図 1 の ATTN) [12], および層の前半部分 (注意機構・残差結合・層正規化) の包括的な

分析 (図 1 の ATTNResLN) [13] はこの方法を用いて実現したものである。本研究では分析対象を FF に広げることで層の包括的な分析を実現する。

4.2 FF の加法分解

分析の第 1 段階として今回の分析対象である Transformer 層全体を加法分解したい。層全体の加法分解は、層を構成する 4 要素の処理をそれぞれ加法分解することで実現できる。このうち注意機構・残差結合・層正規化の 3 要素についてはそれぞれ近似なしで加法分解できることを以前の研究 [12, 13] で示している。しかし本研究で新たに組み込む FF には活性化関数が含まれている¹⁾。非線形な作用には分配則が適用できないため近似なしで加法分解できない。そこで活性化関数部分を近似的に加法分解する方法について検討した。分解前後で合計値が変化しないことを要件とし、これを満たす近似手法として (1) 勾配に基づいた近年代表的な特徴量帰属手法 *Integrated Gradients* [17] と (2) ゲーム理論に基づいて出力への寄与度を分配する *Shapley Value* [18] を候補に挙げた。両手法で試したところ、*Shapley Value* には系列長に応じて急激に計算速度が遅くなる問題があったため、本研究では *Integrated Gradients* を採用した。他の方法も踏まえて近似分解手法を比較検討することは今後の展望の一つである。

4.3 提案手法まとめ

以上によりすべての構成要素の加法分解が可能となり、Transformer 層での埋め込みの更新を入力に紐づくベクトルの和 $y_i \approx \sum_j F(x_j)$ で表すことができた。あとは分析の第 2 段階として分解した各ベクトルのノルム $\|F(x_j)\|$ を測ることで、埋め込み x_i を y_i へと更新する際の各入力埋め込み x_j の貢献を測る。測定した値はその層における j 番目の単語から i 番目の単語への混ざ度合いに相当しており、その値の大きさからモデルが周囲の各単語情報をどれほど強く混ざらせているかを観察できる。

5 実験

Transformer を用いた代表的なモデルである BERT の内部挙動を FF の作用に注目しながら分析する。5.1 節では FF が文脈の混ざり合わせを変化させることをデータ全体のマクロな分析から確かめる。5.2 節では個別の単語ペアに注目したミクロな分析によ

1) ReLU や GELU[16] が一般的に使われる。

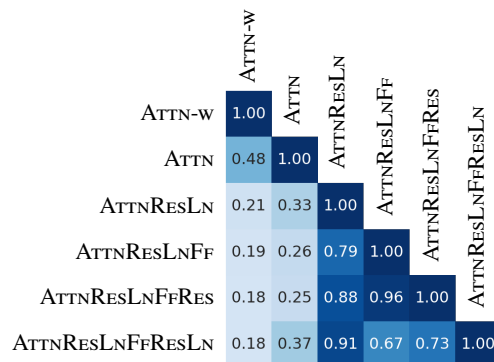


図 3 分析範囲の異なる手法間の結果の類似度。

り、FF が言語的に興味深い単語ペアの混ぜ合わせを大きく変化させる傾向を明らかにする。

モデル 事前学習済み BERT-base (uncased) を用いた。BERT-base は 12 層から構成されており、その各層が注意機構、フィードフォワードネットワーク、残差結合、層正規化から構成されている (図 1)。

データ 既存研究 [4, 12, 13] に従い、モデルへ入力するテキストデータに Wikipedia から抽出された 992 系列 [4] を用いた。BERT の事前学習の設定に合わせ、各系列の 12%²⁾ のトークンは [MASK] トークンに置換してからモデルに入力した。

分析手法 実験では分析範囲の異なる 6 種類の手法で分析し、結果を比較することで各モジュールの混ぜ合わせへの影響を調査する。各手法の分析範囲は図 1 に示す通りである。ATTN-w は注意重み α を分析する最も典型的な方法で、他 5 手法はベクトルノルムに基づく分析 (4.1 節) を採用している。

5.1 各モジュールの混ぜ合わせへの影響

分析に考慮するモジュールを増やすことで混ぜ具合の観察結果がどの程度変化するかを確認する。

手順 BERT に各系列 (トークン数 n) を入力し、6 種類の分析手法それぞれで入力単語間の混ぜ度合い ($n \times n$ の行列) を算出する。続いて系列毎に手法間で結果の類似度を測り、全系列で平均する。その後、モデル全体での傾向を調べるために各層から算出された値を平均する。行列間の類似度を測定する方法は、多変量解析法である表現類似性分析 [19] に従う。具体的には各行列 ($\mathbb{R}^{n \times n}$) をベクトル (\mathbb{R}^{n^2}) に平坦化してスピアマン順位相関係数を計算する。

結果・考察 図 3 に 6 種類の分析手法で算出された混ぜ合わせの結果同士の類似度を示す。

2) BERT の事前学習では入力系列の 15% が選ばれ、そのうち 80% (全体の 12%) が [MASK] トークンに置換される。

表1 FFが混ぜ度合いを増幅した単語ペア.

層	単語ペア例
1	(det -> ##our), (week -> ##night), (phones -> cell)
6	(behalf -> on), (tasked -> with), (impact -> on)
9	(never -> despite), (she -> juliet), (died -> 1993)
10	(maison -> maison), (deeds -> deeds), (bela -> bela)

表2 FFが混ぜ度合いを減衰した単語ペア.

層	単語ペア例
3	(june -> october), (february -> may), (april -> may)
5	(hard -> hard), (choice -> choice), (ties -> ties)
10	(moldova -> [SEP]), (kenya, [SEP])
12	(juventus -> [CLS]), (apple -> [CLS]), (gray -> [CLS])

FFは混ぜ合わせを変化させる：FFを分析に組み込む前後で混ぜ合わせの類似度は0.79で1.0より確かに低く、FFが混ぜ合わせに変化をもたらしていることが分かった。FFは埋め込み間に相互作用を生まないため一見混ぜ合わせには寄与しないように思えるが、その各入力直前の注意機構などを通して埋め込みが混ぜられたものであり、FFはベクトル変換を通してその混ぜ具合を確実に変化させている。

残差結合・層正規化も混ぜ合わせを変化させる：FFの後にある残差結合の前後では類似度が0.96、層正規化の前後では類似度が0.73であり、残差結合と層正規化も混ぜ合わせに変化をもたらしていることが分かった。特に層正規化による変化は大きく、アフィン変換が強く働いていると考えられる。この傾向を詳細に調べることは今後の展望の一つである。

5.2 フィードフォワードネットの仕事

前節ではFFが単語埋め込み間の混ぜ具合を変化させていることを明らかにした。本節ではこのFFによる混ぜ合わせへの影響についてマイクロに調査し、どのような単語ペアに対して大きな変化をもたらすのかを明らかにする。

手順 BERTに各系列を入力し、FFを分析に組み込む前後 (ATTNRESLN と ATTNRESLNFF) でそれぞれ単語間の混ぜ度合い ($n \times n$ の行列) を算出する。それぞれの結果に対して各行の合計値が1になるよう正規化を行って³⁾から差分を取り、FFが各2単語間での混ぜ合わせ ($w_j \rightarrow w_i$) にもたらす変化量を計算する。これを単語の組み合わせ毎にデータ全体で平均し、変化が大きい傾向にある単語ペアを特定する。このとき周囲文脈との関係に注目する目的で、同じ位置の単語への混ぜ合わせ ($w_j \rightarrow w_j$) は除外した。さらに外れ値の影響に対処するため、データ中に一度しか出現しないペアも集計から除外した。

結果・考察 各層のFFが混ぜ度合いを平均的に大きく増幅もしくは減衰した単語ペアを特定し、そのうち顕現的なペアを表1および表2にまとめた。⁴⁾

3) 手法間で値のスケールを統一するために正規化を行った。

4) 平均的に変化量が大きい上位10ペアから興味深い傾向のあった3ペアを列挙した。10ペアすべてを付録Aに示す。

関係深い単語ペアの混ぜ度合いの増幅 (表1)：1層目では2トークンで1つの単語をなすサブワードのペア (detour や weeknight) や熟語をなす2単語 (cell phones) といった、意味の繋がりが強く、隣り合ったペアで混ぜ合わせが増幅されていた。6層目では副詞を伴うコロケーションを成す単語ペア、9層目では文内において意味的な接続があると予想される単語ペアでの混ぜ合わせが増幅されていた。10層目では同系列内に複数個出現する同じ単語同士での混ぜ合わせが増幅される傾向にあった。以上により、FFは何らかの意味で関係深い単語ペアの混ぜ度合いを増幅する傾向にあることが示唆される。

同グループに属する単語同士の混ぜ合わせと特殊トークンへの混ぜ合わせの減衰 (表2)：3層目では月名単語同士の混ぜ合わせが減衰され、5層目では同系列内に複数回登場する同じ単語同士での混ぜ合わせが減衰されていた。これより、FFが同グループに属する単語同士に対して不要な混ぜ合わせを抑制する働きを持つことが示唆される。10, 12層目では特殊トークンである [CLS] と [SEP] への混ぜ合わせが減衰される傾向にあった。特に最終層である12層目で [CLS] への混ぜ合わせを減衰していることは、[CLS] の埋め込みを用いて2文間の連続性について予測する事前学習タスクを解くために重要な情報を抑制している可能性が考えられる。

6 おわりに

本稿では、Transformerの各層における単語埋め込み間の混ぜ合わせを包括的に分析する方法を提案した。BERTを対象とした実験結果から、FFがベクトル変換を通じて混ぜ具合を変化させていることを明らかにした。特に、FFは意味的な接続や事前学習タスクに紐づいた混ぜ合わせに大きく変化をもたらすことが示唆された。

今後は、PMIなどの言語的性質と照らし合わせながらFFの働きについて更に深掘りしていく。また、これまでのように各層を独立に分析するのではなく層を跨ぐ拡張を導入することで、複数層もしくはモデル全体を包括的に分析する方向性も興味深い。

謝辞

本研究は JSPS 科研費 JP20J22697 の助成を受けたものです。また本研究は、JST, ACT-X, JPMJAX200S および JST, CREST, JPMJCR20D2 の支援を受けたものです。

参考文献

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In **Advances in Neural Information Processing Systems 30 (NIPS)**, pp. 5998–6008, 2017.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)**, pp. 4171–4186, 2019.
- [3] Anna Rogers, Olga Kovaleva, and Anna Rumshisky. A Primer in BERTology: What We Know About How BERT Works. **Transactions of the Association for Computational Linguistics**, Vol. 8, pp. 842–866, 2021.
- [4] Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D Manning. What Does BERT Look At? An Analysis of BERT’s Attention. In **Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP**, pp. 276–286, 2019.
- [5] Olga Kovaleva, Alexey Romanov, Anna Rogers, and Anna Rumshisky. Revealing the Dark Secrets of BERT. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 4364–4373, 2019.
- [6] Emily Reif, Ann Yuan, Martin Wattenberg, Fernanda B Viegas, Andy Coenen, Adam Pearce, and Been Kim. Visualizing and Measuring the Geometry of BERT. In **Advances in Neural Information Processing Systems 32 (NeurIPS)**, pp. 8594–8603, 2019.
- [7] Yongjie Lin, Yi Chern Tan, and Robert Frank. Open Sesame: Getting Inside BERT’s Linguistic Knowledge. **Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP**, pp. 241–253, 2019.
- [8] David Mareček and Rudolf Rosa. From Balustrades to Pierre Vinken: Looking for Syntax in Transformer Self-Attentions. In **Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP**, pp. 263–275, 2019.
- [9] Phu Mon Htut, Jason Phang, Shikha Bordia, and Samuel R. Bowman. Do Attention Heads in BERT Track Syntactic Dependencies? **arXiv preprint arXiv:1911.12246**, 2019.
- [10] Alessandro Raganato and Jörg Tiedemann. An Analysis of Encoder Representations in Transformer-Based Machine Translation. In **Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP**, pp. 287–297, 2018.
- [11] Gongbo Tang, Rico Sennrich, and Joakim Nivre. An Analysis of Attention Mechanisms: The Case of Word Sense Disambiguation in Neural Machine Translation. In **Proceedings of the 3rd Conference on Machine Translation (WMT): Research Papers**, pp. 26–35, 2018.
- [12] Goro Kobayashi, Tatsuki Kuribayashi, Sho Yokoi, and Kentaro Inui. Attention is Not Only a Weight: Analyzing Transformers with Vector Norms. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 7057–7075, 2020.
- [13] Goro Kobayashi, Tatsuki Kuribayashi, Sho Yokoi, and Kentaro Inui. Incorporating Residual and Normalization Layers into Analysis of Masked Language Models. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 4547–4568, 2021.
- [14] Mor Geva, Roei Schuster, Jonathan Berant, and Omer Levy. Transformer Feed-Forward Layers Are Key-Value Memories. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 5484–5495, 2021.
- [15] Daisuke Oba, Naoki Yoshinaga, and Masashi Toyoda. Exploratory Model Analysis Using Data-Driven Neuron Representations. In **Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP**, pp. 518–528, 2021.
- [16] Dan Hendrycks and Kevin Gimpel. Gaussian Error Linear Units (GELUs). 2016.
- [17] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic Attribution for Deep Networks. In **Proceedings of the 34th International Conference on Machine Learning**, Vol. 70 of **Proceedings of Machine Learning Research**, pp. 3319–3328. PMLR, 2017.
- [18] Shapley L. S. A Value for n-person Games. **Contributions to the Theory of Games**, pp. 307–317, 1953.
- [19] Nikolaus Kriegeskorte, Marieke Mur, and Peter Bandettini. Representational similarity analysis - connecting the branches of systems neuroscience. **Frontiers in Systems Neuroscience**, Vol. 2, p. 4, 2008.

A 5.2 節の結果の詳細

5.2 節ではフィードフォワードネットによる混ぜ合わせへの影響が大きい単語ペアの中から顕現的なものを列挙した (表 1 および表 2). 本節ではそれらを列挙する前の, 影響が大きかった上位 10 ペア全てを表 3 および表 4 に示す. ここで上位ペアを算出する際の計算量の問題により, データ全体で 1 度しか出現しないペアが上位に多すぎる層では表の結果に 10 ペア未満しか載っていない.

表 3 フィードフォワードネットによって混ぜ度合いが増幅した上位単語ペア.

層	増幅された単語ペア
1	(det -> ##our), (space -> outer), (res -> ##ili), (##imi -> ##nate), (art -> deco), (week -> ##night), (phones -> cell), (spear -> ##head), (missile -> ballistic), (##ush -> ##chev)
2	(national -> guard), ([SEP] -> healthcare), ([SEP] -> pumped), (f -> ##oy), (could -> easily), (s -> '), (latin -> america), (anderson -> anderson), (##lard -> ##lard), ([SEP] -> ahmed)
3	(##k -> paso), (##om -> sans), (##p -> ##la), (##ant -> avoid), (though -> —), (##mel -> water), (##wind -> nord), (##oh -> rabbit), (##ci -> ##pres), (alcoholism -> alcoholism)
4	(ds -> ##l), (ter -> ter), (##ili -> ##ent), (##tera -> ##l), (spa -> ##ll), (hansen -> hansen), (mn -> ##et), (baker -> baker), (##lea -> ##r), (7 -> 151)
5	(##ili -> ##ent), (##jo -> ##rative), (##ili -> res), (ds -> ##l), (##ense -> ##r), (nam -> ##ur), (##vor -> ##able), (eli -> ##el), (##pres -> ##ci), (##ear -> ##ment)
6	(nam -> ##ur), (behalf -> on), (benefit -> from), (denoted -> by), (##pres -> ##ci), (considers -> be), (paso -> ##k), (##lk -> ##o), (tasked -> with), (impact -> on)
7	(st -> dun), (st -> vi), (##car -> pre), (among -> vessels), (##ife -> ##l)
8	(##lz -> ##ano), (against -> serbia), (than -> accidents), (- -> 1643)
9	(never -> despite), (##ef -> sha), (she -> juliet), (died -> 1993), (and -> swear)
10	(##nen -> jan), (##nen -> ##tu), (maison -> maison), (deeds -> deeds), (fame -> fame), (chemicals -> chemicals), (bela -> bela), (fruit -> fruit), (##oping -> link), (##rmi -> fe)
11	(czech -> ve), (catalonia -> ##ll), (hack -> hack), (##water -> ##water), (czech -> ##ka), (##nen -> jan), (tiny -> tiny), (khmer -> ##k), (singer -> fine), (juventus -> juventus)
12	(luis -> luis), (france -> montpellier), (administrative -> ##c), (" -> ballad), (" -> operators), (同 -> tong), (index -> ##i), (base -> near), (##re -> ##ide), (analysis -> ##ss)

表 4 フィードフォワードネットによって混ぜ度合いが減衰した上位単語ペア.

層	減衰した単語ペア
1	(spa -> ##l), (c -> &), (e -> ##qui), (f -> ##oy), (##tree -> ##t), (dr -> ##iest), (tr -> ##ai), (seven -> months), (red -> sandstone), (red -> ##ef)
2	([CLS] -> "), ([CLS] -> -), ([CLS] -> "), ([CLS] -> —), ([CLS] -> ') ([CLS] -> chapels), ([CLS] -> ##zza), (##ffe -> ##uli), (##ef -> sha), (##op -> ##hr)
3	(june -> october), (february -> may), (##pre -> [SEP]), (15 -> 21), (april -> may), ([CLS] -> [SEP]), (excuses -> [SEP]), (july -> august), (breakup -> [SEP]), (january -> march)
4	(l -> /), (kenya -> central), (festival -> prize), (but -> unable), (1948 -> war), (##rmin -> ##rmin), (ld -> ld), (td -> td), (but -> merely), (##hy -> ##hy)
5	(hard -> hard), (choice -> choice), (##zo -> ##zo), (##hu -> ##hu), (##vis -> ##vis), (##ap -> ##ap), (##aa -> ##aa), (ties -> ties), (##och -> ##och), (ti -> ti)
6	(also -> addition), (nominee -> republican), (differences -> between), (##ulus -> ##nus), (appeared -> having), (##or -> no), (obe -> ##rst)
7	(lead -> 0), (households -> 47), (##ant -> avoid), (dal -> "), (set -> 3), (rugby -> national), (ethel -> ')
8	(##fers -> upon), (km ² ->), (bought -> from), (44 -> 25), (##hen -> ##hai), (tournament -> ;), (africa -> my), (inherently -> not), (reserves -> last)
9	([CLS] -> ##kill), (relative -> .), (recruited -> .), ([CLS] -> pr), (voiced -> '), ([CLS] -> thai), (debated -> .), ([CLS] -> hockey), (beetle -> the), (##by -> ka)
10	(##uke -> ##t), (moldova -> [SEP]), (##bo -> ##u), (moroccan -> [SEP]), (##by -> ##lie), (kenya -> [SEP]), (##n -> ##ja), (##ari -> ##nen), (judo -> [SEP]), (taiwanese -> [SEP])
11	(38 -> 8), (##cet -> .), (##rmin -> ##rmin), (##oum -> ##oum), (erebidae -> the), (erebidae -> fa,o;u), (##rgeon -> ##rgeon), (discuss -> discuss), (comprise -> was), (religion -> nationality)
12	(juventus -> [CLS]), (apple -> [CLS]), (gray -> [CLS]), (##uman -> [CLS]), (jc -> [CLS]), (henley -> [CLS]), (bulge -> [CLS]), (rum -> [CLS]), (fungus -> [CLS]), (1612 -> [CLS])