

音響・音韻特徴に基づく 言語類似度マッピングシステムの構築と 言語地理学的考察

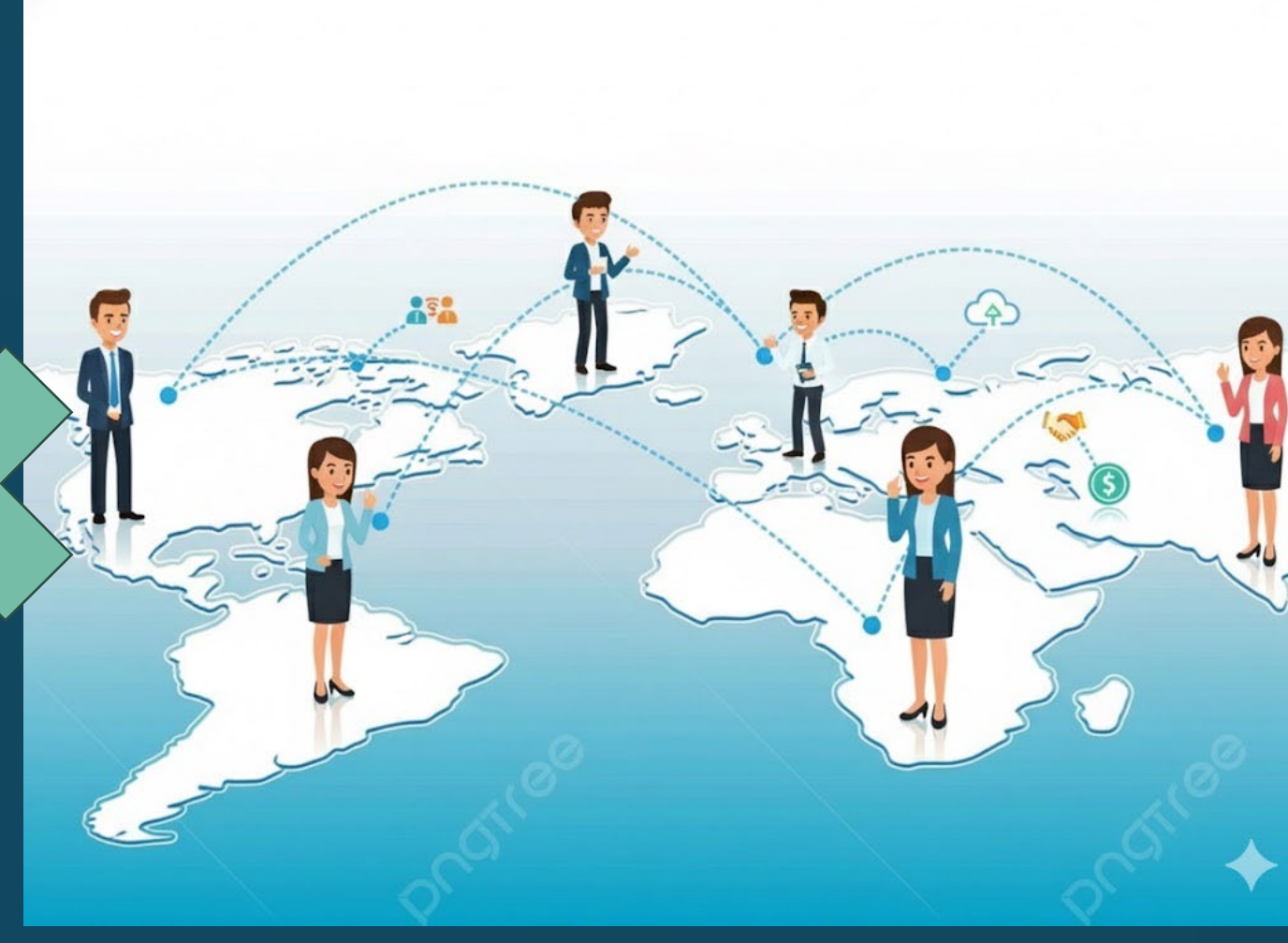
中央大学理工学部ビジネスデータサイエンス学科自然言語処理研究室
西部実堯、福田悟志、難波英嗣

背景・目的

コミュニケーションの壁

学習の効率化

グローバルビジネス



言語の構造や
使用状況を
客観的に
分析・可視化

提案手法

- ① 特定の単語をGoogle Translateを用いて各国の言語に翻訳(約80言語)
- ② 翻訳結果をespeak-ngを用いて発音記号(IPA)に変換(約64言語)
- ③ 翻訳結果を基にespeak-ngが音声を生成
- ④ IPAを基に類似度の距離行列を計算し3次元のMDSにRGBを当てはめる
- ⑤ Base64が③の結果をデータ化しHTML向けに整える
- ⑥ HTML用の白紙の地図を生成
- ⑦ ③と④の結果を地図上に提示

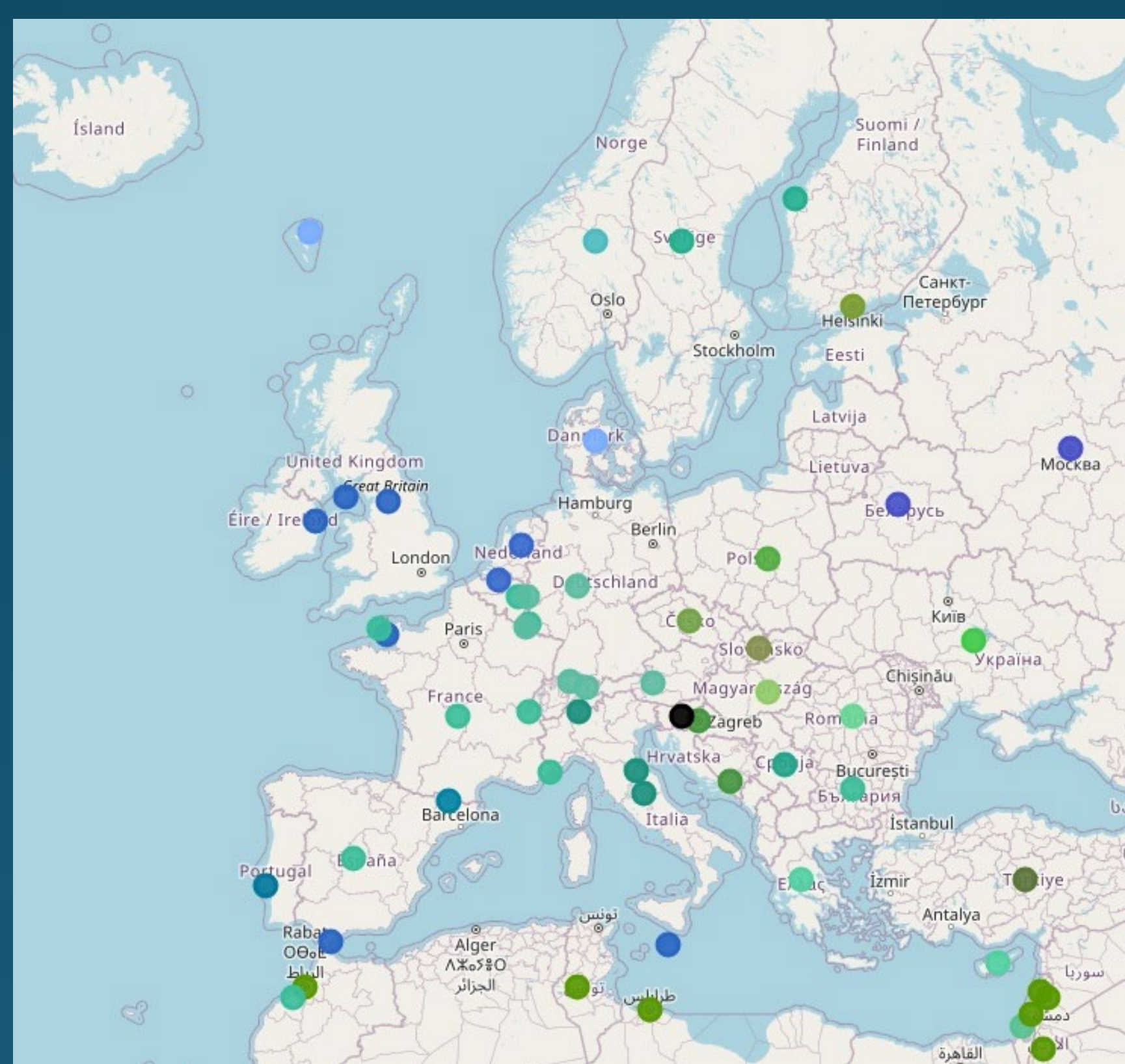
結果

「お茶」の分布



寒色 ⇒ 陸のルート
暖色 ⇒ 海のルート

「コーヒー」の分布



青色 ⇒ "coffee"系の発音が多い
緑色 ⇒ "cafe"系の発音が多い

結論・今後の課題

提案手法によって、言語の構造や使用状況を可視化することが可能となった。今後の課題としてさらに精度を上げていくため以下の手法を試みたい。

- ✓ 翻訳言語の増加
 - ✓ 音声からの音素抽出
 - ✓ 同言語でも地域ごとのなまりを反映
- また、他にも単語ごとの類似度分布の比較も行っていきたいと考えている。

参照

・【中国茶】中国から世界へ！お茶の伝わり方(陸路のchaと海路のtea), <https://kimama-yakuzen.com/totheworld/>
 ・ Maps on the Web, <https://mapsontheweb.zoom-maps.com/post/752332291764797440/coffee-in-different-languages-by>