対象言語・対象単語を選ばない汎用的な文法化度の定量化手法

永田亮 [1] 持橋大地 [2] 井戸美里 [3] 窪田悠介 [3] 高村大也 [4] 川崎義史 [5] 大谷直輝 [6]
[1] 甲南大学 [2] 統計数理研究所 [3] 国立国語研究所
[4] 産業技術総合研究所 [5] 東京大学 [6] 東京外国語大学
nagata-nlp2025 @ ml.hyogo-u.ac.jp. daichi@ism.ac.jp {ido,kubota}@ninjal.ac.jp
takamura.hiroya@aist.go.jp ykawasaki@g.ecc.u-tokyo.ac.jp otani@tufs.ac.jp

概要

文法化とは内容語が機能語に変わる通時変化のことである．本稿では，文法化の度合いをコーパスデータに基づいて定量化する三種類の手法を提案する．提案手法は従来手法と異なり汎用性が高いにもかかわらず，人手の分析で得られた文法化度と中程度〜高い相関を示す．また，提案手法を歴史コーパスに適用して文法化の一方向仮説の吟味を行う．

1 はじめに

文法化とは，内容語が機能語に変化する現象を指す [1]．例えば，英語の can は，14 世紀ごろまで「知る」という意味の本動詞で使われていたが，現代英語では助動詞に変化している．

文法化について盛んに研究されているトピックの一つに文法化度の定量化がある．文法化を段階的な変化と捉え，文法化の程度を数値化する試みである．なかでも，英語動詞派生前置詞（例：following）について研究例 [2, 3, 4] が多い．しかしながら，従来研究は事例分析や内省に基づくため研究規模を大きくすることは容易でない．自動的に文法化度を定量化する手法 [5] も存在はするが，英語動詞派生前置詞に特化しており適用範囲が限られている．文法化は通言語的で，かつ，様々な単語に見られる [6, 7] ため，汎用的な手法の考案が求められる．

そこで，本稿では，汎用的な文法化度の定量化手法を三種類提案する．一つ目は，単語の出現間隔に着目した手法で，必要とするのは対象言語のコーパスのみである．二つ目の手法は文法化度の定量化を確率的な分類問題として解く．すなわち，機能語である確率を文法化度とみなす．その際，分類器の学習のための訓練データが存在しないということが問題となるが，Positive-Unlabeled learning（PU-learning）[8] という機械学習手法を利用してこの問題を解決する．最後の手法は，二番目の手法を品詞タガーを用いて拡張する．品詞タグ付けの結果から，典型的な機能語と内容語のリストを作成し，その結果に，Cross-Validation を応用した分類器の訓練方法を適用することで，更に性能を向上させる．

本研究の貢献は次の三つである：(i) 三種類の新たな文法化度定量化手法を提案し，汎用性を大幅に改善した；(ii) 汎用性が高いにもかかわらず，人間の判断との相関が高いことを英語動詞派生前置詞と日本語名詞を対象にして示した；(iii) 提案手法を利用して，文法化の一方向仮説 — 内容語から機能語への一方向にしか変化は起こらないという仮説 — の検証を大規模な歴史コーパス中の 1459 語を対象にして行った．

2 関連研究

文法化に関する言語学的研究は多く，文献 [1, 9, 6, 10, 11] などがある．従来研究では，文法化を内容語から機能語への段階的変化と捉えることが多い．例えば，文献 [2, 3, 4] は，動詞派生前置詞の文法化を段階的に捉え定量化してる．なかでも，Hayashi [4] は 37 種類の動詞派生前置詞に対して 0〜10 の文法化度スコアを与えており，研究の規模が大きい．日本語については，寺村 [12] が名詞の接続詞化および助動詞化を判定する一連のテストを提案している．また，それぞれ 39 および 16 の名詞に対して，テスト結果を示している．

文法化度を自動的に定量化する研究は非常に限られている．我々が知る限り，唯一の例外は Nagata ら [5] の研究である．この研究では，動詞派生前置詞の分散表現と前置詞／動詞の分散表現間の余弦類似度に基づいて文法化度を決定する．ただし，この手法は動詞派生前置詞に特化した手法である．Nagata ら [5] は，対象単語の文脈の多様性を定量化する単語ベクトルの集中度という指標に基づいた汎

用的な手法も提案しているが，人間の判断との相関は低いと報告している．

文法化の一方向仮説も盛んに研究が行われている．多くの言語学者（例：文献 [13]）は一方向仮説を支持する一方で，反例を示す研究者 [14, 15] も存在する．本稿では，提案手法を歴史コーパス CCOHA [16] に適用し，一方向仮説の吟味を行う．

3 文法化度定量化手法

3.1 間隔分布に基づいた手法

本手法の理解のために，まず，図 1 に示す内容語 effects と機能語 in の出現間隔を吟味する．横軸は CCOHA 上の単語位置に対応し，上述の単語が出現した位置に青い縦線を描画している．effects は出現頻度は低いが，一旦出現するとしばらく短い間隔で出現することが分かる．一方で，in は，コーパス全体を通して出現間隔が短く，より一様である．もし，内容語および機能語一般に同様な傾向が見られるのであれば，間隔分布を通じて，文法化度が定量化できる可能性がある．

以上の観察に基づき，間隔分布に基づいた文法化度を提案する．幸い，Altmann [17] は，単語の間隔分布は Weibull 分布でモデル化できることを示している．同分布は，$f(\tau) = \frac{\beta}{\eta}(\frac{\tau}{\eta})^{m-1}\exp\{-(\frac{\tau}{\eta})^{\beta}\}$ で定義される．ただし，τ は単語の出現間隔（ある単語が出現してから次に出現するまでの間に出現した別の単語の数+1 と定義）を表す[1)]．また，β と η は分布のパラメータである．β は分布の形状を決めるパラメータであり，Altmann [17] は，機能語では値が大きくなることを経験的に示している．

以上より，Weibull 分布のパラメータ β を文法化度の指標とすることを提案する．具体的な処理手順は次の 3 ステップに従う：各対象単語について，(i) 入力コーパス中の出現間隔を算出；(ii) その結果から β を推定；(iii) 推定値を文法化度として出力．以上の通り，この手法で必要となるのはコーパスデータだけであり，言語と単語に非依存な手法である．

3.2 PU-Learning に基づく手法

本手法は，対象単語 w が文法化している確率 $p(w)$ を確率的分類問題として直接モデル化する．言い換えれば，$p(w)$ の推定値を単語 w の文法化度と見なす．確率 $p(w)$ のモデルとしてロジスティック回帰など様々な確率的分類器を用いることができる．特に，ベクトルに対する分類器であれば，対象単語の分散表現を入力として用いることができる．

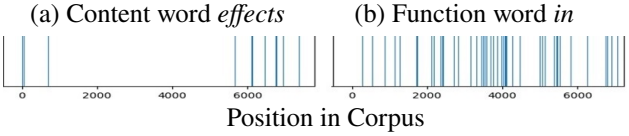

図 1 CCOHA の 2000 年の代文書における内容語 effects と機能語 in の間隔分布．青線が出現位置を示す．

ここで，訓練データをどのように得るかということが問題となるが，本手法では，PU-learning [8] という枠組みを用いて解決する．PU-learning では，部分的な正例集合とラベルなしデータ集合から分類器の学習を行う．本手法では，部分的な機能語のリストを用いて正例集合を得る．すなわち，一部の機能語とその他ラベルが未知である語，それぞれの分散表現から分類器を学習する．

処理手順は次の 5 ステップからなる：(i) 入力コーパスから単語分散表現を獲得；(ii) 機能語リスト中の語を正例として訓練データを作成；(iii) PU-learning を適用し，分類器を学習；(iv) 対象単語について $p(w)$ を推定；(v) 推定値を文法化度として出力．大抵の言語で機能語の一部のリストは存在すると考えられるため，この手法も汎用性が高い．

3.3 CV-Learning に基づいた手法

PU-learning に基づいた手法は汎用性の高い手法であるが，機能語の部分的なリストから文法化度を推定するという難しい問題を解かなければならない．そこで，品詞タガーと Cross-Validation を応用した訓練を用いて問題の緩和を行う（以降，CV-learning に基づいた手法と略記する）．品詞タガーを入力コーパスに適用することで，各単語における品詞の分布を推定することができる．大部分が機能語または内容語として使用される語を，それぞれ正例と負例として分類器の学習を行うというのが基本的なアイデアである．ただし，品詞タガーでは文法化の過程にある単語について正しい品詞情報が得られない可能性がある．そのため，負例については，仮の訓練データとして分類器の学習を行う．

処理手順は次の 7 ステップに従う．(i) 入力コーパスから単語分散表現を獲得；(ii) 入力コーパスに品詞タガーを適用し，θ 以上（本稿では $\theta = 0.95$）の割合で機能語または内容語として使われる語の分散表現をそれぞれ正例と負例とする；(iii) 負例を N

1) Weibull 分布は連続変数を対象にした分布である．本稿では，単語出現間隔を近似的に連続変数とみなす．

セットに分割（同じく $N = 10$）；(iv) 分割した負例それぞれと正例を合わせて分類器を N 回学習；(v) N 個の分類器を対象単語の分散表現に適用し，$p(w)$ を推定（ただし，対象単語が訓練データに含まれる分類器は除外して推定）；(vi) 得られた $p(w)$ の推定値を平均；(vii) $p(w)$ の平均値を文法化度として出力．現在では，多くの言語で品詞タガーが利用可能であるため，この手法も適用範囲は広い．

4 評価実験

2 節で述べた動詞派生前置詞 [4] と接続詞化／助動詞化した日本語名詞 [12] の文法化度を用いて評価を行った（詳細は付録 C に記す）．評価尺度は，文献 [5] と同様にスピアマンの順位相関係数とした．コーパスは CCOHA の 2000 年代の文書（英語）と BCCWJ [18]（日本語）を利用した（詳細は付録 A を参照のこと）．単語の分散表現は word2vec [19] を用いて得た．PU-learning には，pulearn package[2] の WeightedElkanotoPuClassifier を用いた．また，分類器は sklearn の LogisticRegression[3] を用いた．付録 B の定義に従い，機能語／内容語で 95%以上使われる語を初期の機能語／内容語と見なした．

比較対象は次の三手法とした．**品詞分布に基づく手法**：各単語について，コーパスにおける機能語での使用比率を文法化度とした．機能語の定義は付録 B に示した．残りの二つは，2 節で述べた**ベクトルの集中度に基づく手法**と**単語依存手法**（動詞派生前置詞の分散表現と前置詞／動詞の分散表現に基づく手法）[5] である．ただし，後者は，動詞派生前置詞に特化した手法であるため，名詞を対象にした評価には使用しない．

表 1 に評価結果を示す．表 1 より，CV-learning に基づいた手法が安定的に高い相関を示すことがわかる．動詞派生前置詞に特化した手法（単語依存手法）よりは相関係数は低いが，それ以外はどの条件でも最も高い相関を示す．いずれのタスクにおいても，言語非依存な既存手法（ベクトルの集中度に基づいた手法）より高い相関を示し，CV-learning に基づいた手法の有効性がうかがえる．一方，PU-learning に基づいた手法については，一部，高い相関を示すが，CV-learning に基づいた手法より性能は低く，正例しか与えられないという問題の難しさがうかがえる．品詞分布に基づく手法の性能から，品詞解析により，ある程度文法化度が推定できることがわかる．ただし，見た目上の相関係数は高いが，品詞分布に基づくと同順位が多数発生する（動詞派生前置詞では 19/25，日本語では 27/37 と 11/15 が同順位であった）．より詳細な文法化度を得るためには，品詞分布以上の情報が必要となる．

間隔分布に基づいた手法は汎用的な手法ではあるが，それだけで文法化度を測ることは難しいこともわかる．統計的に有意ではないが，動詞派生前置詞については負の相関を示す．言い換えれば，文法化が進むほど内容語のような出現間隔を示す傾向が見られる．今回対象にした動詞派生前置詞は，文法化の初期段階にあるものが多く，限定された文脈（例えば，文頭の following や considering）で出現している可能性がある．その場合，機能語のように一様かつ頻繁に出現することはなく，Weibull 分布のパラメータ β は小さい値（すなわち内容語に近い値）として推定される．このように単体での有効性は低いものの，分類器に基づいた二手法やベクトルの集中度に基づく手法とは異なった文法化の側面を捉えているため，間隔分布とその他の手法をうまく組み合わせることで更に性能を高められる可能性がある．

5 一方向仮説の検証

本節では，最も性能が良かった CV-learning に基づいた手法を用いて一方向仮説を吟味する．同手法を，CCOHA の 1800 年代と 2000 年代の文書に適用し，各単語の文法化度を推定した．ただし，文法化が起こるのはある程度の頻度以上の単語であると予想されることから，両コーパスいずれにおいても頻度 3000 以上である単語 1459 語を対象にした．分類器の訓練もこれらの単語を対象とした（動詞派派生前置詞における順位相関係数 0.714（p 値 0.047）．ただし，頻度 3000 という閾値のため対象単語数は 8 である）．

結果を散布図として図 2 に示す．横軸，縦軸それぞれが，1800 年代，2000 年代の文書で推定された文法化度に対応する．可読性のため，単語のラベルについては 1/3 に間引いている．図 2 より，右下の領域には点が少ないことがわかる．言い換えれば，内容語から機能語への変化は非常に少なく，一方向仮説を支持している．対角線の上側では，during, among, due など文法化が進む語が見られる．一方，その近くに cause や social など機能語とは思われない語も出現している．前者については，2000 年代の

2) https://github.com/pulearn/pulearn
3) https://scikit-learn.org/1.5/modules/linear_model.html

表 1　人手の分析と提案手法との間のスピアマンの順位相関係数.

	英語動詞派生前置詞		日本語接続詞化名詞		日本語助動詞化名詞	
手法	順位相関係数	p 値	順位相関係数	p 値	順位相関係数	p 値
間隔分布	-0.301	0.144	0.282	0.090	0.313	0.255
PU-learning	0.217	0.298	0.388	0.018	0.754	0.001
CV-learning	0.415	0.039	**0.642**	1.8×10^{-5}	**0.768**	0.001
品詞分布	0.312	0.128	0.451	0.005	0.595	0.019
ベクトルの集中度 [5]	0.362	0.075	0.297	0.099	0.333	0.225
単語依存手法 [5]	**0.650**	4.0×10^{-4}	—	—	—	—

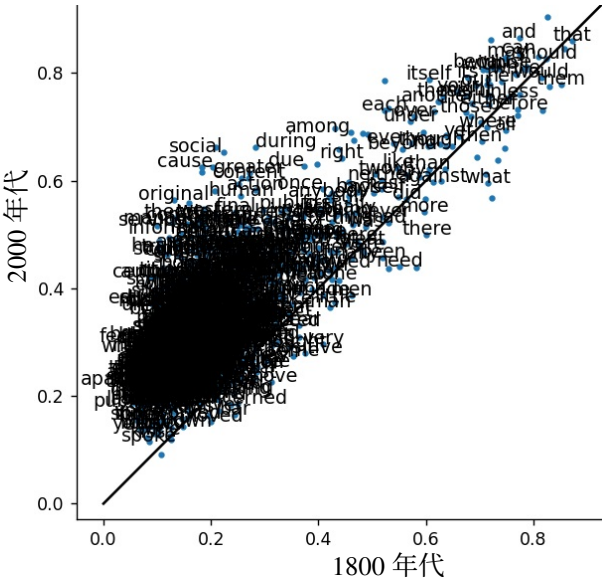

図 2　CCOHA の 1800 年代および 2000 年代の文書から推定された文法化度の散布図.

文書では機能語としての使用（because の省略）が他の品詞の約 10 倍に増えていたが，文法化というよりは同形異義語とみるべきであろう．後者については，2000 年代では固有名詞（の一部）としての使用が 4 倍に増えていた．小規模ではあるが同様な分析を他の単語についても行ったところ，使用品詞の種類が増え，一様な分布に近づくほど文法化度が高くなる傾向が見られた．すなわち，使用文脈が多様なほど文法化度が高く見積もられる．文法化の主要因とされる類推と再分析 [1] は，使用文脈が多様になるほど起こりやすいと考えられることから，この結果は文法化の一面を捉えている可能性もある．ただし，文脈の多様性を定量化するベクトルの集中度に基づいた手法の性能がそれほど高くないことから，CV-learning に基づいた手法は，文脈の多様性以外の要因も考慮している可能性が高い．

以上は一方向仮説を支持はしているが，暫定的な結果であることに注意が必要である．用いた手法と人間の判断との相関係数は 0.415〜0.768 であり非常に高いわけではない．また，CCOHA では，1800 年代の文書の方がレジスタの種類が少ないという影響もあるかもしれない（「映画 TV」レジスタは含まれない）．実際，上述の通り，必ずしも文法化とはいえない現象に対しても高い文法化度を与える場合もある．仮に，手法やデータが完全だとしても，1800 年代と 2000 年代という短い期間の比較でしかない．期間が長くなれば，逆方向の変化も観測されるかもしれない．したがって，図 2 は，定量化誤差を含む条件付きの結果であると認識することが重要である．それを踏まえた上でも，図 2 の極端な非対称性は，何らかの要因が存在することを示唆する．もし手法の限界などで，一定の確率で文法化度の定量化に失敗するのであれば，右下にも同様に点が分布するはずである．なぜこのような非対称性が生じるのか，言語学的にも計算科学的にも興味深い．その一つの要因として，文法化の一方向性がかかわっている可能性は十分にある．

6　おわりに

本稿では，汎用性の高い文法化度定量化手法の提案を行った．間隔分布に基づく手法，PU-learning に基づく手法，Cross-Validation を応用した手法の三種類を提案した．このうち三つ目の手法は，英語の動詞派生前置に特化した従来手法に比べ大幅に汎用性を向上させたにもかかわらず，人手の分析に基づいた文法化度と中程度から高い相関を示すことが確かめられた．更に，この手法を用いて，文法化の一方向仮説の吟味を行った．5 で述べたとおり条件付きの暫定的な結果ではあるが，一方向仮説を支持した．大規模な歴史コーパス中の 1000 種類以上の単語を同時に対象にして一方向仮説の検証を行ったのは，恐らく本研究が世界で初めてであり，その意義は大きい．今後は，手法および使用データを拡張することで検証の信頼性を高めることが望まれる．

参考文献

[1]Paul J. Hopper and Elizabeth Closs Traugott. **Grammaticalization**. Cambridge University Press, New York, second edition, 2003.

[2]Bernd Kortmann and Ekkehard König. Categorial reanalysis: The case of deverbal prepositions. **Linguistics**, Vol. 30, pp. 671–698, 1992.

[3]Teruhiko Fukaya. **The Emergence of *-ing* Prepositions in English: A Corpus-Based Study**, pp. 285–300. Taishukan, Tokyo, 1997.

[4]Tomoaki Hayashi. Prepositionalities of deverbal prepositions: Differences in degree of grammaticalization. **Papers in Linguistic Science**, Vol. 21, pp. 129–151, 2015.

[5]Ryo Nagata, Yoshifumi Kawasaki, Naoki Otani, and Hiroya Takamura. A computational approach to quantifying grammaticization of English deverbal prepositions. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 211–220, Torino, Italia, May 2024. ELRA and ICCL.

[6]Joan Bybee. From usage to grammar: The mind's response to repetition. **Language**, Vol. 82, pp. 711–733, 12 2006.

[7]Bernd Heine and Tania Kuteva. **World Lexicon of Grammaticalization**. Cambridge University Press, Cambridge, 2002.

[8]Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. In **Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, pp. 213–220, 2008.

[9]Paul Hopper. **On some Principles of Grammaticalization**, pp. 17–35. John Benjamins Publishing Company, 1991.

[10]Antoinne Meillet. L'évolution des formes grammaticales. **Scientia**, Vol. 6, No. 12, pp. 130–148, 1912.

[11]Christian Lehmann. **Thoughts on grammaticalization**. Lincom Europa, Munich, 1995.

[12]Hideo Teramura. **Syntax and semantics of noun modification – part 4**, pp. 1–34. Kuroshio, Tokyo, 1992.

[13]Martin Haspelmath. Why is grammaticalization irreversible? **Linguistics**, Vol. 37, pp. 1043–1068, 01 1999.

[14]Lyle Campbell. What's wrong with grammaticalization? **Language Sciences**, Vol. 23, pp. 113–161, 2000.

[15]Brian Joseph. Is there such a thing as "grammaticalization?". **Language Sciences - LANG SCI**, Vol. 23, pp. 163–186, 2000.

[16]Reem Alatrash, Dominik Schlechtweg, Jonas Kuhn, and Sabine Schulte im Walde. CCOHA: Clean corpus of historical American English. In **Proc. of the 12th Language Resources and Evaluation Conference**, pp. 6958–6966, 2020.

[17]Eduardo G. Altmann, Janet B. Pierrehumbert, and Adilson E. Motter. Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. **PLOS ONE**, Vol. 4, pp. 1–7, 11 2009.

[18]Kikuo Maekawa, Makoto Yamazaki Makoto, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. Balanced corpus of contemporary written japanese. **Language Resources and Evaluation**, Vol. 48, pp. 345—-371, 2014.

[19]Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In **Advances in Neural Information Processing Systems 26**, pp. 3111–3119, 2013.

[20]Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

A　使用したコーパスの詳細

英語コーパスとして CCOHA の 1800 年代と 2000 年代の文書を利用した．日本語については，BCCWJ [18] を利用した．トークン分割と品詞タグ付けは，英語については SpaCy[4]，日本語については Mecab[5] を用いた．更に，英語についてはトークン分割後，全て小文字に変換した．CCOHA に次のような前処理を行った．ノイズと思われる文書は除外した．具体的には，「@@年.txt」（例：@@1525.txt）のように年とファイル名と思われる文字列を含む文書は分析対象外とした．また，文書中のタグ（<P></P> など）は除去した．更に，伏字が含まれている文（CCOHA では，著作権の制限により，一定の割合で文章の一部が伏字になっている）も除外した．

B　実験条件の詳細

分散表現は Gensim の word2vec[6] を利用して得た．ハイパーパラメータの設定は文献 [5] を参考にし，200 次元，窓幅 10，エポック数 100，その他はデフォルトの値を用いた．また，各次元，平均 0，分散 1 となる標準化とノルム 1 となる正規化をこの順で行った．PU-learning には，pulearn package[7] の WeightedElkanotoPuClassifier を用いた．また，分類器は sklearn の LogisticRegression[8] を用いた．Weibull 分布のパラメータの推定には Fit_Weibull_2P[9] を用いた．いずれもデフォルトのハイパパラメータで用いた．

分類器に基づいた手法で必要となる，初期の機能語および内容語のリストは次の通り得た．品詞解析の結果，次の品詞のいずれかで 95%以上使われる語を初期の機能語／内容語と見なした[10]．英語：機能語：CC, DT, IN, MD, PDT, PRP, PRP$, RP, TO, WDT, WP, WP, WRB；内容語：JJ, JJR, JJS, NN, NNS, NP, NPS, RB, RBR, RBS, VB, VBD, VBG, VBN, VBP, VBZ；日本語：機能語：代名詞，助動詞，助詞，感動詞，接続詞；内容語：副詞，動詞，名詞，形容詞，形状詞．なお，頻度 100 以上の単語を対象にして分類器の訓練行った．この閾値は，付録 C に記載した対象単語が全て含まれ，かつ，きりのよい頻度ということで選択した．

従来手法の実装は文献 [5] に従い行った．意味の集中度に基づいた手法は，BERT[20] を利用するため，一部の対象名詞は複数のサブワードに分割された．そのような対象単語は評価の対象外とした．具体的には，表 1 に掲載した，日本語接続詞化名詞に対するこの手法の順位相関係数は 32 種類についてのものとなっている．

C　文法化度の正解データ

動詞派生前置詞については，文献 [5] と同様に，Hayashi [4] のデータを使用した．文献 [5] で対象にしている動詞派生前置詞のうち一語からなるもの 25 種類を対象とした．表 2 に，対象動詞派生前置詞と文法化度を示す．

4） https://spacy.io/
5） https://taku910.github.io/mecab/，unidic を利用．
6） Gensim 4.3.1: https://radimrehurek.com/gensim/models/word2vec.html
7） https://github.com/pulearn/pulearn
8） https://scikit-learn.org/1.5/modules/linear_model.html
9） https://reliability.readthedocs.io/en/latest/API/Fitters/Fit_Weibull_2P.html
10） PU-learning に基づく手法でも，このように自動的に作成した機能語のリストを使用した．

表 2　Hayashi[4] の動詞派生前置詞の文法化度．括弧内は文法化度のスコア（値が大きいほど文法化が進んでいる）

対象単語（文法化度スコア）
facing (1.5), saving (2.1), granted (2.3), wanting (2.3), covering (2.5), failing (2.7), considering (2.8), notwithstanding (3.1), confronting (3.2), granting (3.3), save (3.3), touching (3.3), lacking (3.4), except (3.6), concerning (3.7), pending (3.7), excluding (3.8), given (3.9), including (5.1), preceding (5.3), regarding (5.8), starting (6.0), following (6.4), during (7.4), past (7.8)

表 3　Teramura[12] の接続詞化している名詞の文法化度．

対象単語（文法化度順位）
きり (1.5)，なり (1.5)，以来 (3.0)，かぎり (5.0)，まで (6.0)，ゆえ (7.0)，だけ (8.0)，ほど (9.0)，たび (10.0)，以後 (11.0)，末 (12.0)，せい (13.0)，あまり (14.0)，くらい (15.0)，うえ (16.0)，あげく (17.0)，くせ (18.5)，わり (18.5)，毎度 (20.0)，以前 (21.0)，よう (22.0)，わけ (23.0)，ため (24.0)，とおり (25.0)，まま (26.0)，場合 (27.0)，限度 (28.0)，程度 (29.0)，結果 (30.0)，とき (32.0)，あいだ (32.0)，ころ (32.0)，原因 (34.0)，ところ (35.0)，様子 (36.0)，目的 (37.5)，理由 (37.5)

Teramura [12] の日本語名詞の文法化度も利用した．同文献では，名詞が接続詞または助動詞の性質を満たすかどうかのテスト結果が示されている．第一，第三，第四著者が協力し，テスト結果を解釈し，文法化度に変換した．具体的な手順は次の通りである：(i) 機能語性（接続詞または助動詞）に関するテストにパスする数を計数する（結果は○，△，×で与えらており，それぞれ 1，0.5，0 として計数）；(ii) 名詞性に関する条件についても同様に計数する（ただし，○，△，×をそれぞれ，0，0.5，1 とする）；(iii) 両者の合計で対象名詞に順位をつける（合計が多いほど文法化度が高いとする）；(iv) 同順位のものについては，テストにパスする名詞が少ないほど重要であるとし，この重要度により解消する；(v) それでも同順位が解消されない場合は，そのまま同順位を認める．その結果，接続詞化した可能性がある名詞 37 種類[11] と助動詞化した可能性のある名詞 15 種類について文法化度の順位の情報を得た．表 3 と表 4 に，それぞれの結果を示す．

表 4　Teramura[12] の助動詞化している名詞の文法化度．

対象単語（文法化度順位）
らしい (1.5), だろう (1.5), そう (3.0), 甲斐 (4.0), 由 (5.0), よう (6.0), ほう (7.0), はず (8.0), つもり (9.5), 気 (9.5), わけ (11.0), 様子 (12.0), 意図 (13.0), 予定 (14.5), 事情 (14.5)

11） 「から」の結果が二通りあり，どちらを採択すべきか判断がつかなかったため除外した．その結果 37 種となった．