ペア類似度に基づく統計的な意味変化検出法

相田太一 [1]　持橋大地 [2,3]　高村大也 [4]　小木曽智信 [3]　小町守 [1]
[1] 一橋大学　[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
taichia@scl.sds.hit-u.ac.jp　daichi@ism.ac.jp
takamura.hiroya@aist.go.jp　togiso@ninjal.ac.jp　mamoru.komachi@r.hit-u.ac.jp

概要

意味変化検出は、単語が意味変化したかを判定する**分類**と、変化の度合いに基づいて単語を並べる**ランキング**の 2 つのサブタスクからなる。従来研究の多くはランキングに注目してきたが、実用上は意味変化の有無を判定する**分類**も重要である。本研究では、単語の用例間類似度の一貫性を直接モデル化する統計的な手法を提案する。SemEval-2020 Task 1 および WUGS データセットでの実験により、本手法が既存の埋め込みベースの手法を上回り、多言語環境でも頑健に意味変化を検出できることを示した。

1　はじめに

単語の意味は、時間の経過や使用されるドメインの違いに応じて変化することがある。このような意味変化を検出することは、言語学および辞書編纂にとって重要であるだけでなく、文化的・社会的動向を分析する上でも不可欠である [1, 2]。さらに、こうした人文学的応用にとどまらず、近年の研究では、情報検索 [3] やマスク化言語モデルの効率的な更新 [4] など、さまざまな目的において意味変化検出の重要性が示されている。

意味変化検出（Semantic Change Detection; SCD）タスクは、意味が変化した単語を自動的に特定することを目的とする。SemEval-2020 Task 1 [5] や WUGS [6] フレームワークといった近年の共有タスクにより、本タスクに対するベンチマークデータセットおよび評価プロトコルが整備されてきた。SCD には 2 つのサブタスクが存在する。一つは、対象語が意味変化を起こしたか否かを予測する分類タスクであり、もう一つは、意味変化の程度に基づいて対象語を並べ替えるランキングタスクである [5]。本問題は教師なし学習の性質を持つため、従来研究の多くはランキングタスクに焦点を当て、2 つの時期またはドメインにおける単語埋め込み間の類似度スコアに基づいて評価してきた [7, 8, 9, 10]。しかし、ランキングに基づく評価には、生の類似度スコアの解釈が困難である点や、順位リストのどの部分が信頼できるのかが不明確である点など、本質的な制約が存在する。これらの問題点から、分類タスクにより強く焦点を当てる必要性が示唆される。

Method	EN	DE	LA	SV
SGNS [11]	73.0	54.2	45.0	61.3
SGNS [12]	62.2	75.0	**70.0**	67.7
BERT [13]	70.3	75.0	55.0	74.2
Pólya (**ours**)	**76.1**	**80.0**	N/A	**88.6**

表 1: SemEval-2020 Task 1 における正解率（%）[5]。提案手法は、現在主流である単語埋め込みベースの手法とは異なり、用例間の類似度の一貫性を直接モデル化することで、高い分類性能を達成する。

本研究では、用例レベルの類似度スコア集合における通時的／共時的な一貫性を評価することにより、語が意味変化を経験したか否かを統計的に判定する新しい手法を提案する。本手法は、SemEval-2020 Task 1 や WUGS といった既存のベンチマークデータセットにおいて、人手アノテーションによって[1)]用例レベルの類似度スコアが与えられているという前提の下で検出を行う。これらの類似度スコアをどのように予測するかについては扱わず、用例レベルの類似度情報が与えられた状況において、意味変化をいかに統計的に判定するかに焦点を当てる。この設計方針により、特定の類似度推定手法に依存することなく、検出部分を独立に検討することが可能となる。本研究では類似度スコアの分布を Pólya 分布 によってモデル化し、2 つの時期におけるスコアが同一の分布から生成された可能性が高いか（意味変化がないことを示す）、あるいは異なる分布から生成された可能性が高いか（意味変化

1）4.2 節で述べるように、提案手法は人手に限らず、LLM などによる自動的なアノテーションの場合でも適用できる。

の存在を示す）を検定する枠組みを提案する。

2 関連研究

SemEval-2020 Task 1 [5] における分類およびランキングの両評価は、WUGS 形式のアノテーショングラフ [6] に基づいて構築されている。これらのグラフは、各対象語について 2 つの異なる時期またはドメインから一定数の用例を収集することにより作成される。その後、人手アノテータが用例対に対して 4 段階の意味的類似度尺度で評価を行い [14, 15]、得られた重み付きグラフを処理することで、意味変化の有無を表す 2 値ラベルと、その程度を表す連続値スコアの双方が算出される。これらのデータは、SCD の評価における標準的な基盤となっている。

SCD の手法は一般に、静的な単語埋め込み [16, 17]、あるいは動的な単語埋め込み [7, 8, 10] の比較に基づいている。これらの進展にもかかわらず、既存手法の多くはランキングタスクに焦点を当てている。その結果、意味変化の分類に関する根本的な課題はいまだに解決されていない。

3 提案手法

本研究では、前節で述べた WUGS 形式のデータセットを利用する。ある語（例：*plane*）に対して、時期 A のテキストから N 個の用例（例：“*in the horizontal plane*”）を収集し、時期 B からは M 個の用例（例：“*the plane was in flight*”）を収集する。

これら $(N+M)$ 個の用例に対するすべての用例対を対象として、$(N+M)\times(N+M)$ 次元の行列 $\mathbf{X}$ を構築する（図 1）。行列の要素 x_{ij} の一部には、用例 i と用例 j の対の意味的類似度として、1 から 4 のスコアが付与されている。

本研究では、この行列に対して次の 2 つのケースを考える。

意味変化がない場合　この場合、行列 $\mathbf{X}$ に含まれる要素 x_{ij} の間に区別は存在せず、すべてのアノテーションスコアは、スコア 1 から 4 に対する同一の確率分布 $\mathbf{p}_0=(p_1,p_2,p_3,p_4)$ から生成されたものと仮定できる。すなわち、以下のように表せる。

$$x_{ij}\sim\mathbf{p}_0 \quad \text{i.i.d.} \quad (1\le i,j\le N+M) \tag{1}$$

意味変化がある場合　時期 A と時期 B の間に意味変化が存在する場合、同一時期内の用例対と、異なる時期間の用例対とでは、付与される類似度スコアが異なると考えられる。そのため、図 1に示す $\mathbf{X}$

$$\mathbf{X}=\begin{array}{c} \\ 1 \\ \\ N \\ N{+}1 \\ \\ N{+}M \end{array}\begin{array}{c} \begin{array}{cccccccc} 1 & & & N & N{+}1 & & & N{+}M \end{array} \\ \left(\begin{array}{cccc|cccc} 4 & & & & & 2 & & \\ & 4 & 3 & & & & 3 & 4 \\ & 2 & & & 2 & & & \\ 1 & & & & 2 & & & \\ \hline 1 & & & 4 & & 4 & 1 & \\ & & 2 & & & & & \\ 2 & & & & 2 & 3 & & \\ 3 & & & & & & & 4 \end{array}\right) \end{array}$$

図 1: 用例間の類似度から構成されるスコア行列 $\mathbf{X}$ の例。一部の用例対には、1 から 4 の類似度スコアが付与されている。

の各ブロックに対応して、要素 x_{ij} は 4 つの異なる分布 $\mathbf{p}_n$（$n=1,\cdots,4$）のいずれかから生成されると仮定する。

$$x_{ij}\sim\begin{cases}\mathbf{p}_1 & (1\le i,j\le N)\\ \mathbf{p}_2 & (1\le i\le N, N{+}1\le j\le N{+}M)\\ \mathbf{p}_3 & (N{+}1\le i\le N{+}M, 1\le j\le N)\\ \mathbf{p}_4 & (N{+}1\le i,j\le N{+}M)\end{cases} \tag{2}$$

生成モデルとしては、$\mathbf{p}=\mathbf{p}_0,\mathbf{p}_1,\cdots,\mathbf{p}_4$ に対する事前分布を定める必要があり、最も単純な選択としてディリクレ分布がある。

$$p(\mathbf{p})=\frac{\Gamma(\sum_k\alpha_k)}{\prod_k\Gamma(\alpha_k)}\prod_{k=1}^{K}p_k^{\alpha_k-1} \tag{3}$$

ここで、本研究では $K=4$ であり、$\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_K)$ はこの事前分布に対するハイパーパラメータである。本研究全体を通して、確率単体上の一様分布に対応する $\boldsymbol{\alpha}=(1,\cdots,1)$ を採用した。

意味変化が存在しない場合（$\theta=0$）と、存在する場合（$\theta=1$）を表す 2 値の潜在変数 θ を導入しよう。このとき、$\theta=0$ の下でのデータ $\mathbf{X}$ の尤度は、次のように与えられる。

$$\begin{aligned} p(\mathbf{X}|\theta{=}0)=p(\mathbf{X})&=\int p(\mathbf{X}|\mathbf{p})p(\mathbf{p})d\mathbf{p}\\ &=\int\prod_{i,j}\prod_{k=1}^{K}p_k^{\mathbb{I}(x_{ij}=k)}\cdot\frac{\Gamma(\sum_k\alpha_k)}{\prod_k\Gamma(\alpha_k)}\prod_{k=1}^{K}p_k^{\alpha_k-1}d\mathbf{p}\\ &=\frac{\Gamma(\sum_k\alpha_k)}{\Gamma(L+\sum_k\alpha_k)}\prod_{k=1}^{K}\frac{\Gamma(\alpha_k+n_k)}{\Gamma(\alpha_k)} \end{aligned} \tag{4}$$

ここで、$n_k=\sum_{ij}\mathbb{I}(x_{ij}=k)$ は、行列 $\mathbf{X}$ においてスコア k が付与された回数を表し、L はアノテーションが与えられている要素の総数である。

式 (4) は、Pólya 分布として知られている [18, 19]。意味変化が存在する場合には、式 (2) に従い、ブロッ

クごとの Pólya 分布の積として尤度が与えられる。

$$p(\mathbf{X}|\theta=1) = \prod_{n=1}^{4} p(\mathbf{X}_n) \tag{5}$$

ここで、各 $\mathbf{X}_n$（$n = 1,\ldots,4$）は式 (2) で定義される $\mathbf{X}$ の部分行列であり、n_k と L も同様に各 $\mathbf{X}_n$ 内で計算される。したがって、事前確率を $p(\theta = 0) = p(\theta = 1) = 1/2$ とすると、θ の事後確率は次のように求められる：

$$p(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta)p(\theta) \propto \begin{cases} p(\mathbf{X}|\theta=0) \\ p(\mathbf{X}|\theta=1) \end{cases} \tag{6}$$

ここで、$p(\mathbf{X} \mid \theta = 0)$ および $p(\mathbf{X} \mid \theta = 1)$ は、それぞれ式 (4) と式 (5) により与えられる。直感的には、この確率は観測されたスコア行列が一様であるか否かを測る指標となっている。

4 実験

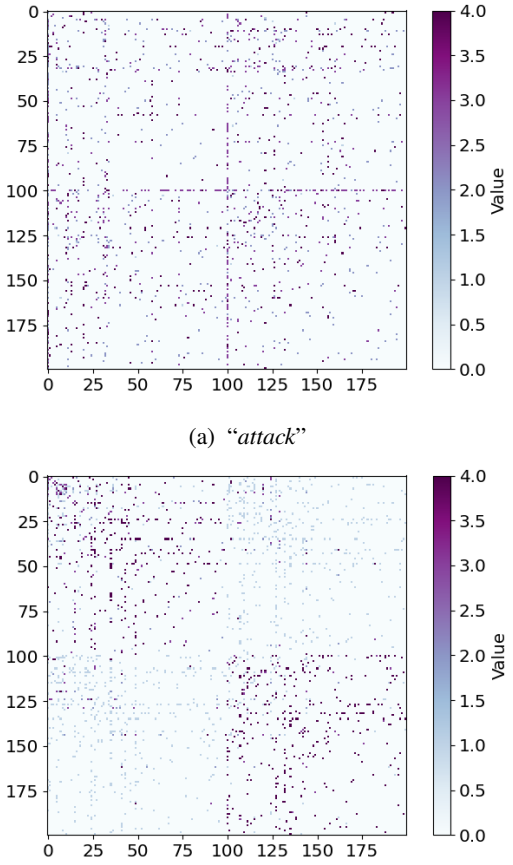
4.1 実験設定

データセット　提案手法の評価には、二つのベンチマークである SemEval-2020 Task 1 と、WUGS データセットの一部を用いる。[2)]

評価　どちらのベンチマークにおいても、正解率を用いてすべての手法を評価する。各対象語は意味変化の有無がラベル付けされているため、予測結果と比較して評価を行った。

比較手法　SemEval-2020 Task 1 に対しては、共有タスクにおいて各言語で最も高い性能を示した 3 つのシステム [11, 12, 13] と比較を行う。これらのベースラインはいずれも、静的または動的な単語埋め込みに基づく手法である。一方、言語やドメインをまたいでシステム出力を直接比較することが困難な WUGS データセットに対しては、単純なベースラインとして各データセットにおける多数派クラスを予測する MostFreq を採用する。

提案手法　図 2 は、意味的に安定な語である *attack* と意味変化を起こした語である *plane* についてのスコア行列を示している[3)]。この図は、意味的に不変



(a) “*attack*”

(b) “*plane*”

図 2: “*attack*” と “*plane*” に対するスコア行列の可視化。*attack* は意味変化が生じていない語を、*plane* は意味変化が生じた語を表す。各行列において、0–99 番の用例は前期、100–199 番は後期の用例を表す。アノテーションは 1（無関係）から 4（同一）までの尺度で付与され、0 は未付与または未知のラベルを示す [6]。

な単語／変化した単語では、スコア行列がそれぞれ均一／不均一になる傾向があるという、本手法の重要な直観を示している。予測は、**意味変化がない場合**と**意味変化がある場合**での対数尤度を比較することで行う（$p(\mathbf{X}|\theta = 1) > p(\mathbf{X}|\theta = 0)$ のとき、**意味変化がある**と予測する）。*attack* に対しては、$\log p(\mathbf{X}|\theta = 0) = -6213.9$ および $\log p(\mathbf{X}|\theta = 1) = -6282.8$ が得られ、モデルは**意味変化がある**と予測した。一方、*plane* に対しては、$\log p(\mathbf{X}|\theta = 0) = -7427.2$ および $\log p(\mathbf{X}|\theta = 1) = -7095.6$ が得られ、モデルは**意味変化がない**と予測した。このいずれも、正解のラベルと等しい。これらの例は、提案手法の定性的な振る舞いを示すとともに、次に示す実験結果の直観的な

2) データセットは https://www.ims.uni-stuttgart.de/en/research/resources/experiment-data/wugs/ から入手可能である。本研究では、用例間類似度アノテーションを含むデータセットのみを選択した。これは、本手法が類似度スコア分布を前提とするためである。

3) 図 2a に見られる十字状のパターンは、WUGS のアノテーション手順に起因する。WUGS では、すべての用例対を網羅的に注釈するのではなく、情報量の高い用例対を優先して注釈する [6]。その結果、ある用例は多くの他用例と比較される一方で、多くの要素は未注釈（0）のままとなり、観察される十字状の構造が生じる。

Data	Language	Grouping 1	Grouping 2	Accuracy (%)	
				MostFreq	Pólya
DWUG EN	EN	1810–1860	1960–2010	54.3	**76.1**
DWUG EN Resampled	EN	1810–1860	1960–2010	60.0	**80.0**
DWUG DE	DE	1800–1899	1946–1990	60.0	**80.0**
DWUG DE Resampled	DE	1800–1899	1946–1990	60.0	**73.3**
DiscoWUG	DE	1800–1899	1946–1990	51.0	**72.0**
RefWUG	DE	1750–1800	1850–1900	**54.5**	45.5
DURel	DE	1750–1800	1850–1900	**63.6**	**63.6**
SURel	DE	general	domain specific	63.6	**68.2**
RuSemShift 1	RU	1682–1916	1918–1990	**77.5**	**77.5**
RuSemShift 2	RU	1918–1990	1991–2016	**62.3**	**62.3**
RuShiftEval 1	RU	1700–1916	1918–1990	**74.8**	**74.8**
RuShiftEval 2	RU	1918–1990	1992–2016	**70.3**	**70.3**
RuShiftEval 3	RU	1700–1916	1992–2016	**68.5**	**68.5**
DWUG ES	ES	1810–1906	1994–2020	55.5	**78.0**
DiaWUG	ES	Spanish variant 1	Spanish variant 2	65.6	**81.3**
DWUG SV	SV	1790–1830	1895–1903	68.1	**88.6**
DWUG SV Resampled	SV	1790–1830	1895–1903	60.0	**73.3**
ChiWUG	ZH	1954–1978	1979–2003	**57.5**	52.5
DWUG IT	IT	1948–1970	1990–2014	**69.2**	N/A
DWUG LA	LA	−200–0	0–2000	**55.5**	N/A
NorDiaChange 1	NO	1929–1965	1970–2013	67.5	**75.0**
NorDiaChange 2	NO	1980–1990	2012–2019	**77.5**	70.0

表 2: WUGS データセットにおける実験結果。各データセットについて、MostFreq ベースラインおよび提案する Pólya 分布に基づく手法の正解率を示す。提案手法は、言語やデータセットの種類に依らず高い性能を示しており、通時的・共時的な意味変化検出の双方において高い頑健性を有することが確認できる。

理解を与える。

4.2 結果と考察

結果 表 1 に、SemEval-2020 Task 1 における結果を示した。提案手法は単語埋め込みに一切依存しないにもかかわらず、4 言語中 3 言語において、すべての埋め込みベースの最先端手法を上回る精度を達成しており、用例間類似度の分布における時間的一貫性をモデル化することの有効性が示されている。SemEval の設定を超えた枠組みの頑健性を評価するため、表 2 には WUGS データセットにおける結果を示した。提案手法は、すべての言語において一貫して高い精度を達成しており、通時的・共時的設定の双方において安定した性能を示す。これらの結果は、提案フレームワークが SemEval データセットにとどまらず、多様な言語およびドメインに対して良好に一般化可能であることを示唆している。

考察 本実験では、人手による用例類似度アノテーションを使用しているが、近年の研究により、高品質なラベルがモデルによって比較的容易に得られることが示されている。DURel Annotation Tool [15] には、すでに XL-LEXEME [8] が統合されており、用例類似度のアノテーションを自動的に提供する仕組みが備わっている。さらに、Periti and Tahmasebi [9] は、動的な単語埋め込みや大規模言語モデルが、人手アノテーションに近い品質のラベルを生成可能であることを示している。これらの進展により、用例類似度ラベルを人手で収集するのではなく、自動的に予測する完全自動の設定においても、本研究で提案した統計的枠組みが将来的に適用可能であることが示唆される。

5 おわりに

本研究では、用例間の類似度スコア分布における一貫性をモデル化することによる、意味変化検出のための統計的枠組みを提案した。SemEval-2020 Task 1 および WUGS データセットに対する実験により、提案手法は言語や設定を問わず、既存手法と同等またはそれ以上の精度を達成することが確認された。これらの結果は、用例レベルの類似度スコアをモデル化することが、意味変化検出を効果的に実現する上で有効であることを示している。さらに、高品質な類似度ラベルを自動的に予測する近年の技術的進展と組み合わせることで、完全自動かつ解釈性の高い意味変化検出が期待できる。

謝辞

本研究は国立国語研究所の共同研究プロジェクト「開かれた共同構築環境による通時コーパスの拡張」および科学技術振興機構（JST）戦略的創造研究推進事業（さきがけ）の研究課題『意思決定のための自然言語処理による未来予測』の研究成果を報告したものである。

参考文献

[1] Elizabeth Closs Traugott and Richard B. Dasher. **Prior and current work on semantic change**, p. 51–104. Cambridge Studies in Linguistics. Cambridge University Press, 2001.

[2] Paul Cook and Suzanne Stevenson. Automatically identifying changes in the semantic orientation of words. In **Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)**, Valletta, Malta, May 2010. European Language Resources Association (ELRA).

[3] Andrey Kutuzov, Lilja Ovrelid, Terrence Szymanski, and Erik Velldal. Diachronic word embeddings and semantic shifts: a survey. In **Proceedings of the 27th International Conference on Computational Linguistics**, pp. 1384–1397, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.

[4] Zhaochen Su, Zecheng Tang, Xinyan Guan, Lijun Wu, Min Zhang, and Juntao Li. Improving temporal generalization of pre-trained language models with lexical semantic change. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 6380–6393, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

[5] Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. SemEval-2020 task 1: Unsupervised lexical semantic change detection. In **Proceedings of the Fourteenth Workshop on Semantic Evaluation**, pp. 1–23, Barcelona (online), December 2020. International Committee for Computational Linguistics.

[6] Dominik Schlechtweg, Nina Tahmasebi, Simon Hengchen, Haim Dubossarsky, and Barbara McGillivray. DWUG: A large resource of diachronic word usage graphs in four languages. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 7079–7091, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.

[7] Guy D. Rosin and Kira Radinsky. Temporal attention for language models. In **Findings of the Association for Computational Linguistics: NAACL 2022**, pp. 1498–1508, Seattle, United States, July 2022. Association for Computational Linguistics.

[8] Pierluigi Cassotti, Lucia Siciliani, Marco DeGemmis, Giovanni Semeraro, and Pierpaolo Basile. XL-LEXEME: WiC pretrained model for cross-lingual LEXical sEMantic changE. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 1577–1585, Toronto, Canada, July 2023. Association for Computational Linguistics.

[9] Francesco Periti and Nina Tahmasebi. A systematic comparison of contextualized word embeddings for lexical semantic change. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, **Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)**, pp. 4262–4282, Mexico City, Mexico, June 2024. Association for Computational Linguistics.

[10] Taichi Aida and Danushka Bollegala. A semantic distance metric learning approach for lexical semantic change detection. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Findings of the Association for Computational Linguistics: ACL 2024**, pp. 7570–7584, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[11] David Rother, Thomas Haider, and Steffen Eger. CMCE at SemEval-2020 task 1: Clustering on manifolds of contextualized embeddings to detect historical meaning shifts. In Aurelie Herbelot, Xiaodan Zhu, Alexis Palmer, Nathan Schneider, Jonathan May, and Ekaterina Shutova, editors, **Proceedings of the Fourteenth Workshop on Semantic Evaluation**, pp. 187–193, Barcelona (online), December 2020. International Committee for Computational Linguistics.

[12] Ondřej Pražák, Pavel Přibáň, Stephen Taylor, and Jakub Sido. UWB at SemEval-2020 task 1: Lexical semantic change detection. In Aurelie Herbelot, Xiaodan Zhu, Alexis Palmer, Nathan Schneider, Jonathan May, and Ekaterina Shutova, editors, **Proceedings of the Fourteenth Workshop on Semantic Evaluation**, pp. 246–254, Barcelona (online), December 2020. International Committee for Computational Linguistics.

[13] Ehsaneddin Asgari, Christoph Ringlstetter, and Hinrich Schütze. EmbLexChange at SemEval-2020 task 1: Unsupervised embedding-based detection of lexical semantic changes. In Aurelie Herbelot, Xiaodan Zhu, Alexis Palmer, Nathan Schneider, Jonathan May, and Ekaterina Shutova, editors, **Proceedings of the Fourteenth Workshop on Semantic Evaluation**, pp. 201–207, Barcelona (online), December 2020. International Committee for Computational Linguistics.

[14] Dominik Schlechtweg, Sabine Schulte im Walde, and Stefanie Eckmann. Diachronic usage relatedness (DURel): A framework for the annotation of lexical semantic change. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)**, pp. 169–174, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[15] Dominik Schlechtweg, Shafqat Mumtaz Virk, Pauline Sander, Emma Sköldberg, Lukas Theuer Linke, Tuo Zhang, Nina Tahmasebi, Jonas Kuhn, and Sabine Schulte Im Walde. The DURel annotation tool: Human and computational measurement of semantic proximity, sense clusters and semantic change. In Nikolaos Aletras and Orphee De Clercq, editors, **Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations**, pp. 137–149, St. Julians, Malta, March 2024. Association for Computational Linguistics.

[16] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. Statistically significant detection of linguistic change. In **WWW 2015**, pp. 625–635, 2015.

[17] Taichi Aida, Mamoru Komachi, Toshinobu Ogiso, Hiroya Takamura, and Daichi Mochihashi. A comprehensive analysis of PMI-based models for measuring semantic differences. In **Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation**, pp. 21–31, Shanghai, China, 11 2021. Association for Computational Lingustics.

[18] Thomas P. Minka. Estimating a Dirichlet distribution, 2000. https://tminka.github.io/papers/dirichlet/.

[19] Kevin P. Murphy. **Probabilistic Machine Learning: An Introduction**. MIT Press, 2022.