理論言語学の知見を応用した多言語クラスタリング

今井咲良 [1]　河原大輔 [1]　折田奈甫 [1]　小田博宗 [2]
[1] 早稲田大学理工学術院　[2] 東京大学大学院総合文化研究科
sakura_imai@toki.waseda.jp, {dkw, orita}@waseda.jp
hiromuneoda@g.ecc.u-tokyo.ac.jp

概要

多言語クラスタリング研究では、言語モデルの埋め込みを用いる手法が主流であり、語族がベースライン手法として扱われている。しかし、語族による分類は粒度が粗く、言語の性質を十分に考慮しているとは言い難い。そのため、より言語の性質に着目した分類を検討する必要がある。本研究では、固有表現認識 (NER) のための多言語クラスタリングを目的として、理論言語学の知見に基づく言語分類方法を検討し、名詞句の形態・統語的特徴による分類と主要部の位置による分類を選定した。これらの分類と埋め込みによるクラスタリング手法を NER において比較した結果、どちらの分類も埋め込みによるクラスタリング手法の精度を上回った。

1　はじめに

低リソース言語への言語間転移を目的とした言語クラスタリングは、固有表現認識 (NER) をはじめとした自然言語処理の様々な分野で利用されている。近年は言語モデルの埋め込みを用いたクラスタリングが主流となっており、言語学の知見をもとにしたクラスタリングは、語族、すなわち系統分類がベースラインとして用いられている程度である [1, 2]。しかし、言語学では系統分類以外に様々な言語の分類方法が提案されており、言語学の知見を生かした言語クラスタリングには改善の余地が多分にある。

本研究は、これまでに検証されてこなかった言語の形態・統語的性質による分類に着目し、このクラスタリングの有用性を調査する。言語の形態・統語的性質による分類とは、言語の系統関係、すなわち語族ではなく、語順や定冠詞の有無といった言語の特徴に基づく分類であり、理論言語学の知見が用いられる。これらの知見を応用すれば、埋め込みでは捉えられない言語的性質が反映されたクラスタリングが可能になると予測する。

形態・統語的性質に基づくクラスタリングの有用性を評価するため、すでに埋め込みによるクラスタリングと系統分類によるクラスタリングが検証されている [2] NER タスクを評価対象として用いる。また、本稿で報告する実験では印欧語族の言語データを用いる。これは、NER タスクの訓練・評価データが存在する言語が多く、理論言語学研究においても最も知見の蓄積がある語族であるためである。

本稿で新たに提案する分類には 2 種類あり、どちらも理論言語学、特に統語理論におけるパラメータを利用している。

第一の分類は、名詞句の形態・統語的性質に関するパラメータに基づく分類である。NER は入力文中の名詞句中で固有表現が表れる位置を推測するタスクであるため、名詞句の形態・統語パラメータに類似性をもつ言語同士をクラスタリングする方が精度が上がると予測する。本研究では、Ceolin ら [3] が提案する、名詞句に関する様々な形態・統語パラメータをもとに作られた言語樹を用いてこのクラスタリングを行う。第二の分類は、句構造内で主要部が現れる位置を示す主要部パラメータによる分類である。各言語の句構造において、例えば場所を表す語（前置詞や後置詞）が似た位置に現れる言語で分類する方が NER の精度が上がると予測する。

本研究では、上記 2 種類のクラスタリング手法を、言語クラスタリングで主流の埋め込みベースの手法および系統分類と比較し、理論言語学的分類の有用性を検証する。

2　関連研究

2.1　言語間転移学習

言語間転移学習 [4] とは、学習データが特定の言語しか存在しない状況下で言語モデルの学習を行い、それを転移先のターゲット言語でのタスクに転用する手法である。言語間転移学習を効率的に

行うための手法として、様々な研究が行われている [5, 6, 7, 8, 9, 10]。しかし、これらの研究において用いられている事前知識は、粒度の粗い系統分類や表層的な文字種などに限られており、言語学の知見を生かす余地が多分にある。

2.2 多言語クラスタリング

多言語間での転移学習をより効果的に実施するための一手法として、言語同士をクラスタリングして言語モデルを学習する多言語クラスタリングが挙げられる。多言語クラスタリングは、特にニューラル機械翻訳の分野において活用されており、翻訳モデルの埋め込みを用いて言語間の距離を測定しクラスタリングを行う手法が提案されている [1]。

Shaffer [2] は、NER において埋め込みと系統分類によるクラスタリングを比較し、埋め込みベースのクラスタリングの有効性を確認している。しかし、多言語クラスタリングにおいて、系統分類以外の言語学的知識を用いた実験は未だ行われておらず、さらなる検証が必要である。

3 理論言語学のパラメータを用いた言語クラスタリング

3.1 対象のタスクおよび言語の選定

本研究では、埋め込みおよび系統分類に基づくクラスタリングによって NER の精度向上を図った先行研究 [2] と比較するため、対象のタスクとして NER を選定する。また、対象言語には、印欧語族に属する 25 言語を用いた。これは、NER のデータが存在する言語が多く、理論言語学でも知見の蓄積が多いためである。

本研究で用いる言語の一覧を表 1 に示す。先行研究 [2] では、データが存在するにもかかわらず、ケルト語派等の下位分類は使用されなかった。本研究では、下位分類の網羅性を可能な限り考慮して言語を選択する。なお、本稿では、各言語を ISO639-1[1] に準拠した言語コードで表す。

3.2 名詞句パラメータクラスタリング

NER は、入力文をトークンとしてモデルに与え、名詞句から構成される固有表現とその属性を推測するトークン分類のタスクである。そのため、名詞句

表 1 本研究で使用した言語と先行研究との比較

下位分類	本研究の対象言語	[2] の対象言語
ロマンス諸語	ro, fr, es, pt, it, scn	fr, es, it
ゲルマン語派	af, nl, de, is, en, da, no, fo	de, en, da
ヘレニック語派	el	–
スラヴ語派	bg, pl, ru, sl, hr	ru
インド・イラン語派	ps, mr, hi	hi
ケルト語派	cy, ga	–

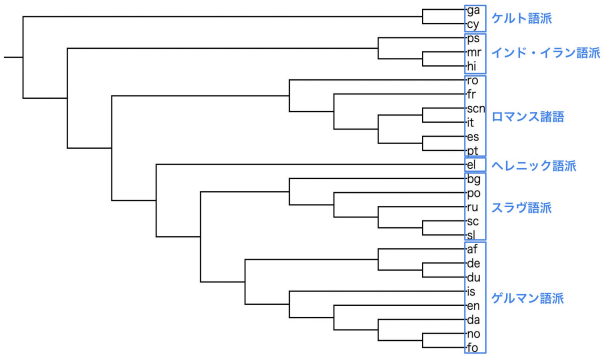

図 1 Ceolin ら [3] による言語樹

の構造が類似した言語同士でクラスタリングをすると学習の効果が高いと考え、定冠詞の有無など名詞句における形態・統語的性質を通言語的に捉えたパラメータに着目する。本研究では、Ceolin ら [3] が提案する、名詞句に関する 94 種類の形態・統語パラメータ[2]（関係節への定冠詞付加の有無や接置詞を用いた属格の有無など）をもとに作られた言語樹をクラスタリングに用いて、名詞句の形態・統語的性質が似た言語で分類する。本研究で用いる Ceolin らの言語樹を図 1 に示す。

言語樹を用いたクラスタリングでは、言語樹上で距離が近い下位分類同士を結合し一つのクラスタを形成するという操作を繰り返し行う。例として、クラスタ数 3 において図 1 を用いてクラスタリングを行った結果を表 2 に示す。クラスタ数は 4.2 節にて述べるエルボー法の結果により決定する。

3.3 主要部パラメータクラスタリング

言語モデルが NER のタスクを解く際は、固有表現が含まれる名詞句全体だけでなく、周辺のトークンの羅列（語順）も暗黙的に学習すると仮定する。固有表現である名詞句は、場所を表す接置詞句などの一部を構成する場合や、目的語として動詞句の一部を構成する場合があるため、主要部の位置が同じ言語同士で分類する方が学習効果が高くなると予測

1） http://www.infoterm.info/standardization/iso_639_1_2002.php

2） https://github.com/AndreaCeolin/Boundaries/blob/main/TableS2.pdf

表 2　図 1 の言語樹によるクラスタリング (クラスタ数 3)

#	下位分類
1	ゲルマン語派, スラヴ語派, ヘレニック語派, ロマンス諸語
2	インド・イラン語派
3	ケルト語派

表 3　主要部パラメータに基づくクラスタリング

主要部パラメータ	下位分類
主に主要部先行	ロマンス諸語, スラヴ語派, ゲルマン語派, ヘレニック語派, ケルト語派
主に主要部後行	インド・イラン語派

する。この仮説をもとに、各言語の句構造においてどの位置に句の主要部が置かれるかを表す主要部パラメータ [11] を用いて言語をクラスタリングする。例えば接置詞句 (PP) では、主要部先行であれば主要部である接置詞 (P) が名詞句 (NP) に先行し、主要部後行であればその逆になる (図 2)。こうした主要部パラメータに基づく分類を表 3 に示す。

4　固有表現認識 (NER) 評価実験

4.1　実験設定

NER による評価実験では、3 節の 2 種類のクラスタリング手法を用いて実験を行う。本研究における全言語は系統分類 (印欧語族) にあたるため、系統分類で Fine-tuning する場合と単言語で Fine-tuning する場合とも比較する。

初めにクラスタ内の全ての言語の訓練セットを結合し、NER で言語モデルを Fine-tuning する。その後、クラスタ内の各言語の評価セットを用いて評価を行い、スコアを取得する。本研究では、Wikiann [4] データセット[3] と XLM-RoBERTa-base[4] [12] を用いる。各クラスタにおける NER 評価実験を 3 回ずつ行い、平均 F1 スコアと標準偏差を算出する。また、すべての実験に関して、バッチサイズを 32、入力の最大長を 512、学習率を 5e-05 と設定し、3 epoch の Fine-tuning を行う。

4.2　埋め込みベースクラスタリング

本研究では、比較対象として Shaffer [2] による埋め込みベースのクラスタリング手法を採用する。埋め込みベースクラスタリングの概略を図 3 に示す。

初めに、Wikiann 訓練セットを用いて、XLM-RoBERTa-base を言語識別タスクで Fine-tuning する。

3)　https://huggingface.co/datasets/wikiann
4)　https://huggingface.co/xlm-roberta-base

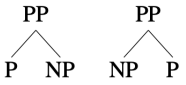

図 2　接置詞句 (PP) の主要部先行 (左) と主要部後行（右）

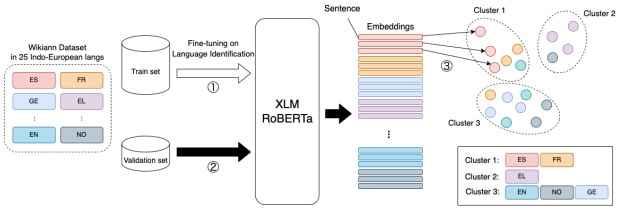

図 3　埋め込みベースクラスタリングの概略

言語識別タスクとは、言語モデルに与えた入力がどの言語で記述されているかを推測するタスクである。本研究では、表 1 の全 25 言語を用いて行う。

次に、言語識別タスクで Fine-tuning した XLM-RoBERTa に Wikiann 検証セットの各文を入力として与え、[CLS] トークンから埋め込みを取得する。そして、得られた埋め込みをもとに、凝集クラスタリングにより再帰的にクラスタリングを行う。そして、各入力に対してクラスタをラベリングし、各言語においてどのクラスタに一番多く割り当てられたかによりその言語が属するクラスタを決定する。

Wikiann データセットの各言語の検証セットから、1,000 サンプルおよび 10,000 サンプルを用いてクラスタリングを行った結果を表 4 に示す。

本研究では、3 節にて述べた名詞句のパラメータによるクラスタリング手法と比較する際は、エルボー法 [13] により決定されたクラスタ数 3 を用いる (その他のクラスタ数 {2, 4, 5} での比較実験の結果は付録参照)。また、主要部パラメータを用いたクラスタリング手法と比較する際はクラスタ数を 2 に揃えてクラスタを生成する。

4.3　結果と議論

表 5 に名詞句パラメータと単言語、埋め込み、系統分類を用いた NER 評価結果を示す。表 6 に主要部パラメータとの比較結果を示す。表では、言語ごとに最高スコアを太字にしている。

はじめに、本研究で用いた形態・統語パラメータによるクラスタリングと埋め込みベースのクラスタリングによる NER 評価結果を比較する。表 4 から、埋め込みによるクラスタリング手法は、対象言語の選定や埋め込みを得るためのサンプル数によって結果が大きく異なり、一様に最も有効な手法とは言い難いことがわかった。名詞句パラメータに基づくク

表 4 埋め込みベースクラスタリング結果 (クラスタ数 3)

#	言語	
	1,000 サンプル	10,000 サンプル
1	cy, ga, ps, mr, hi, ro, fr, bg, pl, ru, sl, hr, af, nl, de, is, en, da, no, fo	ga, ro, fr, es, pt, it, scn, pl, sl, hr, de, en
2	es, pt, it, scn	mr, hi, ru, af, nl, is, da, no, fo
3	el	cy, ps, el, bg

表 5 名詞句パラメータクラスタリング評価結果 (F1)

言語	#train	単言語	3 クラスタ			
			名詞句	#1000	#10000	系統分類
cy	10,000	0.911	0.916	0.917	0.924	**0.930**
ga	1,000	0.765	**0.857**	0.841	0.844	0.849
ps	100	0.000	**0.559**	0.547	0.533	0.522
mr	5,000	0.855	0.870	0.879	**0.886**	0.883
hi	5,000	0.861	0.869	0.892	**0.899**	0.895
ro	20,000	0.926	**0.943**	0.940	0.940	0.942
fr	20,000	0.890	0.910	0.907	0.905	**0.910**
es	20,000	0.892	0.915	0.913	0.905	**0.916**
pt	20,000	0.902	0.921	0.918	0.914	**0.922**
it	20,000	0.908	**0.922**	0.919	0.915	0.921
scn	100	0.012	0.801	0.756	0.771	**0.810**
el	20,000	0.901	**0.912**	0.904	0.901	0.910
bg	20,000	0.925	0.933	0.926	0.933	**0.934**
pl	20,000	0.899	0.913	0.911	0.912	**0.914**
ru	20,000	0.885	0.900	0.893	**0.900**	0.899
sl	15,000	0.930	**0.939**	0.937	0.939	0.939
hr	20,000	0.909	0.921	0.919	**0.921**	0.920
af	5,000	0.891	0.912	0.915	0.907	**0.918**
nl	20,000	0.906	**0.926**	0.917	0.922	0.925
de	20,000	0.875	0.886	0.881	0.883	**0.887**
is	1,000	0.740	0.875	0.868	0.874	**0.883**
en	20,000	0.823	0.841	**0.842**	0.840	0.840
da	20,000	0.917	**0.931**	0.926	0.930	0.930
no	20,000	0.920	0.933	0.931	0.932	**0.935**
fo	100	0.000	0.866	0.863	0.874	**0.877**

ラスタを用いて NER 評価を行った結果（表 5）、7 割の言語において埋め込みによるクラスタリングを上回った。また、主要部パラメータに基づくクラスタを用いた結果（表 6）では、8 割の言語において埋め込みベースの結果を上回った。これらの結果は、理論言語学の通言語的パラメータが自然言語処理タスクにおいて有用であることを示している。

つぎに、系統分類と形態・統語パラメータによるクラスタを用いた結果を比較する。本研究の対象言語はすべて印欧語族に属することから、全言語を用いた結果は系統分類によるクラスタリングを行った結果と等しい。したがって、系統分類は他の分類に比べ訓練サンプル数が圧倒的に多い。本研究が提案する形態・統語パラメータを用いた分類はデータ量

表 6 主要部パラメータクラスタリング評価結果 (F1)

言語	#train	単言語	2 クラスタ			
			主要部	#1000	#10000	系統分類
cy	10,000	0.911	**0.931**	0.922	0.919	0.930
ga	1,000	0.765	**0.854**	0.841	0.844	0.849
ps	100	0.000	**0.559**	0.553	0.550	0.522
mr	5,000	0.855	0.870	**0.887**	0.883	0.883
hi	5,000	0.861	0.869	**0.895**	0.894	0.895
ro	20,000	0.926	**0.944**	0.940	0.942	0.942
fr	20,000	0.890	**0.911**	0.907	0.906	0.910
es	20,000	0.892	**0.917**	0.913	0.905	0.916
pt	20,000	0.902	0.920	0.918	0.914	**0.922**
it	20,000	0.908	0.920	0.919	0.915	**0.921**
scn	100	0.012	0.770	0.756	0.771	**0.810**
el	20,000	0.901	**0.915**	0.909	0.913	0.910
bg	20,000	0.925	**0.936**	0.932	0.933	0.934
pl	20,000	0.899	0.914	0.911	0.913	**0.914**
ru	20,000	0.885	0.899	0.898	**0.900**	0.899
sl	15,000	0.930	**0.940**	0.937	0.939	0.939
hr	20,000	0.909	**0.923**	0.919	0.921	0.920
af	5,000	0.891	0.917	0.913	0.917	**0.918**
nl	20,000	0.906	**0.926**	0.919	0.922	0.925
de	20,000	0.875	**0.891**	0.881	0.886	0.887
is	1,000	0.740	0.880	0.873	0.876	**0.883**
en	20,000	0.823	**0.844**	0.842	0.840	0.840
da	20,000	0.917	**0.934**	0.928	0.929	0.930
no	20,000	0.920	0.935	0.931	0.933	**0.935**
fo	100	0.000	0.882	0.876	**0.887**	0.877

で大きく劣るにも関わらず、系統分類と比べて同等かそれ以上のスコアを達成した。最も良いスコアを出した分類の全言語での割合を比較すると、系統分類の 4 割に対して名詞句パラメータは 3 割に達している（表 5）。訓練サンプル数が格段に少ない名詞句パラメータの分類が全言語での学習に匹敵している。さらに、約 7 割の言語において主要部パラメータが最も良いスコアを出した（表 6）。これらの結果は、最近のデータ偏重の自然言語処理において、言語学が役立つ可能性を示唆している。

5 おわりに

本稿では、言語の形態・統語的性質による分類を活用した多言語クラスタリングを提案し、NER タスクにおける本手法の有用性を示した。

本研究でクラスタリングに用いた形態・統語パラメータ以外にも、様々なパラメータが提案されている [14]。対象とするタスクによってクラスタリングに最適な言語パラメータが異なる可能性がある。また、考慮するパラメータをさらに増やして複合的なクラスタリングを行うことでより効果的に言語間転移が行われる可能性もあり、今後の課題としたい。

謝辞

本研究は JSPS 科研費 JP21H04901 の助成を受けて実施した。

参考文献

[1] Xu Tan, Jiale Chen, Di He, Yingce Xia, Tao Qin, and Tie-Yan Liu. Multilingual neural machine translation with language clustering. 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. 963–973, Hong Kong, China, November 2019. Association for Computational Linguistics.

[2] Kyle Shaffer. Language clustering for multilingual named entity recognition. In **Findings of the Association for Computational Linguistics: EMNLP 2021**, pp. 40–45, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.

[3] Andrea Ceolin, Cristina Guardiano, Giuseppe Longobardi, Monica Alexandrina Irimia, Luca Bortolussi, and Andrea Sgarro. At the boundaries of syntactic prehistory. **Philosophical Transactions of the Royal Society B**, Vol. 376, , 2021.

[4] Mikel Artetxe and Holger Schwenk. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. **Transactions of the Association for Computational Linguistics**, Vol. 7, pp. 597–610, 2019.

[5] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1715–1725, Berlin, Germany, August 2016. Association for Computational Linguistics.

[6] Vaidehi Patil, Partha Talukdar, and Sunita Sarawagi. Overlap-based vocabulary generation improves cross-lingual transfer among related languages. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 219–233, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[7] Ryokan Ri and Yoshimasa Tsuruoka. Pretraining with artificial language: Studying transferable knowledge in language models. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 7302–7315, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[8] Telmo Pires, Eva Schlinger, and Dan Garrette. How multilingual is multilingual BERT? In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 4996–5001, Florence, Italy, July 2019. Association for Computational Linguistics.

[9] Yoshinari Fujinuma, Jordan Boyd-Graber, and Katharina Kann. Match the script, adapt if multilingual: Analyzing the effect of multilingual pretraining on cross-lingual transferability. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1500–1512, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[10] Wietse de Vries, Martijn Wieling, and Malvina Nissim. Make the best of cross-lingual transfer: Evidence from POS tagging with over 100 languages. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 7676–7685, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[11] Noam Chomsky. **Lectures on Government and Binding**. De Gruyter, Berlin, Germany, 1981.

[12] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 8440–8451, Online, July 2020. Association for Computational Linguistics.

[13] Robert L. Thorndike. Who belongs in the family? **Psychometrika**, Vol. 18, pp. 267–276, 1953.

[14] Ian Roberts. **Parameter Hierarchies and Universal Grammar**. Oxford University Press, Jun 2019.

表 7　埋め込みベースクラスタリング結果 (サンプル数 10,000、クラスタ数 {2, 3, 4, 5})

#	言語			
	クラスタ数 2	クラスタ数 3	クラスタ数 4	クラスタ数 5
1	cy, ga, ps, mr, hi, ro, fr, el, bg, pl, ru, sl, hr, af, nl, de, is, en, da, no, fo	cy, ga, ps, mr, hi, ro, fr, bg, pl, ru, sl, hr, af, nl, de, is, en, da, no, fo	cy, ga, ps, mr, hi, ro, fr, pl, ru, sl, hr, af, nl, de, is, en, da, no, fo	cy, ga, ps, mr, hi, ro, fr, pl, sl, hr, af, nl, de, is, en, da, no, fo
2	es, pt, it, sc	es, pt, it, sc	es, pt, it, sc	es, pt, it, sc
3	−	el	el	el
4	−	−	bg	bg
5	−	−	−	ru

表 8　クラスタ数 {2, 3, 4, 5} における名詞句パラメータクラスタリング評価結果 (F1)

		2 クラスタ			3 クラスタ			4 クラスタ			5 クラスタ		
言語	#train	名詞句	#1000	#10000	名詞句	#1000	#10000	名詞句	#1000	#10000	名詞句	#1000	#10000
cy	10,000	0.916	**0.922**	0.919	0.916	0.917	**0.924**	0.916	0.917	**0.920**	0.916	0.913	**0.926**
ga	1,000	**0.857**	0.841	0.844	**0.857**	0.841	0.844	**0.857**	0.841	0.845	**0.857**	0.841	0.851
ps	100	0.540	**0.553**	0.550	**0.559**	0.547	0.533	**0.559**	0.547	0.554	**0.559**	0.530	0.535
mr	5,000	0.883	**0.887**	0.883	0.870	0.879	**0.886**	0.870	0.874	**0.881**	0.870	0.874	**0.881**
hi	5,000	**0.901**	0.895	0.894	0.869	0.892	**0.899**	0.869	0.887	**0.897**	0.869	0.887	**0.890**
ro	20,000	**0.943**	0.940	0.942	**0.943**	0.940	0.940	0.937	**0.940**	0.940	0.937	0.940	**0.941**
fr	20,000	**0.910**	0.907	0.906	**0.910**	0.907	0.905	0.904	**0.907**	0.905	0.904	**0.907**	0.903
es	20,000	**0.914**	0.913	0.905	**0.915**	0.913	0.905	0.910	**0.913**	0.905	0.910	**0.913**	0.905
pt	20,000	**0.921**	0.918	0.914	**0.921**	0.918	0.914	0.916	**0.918**	0.914	0.916	**0.918**	0.914
it	20,000	**0.922**	0.919	0.915	**0.922**	0.919	0.915	0.915	**0.919**	0.915	0.915	**0.919**	0.915
scn	100	0.765	0.756	**0.771**	**0.801**	0.756	0.771	0.768	0.756	**0.771**	0.768	0.756	**0.771**
el	20,000	0.912	0.909	**0.913**	**0.912**	0.904	0.901	**0.912**	0.904	0.901	**0.901**	0.901	**0.901**
bg	20,000	**0.934**	0.932	0.933	0.933	0.926	**0.933**	**0.932**	0.926	0.925	**0.932**	0.926	0.925
pl	20,000	**0.915**	0.911	0.913	**0.913**	0.911	0.912	0.912	0.911	**0.912**	0.912	0.911	**0.912**
ru	20,000	**0.900**	0.898	0.900	0.900	0.893	**0.900**	**0.900**	0.892	0.897	**0.898**	0.892	0.885
sl	15,000	0.938	0.937	**0.939**	0.939	0.937	**0.939**	**0.939**	0.937	0.936	0.938	0.937	**0.938**
hr	20,000	**0.921**	0.919	0.921	0.921	0.919	**0.921**	0.919	0.919	**0.921**	**0.920**	0.919	0.919
af	5,000	0.912	0.913	**0.917**	0.912	**0.915**	0.907	**0.915**	0.907	0.912	**0.914**	0.907	0.911
nl	20,000	**0.926**	0.919	0.922	**0.926**	0.917	0.922	**0.923**	0.909	0.921	0.921	0.909	**0.922**
de	20,000	0.885	0.881	**0.886**	**0.886**	0.881	0.883	0.882	0.881	**0.883**	0.883	0.881	**0.884**
is	1,000	**0.877**	0.873	0.876	**0.875**	0.868	0.874	**0.879**	0.865	0.878	0.875	0.865	**0.877**
en	20,000	0.841	**0.842**	0.840	0.841	**0.842**	0.840	0.838	**0.842**	0.839	0.838	**0.842**	0.839
da	20,000	**0.931**	0.928	0.929	**0.931**	0.926	0.930	**0.930**	0.924	0.928	0.929	0.924	**0.930**
no	20,000	**0.935**	0.931	0.933	**0.933**	0.931	0.932	**0.933**	0.928	0.933	**0.932**	0.928	0.932
fo	100	0.870	0.876	**0.887**	0.866	0.863	**0.874**	**0.887**	0.877	0.878	0.868	0.877	**0.883**

A　クラスタ数を変化させた場合の NER 評価結果

3 節および 4 節では、クラスタ数を 3 とした場合における名詞句パラメータクラスタリングと埋め込みベースクラスタリングの比較を行った。本節では、クラスタ数を {2, 4, 5} と変化させた場合のクラスタリング結果、またそれらを用いた NER による評価実験結果について述べる。はじめに、例として Wikiann データセットの検証セット 10,000 サンプルを用いて作成したクラスタを表 7 に示す。これらのクラスタを用いた NER 評価実験の結果を表 8 に示す。

表 8 から、クラスタ数が減少するにつれて名詞句パラメータクラスタリングが埋め込みベースクラスタリングを大幅に上回ることがわかった。クラスタ数を 2 に設定した場合は、全体の 6 割の言語で最高スコアを記録している。

一方で、クラスタ数 5 の結果では、1,000 サンプルと 10,000 サンプルの埋め込みベースクラスタリングでそれぞれ 5 言語と 13 言語、名詞句パラメータクラスタリングでは 8 言語が最高スコアを記録した (表 8)。表 7 を参照すると、埋め込みベースのクラスタ数 5 において各クラスタに属する言語数に大きな偏りがあることがわかる。特に、10,000 サンプルを用いたクラスタリングでは 1 つのクラスタに言語が集中していることがわかる (表 7)。したがって、各クラスタのデータ量にも大きな差が生じる。そのような状況下でも、NER 評価結果において名詞句パラメータクラスタリングは埋め込みベースクラスタリングに NER スコアにおいて匹敵している。この結果は、クラスタ数を変化させてもクラスタリングに形態・統語的性質を用いる有効性があることを示している。