

[文献] Martin-Caughey, Ananda. 2021. “What’s in an Occupation? Investigating Within-Occupation Variation and Gender Segregation Using Job Titles and Task Descriptions.” *American Sociological Review* 86(5):960–99.

麦山 亮太（学習院大学）

問題設定

職業 *Occupation* は社会学、社会階層・社会移動、不平等研究においてきわめて重要な変数として位置づけられてきた。社会調査を行ったとき、研究者やコーダーは回答者の行っている仕事の内容 *Job description* や職名 *Job title* の記述を参照して、職業カテゴリを与えている。このとき、ある職業は類似する *Job* (task, skill) をまとめたカテゴリと想定される。

しかし、実際のところ同じ職業に就いている人のやっている *Job* がどれくらい「似ている」のか、あるいは、その類似度が時代によって変わっているのではないか、ということについてはよくわかっていない。このことは、職業を主たる変数として用いている社会学にとって重要な問題であるとともに、近年職業内の賃金格差が増加しているという事実を理解するうえでも重要である。

さらに、職業内の違いを無視することは本来あるはずの不平等を見えなくしてしまっているかもしれない。本論文では性別職域分離を事例としてこれを検証する。たとえば同じ職業でも、男性と女性は異なる職務に就いているのではないだろうか？もしそうならば、職業レベルで性別職域分離を測定すると、本来の性別職域分離を過小評価してしまうだろう。

本研究における問い

- (1) 職業内における職名や仕事の内容にどれくらいばらつきがあるのか？
- (2) どのような職業の特性がばらつきの大きさと関連するのか？
- (3) 職業レベルの性別職域分離は、職名レベルの性別職域分離を過小評価しているのか？

方法

データ

General Social Survey の 1972–2018 年データ（自由記述が残った状態のもの）を用いる。本調査ではそれぞれ以下のような質問で職名とタスクを聴取している。

- 職名 *Job name* : “What kind of work do you normally do? That is, what is your job called?” という質問によって聴取される。ステータスやアイデンティティをより良く捉えていると考えることができる。回答例) “I am a sociologist.”
- タスク *Task description* : “What do you actually do in that job? Tell me what are some of your

main duties?” という質問によって聴取される。業務内容そのものをより良く捉えていると考えることができる。回答例) “I conduct research and teach students at university.”

職業内の類似性の測定

ある回答者 i の自由記述 (bag of words) のベクトルを x_i とする。このとき回答者 i, j の記述の類似度をコサイン類似度で測定する。 i, j がまったく同一の (単語を用いた) 記述をしていれば 1、全く異なる記述をしていれば 0 をとる。同一職業内のすべての組み合わせについて計算したコサイン類似度の平均値を取り、当該の職業内類似度とみなす。

職名の特定と職域分離の計算

Census から作られ、全部で 30000 以上の職名が記載された辞典である *Alphabetical Index of Industries and Occupations* を参照し、回答者の職名の記載がこの辞典に記載された職名のどれに当たるかその距離を計算したうえで、もっとも近い距離の職名を割り当てる。

職域分離の程度は分離指数 (D-index) で測定。

結果

職業内類似性

職業内の職名およびタスクの類似度は 1 よりもかなり低い。しかし同時に職業によって類似度の高さにはばらつきがある (Figure 1)。たとえば、Pharmacists; travel agents; tellers; postal service mail carriers などの類似性はかなり高い。一方で、compliance officers; other installation, maintenance, and repair workers; management analysts; miscellaneous entertainment attendants and related workers などの類似性はかなり低い。

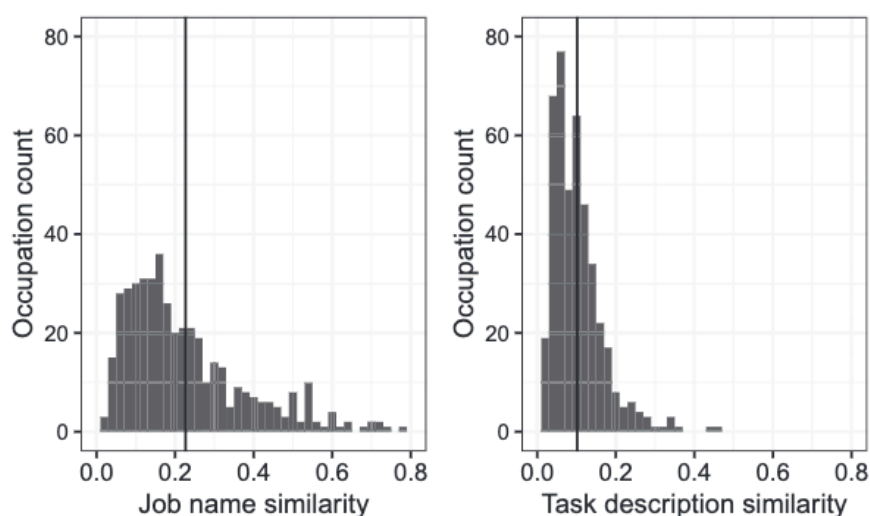


Figure 1. Distribution of Within-Occupation Similarity, by Job Names and Task Descriptions

Note: Data come from the General Social Survey, 1972 to 2018. $N = 435$ occupations with at least 10 observations for job name, and $N = 430$ occupations with at least 10 observations for task description. Similarity is measured based on the average cosine similarity between job names or task descriptions within occupations and can range from 0 to 1, from least to most internally similar. Vertical lines indicate mean similarity.

職業内類似度は近年になるほど下がっている、つまり、職業内の職名やタスクのばらつきは近年ほど大きくなっている（Figure 3）。

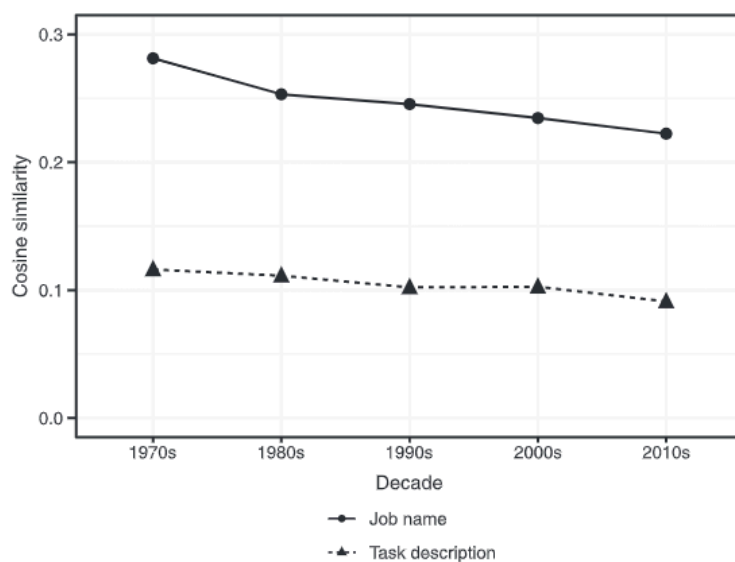


Figure 3. Occupational Similarity (Weighted Mean) by Job Name and Task Description, 1972 to 2018

Note: Data come from the General Social Survey, 1972 to 2018. Each point represents the mean similarity across occupations, weighted by occupation size. Similarity is measured based on the average cosine similarity between job names or task descriptions within occupations within decades. The measure ranges from 0 to 1, from least to most internal similarity. Occupations with fewer than two observations in a given decade are excluded. $N = 2,119$ occupation-decades for job name. $N = 2,105$ occupation-decades for task description.

SEI スコアが高い職業ほど職名の類似度が高い一方で、タスクの類似度は低い (Table 4)。たとえば大学教員や医師、弁護士などを想像するとわかりやすいかも。こういった職業は職名がはっきりしており、かつ幅広い業務に従事しているだろう。

また、男性割合と女性割合が均等に近いほど、職業内の類似度は低い傾向がある (Gender balance の係数を参照)。

Table 4. Predicting Job-Name and Task-Description Similarity with Occupational Prestige

	(1)	(2)	(3)	(4)	(5)	(6)
	Job-Name Similarity			Task-Description Similarity		
SEI (scaled)	.187*** (.039)	.247*** (.040)	.269*** (.042)	-.067*** (.014)	-.054*** (.016)	-.048** (.018)
Prop. men		-.071** (.026)	-.073* (.028)		-.013 (.010)	-.022* (.011)
Gender balance		-.774*** (.097)	-.749*** (.101)		-.170*** (.040)	-.159*** (.043)
Prop. White		.092 (.267)	.075 (.270)		-.234 (.173)	-.241 (.177)
Racial balance		.077 (.460)	.057 (.464)		-.413 (.273)	-.427 (.278)
Med. age		-.001 (.001)	-.001 (.001)		-.000 (.001)	-.000 (.001)
Prop. union			.036 (.045)			.041 (.034)
Prop. public			-.094* (.047)			-.021 (.018)
Size (scaled)			.001 (.003)			-.001 (.001)
Constant	.138*** (.018)	.243 (.294)	.246 (.296)	.134*** (.008)	.421* (.187)	.429* (.190)
Observations	435	435	435	430	430	430
Adjusted R^2	.064	.195	.200	.048	.084	.089

Note: Data come from the General Social Survey, 1972 to 2018. The unit of analysis is detailed occupations. In Models 1, 2, and 3, the dependent variable is the average pairwise cosine similarity based on job names. In Models 4, 5, and 6, the dependent variable is the average pairwise cosine similarity based on task descriptions. *Gender balance* equals (prop. men) \times (prop. women). *Racial balance* equals (prop. White) \times (prop. non-White). *Size* refers to the number of observations in the GSS analytic sample for each occupation, divided by 100. All regressions use ordinary least squares and robust standard errors. Standard errors are in parentheses.

* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed tests).

性別職域分離

職業小分類で測定するよりも、職名で測定したほうが性別職域分離の程度は大きい (Figure 4)。つまり、職業を使うことで、性別職域分離を過小評価してしまうかもしれない。

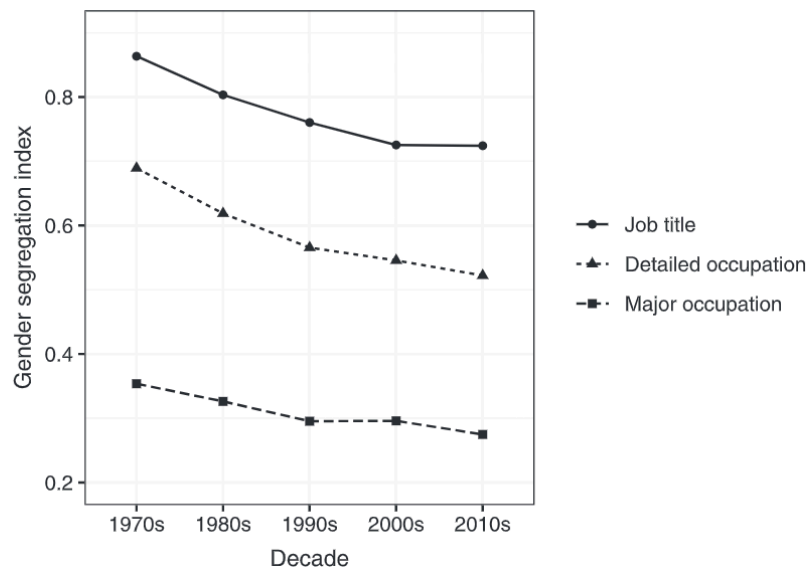


Figure 4. Gender Segregation at Three Levels, 1972 to 2018

Note: Data come from the General Social Survey, 1972 to 2018. Segregation is measured using the Duncan Segregation Index, ranging from 0 to 1 (least to most segregated). The index indicates the proportion of men (or women) who would have to switch categories (job titles, detailed occupations, or major occupations) for the distribution across categories to be equal. $N = 55,343$ individuals. The data contain 5 major occupations, 525 detailed occupations, and 10,390 job titles.

職業内類似度が低いほど、職名レベルの性別職域分離は大きい (Figure 6)。いいかえれば、同じ職業内で異なる職名がたくさん存在する職業においては、男性と女性が異なる職名に就く傾向が強い。

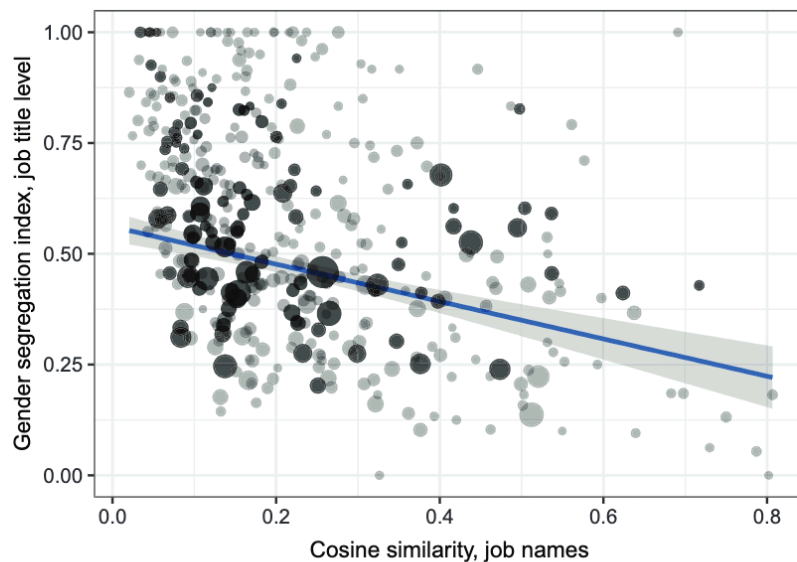


Figure 6. Job Title–Level Gender Segregation by Occupational Similarity

Note: Data come from the General Social Survey, 1972 to 2018. $N = 409$ occupations with at least 10 observations and a combination of men and women. Dot sizes correspond to the number of observations per occupation. Darker shaded dots ($N = 103$) indicate the gender segregation index at the job-title level is significant at $p < .05$ based on permutation tests. See Appendix C for details. The solid line and the shaded area represent the linear fit, weighted by occupation size, and the 95 percent confidence interval.

今後の展望

- なぜ職業の類似性が生じるのか？職業の類似性はいかなる結果をもたらすのか？
- 現今の職業分類は妥当なのか？とくに類似性が著しく低い職業は分類を改訂すべきであることを示唆するだけでなく、誤った分類をもたらすかもしれないため問題。
- 自由記述を残しておくことはデータの分析可能性を広げる。

感想

- 看護師はコーディングしやすいけど事務職はコーディングしにくいよね、とか、おなじ事務職でも一般職と総合職ってぜんぜん違うよね、とか、おなじ学校の先生でも校長先生って男性ばかりだよ、といったような、普段なんとなく思っていることを実際にデータをもとに示している点がすごい。必ずしも複雑な分析はしていないが、まさに新たな研究領域を切り拓くような研究。これが ASR に載る論文か……
- 日本の社会調査（国勢調査、就業構造基本調査、SSM、JGSS など）では **task description** は聞いているけど **job title** は聞いている（なぜだろう？）。データさえ手に入れば、**task description** から職業の類似性の指標を作ることができる。ただ、別の国でこれを追試してどれくらい意味があるのかはわからない。
- 職業に限らず、同質な集団に所属することと異質な集団に所属することはそれぞれ個人の行動や意識に対して何らかの影響を持ちそう。そういった研究にもこの研究のアプローチ（テキスト分析による類似度の定量化）は応用可能性があるかもしれない。
- RC28 でこの論文とは別の報告を聞いたことがあるが、そのときは自由記述から **analytical task** に関する記述が増えていて、同じ職業内でも徐々に異なるスキルが求められるようになってきているということを示していた（うろおぼえ）。