

独立成分分析は埋め込み空間の普遍的な幾何学的形状を明らかにする

山際 宏明*1 大山 百々勢*1,2 下平 英寿1,2 1京都大学 2理研AIP



目的

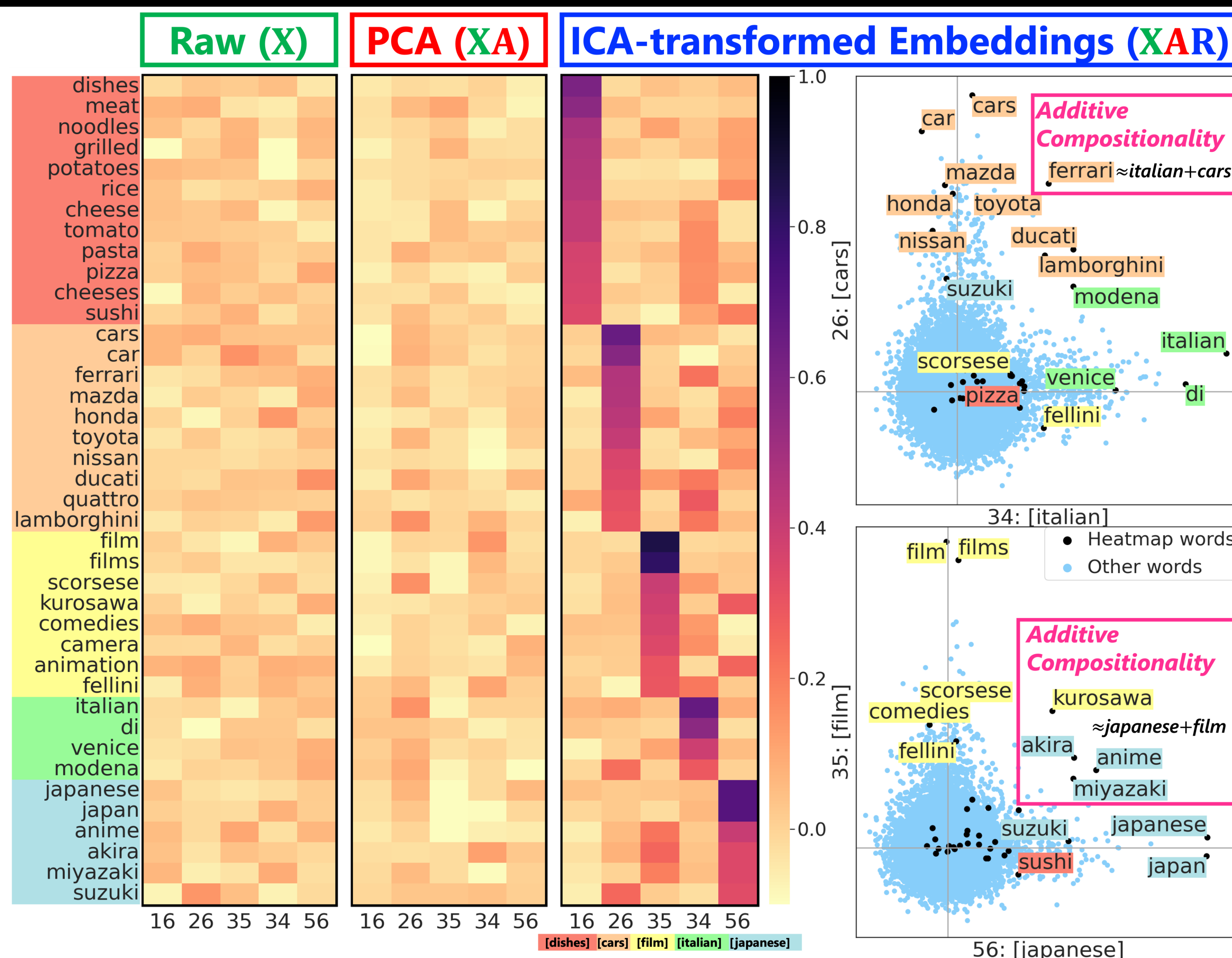
1. 埋め込みがどのように意味情報をエンコードするかを理解する。
2. 埋め込み空間における幾何学的な意味の関心の普遍性を探求する。

ICA を使用

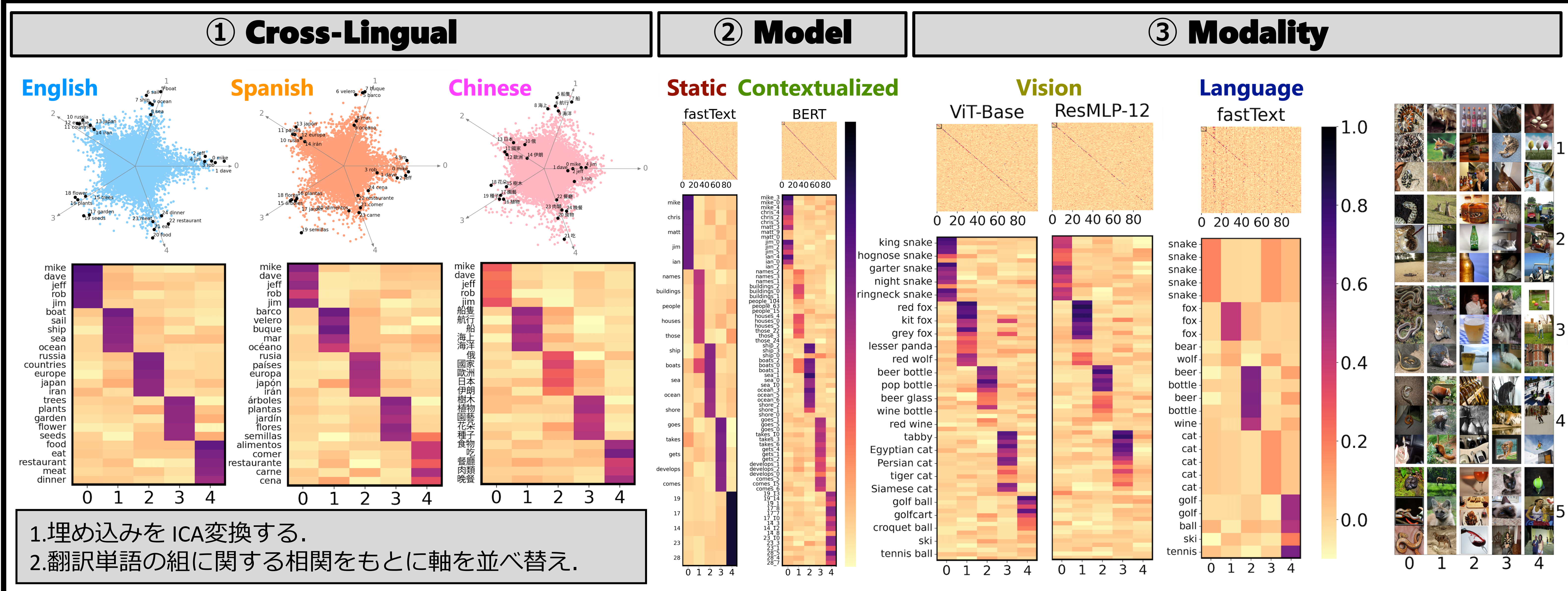
独立成分分析 (ICA) は統計的独立な軸を見つける。

結果

1. ICA 変換後の埋め込みの軸の成分は **尖っている & 解釈可能**
2. **尖っている & 解釈可能** な軸は様々な埋め込みで普遍的



埋め込みにおける普遍的な幾何学的形状



ICA は尖った軸を見つける

● 独立な軸を見つけるための **ICA** の手順:

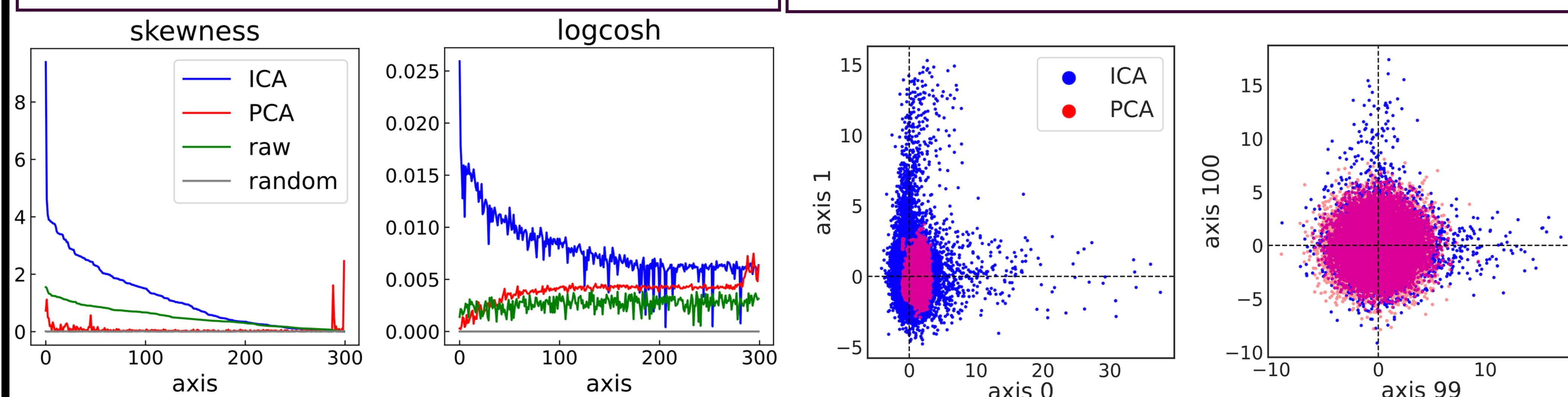
1. **白色化 (PCA)**: それぞれの軸を無相関化。
2. **直交変換**: 軸の成分がガウス分布との違いの指標を最大化。

$$S = XAR$$

● 独立な軸の成分の分布は尖っている:

- **PCA** は尖った軸を見つけることが**不可能**。等方的な埋め込みが得られる。
- **ICA** は尖った軸を見つけることが**可能**。異方的な情報に着目。

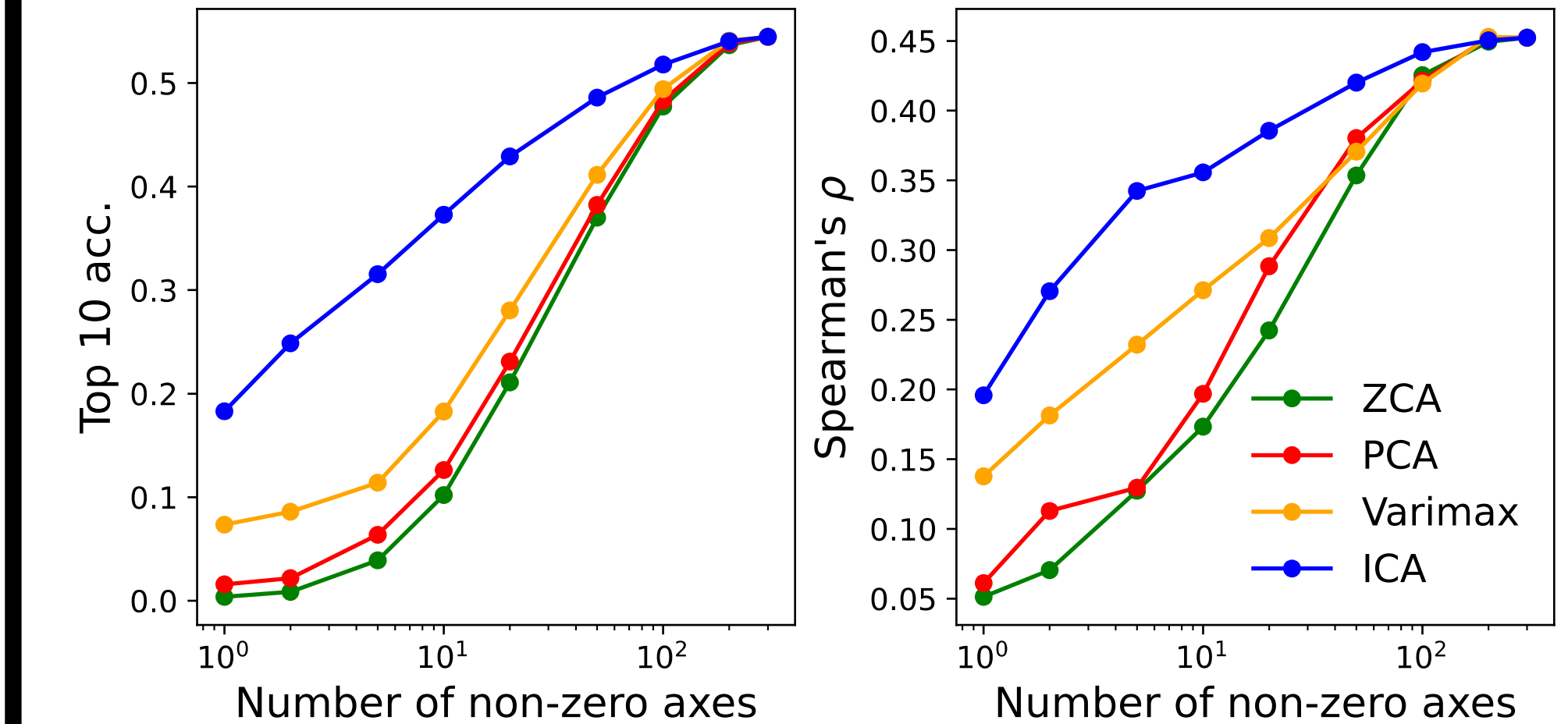
ICA は軸のガウス分布との違いの指標を最大化 ICA 後の埋め込みの軸成分の分布は尖っている



スパース性

- ICA変換後の埋め込みは**少ない軸**で意味を表す
- 評価タスク:
 1. アナロジータスク
 2. 類似度タスク

● いくつかの軸の成分をゼロに置換。非ゼロの軸数を変えて評価を行う。



Ref. [1] Aapo Hyvärinen and Erkki Oja. Independent Component Analysis: Algorithms and Applications. (Neural Networks 2000) [2] Tomáš Musil and David Mareček. Independent Components of Word Embeddings Represent Semantic Features (arXiv 2022)