

博士論文

文化差認識による
多言語非対称コミュニケーション支援
(Culturally-Aware Support
for Multilingual Asymmetric Communication)

2025年3月

立命館大学大学院 情報理工学研究科
情報理工学専攻 博士課程後期課程

西村 一球

立命館大学審査博士論文

文化差認識による
多言語非対称コミュニケーション支援
(Culturally-Aware Support
for Multilingual Asymmetric Communication)

2025年3月
March 2025

立命館大学大学院 情報理工学研究科
情報理工学専攻 博士課程後期課程
Doctoral Program in Advanced
Information Science and Engineering
Graduate School of Information Science and Engineering
Ritsumeikan University

西村 一球
NISHIMURA Ikkyuu

研究指導教員：村上 陽平教授
Supervisor : Professor MURAKAMI Yohei

内容梗概

近年、グローバル化に伴い、人やモノの往来が盛んになり、異文化の人々との交流が増えつつある。そうした中で、異文化の人々との交流において、障害となるのが言語の違いである。ニューラル機械翻訳など翻訳技術の発展により、これらの障害はなくなりつつあるが、いまだにコミュニケーションに齟齬が生じる場合がある。その要因の一つとして、文化の違いがあげられる。例えば、日本語の“団子”を機械翻訳で英訳した結果は“dumpling”であるが、これら二つの単語が表現する概念は少し異なっている。団子に比べて、dumpling はより広義な概念を指し、団子に加えて餃子や小籠包など小麦粉を練ったもの全般を指す言葉である。このケースは機械翻訳の誤訳ではなく、英語には団子と餃子、小籠包を細かく区別する文化がなく、それに該当する単語がないことからより広義な dumpling が団子の対訳結果として出力されている。また、単語間で表現する概念は同じ場合でも、話者間で単語から想起するイメージが異なる場合がある。例として、“ごぼう”とその英訳結果である“burdock”は、同じ植物を指す言葉である。しかしながら、この植物の根を食べる文化を持つ日本人はごぼうから根っこをイメージし、そのような食文化を持たないアメリカ人は burdock から花のイメージを想起する。このように、翻訳を通して異文化の話者と会話を行った場合でも、各話者でイメージや受け取る意味が対称ではない。

このような文化差を解決するためには、どちらかの話者が違和感に感じた際に相手に確認を行う。それによって、お互いの話者が認識のギャップを修復することを試みる。しかしながら、多言語コミュニケーションでは、機械翻訳を使うため、修正を図る発話にも文化差があり、認識のギャップが埋まりづらい。そこで、本研究ではその文化差を検出し、話者に注意を与える。それによって、話者自身に気付かせ、発話時に文化差を

考慮した発話を促すことで、非対称性のある多言語コミュニケーションを支援することを目的とする。

本論文では、文化差認識による多言語コミュニケーションの支援において以下の課題に取り組む。

1. 概念レベルの文化差の検出手法の構築

言語はその土地の文化や歴史に影響を受けながら独自に発展するため、その概念体系は言語によって異なる。これにより、ある単語を翻訳した時に対訳関係とされる単語でも、それぞれの単語間で表現する概念の範囲は異なる。このような単語間の概念差は多言語コミュニケーションに齟齬を生む原因となる。そのため、本研究では単語の概念差を“概念レベルの文化差”と定義し、概念辞書を用いて概念レベルの文化差を検出する。具体的には、WordNetとOpen Multilingual WordNetという概念辞書を用いて、単語が表現する概念の範囲を比較する。WordNetとは同義の英単語を集めて一つの概念(以下 synset)とした英語の概念辞書であり、その各 synset の英単語を翻訳などによって半自動的に多言語を対応づけたものがOpen Multilingual WordNetである。まず、単語が表現する概念の範囲を定量化するために、単語が紐づいている synset の集合を単語が表現する概念の範囲とする。そして、異言語の単語間でその概念の範囲を比較することで、概念レベルの文化差を検出する。検出精度を測るために、人手で画像と単語を用いた評価データセットを作成し、その評価データセットを用いて検出手法を評価した。その結果、日本語と中国語間で正解率:81.0%、日本語とインドネシア語間で正解率:70.0%、中国語とインドネシア語間で69.0%であった。また、本手法の再現率は日本語-中国語で87.0%、日本語-インドネシア語で80.0%、中国語-インドネシア語で71.0%であり、これまでの先行研究を上回る結果となった。

2. 想起イメージレベルの文化差の検出手法の構築

普段から見るものや使い方、印象などに影響を受けるため、文化によって単語から想起するイメージは異なる。このような単語のイメージの違いは多言語コミュニケーションに齟齬を生む原因とな

る。そのため、本研究では、単語から想起されるイメージの差を“想起イメージレベルの文化差”と定義し、単語から画像検索によって取得される画像の類似性に基づいてイメージの文化差を検出する。具体的には、対象となる単語をキーワードに画像検索を行い、取得された上位 10 件の画像から、VGG16 という畳み込みニューラルネットワークを用いて画像の特徴ベクトルを生成する。各言語ごとに特徴ベクトルを平均することで、その単語が持つイメージの特徴ベクトルを作る。そして、言語間でイメージの特徴ベクトル間の類似度を算出し、閾値を下回った場合を想起イメージレベルの文化差があると判定する。検出精度を測るために、人手で画像と単語を用いた評価データセットを作成し、その評価データセットを用いて検出手法を評価した。その結果、正解率は 76.0%で、再現率は 75.2%であった。本研究では、閾値を定めることで、それまでヒューマンベースであった想起イメージレベルの文化差検出の自動化に成功した。

3. 多言語コミュニケーションにおける文化差検出手法の効果検証

コミュニケーションは非対称であり、一方の話者が伝えたい内容がそのまま他方の話者に伝わることは少ない。特に多言語コミュニケーションでは、翻訳を挟むことによって言語や文化の違いからその溝はより大きくなる。そこで、私は概念レベルの文化差と想起イメージレベルの文化差の検出手法を提案した。これらの提案手法の評価を行うために、多言語コミュニケーションの実験をデザインし、実際のコミュニケーションで提案手法の効果を評価する。具体的には、多言語チャットツールを用いて、日本語話者とインドネシア語話者にサバイバルタスク(協働タスク)を2回行ってもらう。一方のタスクでは、文化差のあるワードが出現した際に、話者らに注意を与える支援を入れる。もう一方のタスクでは、支援を入れずに実験を行う。効果検証として、支援の有無によって、話者の文化差の認識、対話の情報量、会話の理解度が変化するかを確認する。実験の結果、支援を入れることで各話者は文化差の認識が向上した。それに伴い、話者の対話の情報量が増え、お互いの発話に対して認識をすり合わせる事が確認された。よって、支援を入

れることで、文化差によるコミュニケーションの齟齬が起こらないように、各話者が振る舞うことがわかった。

目次

第1章	序論	1
1.1	研究背景	1
1.2	コミュニケーションの非対称性	2
1.2.1	概念レベルの文化差	3
1.2.2	想起イメージレベルの文化差	4
1.3	研究目的	6
1.4	論文の構成	7
第2章	多言語コミュニケーション支援	9
2.1	言語支援	10
2.1.1	機械翻訳	10
2.1.2	機械翻訳を活用した応用研究	11
2.2	文化支援	14
2.3	インタラクション支援	16
第3章	概念レベルの文化差検出	19
3.1	単語が表現する概念	19
3.2	自然言語処理における単語の意味の対応づけ	20
3.3	概念レベルの文化差検出手法	21
3.3.1	WordNet	21
3.3.2	概念の範囲の定量化	25
3.3.3	概念範囲の比較	26
3.4	人手による概念レベルの文化差の判定	29
3.4.1	単語三つ組の作成	30
3.4.2	アンケートの設計	33
3.4.3	アンケートに基づく文化差判定	35

3.5	評価	37
3.6	分析	39
3.6.1	既存手法との比較	39
3.6.2	概念レベルの文化差の分類と精度の関連性	40
3.6.3	コーパスを対象とした概念レベルの文化差検出の適用	43
3.7	まとめ	44
第4章	想起イメージレベルの文化差検出	45
4.1	単語によって想起されるイメージ	45
4.2	画像の類似性の算出	46
4.3	想起イメージレベルの文化差検出手法	46
4.3.1	画像認識技術	47
4.3.2	特徴ベクトルに基づく類似度算出	47
4.3.3	閾値を用いた文化差判定	50
4.4	評価実験	51
4.4.1	実験手順	51
4.4.2	評価	53
4.4.3	既知の文化差への適用	58
4.5	考察	60
4.5.1	提案手法で判定に成功したケース	60
4.5.2	提案手法で判定に失敗したケース	62
4.6	まとめ	66
第5章	文化差検出によるコミュニケーション支援	67
5.1	多言語コミュニケーションにおける文化差	67
5.2	多言語コミュニケーションの非対称性	68
5.2.1	予備実験	68
5.2.2	仮説	71
5.3	実験	72
5.3.1	参加者	72
5.3.2	実験で使用するコミュニケーションツール	73
5.3.3	実験手順	75

5.3.4	実験タスク	76
5.3.5	アンケート	81
5.4	定量的分析	82
5.4.1	文化差の認識	82
5.4.2	対話の情報量	83
5.4.3	ユーザの会話の理解度	88
5.5	定性的分析	90
5.5.1	認識の乖離	90
5.5.2	文化差によって会話に齟齬が起こったケース	91
5.5.3	支援後の認識のすり合わせ	94
5.5.4	支援効果の継続性	95
5.6	まとめ	96
第6章	結論	98
6.1	研究の貢献	98
6.2	今後の展望	100
	参考文献	102
	研究業績	120

目次

1.1	概念レベルの文化差の例	3
1.2	想起イメージレベルの文化差の例	5
3.1	グラフ G での V と W のイメージ例	26
3.2	水平 (部分共通)	30
3.3	水平 (包含)	30
3.4	垂直 (下位包含)	30
3.5	垂直 (上位包含)	30
3.6	垂直 (部分共通)	31
3.7	垂直 (包含)	31
3.8	垂直 (上位共通)	31
3.9	日本語 WordNet における Synset(“abstraction”) の情報	32
3.10	アンケートに使用された画像とクエリの例	33
4.1	想起イメージレベルの文化差検出手法	49
4.2	文化差判定のアンケート	52
4.3	WordNet の上位 3 層	55
5.1	LangridChat	73
5.2	LangridChat のシステム構成	74
5.3	実験の流れ	75
5.4	アイテムセットと支援の組み合わせ	76
5.5	総単語数	84
5.6	1 発話の単語数	85
5.7	対話の発話数	87

表目次

3.1	Multilingual WordNet のデータ数 1	23
3.2	Multilingual WordNet のデータ数 2	24
3.3	検出された単語ペア数	32
3.4	(ボート, 船, kapal) の synset とクエリ	34
3.5	(ボート, 船, kapal) のアンケートの解答例	37
3.6	表 3.5 の解答例から作成される 「“craft” -“vessel ship” - “boat”」 の集計表	38
3.7	表 3.5 の解答例から作成される 「“vessel ship” -“boat”」 の 集計表	38
3.8	表 3.5 の解答例から作成される 「“boat”」 の集計表	38
3.9	既存手法との比較結果	39
3.10	(日中) 文化差の分類とデータの内訳	40
3.11	(日尼) 文化差の分類とデータの内訳	41
3.12	(中尼) 文化差の分類とデータの内訳	42
3.13	出現頻度上位 10,000 件の名詞への文化差検出結果	43
4.1	Synset(“burdock”) の日本語と英語の情報	48
4.2	Synset(“burdock”) の単語群とその画像検索のクエリ	48
4.3	人と提案手法の一致	50
4.4	交差検証での各プロセスのテスト結果	53
4.5	Synset(“entity”) の日本語と英語の情報	56
4.6	Synset(“physical entity”) の日本語と英語の情報	56
4.7	Synset(“abstraction”) の日本語と英語の情報	57
4.8	2 層目の synset による分類の Accuracy	57
4.9	3 層目の synset ごと分類した時の Accuracy	58
4.10	各層の検出精度の検定結果	59

4.11	既知の文化差に対する提案手法の判定結果	59
4.12	文化差判定成功 - 事例: フットボール	60
4.13	文化差判定成功 - 事例: ワイン	61
4.14	文化差判定に失敗した事例: 水	62
4.15	文化差判定に失敗した事例: ザクロ	63
4.16	文化差判定に失敗した事例: タピオカ	64
4.17	日本語話者が英語で取得された画像のみを選択したケース	64
4.18	パターン4の事例	65
5.1	予備実験で文化差により会話に齟齬が起こったケース .	69
5.2	予備実験で支援を入れたケース	70
5.3	パターンAのアイテム	78
5.4	パターンBのアイテム	78
5.5	Q3(アイテムに対して, 概念が違うと感じたアイテムの 数)の集計表	83
5.6	アイテムについて言及した発話数	87
5.7	1アイテムあたりの発話数	88
5.8	Q1(相手の言っていることがよくわかったか?)の集計表	89
5.9	Q2(相手はあなたの言っていることを理解していました か?)の集計表	89
5.10	話者の認識が異なっているケース	91
5.11	文化差によって齟齬が起こったケース	92
5.12	表5.11のID12, 13の発話の各言語の表示結果	93
5.13	支援後に認識のすり合わせが見られる会話ログ	94
5.14	支援の効果が次のタスクでも続いている例	96

第1章 序論

1.1 研究背景

近年, グローバル化により国境を超えて, 異なる言語・文化の人との交流が増えつつある. このような異文化コラボレーションは社会や技術の発展を促進する. 2015年に国連によって定められたSDGs*(Sustainable Developments Goals)では, 持続可能でよりよい世界を目指すをコンセプトに17のゴール, 169のターゲットの開発目標が定められている. その中の一つに, 地球規模の課題に取り組むために, 言語や文化など様々な多様性を考慮できる人材を育成することを目標とする地球市民教育の実現がある. 例えば, NPO Pangaeaでは, 異なる言語・文化を持つ児童を集めて, 協働で世界の諸問題の解決をデザインする異文化サマースクール(KISSY:Kyoto Intercultural Summer Schools for Youths[†])を開催し, 地球市民教育の実現を目指している. このように, 社会的に異文化コラボレーションに対する機運が高まっている.

このような異文化コラボレーションでは, 言語による障壁が存在するが, 近年の機械翻訳技術の進歩は言語面の問題を解決しつつある. 特に, ニューラル機械翻訳の登場により日英の翻訳精度がTOEIC900点の話者と同程度まで向上するなど, 多言語コミュニケーションの支援環境が構築されつつある.

しかしながら, いまだにコミュニケーションに齟齬が起こる場合がある. 例えば, KISSYでは話者間の多言語コミュニケーションを機械翻訳ツールを用いてサポートしているが[1, 2], 日本語話者が“あんこ”を翻

*<https://sustainabledevelopment.un.org/?menu=1300>

[†]https://www.pangaeaan.org/web/japanese/general/aboutpangaeaact_jp.html

訳した際に、得られた英語の翻訳結果が“red beans paste”であった。この“red beans paste”という単語を用いて、クメール語話者にあんこを説明したが、それぞれの話者で想起しているイメージが異なった。日本語話者は、あんこから小豆をペースト状にしたものを想起していた。一方で、クメール語話者はred beans pasteからスティック状に押し固められた固形物を想起していた。それによって、翻訳結果は正しいのにも関わらず、クメール語話者は話の内容を理解することができなくなった。

このような齟齬が起こる要因の一つに、コミュニケーションの非対称性がある。非対称性とは各話者の使用する言語や文化背景の違いから、翻訳が正しい場合でも話者ごとに受け取る意味や想起するイメージが対称ではないことである。これによって、話者たちは翻訳を介して会話が円滑に進んでいるように感じるが、実際には話者間でわずかな認識のずれが生じていることに気づいておらず、それがコミュニケーションに齟齬を生む要因となりうる。現状の多言語コミュニケーションでは、非対称性による認識のずれはどちらかの話者が違和感を感じた際に相手に確認を行い、お互いの話者が認識のずれがあることを理解する。その後、お互いの話者が認識のギャップを修復することを試みるが、多言語コミュニケーションでは、機械翻訳を使うため、修正を図る発話にも非対称性があり、認識のギャップが埋まりづらい。そのため、非対称性を検出し、各話者に注意を与えることで、話者自身に非対称性を考慮した発話を促すことが必要となってくる。

1.2 コミュニケーションの非対称性

多言語コミュニケーションでは、対訳関係にある単語同士でも言語によって表現する概念が違う。これによって話者間で受け取る意味が異なることがある。また、表現する概念が同じ場合でも話者ごとに想起するイメージが異なる場合がある。このように、翻訳によって流れを汲んだコミュニケーションが行われているように感じても、話者ごとに受け取る意味や想起するイメージが対称ではない。このようなコミュニケーションの非対称性の要因の一つが文化の違いである。本研究では、単語によって表現する概念が異なる非対称性を“概念レベルの文化差”、単語



図 1.1: 概念レベルの文化差の例

から想起するイメージが異なる非対称性を“想起イメージレベルの文化差”と定義する。

1.2.1 概念レベルの文化差

言語によって、表現できる概念体系は異なる。これは言語がその地域の文化に影響を受けながら独自に発展することに起因する。ある文化ではよく使われ、様々な形や種類があるものは、その言語において、細かく区分され、一つ一つにそれを表現する単語が定義されている。一方で、違う文化ではそれほど重要ではない場合、その言語では一つの単語で定義される場合や、それを表す単語がない場合がある。

具体的な例として、日本語の“団子”と英語の“dumpling”がある。この二つの単語は機械翻訳において、正しい対訳とされる単語のペアである。しかしながら、それぞれが表現する概念の範囲は異なる。日本語の団子は図 1.1 の真ん中のような、穀物の粉などを丸めて蒸したもので、一般的には甘いものを表す。一方で、dumpling は図 1.1 のように、団子の他に餃子や小籠包[‡]などをまとめて表す言葉である。これらは、誤訳などでは

[‡]図 1.1 の右の画像：立命館大学大学院 社会知能研究室 森叶葉 提供

なく、日本語では団子と餃子と小籠包は異なるものとして細かく定義され、小麦粉などを丸めたもので甘いものは団子、肉や野菜などを入れたもののうち形や調理方法によって餃子や小籠包と呼び分けている。しかし、英語ではこれらと呼び分けるような概念体系がないため、より広義な *dumpling* が団子の翻訳結果として出力される。

このような違いはコミュニケーションにおいて、英語話者が“Please get me some dumplings.”と日本語話者に伝えた時、日本語話者には翻訳結果の“団子を買ってきてください。”と伝わる。この場合、日本語話者が団子を買って、英語話者に渡しても団子は *dumpling* の一種であるため、英語話者にとって問題ではない。しかし、日本語話者から英語話者に“団子を買ってきてください。”と言った時、英語話者には“Please get me some dumplings.”と伝わる。この場合、*dumpling* は団子や餃子、小籠包など広義な意味を持っているため、英語話者は団子以外のものを買う可能性がある。このケースでは、英語話者が発信し、日本語話者が受信する向きでの会話ではコミュニケーションに齟齬は起こりづらいが、日本語話者が発信し、英語話者が受信する向きでの会話ではコミュニケーションに齟齬が起こる可能性が高い。このように、概念の文化差では、ある方向の向きでは会話に問題はないが、逆の向きでは齟齬が起こる可能性を含んでおり、話者が概念の範囲が異なることを知らない場合、気づかずに話者間で認識のずれが生じており、会話に齟齬が起こる。たとえ、認識のギャップに気付いたとしても、それを修復する言葉にも概念の文化差があるため、修復が容易ではない。

1.2.2 想起イメージレベルの文化差

単語からイメージするものは話者の文化背景や話者間の共通認識などで作られる。ある単語や会話で話者達に共有できる思い出があれば、その場面や風景を想起する。同じ文化の話者同士ならあるものについての認識はあまり差がないものになる。しかしながら、多言語コミュニケーションにおいて、異なる言語・文化の話者同士では単語から想起するイメージは大きく違うことがある。

例えば、日本語の“ごぼう”とその英訳結果の“burdock”では、それぞ

ごぼう



burdock



図 1.2: 想起イメージレベルの文化差の例

れの単語から各話者が想起するイメージは図 1.2 のように異なる。日本語話者は、根っこの見た目の食べ物を想起するが、英語話者は burdock から赤く特徴的な花[§]のイメージを想起する。このようなイメージの差は誤訳や先ほどの概念レベルの文化差から起こるものではなく、ごぼうと burdock は同じ植物を指す言葉であるが、日本人はこの植物の根の部分を食べる食文化を持っている。そのため、ごぼうと聞いたら根っこの野菜のイメージを想起する。一方で、そのような食文化を持たない英語圏では burdock は根の部分想起せず、より特徴的な花の部分想起する。

このようにそれぞれの話者で単語から想起するイメージが違うため、実際のコミュニケーションで日本語話者がごぼうについて英語話者と話すとお互いが会話の流れがわからなくなる。日本語話者が“ごぼ天うどんが美味しいので、ぜひ食べてみてください。”と英語話者に伝えた場合、“The burdock tempura udon is delicious. Please try it.”と英語話者には伝わる。しかし、それぞれの話者が想起しているイメージは異なるため、もし英語話者がごぼ天うどんを注文した時には、期待していたものとはイメージが異なる。

[§]図 1.2 の右の画像：東京生薬協会 山上勉 提供

1.3 研究目的

本研究では文化差を自動的に検出することで、話者自身に非対称性があることを気付かせ、発話時にそれを考慮した発話を促す。それによって、非対称性のある多言語コミュニケーションを支援することを目的とする。そのために、“概念レベルの文化差”と“想起イメージレベルの文化差”の検出手法を提案する。

まず、概念レベルの文化差を検出するために、概念辞書を用いて単語が表現する概念の範囲を定量化し、単語間の概念を比較する。具体的には、人手で同義の英単語を一つの概念 (synset) として構成された WordNet という概念辞書を用いる。WordNet は、各 synset 間の階層的な is-a 関係を反映した木構造を持つ。この構造を利用して、単語が表現する概念の範囲をその単語が対応している synset の集合で定義する。また、WordNet は英単語の情報しかないので、その多言語版である Open Multilingual WordNet を活用して多言語の情報を取得する。これによって、言語間で単語の概念を比較する。

次に、想起イメージレベルの文化差を検出するために、単語から画像検索で取得できる画像の類似性に基づいて、単語が持つイメージを比較する。具体的には、単語から取得できる画像をその単語がもつイメージとして捉え、画像の特徴ベクトルを抽出する。各言語で複数の画像の特徴ベクトルから一つの特徴ベクトルを作り出し、言語間で特徴ベクトルのコサイン類似度を算出する。この時、類似度が一定の閾値を下回っていたときに、その単語間で想起イメージに差があるとして、文化差ありと判定する。

これらの文化差検出による支援が、効果的かを検証するために多言語チャットツールを用いた実験を行った。具体的には、異なる言語の話者間で文化差によって会話に齟齬が起るようなコミュニケーション実験をデザインする。このとき、発話に文化差のある単語が含まれていたときに、話者に注意を与えるエージェントを用意し、支援の有無で話者の振る舞いを比較する。

1.4 論文の構成

本論文は, 2章から6章までで構成される.

2章では, 多言語コミュニケーション支援を行っている関連研究について述べる. 最初に, 多言語コミュニケーションで最初に問題となる言語の問題に取り組む研究を述べる. その中で, 言語面の壁を取り除くために用いられる機械翻訳がどのように変遷してきたかを紹介する. そして, 機械翻訳を応用・活用して多言語コミュニケーションを支援する研究を紹介する. 次に, 言語の壁を超えても問題となる文化の問題に取り組む研究を述べる. 文化差によって生まれる認識や印象などの違いを検出することが先行研究で行われており, それらを紹介する. そして, 最後に異文化の話者とのインタラクションを支援する研究について述べる. 実際の多言語コミュニケーションではどのような問題があり, 会話を促進するためにどのような振る舞いが有効かが分析されている. それらの知見をもとに行われている異文化コラボレーション支援の研究について述べる.

3章では, 概念レベルの文化差の検出手法について述べる. 概念レベルの文化差を検出するために, WordNetを用いた単語の概念の定量化と, 概念の範囲の比較について説明する. また, 提案した概念の文化差の検出手法を評価するための評価データセットの作成方法について説明する. 評価データセットの作成では, 人手で単語に概念差があるかを判定する. そのために, 異言語の話者間で概念差が抽出できるような実験をデザインする. そして, 作成した評価データセットを用いて, 概念レベルの文化差検出手法の精度評価を行う. 加えて, 検出精度を分析するために, 単語が対応づいている synset のトポロジーに基づいて, 概念レベルの文化差を7つに分類し, どのようなデータに対して概念レベルの文化差検出は有効であったかを調べる.

4章では, 想起イメージレベルの文化差検出手法について述べる. 想起イメージレベルの文化差を検出するために, 画像の特徴ベクトルを抽出し, その類似性によって文化差検出を行う. この時, 文化差の有無の判断基準となる閾値の最適化を行った. そして, 評価のための評価データセットの作成を行う. 人手で文化差の判定を行うためのアンケートをデザインする. 作成した評価データセットを用いて, 想起イメージレベル

の文化差検出手法を評価する。また、WordNet の synset の種類ごとに、本手法がどのようなデータに対して有効かを検証する。

5章では、作成した二つ(概念レベルと想起イメージレベル)の文化差検出手法が実際の多言語コミュニケーションでどの程度有効かを検証する。検証では、多言語チャットツールを介した多言語コミュニケーションの実験を行う。実験において、被験者が文化差のある単語を使うように協働タスクをデザインする。文化差がある単語が使われた場合、話者らにアラートを出す支援を行うグループと支援を行わないグループの2群に分け、2群間で文化差の認識や発話の情報量、会話の理解度を比較する。

6章では、本研究の貢献と今後の展望について述べる。

第2章 多言語コミュニケーション支援

本章では、関連研究として、多言語コミュニケーションを支援するための研究について述べる。

多言語コミュニケーションでは、話者間に二つの障壁があるため、会話に齟齬が生まれやすい。最初に問題となるのが、言語の壁である。一般的に、異なる言語話者同士では共通の言語を用いて、会話を行うが、各話者の共通言語の習熟度によってはスムーズに意思疎通を行うことが難しい。このような問題に対して、ニューラルネットワークを用いて、ある言語の文章を他言語に翻訳する機械翻訳の研究がある。このような機械翻訳技術の研究では、大きく二つの研究があり、一つは機械翻訳の翻訳エンジンの研究である。これにより、各言語話者がそれぞれの母国語で会話を行うことを支援している。もう一つは、それらの機械翻訳技術を活用して、多言語での人のインタラクションを支援するものである。

次に、多言語コミュニケーションで問題となるのが文化の壁である。機械翻訳によって、言語による問題はなくなりつつあるが、翻訳が正しい場合でも、各話者の文化背景の違いによって、齟齬が生まれる場合がある。そのため、文化の違いを分析し、それらを検出することを目的とする取り組みがなされている。

以上のような、二つの障壁があるため、多言語コミュニケーションでは話者間に認識の不一致や食い違いが起りやすい。そのため、これらの言語・文化を考慮して、異言語の話者達がコミュニケーションを行うための多言語コミュニケーションツールの構築がされている。さらに、実際の多言語コミュニケーションにおいて、どのような問題があるかを分析する取り組みがされている。これらは、異なる言語話者が会話を行

うときの問題を分析し、よりコミュニケーションを促進させるためには、どのような振る舞いが効果的かを分析している。それらの分析で得られた知見をもとに、会話促進に効果的な振る舞いを行うエージェントを構築し、多言語コミュニケーションを支援することが取り組まれている。

2.1 言語支援

2.1.1 機械翻訳

多言語コミュニケーションを支援するために、機械翻訳ではさまざまな手法が開発されてきた。最初に登場したのは、大きくルールベースの機械翻訳である。これらは対象となる言語の文法ルールをプログラムし、そのルールに沿って、翻訳元の言語の文章を翻訳先の言語の文章に置き換えるものである。ルールベースの問題点として、単語や文法の大量のデータが必要であり、文法ルールを一つ一つプログラムしなければならない。次に、登場したのが大きく用例ベースと言われる手法である。この手法は、翻訳したい文章と類似している翻訳事例を対訳コーパスから参照して翻訳を作成する。用例ベースではルールの作成を必要としないため、ルールベースよりも低コストで翻訳モデルを作成することができる。次に登場したのが統計ベースの機械翻訳である。統計ベースでは、単語の対訳の確立や次に来る単語の確率などを対訳コーパスから算出し、その確率に基づいて、翻訳文を作成する。この手法によって、ルールベースや用例ベースの機械翻訳よりも流暢な翻訳結果を出力することができるようになった。

2010年頃には、それまで画像や音声などの処理に用いられていたニューラルネットワークを自然言語処理の分野でも取り入れられた。Mikolovらは、RNN(Recurrent Neural Network)を自然言語処理に用いたRNNLM(Recurrent Neural Network based Language Model)が提案された[3, 4]。それに伴い、機械翻訳にもそれまでの統計ベースの機械翻訳に対して、ニューラルネットワークの技術を用いる手法が提案された。Kalchbrennerらは翻訳元の言語の単語にCNN(Convolutional Neural Network)を利用し、ターゲット言語に翻訳するモデル(RCTM)を提案し

ている [5, 6]. このモデルの翻訳精度を機械翻訳を評価する BLEU スコア[¶]という指標によって評価した結果 [7], BLEU スコアは 0.2 前後であった. 2014 年には, Sutskever らが単語文に対して RNN を利用して, 文のベクトルを作る翻訳モデル (seq2seq) を提案している [8]. seq2seq は, それまでの統計ベースの翻訳モデルに比べて, BLEU スコアが向上したと述べている. 同様に, Cho らも RNN を用いた機械翻訳の手法を提案している [9, 10]. この手法では, Encoder-Decoder に二つの RNN を利用することで, 統計ベースの翻訳モデルよりも BLEU スコアが向上したと述べている. Encoder-Decoder モデルを拡張し, Decoder に注意機構を取り入れた翻訳モデルが Bahdanau らによって提案されている [11]. これによって, Cho らの提案した RNN を用いた Encoder-Decoder モデルよりも BLEU スコアは向上した. しかし, 注意機構では単語によって特徴量の抽出に差が出る. それによって, ある単語が翻訳されないといった問題がある. Tu らはこのような問題に対して, カバレッジベースのモデルを提案している [12].

2017 年には, これまでの翻訳モデルの主流であった RNN を用いない翻訳モデルが提案されている. 近年のさまざまな自然言語処理モデルに用いられている Transformer モデルが Vaswani らによって提案された [13]. Transformer モデルでは, 注意機構と呼ばれる入力シーケンスの各要素に重みをつけることで, それまでの翻訳モデルよりも高い BLEU スコアとより少ない計算コストを誇っている. これらのような自然言語処理の技術の向上によって, 機械翻訳の精度は向上している.

これらの研究では, 多言語コミュニケーションで問題となる言語の障害を取り除くために, 機械翻訳の精度を向上を目的としている. しかし, 機械翻訳の翻訳精度が向上したとしても文化差によるコミュニケーションの齟齬は起こりうるため, 本研究ではこのような文化差を検出する.

2.1.2 機械翻訳を活用した応用研究

言語の異なる話者と会話を行うために, 様々な言語サービスが構築されてきた. 具体的には, 前項で説明した翻訳や形態素解析などの自然言

[¶]BLEU スコアは 0~1 で示され, 一般的に 0.4 を超えると高品質と言われる

語処理技術のほかにも、対訳コーパスや概念辞書などの言語資源を構築することが挙げられる。石田らは異文化コラボレーションを支援するために、これらのような言語サービスを連結した多言語プラットフォーム(言語グリッド)を提案している [14, 15, 16]。言語グリッドでは、独立していた様々な言語サービスを組み合わせて異なるサービスを提供することが可能である。例えば、一般的な翻訳サービスでは、その地域・文化の固有名詞などに適した翻訳を作ることができないため、地域に応じた翻訳サービスが必要になる。一方で、言語グリッドでは翻訳サービスとその地域の辞書データを組み合わせることで、適した翻訳サービスを提供できることを可能にした。さらに、稲葉らは言語グリッドを用いた機械翻訳の精度を大幅に改善することに成功している [17]。石田らの言語グリッドは、四つのアジアの言語(日本語・中国語・韓国語・マレー語)に英語を加えた五つの言語でプロジェクトが始められた。これに対し、Rehmらは、ヨーロッパ版の多言語プラットフォームを提案している [18, 19]。

機械翻訳を活用して、異なる言語の話者とのインタラクションを支援する研究について述べる。前項で紹介した機械翻訳技術によって、多言語でのコミュニケーションは円滑になりつつある。しかしながら、翻訳精度にも限界があり、不適切な翻訳が含まれることがある。このような問題に対して、機械翻訳のユーザー自身が適切な翻訳文になるような入力文を作成することが必要となる。これを翻訳リペアといい、宮部らは入力文を翻訳した結果をもう一度入力文の言語に翻訳すること(折り返し翻訳)で、ユーザー自身で翻訳リペアを行うように促す支援を提案している [20]。一方で、折り返し翻訳を用いることで、翻訳前の文章と異なった意味になる場合がある。具体的には、原文から対象の言語への翻訳文では意味が異なっているが、翻訳先の文章から折り返し翻訳で表示された文章は原文と意味が同じ場合があり、話者は伝わっていると思っ

ている内容でも誤解が起きる。そのため、宮部らはこのような問題を解決するために、複数の翻訳機を用いて折り返し翻訳の不確かさを検出する手法を提案している [21, 22, 23]。藤井らは、多言語コミュニケーションを支援するために、異文化間コミュニケーションツール(AnnoChat)を提案している [24]。AnnoChatはチャット型のツールとなっており、各話者の発話は機械翻訳を用いて、相手話者に伝えることができる。加えて、二つの特徴的な機能を持っており、一つは翻訳による意味の変化に気づ

けるように折り返し翻訳を表示する機能がある。もう一つは固有名詞などに対して、注釈を追加するアノテーション機能がある。その他にも、宮部らは医療現場での言語の問題を解決する研究を行っている。医療分野において、治療や薬の説明が医師と患者で正確に共有されなければ、致命的なミスを生む原因となる。しかしながら、多言語コミュニケーションでは、発話者の意図が受け手に正確に伝わりづらい。そのため、外国人の診療時には、より精度の高い翻訳が求められるため、宮部らは用例対訳を用いた多言語医療受付支援システムを構築している [25]。同様に、福島らは用例対訳を用いた問診票記入システムを提案している [26]。一方で、用例対訳は実際に使われている言葉を人手で正確に翻訳した対訳コーパスであるため、データの作成にコストがかかる。そのため、福島らは用例対訳共有システムを作成し、用例対訳の拡張に取り組んでいる [27]。同様に、山下らは、クラウドソーシングを用いた用例対訳の作成手法を提案している [28]。Vanjani らは多言語に対応したチャットボットを提案している [29]。彼らの提案したチャットボットでは 103 の言語に対応しており、実験では英語、ドイツ語、スペイン語、韓国語の話者でチャットを行った。分析では、韓国語からの翻訳はぎこちなさがあることが述べられている。福島らは対面型の会話支援ツールを構築している。多言語コミュニケーションでは共通言語を用いて会話が行われることが多い。その際に、非母語の話者は意味を正しく読み取りづらい。それを解決するために、共通言語と母語言語の二つを表示する対面型の会話支援ツール (PaneLive) を提案している [30]。岡本らは、異文化間のコミュニケーションの支援システムを構築している [31]。この支援システムでは、会話中の名詞に対して、その写真などの情報を加える支援を行っている。これらの研究は、テキストベースで翻訳を行い、異文化コラボレーションを支援するための研究である。

次に、音声データを翻訳し、異文化コラボレーションを支援する研究を述べる。中口らは、国際シンポジウムのための言語支援の研究を行っている [32]。国際シンポジウムでは、様々な言語話者が集まるため、複数の翻訳者が入れ替わりで翻訳を行うため、技術的にも集中力的にも翻訳者の負荷が高くなる。一方で、音声の認識から翻訳までを機械に置き換えると翻訳精度が低くなるため、彼らは人と機械が強調した多言語支援システムを構築した。笹島らは、音声翻訳に着目した研究を行っている。

音声翻訳では、音声認識による誤認識と機械翻訳による誤訳の二つの問題を抱えている。そのため、笹島らは音声認識後に認識結果の候補文をユーザー自身に選ばせるステップを組み込んだシステムを提案している [33].

これらの研究では、機械翻訳を活用して、多言語コミュニケーションならびにインタラクションを支援する研究である。しかしながら、言語による問題を解消するだけでは、話者間による文化の違いからコミュニケーションに齟齬が起こる。そのため、本研究では文化差を検出することでコミュニケーションの齟齬を解消する。

2.2 文化支援

言語の壁がなくなりつつあるが、いまだに文化の壁によって多言語コミュニケーションに齟齬が起こりうる。このような文化差を分析・検出する研究を紹介する。

最初に、画像やイメージなど、視覚的な情報の文化差に関する研究を述べる。柳井らは写真共有サイト上の画像を利用する文化差検出手法を提案している。概念の代表的な写真を地域ごとに分類することによって、同一概念の対象物が地域ごとにどのような文化差があるのかを視覚的に明らかにしている [34]. Pituxcoosuvam らによる研究では、日本語と英語の単語をもとに画像検索を行い、それらの画像の特徴ベクトル間の類似度を算出する。算出された類似度をもとに人の主観で文化差を判定している [2]. この研究では、本研究と同様に、画像の特徴ベクトル間の類似度によって文化差を判定しているが、算出された類似度は連続値であるため、最終的には人手によって文化差を判定している。本研究では算出された類似度をもとに閾値を使って文化差を自動的に判定するとともに、ノイズとなる画像の影響をあまり受けずに文化差判定する手法を提案している。神田らはネットワークコミュニケーションでよく利用されるアバタに着目して、アバタの表情の解釈において文化差が存在するかを分析している。その結果、アジアと欧米 8 カ国間では、否定的な内容の解釈には文化差がなかったが、肯定的な内容の解釈には文化差が存在することを発見している [35, 36]. Xu らはテレビの映像を用いた異文化

研究を行っている [37]. この研究では, 映像内の出演者の表情と字幕などのテキストデータをもとに各国におけるメディアの文化を分析している.

次に, 単語や絵文字などのテキストに基づいた文化差研究を紹介する. 吉野らは Wikipedia を用いた単語の文化差を検出する手法を提案している [38]. この研究では, 事前アンケートによって文化差が存在すると人手で判断された言葉に対して, Wikipedia 上のカテゴリなどの情報を用いて文化差を検出できるかを確認するものである. さらに, 語句の重要度を計算することで文化差の検出精度が向上することが示されている [39]. Pituxcoosuvam らは, 単語における印象の違いについて日本語話者と英語話者の違いを検出する手法を提案している [40]. Livshits らは, 機械翻訳は文化を考慮した翻訳ができないことを指摘し, テキスト間の文化差を特定しようとしている [41]. この研究では, テキストの各単語をノード, 単語間の関係をエッジとしたグラフに落とし込み, そのグラフの構造の違いによって文化差を特定する手法を提案している. Marutschke らは, Twitter の日本語とドイツ語のテキストデータを用いて, 色の認識に対して文化の違いがあるかを調査している [42]. また, 単語以外にも絵文字は多言語コミュニケーションにおいて, 各話者が言語を使わずに意思伝達を行えるものとして使用されている. その絵文字において, Cho らは文化差の検出を試みている [43]. この研究では, アメリカ人と日本人にアンケートで 120 種類の絵文字の印象について調査したところ, 絵文字は文化によって解釈のされ方が違うことが明らかにされている [44]. さらに, Cho らは絵文字の解釈を 5 つのカテゴリに分類し, 意味的関連性を図るための尺度を提案している [45]. これによって, 意味的関連性から最適な絵文字の検索を行うことを目指している.

また, 文化によって, 人のメンタリティーや性格・振る舞いに違いが出ることがある. 西田らは, 日本と海外では労働者のメンタリティーに文化の違いが問題であったと述べている [46]. この研究では, 海外に進出した日系企業の従業員にアンケートを用いて, 異文化の話者間のコミュニケーションでの問題点について分析している. 照井らは, 多言語コミュニケーションでの文化差の分析を行っている [47]. この研究では, 日本人, アメリカ人, 韓国人でのディスカッションのログを分析することで文化ごとに性格的な差があるかを分析している. Garcia らは Twitter 上

での振る舞いがその国の文化と関連しているかを調査している [48]. その評価として, Hofsted の 4 つの次元と関連していることが述べられている.

視覚情報やテキスト, 振る舞い以外では, 知識に基づいて文化差を分析する研究がある. Chung らは文化的なコモンセンスデータベースを作成し, そのデータベースの情報をもとにした推論により文化差を検出しようと試みている [49]. Thakker らは文化的な違いを見つけるために, 文化バリエーションでオントロジーを構築している [50]. Laufer らは Wikipedia 上の記述は文化によって違いがあることを調査している [51]. 具体的には, 各国の食べ物を対象に, Wikipedia の各言語での記述に差があるかをみている. 同様に, Ulrike らは Wikipedia を利用した文化差の研究を行っている. この研究では, フランス, ドイツ, 日本, オランダの文化的多様性を Wikipedia に適用し, それぞれの編集操作から国の文化とコンピューターを介したコミュニケーションの関係を調査している [52]. 研究の結果, Hofstede が明らかにした人の価値観の文化差を理解するための 4 つの次元 (権力の格差・個人主義対集団主義・男性らしさ対女性らしさ・不確実性の回避) と関連していることがわかった [53]. 同様に, Silva らはソーシャルネットワークの位置情報に基づいて文化的な差を分析している [54]. 具体的には, Twitter から食べ物と位置情報を取得し, 位置情報に基づいて食べ物をマッピングしている. この食べ物の情報を文化的な特徴と捉え, 文化的な距離を図る研究を行っている.

これらの研究では, 様々なデータから文化差検出を試みる研究である. 一方で, 本研究では多言語コミュニケーションで問題となる単語に関する文化差を概念の文化差・イメージの文化差とし, 検出することで非対称性のあるコミュニケーションを支援するものである.

2.3 インタラクション支援

多言語の話者とのインタラクションを支援する研究を紹介する. 元澤らは多言語コミュニケーションにおいて, 低資源言語話者を支援するための研究をしている [55]. 低資源言語話者とは, 機械翻訳に使用される言語資源が乏しい言語を母語とした話者で, 機械翻訳の品質が悪いため,

低資源言語話者は会話の理解ができず、会話に参入することが難しい [1]. 元澤らはこの問題に着目し、低資源言語話者の発話を促進させるために、ファシリテータがどのような振る舞いをするべきかを分析している。Pituxcoosuvamらは、その分析をもとに多言語コミュニケーションにおいて、発話を促進させるファシリテータエージェントを提案している [56]. 宮部らはシステムを介したテキストでのコミュニケーションを分析している。この研究では、リアルタイムのコミュニケーションにおいて人が相手の返答を許容できる時間 (対人許容応答時間) を分析している [57]. 彼女らは分析の結果、チャット形式での対人許容応答時間は平均1分51秒であり、入力中などのシステム状態を提示することで、平均2分35秒まで伸びることを述べている。

船越らは機械翻訳を用いた協調作業を分析し、今後の異文化コラボレーションを支援するシステムに求められる要件を分析している [58]. この研究では、日本語・中国語・韓国語・マレーシア語話者でソフトウェア開発を行った時の分析をした。分析から、話者らは他言語の話者には自分の意図が伝えようと振る舞うことがわかった。一方で、相手にどう伝わっているかを考慮した振る舞いがなかったことを述べている。鈴木らは多言語コミュニケーションによって、専門知識を伝える伝達方法を提案している [59]. 高崎らによる研究で他言語の話者に知識を伝える際のコミュニケーションに着目したものがあある。この研究では、ベトナムの児童を介して日本の稲作の専門家からベトナムの農業事業者に農業の知識を伝えるプロジェクトをもとにしている。知識伝達の際に、仲介者を挟むことにより、専門家が意図していないニュアンスでベトナムの事業者に伝達される問題があったことを述べている [60].

次に、高崎らは異文化コラボレーションのために、英語に頼らないコミュニケーションツール (Communicator) を提案している [61]. Communicatorでは、子供の多言語コミュニケーションを支援するために絵文字のみを用いたコミュニケーションが行われている。さらに、高崎らはCommunicatorを改良したPictNetを提案している [62]. PictNetでは、それぞれのユーザーは自身の絵文字セットを登録しており、絵文字セットにはオリジナルの絵文字を追加することができる。そして、絵文字セットの絵文字にはそれが表す概念を紐づけることで、送信者の意図を受信者は絵文字から読み取ることが可能である。同様に、宗森らも絵文字の

みを用いたコミュニケーションツールを提案している [63].

Abou らは留学生などが異文化理解を行うための対話戦略を提案している [64]. その文化のものを自国のものと関連づけることを CSA (Culturally Situated Associations) といい, これを行うことで異文化理解が促進される. しかしながら, 留学生が理解できる CSA を提供するには, それぞれの文化に関する知識が必要となる. そのため, Abou らは留学生に CSA を提供できるような対話戦略を提案している. Li らは, 英語のネイティブ話者と非ネイティブ話者の会話において, 会話のターンに着目したエージェントを提案している [65]. 次に, 角田らは誤訳によるコミュニケーション・ブレイクダウンを防ぐためのエージェント支援を提案している [66]. コミュニケーション・ブレイクダウンとは, 仁科らが提唱したコミュニケーションにおける齟齬である [67]. 徳井は, 仁科らのコミュニケーション・ブレイクダウンをもとに日本人と留学生の誤解について分析を行なっている [68]. その結果, お互いの話者がブレイクダウンに気づくことが重要であると述べている. 仁科浩美らは, コミュニケーション・ブレイクダウンをどのように修復するかを研究している [69]. 留学生の発表時を対象にコミュニケーション・ブレイクダウンの原因を分析し, その修復の方法を提案している.

これらの研究では, 異なる言語の話者とのインタラクションを支援している. 本研究では, これらの研究と同様に, 異文化コラボレーションを支援することを目的としており, その中でも, 話者らの文化差の認識を高めることで多言語コミュニケーションを支援する.

第3章 概念レベルの文化差検出

3.1 単語が表現する概念

多言語コミュニケーションでは機械翻訳の登場により言語の問題は解消されつつある。しかしながら、言語はその土地の文化や歴史と密接に関係して独自に発展するため、その概念体系は言語によって異なる。特に、アジアの言語はヨーロッパの言語に比べて、英語との類似性が低い傾向にある [70]。このような概念体系の違いによって、異言語間で対訳関係にあるとされる単語間においても単語の表現する概念が完全に一致するとは限らない。Fellbaum らも異言語間で単語が表す概念に差があることを述べている [71]。彼らは異なる言語間で同じ意味とされた単語同士でも、単語の指す語義・概念が部分的にしか重複していないことを指摘している。

このような概念の差異は多言語コミュニケーションに齟齬を引き起こす要因となる。山下らの研究では、多言語コミュニケーションを分析するために、異なる言語の話者で協調作業を行う実験をしている。その研究では、機械翻訳を用いたコミュニケーションでは、多義語の翻訳によって意味が変わったことが対話を破綻させたことを述べている [72, 73]。このケースでは単語が複数の意味を持つ多義語のため、単語間で表現する概念に違いがあったが、同じ意味を持つ単語間でも表現する概念に違いが出る場合がある。先ほどの“団子”と“dumpling”の例では、日本人は団子と餃子と小籠包を異なるものとして区別している [74]。しかし、英語圏ではこれらを区別する文化がないため、団子を表現する単語がなく、より広義な dumpling に翻訳される。このように言語は文化によって違う概念体系を持っているため、単語の翻訳結果が同じ概念を表すとは限らない。

そのため、私はこのような概念レベルの文化差を検出する。これを達成するために、以下のような課題がある。

単語が表現する概念の定量化

単語が表現する概念や意味を比較するためには、それらを定量化する必要がある。しかしながら、単語の意味や概念は抽象的であり、明確に定義することが困難である。そのため、それらを定量化することは容易ではない。

異言語の概念の比較

言語によって概念体系は異なるため、単語の概念を定量化できたとしても、異言語間で概念間の対応をとることは容易ではない。

私はこれらの課題を解決するために、WordNet と Open Multilingual WordNet を用いて、単語の概念を集合で表現し、その集合に基づいて概念レベルの文化差を検出する手法を提案する。

3.2 自然言語処理における単語の意味の対応づけ

関連研究として、自然言語処理での単語の語義を特定する語義曖昧性解消の研究を述べる。機械翻訳では単語を他言語に変換するために語義を正しく認識する必要がある。このとき、多義語によって単語が持つ語義の曖昧性が問題となる。このような語義曖昧性を解消する研究を述べる。

一般に、機械翻訳モデルや自然言語処理モデルでは、文章や単語がベクトルで表現される。Mikolov らが提案した skip-gram モデルと CBOW モデルでは、単語分散表現において各単語が単一のベクトルで表現される [75]。Chen らは WordNet を用いて語義ごとに異なる単語表現を割り当てる語義表現を行い、語義表現と語義曖昧性解消を統一したモデルを提案している [76]。また、Patwardhan らは語義曖昧性解消に用いられる単語の意味的関連性を測るための研究を行っている [77]。この研究で

は, WordNet を用いて 2 つの単語間で概念の意味的関連性の尺度を提案している. Apidianaki らは多言語での語義曖昧性解消について研究を行なっている [78]. 一般的な多言語での語義曖昧性解消では曖昧な単語の語義をその単語の翻訳結果の語義を参照することで決定しており, この研究ではその手法について分析している. 國府らは多義語の分散表現獲得には, 用いられたコーパスの影響を受けるという問題を述べ, そのために WordNet から概念情報を用いて分散表現を獲得する手法を提案している [79]. Montoyo らは, 知識ベースの語義曖昧性解消を提案している [80]. この研究では, WordNet の情報に基づいて, 名詞の語義を特定している. Agirre らも同様に, 知識ベースの語義曖昧性解消を特定ドメインへの適用を試みている [81]. これらのアプローチでは, WordNet などの知識データベースを用いた手法によって, 語義曖昧性解消を行なっている. 一方で, Lefever らは, Europarl parallel corpus というコーパスを用いた手法を提案している [82].

これらの研究は, 機械翻訳で問題となる語義の曖昧な多義語などの語義曖昧性を解消する研究である. 一方, 本研究では同じ語義を表現する単語間でも概念の範囲に差異があるため, その差異を検出することを行う.

3.3 概念レベルの文化差検出手法

3.3.1 WordNet

提案手法で用いる WordNet について説明する. WordNet は, 人手で英語の同義語の集合を一つ概念 (以下 synset) とした概念辞書である [83, 84, 85]. 各 synset には, 大きく 4 つの情報が対応づけられている.

1. synset 番号
2. 同義語
3. 定義文・例文
4. 関連 synset とのリンク

4.1. hypernym

4.2. hyponym

各 synset を識別するための synset 番号や、同義語、そして、その synset を説明する定義文などの情報を持っている。その他に、関連する synset とのリンク情報があり、その種類として、hypernym と hyponym がある。これらは、synset 間の is-a 関係を表現したもので、hypernym は下位 synset から上位 synset を特定するリンク、hyponym は上位クラスから下位クラスを特定するリンクとなっている。

これらの情報は全て英語で構成されており、他言語の情報が対応づいていない。そのため、WordNet に様々な言語を対応づける研究が行われている。

日本語では、井佐原らによって、WordNet 内の単語を日本語に翻訳する手法で日本語 WordNet の開発をしている [86]。さらに、SUMO や GoiTaikai など、他のリソースを活用して synset の定義文に日本語の情報を追加するなどの拡張が行われている [87]。中国語では、WordNet に中国語を対応づける手法によって、BOW(Bilingual Ontological WordNet)、CWN(Chinese WordNet)、SEW(Southeast University Chinese WordNet) の 3 つの中国語 WordNet が構築されている [88, 89]。Xu らによって構築された SEW では、英語の WordNet を中国語に翻訳する自動的手法によって、英語と簡体字中国語を対応づけている [90]。一方、Huang らが構築した BOW や Chu らによって構築された CWN では、英語と繁体字中国語を対応づけている。Wang らはこのうちの SEW をさらに拡張した新たな中国語 WordNet(COW) を構築している [91, 92]。次に、Putra らは WordNet にインドネシア語を対応づけることでインドネシア語 WordNet の開発している [93]。マレー語では、Tze らによって、マレー語 WordNet の構築が提案されており [94]、Noor らはインドネシア語とマレー語に対応した Open WordNet Bahasa を開発している [95]。その他にも、韓国語やタイ語などのアジア系の言語に適応した WordNet が構築されている [96, 97]。

一方で、ヨーロッパ系の言語でも同様の取り組みが行われている。Vossen は英語と 7 つの言語 (オランダ語、スペイン語、イタリア語、ドイツ語、フランス語、チェコ語、エストニア語) を対応づけた EuroWordNet を提案している [98]。さらに、オランダ語やフランス語では個別にオランダ語 WordNet やフランス語 WordNet が構築されている [99, 100]。こ

表 3.1: Multilingual WordNet のデータ数 1

言語	単語が対応のある synset 数	単語数	1synset あたりの 平均単語数
アルバニア語	3,201	4,257	1.89
アラビア語	6,329	9,443	2.10
ブルガリア語	3,278	3,611	1.47
カタルーニャ語	32,754	34,771	1.44
中国語	27,363	38,408	1.67
デンマーク語	3,104	3,255	1.38
ギリシャ語	13,361	13,756	1.29
英語	74,374	105,000	1.76
バスク語	25,388	22,333	1.53
フィンランド語	74,111	87,412	1.57
フランス語	37,480	41,073	1.71
ガリシア語	10,839	13,930	1.41

れら以外の言語では、ポーランド語の WordNet が構築されており [101], Rudnicka らはポーランド語 WordNet を英語の Princeton WordNet にマッピングするための研究を行っている [102]. この研究を行うために、彼らはポーランド語と英語の WordNet 間でギャップやミスマッチを検出するシステムを提案している [103].

これらのように、様々な言語の WordNet が構築されており、一般的には、WordNet の synset を各言語に翻訳することで半自動的に構築する手法を使っている. Fellbaum らはこれらの各 WordNet を統合することで 26 以上の言語に対応した Open Multilingual WordNet を提案している [71, 104, 105]. Open Multilingual WordNet では、WordNet を基準に各 WordNet を対応づけているため、単語数と単語が対応づいている synset の数に各言語で違いがある.

表 3.2: Multilingual WordNet のデータ数 2

言語	単語が対応づいている synset の数	単語数	1synset あたりの 平均単語数
ヘブライ語	3,201	3,766	1.24
クロアチア語	14,425	17,893	1.75
インドネシア語	18,386	15,780	2.02
アイスランド語	3,270	6,972	3.02
イタリア語	24,474	30,691	1.80
日本語	37,201	60,910	2.51
リトアニア語	6,072	7,196	1.61
オランダ語	23,509	34,818	2.04
ポーランド語	28,348	38,510	1.56
ポルトガル語	27,588	35,213	1.68
ルーマニア語	35,447	34,208	1.45
スロバキア語	12,186	16,680	1.94
スロベニア語	28,431	28,327	1.45
スペイン語	23,224	24,430	1.49
スウェーデン語	4,578	3,915	1.02
タイ語	49,870	56,538	1.27
マレー語	17,736	15,033	2.17
ニーノシュク*	2,954	2,781	1.32
ブークモール*	3,089	3,025	1.31

*: ノルウェーの方言の一種

表 3.1 と表 3.2 は Open Multilingual WordNet の言語ごとのデータである。Open Multilingual WordNet では、固有名詞などを除く名詞の synset が 74,374 個ある。WordNet を基準にされているため、英語では全ての synset に単語が対応づいており、その英語の単語数の総和が 105,000 個で、1 つの synset に平均 1.76 個の単語が対応づいている。一方で、他の言語は synset によって、単語が対応づいていない synset がある。表内の

“単語が対応づいている synset の数”は各言語で差がある。また、同様に単語数も言語によって差がある。

本研究では、これらの WordNet と Open Multilingual WordNet を利用した概念レベルの文化差検出手法を提案する。

3.3.2 概念の範囲の定量化

概念レベルの文化差を検出するために各単語の概念の範囲を定量的に表す必要がある。そのために、WordNet と Open Multilingual WordNet の synset を基準に単語の概念の範囲を定量化する。

(3.1) 式のように、WordNet 内の synset と各 synset 間の is-a 関係をグラフ G で表す。

$$G = (V, E) \quad (3.1)$$

グラフ G はノードの集合 V とエッジの集合 E を持つ。グラフ G の各ノードは synset、各エッジは synset 間の is-a 関係とする。

$$V = \{v_1, v_2, v_3, \dots, v_l\} \quad (3.2)$$

このとき、is-a 関係には hypernym 用いて、下位から上位に向かう有向グラフとして定義する。例として、 v_i を上位の synset、 v_j を下位の synset とすると、エッジは (3.3) 式と表現する。

$$(v_j, v_i) \in E \quad (3.3)$$

また、synset は同義の単語を集めて定義されており、各 synset にはその概念を表現する単語が紐付いている。そのため、(3.4) 式のように単語の集合 W を定義する。図 3.1 はグラフ G のイメージ例で、それぞれの synset (v_i, v_j) は単語 $(w_k \sim w_m)$ を持っており、各 synset 間のエッジは hypernym で定義される。

$$W = \{w_1, w_2, w_3, \dots, w_n\} \quad (3.4)$$

ここで、(3.5) 式のようにある synset v_i を定義する。

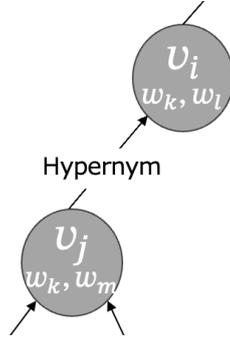


図 3.1: グラフ G での V と W のイメージ例

$$v_i \in V \quad (3.5)$$

(3.6) 式のように, synset v_i に含まれるある単語 w_k を定義する.

$$v_i \subset W, w_k \in v_i \quad (3.6)$$

次に, 単語 w_k が表現する概念の範囲を synset の集合 C_{w_k} として定義する. まず, C_{w_k} は v_i が w_k を含む ($w_k \in v_i$) ため, synset v_i を含む ($v_i \in C_{w_k}$). さらに, (3.7) 式で説明するように, C_{w_k} に含まれる各 synset v ($\forall v \in C_{w_k}$) の親ノード v_p , 子ノード v_c , 兄弟ノード v_b のうち, 単語 w_k が付随した synset は C_{w_k} に含まれる. この synset の集合 C_{w_k} を単語 w_k の表現する概念とし, 本論文ではこれを単語の**概念範囲**と定義する.

$$\begin{aligned} C_{w_k} = & C_{w_k} \cup \{v_p \mid v \in C_{w_k}, (v, v_p) \in E, w_k \in v_p\} \\ & \cup \{v_c \mid v \in C_{w_k}, (v_c, v) \in E, w_k \in v_c\} \\ & \cup \{v_b \mid v \in C_{w_k}, (v, v') \in E, (v_b, v') \in E, w_k \in v_b\} \quad (3.7) \end{aligned}$$

3.3.3 概念範囲の比較

異言語の単語同士で概念範囲を比較し, 概念範囲が異なる単語ペアを検出する. 同じ synset に含まれる単語同士は表現する概念が重なっているとし, 同じ synset に含まれる言語 l_1 の単語 $w_k^{l_1}$ と言語 l_2 の単語 $w_k^{l_2}$ の

概念範囲 ($C_{w_k^{l_1}}$ と $C_{w_k^{l_2}}$) を比較する. このとき, $C_{w_k^{l_1}} \neq C_{w_k^{l_2}}$ となるような $w_k^{l_1}$ と $w_k^{l_2}$ を概念範囲が違う単語ペアとして検出する. 日本語, 中国語, インドネシア語の3言語を対象に, それぞれの言語間で概念範囲の違う単語ペアを取得する. 次に, 検出される単語ペアを概念のトポロジーに基づいて7つのパターンに分類する. 分類を行うために, 言語 l_1 の単語 $w_k^{l_1}$ と言語 l_2 の単語 $w_k^{l_2}$ の概念範囲の共通集合 $ComSet$ を (3.8) 式と定義する.

$$ComSet(w_k^{l_1}, w_k^{l_2}) = C_{w_k^{l_1}} \cap C_{w_k^{l_2}} \quad (3.8)$$

次に, 単語 $w_k^{l_1}$ と $w_k^{l_2}$ の概念範囲の対称差集合 $DiffSet$ を (3.9) 式とする. 概念範囲が異なる単語ペアでは $C_{w_k^{l_1}} \neq C_{w_k^{l_2}}$ であるため, 対称差集合の要素は必ず存在する ($DiffSet \neq \emptyset$).

$$DiffSet(w_k^{l_1}, w_k^{l_2}) = (C_{w_k^{l_1}} - C_{w_k^{l_2}}) \cup (C_{w_k^{l_2}} - C_{w_k^{l_1}}) \quad (3.9)$$

(3.10) 式のように, 共通集合に含まれる synset と対称差集合に含まれる synset のうち少なくとも一つが兄弟関係の場合を水平タイプとする.

$$\begin{aligned} \forall v_x \forall v_y ((v_x \in ComSet(w_k^{l_1}, w_k^{l_2}) \wedge v_y \in DiffSet(w_k^{l_1}, w_k^{l_2})) \\ \Rightarrow \exists v (v \in V \wedge (v_x, v) \in E \wedge (v_y, v) \in E)) \end{aligned} \quad (3.10)$$

この水平タイプのうち, (3.11) 式を満たすものを水平 (部分共通) タイプと定義する.

$$((C_{w_k^{l_1}} - C_{w_k^{l_2}}) \neq \emptyset) \wedge ((C_{w_k^{l_2}} - C_{w_k^{l_1}}) \neq \emptyset) \quad (3.11)$$

図3.2に示すのが水平 (部分共通) タイプの例である. 言語 l_1 の単語 $w_k^{l_1}$ の概念範囲 $C_{w_k^{l_1}}$ は, synset v_2 と synset v_3 を要素に持っており, ノード v_2 とノード v_3 を囲む輪は概念範囲 $C_{w_k^{l_1}}$ を表している. 言語 l_2 の単語 $w_k^{l_2}$ の概念範囲 $C_{w_k^{l_2}}$ は synset v_3 と synset v_4 を持っており, v_3 と v_4 を囲む輪が $C_{w_k^{l_2}}$ を表現している. 図3.2の場合, $C_{w_k^{l_1}} = \{v_2, v_3\}$ と $C_{w_k^{l_2}} = \{v_3, v_4\}$ の共通集

合は $ComSet(w_k^{l_1}, w_k^{l_2}) = \{v_3\}$, 対称差集合は $DiffSet(w_k^{l_1}, w_k^{l_2}) = \{v_2, v_4\}$ となる. このように, 水平 (部分共通) タイプでは共通集合の要素と対称差集合の要素は兄弟関係にあり, かつそれぞれの差集合は要素を持っている ((3.10) 式と (3.11) 式を満たす). ここでそれぞれの単語 ($w_k^{l_1}$ と $w_k^{l_2}$) は表現する概念の一部が共通する (ノード v_3) が, それぞれの単語独自の概念 (ノード v_2 やノード v_4) を有しているため部分共通である.

水平タイプのうち, (3.12) 式を満たすものを水平 (包含) タイプとする. 図 3.3 が水平 (包含) タイプの例で, 単語 $w_k^{l_1}$ の概念範囲 $C_{w_k^{l_1}} = \{v_2, v_3\}$ に対して, 単語 $w_k^{l_2}$ の概念範囲 $C_{w_k^{l_2}} = \{v_2, v_3, v_4\}$ である. この時, $C_{w_k^{l_2}}$ が $C_{w_k^{l_1}}$ を包含している ($C_{w_k^{l_2}} \supset C_{w_k^{l_1}}$) ため, $w_k^{l_1}$ と比べて $w_k^{l_2}$ は広義な意味を表現している.

$$((C_{w_k^{l_1}} - C_{w_k^{l_2}}) = \emptyset) \vee ((C_{w_k^{l_2}} - C_{w_k^{l_1}}) = \emptyset) \quad (3.12)$$

一方で, (3.13) 式を満たすような場合を垂直タイプと分類する. 垂直タイプでは, 共通集合に含まれる synset と対称差集合に含まれる synset に少なくとも 1 つの親子関係が存在する.

$$\begin{aligned} \exists v_x \exists v_y ((v_x \in ComSet(w_k^{l_1}, w_k^{l_2}) \wedge v_y \in DiffSet(w_k^{l_1}, w_k^{l_2})) \\ \Rightarrow ((v_x, v_y) \in E \vee (v_y, v_x) \in E)) \quad (3.13) \end{aligned}$$

さらに, (3.13) 式を満たし, かつ (3.14) 式を満たすもの垂直タイプを垂直 (下位包含) タイプとする. 垂直 (下位包含) タイプでは, 図 3.4 のように, 対称差集合に含まれる synset が共通集合に含まれる synset の上位 synset になる.

$$\begin{aligned} \forall v_x \in ComSet(w_k^{l_1}, w_k^{l_2}), \forall v_y \forall v_z (((v_x, v_y) \in E \wedge (v_z, v_x) \in E) \\ \Rightarrow (v_y \in (ComSet(w_k^{l_1}, w_k^{l_2}) \cup DiffSet(w_k^{l_1}, w_k^{l_2})) \\ \wedge v_z \notin DiffSet(w_k^{l_1}, w_k^{l_2}))) \quad (3.14) \end{aligned}$$

(3.15) 式を満たすような垂直タイプを分類する. 対称差集合に含まれる synset が共通集合に含まれる synset の下位 synset になるもの ((3.15)

式を満たす)のうち,それぞれに独自の概念範囲の要素を持つもの((3.11)式を満たす)を垂直(上位共通)タイプとする. 図 3.8 は垂直(上位共通)タイプの例である. 一方, 図 3.5 の例のような, (3.11)式を満たさないものを垂直(上位包含)タイプとする.

$$\begin{aligned} \forall v_x \in \text{ComSet}(w_k^{l_1}, w_k^{l_2}), \forall v_y \forall v_z ((v_x, v_y) \in E \wedge (v_z, v_x) \in E) \\ \Rightarrow (v_y \notin \text{DiffSet}(w_k^{l_1}, w_k^{l_2}) \wedge \\ v_z \in (\text{ComSet}(w_k^{l_1}, w_k^{l_2}) \cup \text{DiffSet}(w_k^{l_1}, w_k^{l_2}))) \quad (3.15) \end{aligned}$$

次に, (3.16)式のように差部分が共通部分に対して, 少なくとも1つの親関係と少なくとも1つの子関係が存在するものを説明する. このうち, 図 3.6 のように (3.11)式を満たすものを垂直(部分共通)タイプとする. 一方で, 図 3.7 のように (3.12)式を満たすものを垂直(包含)タイプと定義する.

$$\begin{aligned} \exists v_w \exists v_x \exists v_y \exists v_z ((v_w, v_x \in \text{ComSet}(w_k^{l_1}, w_k^{l_2}) \\ \wedge v_y, v_z \in \text{DiffSet}(w_k^{l_1}, w_k^{l_2})) \Rightarrow ((v_w, v_y) \in E \wedge (v_z, v_x) \in E)) \quad (3.16) \end{aligned}$$

これらの7つの分類のうち, 包含関係では一方の単語の概念範囲がもう一方の単語の概念範囲よりも広義である. そのため, 団子と dumpling のようにある言語からもう一方の言語に翻訳する場合にはコミュニケーションに齟齬が起こる可能性があるが, 反対の向きの翻訳では齟齬が起こらない可能性が高い. 一方, 部分共通や上位共通では表現する概念が一部では共通しているが, どちらか一方の単語しか表現できない概念がある. そのため, どちらの言語からの翻訳においても会話に齟齬が生まれる可能性がある.

3.4 人手による概念レベルの文化差の判定

提案手法の精度を評価するために, 人手で正解ラベルを作成する. そのため, アンケートを用いた概念レベルの文化差を判定する方法を説明する.

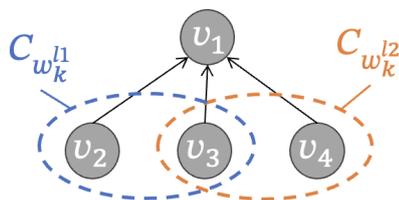


図 3.2: 水平 (部分共通)

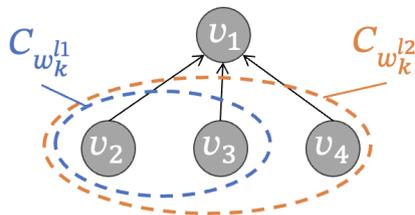


図 3.3: 水平 (包含)

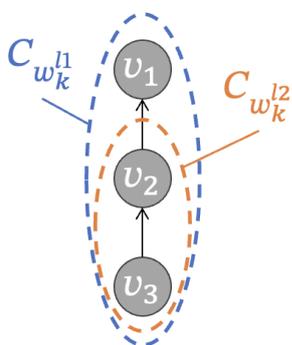


図 3.4: 垂直 (下位包含)

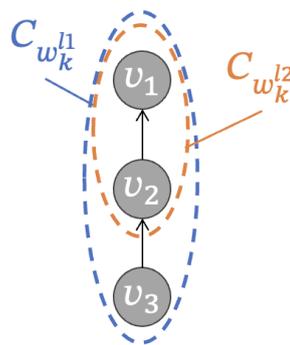


図 3.5: 垂直 (上位包含)

3.4.1 単語三つ組の作成

提案手法を用いて、日本語-中国語-インドネシア語のそれぞれの言語間で概念範囲の異なる単語ペアを検出した。その結果、日中で概念範囲を比較した単語ペア 236,590 個のうち 27,005 個の単語ペアで概念範囲が異なった。同様に、日尼で単語ペア 398,048 個のうち 60,581 個、中尼で単語ペア 111,450 個のうち 14,175 個の単語ペアで概念範囲が異なった (表 3.3)。

次に、表 3.3 の単語ペアから単語三つ組を作成する。単語三つ組とは日本語の単語と中国語の単語、インドネシア語の単語それぞれ一つずつの組み合わせである。ある日本語の単語 w^{ja} 、中国語の単語 w^{zh} 、インドネシア語の単語 w^{id} とした時、 $(w^{ja}-w^{zh}-w^{id})$ とする。これらは検出

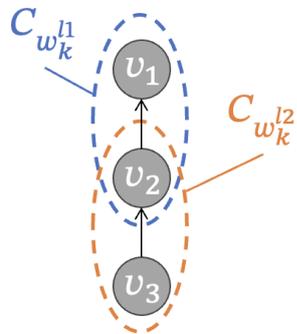


図 3.6: 垂直 (部分共通)

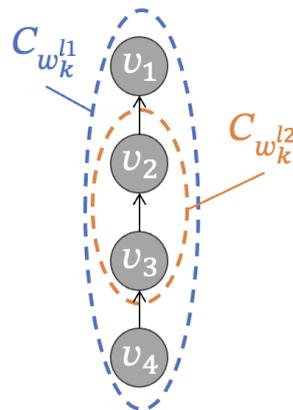


図 3.7: 垂直 (包含)

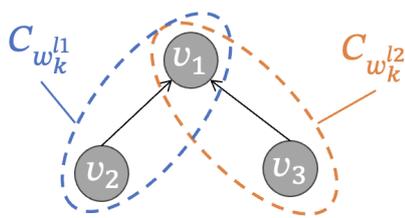


図 3.8: 垂直 (上位共通)

された単語ペアに対して、2段階でフィルタリングし、残った単語ペアから単語三つ組を作成する。1つ目のフィルタリングでは、人手で概念差を判定しづらい抽象的な概念を省く。具体的には、グラフ G の木構造の最上位に `entity` という英単語が対応づいている synset がある。この synset を `Synset("entity")` としたとき、`Synset("entity")` に対して子関係にある synset が 2 つ存在する。それぞれ対応づいている英単語から `Synset("physical entity")`, `Synset("abstraction")` とし、これらの synset のうち、`Synset("abstraction")` の下位から検出された単語ペアを対象外とする。図 3.9 は `Synset("abstraction")` の日本語 WordNet での情報である。それぞれ、英語と日本語でこの概念を表す単語と概念の説明が表記されている。`Synset("abstraction")` は抽象的なものを定義した synset であり、人

表 3.3: 検出された単語ペア数

言語	総単語ペア数	概念範囲が異なる単語ペア
ja - zh	236,590	27,005
ja - id	398,048	60,581
zh - id	111,450	14,175
単語三つ組	11,571	1,375

※ja: 日本語, zh: 中国語, id: インドネシア語

00002137-n 'a general concept formed by extracting common features from specific examples';

English Japanese

English	abstraction (▶), abstract entity	📄
Japanese	抽象的実体	📄

Definitions

English
a general concept formed by extracting common features from specific examples

Japanese
特定の例から共通点を抽出することによって形成された一般概念

図 3.9: 日本語 WordNet における Synset(“abstraction”) の情報

手で概念差を判定することは同じ言語の話者達でも評価がぶれやすいため、今回は Synset(“abstraction”) の下位で検出されたものは対象外とした。

2つ目のフィルタリングでは、単語ペアの対訳関係を確認する。機械翻訳が翻訳結果として表示する可能性がある単語ペアに絞るため、各単語を機械翻訳で翻訳した時にいずれかの方向で対訳関係になったペアのみを対象とする。具体的には、単語 $w_k^{l_1}$ と単語 $w_k^{l_2}$ では、 $w_k^{l_1}$ を機械翻訳した結果が $w_k^{l_2}$ になる。もしくは、 $w_k^{l_2}$ を翻訳した結果が $w_k^{l_1}$ になる場合を対訳関係が成り立つ単語ペアとする。

フィルタリングで対象外となった単語ペアを除いた、日中、日尼、中尼の単語ペアから日中尼の単語三つ組を作成する。日中、日尼、中尼の2つの言語ペアにおいて、共通する単語を持つ単語ペアを取得する。それら

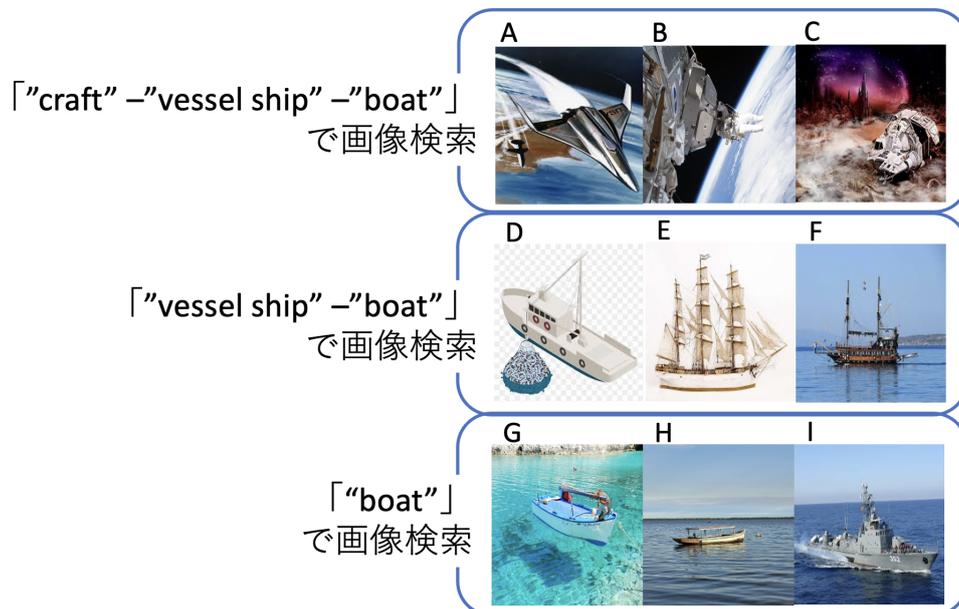


図 3.10: アンケートに使用された画像とクエリの例

2つの単語ペアに含まれる日本語, 中国語, インドネシア語の単語の概念範囲を確認する. 3つの概念範囲の共通集合に要素がある場合, その3つの単語を単語三つ組とする. 例えば, 日中で w_k^{ja} と w_k^{zh} の単語ペア, 日尼で w_k^{ja} と w_k^{id} の単語ペアがあり, それぞれの概念範囲 $C_{w_k^{ja}}, C_{w_k^{zh}}, C_{w_k^{id}}$ とする. これらの概念範囲の共通集合が要素を持つ ($C_{w_k^{ja}} \cap C_{w_k^{zh}} \cap C_{w_k^{id}} \neq \emptyset$) 場合, $w_k^{ja}, w_k^{zh}, w_k^{id}$ を単語三つ組とする. 各言語間で概念範囲を比較した単語ペア (表 3.3 の総単語ペア数) から単語三つ組を作成した結果, 11,571 件の三つ組が作成された. そのうち, 1,375 件の単語三つ組は概念範囲が異なる単語ペアから作成された三つ組である.

3.4.2 アンケートの設計

単語三つ組1つにつき, 1タスクのアンケートを行う. アンケートでは, 単語と一致する画像を画像群から選択するタスクを行なった. この時, 与える単語は単語の三つ組の中のひとつで, 被験者の母語の単語を用

いる。実験では、日本語話者、中国語話者、インドネシア語話者を被験者とし、アンケートを行った。言語間でアンケートの回答に差があるかを確認した。

アンケートの画像群は、単語三つ組の3つの単語の概念範囲に含まれる synset を用いて取得される画像を使用する。具体的には、単語三つ組を $(w_k^{ja}, w_k^{zh}, w_k^{id})$ としたとき、概念範囲の和集合 $C_{w_k^{ja}} \cup C_{w_k^{zh}} \cup C_{w_k^{id}}$ の要素の synset からランダムで英語の単語を検出し、その単語から画像検索のクエリを作成した。クエリは上位概念と下位概念の違いをつけるために、和集合 $(C_{w_k^{ja}} \cup C_{w_k^{zh}} \cup C_{w_k^{id}})$ の synset の関係性を元に上位概念の単語から下位概念の単語の要素を取り除く。そのように、単語を組み合わせた AND 検索のクエリを作成した。例として、単語三つ組を (ボート,

表 3.4: (ボート, 船, kapal) の synset とクエリ

synset	ランダムで取得した単語	生成されたクエリ
v_1	craft	“craft” -“vessel ship” -“boat”
v_2	vessel ship	“vessel ship” -“boat”
v_3	boat	“boat”

船, kapal), これらの単語の概念範囲の和集合を $\{v_1, v_2, v_3\}$ とする。このとき、 v_1, v_2, v_3 に含まれる英語の単語からランダムで1単語ずつ抽出する。表 3.4 は各 synset と単語とクエリの例である。表 3.4 のように、それぞれの synset からランダムで取り出した単語が v_1 :craft, v_2 :vessel ship, v_3 :boat である。 v_1, v_2, v_3 の順に親関係 $((v_2, v_1) \in E) \wedge ((v_3, v_2) \in E)$ が成り立つとしたとき、表 3.4 のように、 v_1 を表すクエリは「“craft” -“vessel ship” -“boat”」、 v_2 を表すクエリは「“vessel ship” -“boat”」、 v_3 を表すクエリは「“boat”」となる。クエリに含まれる“-”は NOT を示し、それらの要素を取り除いている。作成したクエリで画像検索を行い、取得された上位3件ずつの画像(図 3.10)をアンケートで表示する。

3.4.3 アンケートに基づく文化差判定

言語ごとにアンケートの回答を集計し、言語間で回答に差がある場合を文化差ありと判定する。具体的には、画像検索に用いたクエリごとに集計表を作成し、各集計表にフィッシャーの独立性検定を行う。いずれかの集計表で統計的に有意な差があったタスクを文化差ありと判定する。例として、(ボート, 船, kapal) の三つ組で日本語と中国語の文化差を判定する場合、「“craft” -“vessel ship” -“boat”」, 「“vessel ship” -“boat”」, 「“boat”」のクエリごとに日本語話者と中国語話者の回答を集計する。回答者がその synset を表現する画像の1枚以上がキーワードに一致すると判断したか否かを基準に、各言語で“一致する画像あり”と判断した人と“一致する画像なし”と判断した人の人数を集計する。図 3.10 では、synset v_1 の画像はクエリ: 「“craft” -“vessel ship” -“boat”」で取得された図 3.10 内の画像 A, 画像 B, 画像 C である。同様に、synset v_2 の画像はクエリ: 「“vessel ship” -“boat”」で取得された(図 3.10 内の)画像 D, 画像 E, 画像 F, synset v_3 の画像はクエリ: 「“boat”」で取得された(図 3.10 内の)画像 G, 画像 H, 画像 I である。このタスクでは、 v_1, v_2, v_3 の順に親関係のため、synset v_1 の下位クラスに v_2, v_2 のさらに下位クラスに v_3 がある。そのため、 v_1 を表現する画像は、画像 A から画像 I までの全ての画像と判断する。このとき、各クエリから一枚以上の画像を選択し、かつ全てのグループで選択されている(3.17 式を満たす)場合を“一致する画像あり”と判定する。

画像の集合を以下のように定義する.

- 「“craft” -“vessel ship” -“boat”」で取得された画像の集合を S_1
- 「“vessel ship” -“boat”」で取得された画像の集合を S_2
- 「“boat”」で取得された画像の集合を S_3

$$\exists x \in S_1, \exists y \in S_2, \exists z \in S_3 \quad (3.17)$$

$$\exists y \in S_2, \exists z \in S_3 \quad (3.18)$$

$$\exists z \in S_3 \quad (3.19)$$

同様に, v_2 を表現するのは画像 D から画像 I までの全ての画像で, 3.18 式を満たす場合を“一致する画像あり”と判定する. v_3 を表現するのは, 画像 G, 画像 H, 画像 I とし, 3.19 式を満たす場合を“一致する画像あり”と判定する. このように, それぞれの synset のクエリで取得された画像と, その下位 synset のクエリで取得された画像から, それぞれ一枚以上選択されたかを基準に集計表を作成する.

表 3.5 は (ボート, 船, kapal) のタスクの日本語話者と中国語話者の解答例である. この回答の場合, (ボート, 船, kapal) のタスクの「“craft” -“vessel ship” -“boat”」の集計表は表 3.6, 「“vessel ship” -“boat”」の集計表は表 3.7, 「“boat”」の集計表は表 3.8 になる. 作成された 3 つの集計表それぞれにフィッシャーの独立性検定で日本語話者と中国語話者の回答に統計的な有意差があるかを確認する.

提案手法で概念差があると判定された単語ペアから作成された単語三つ組 1,375 個 (表 3.3 参照) を分析データとし, そのうち, 何件が実際に人手で概念差を感じるかを判定した. 計 15 人 (日本語話者:5 人, 中国語話者:5 人, インドネシア語話者:5 人) の被験者による実験の結果, 日本語-中国語間で 528 件, 日本語-インドネシア語間で 476 件, 中国語-イン

表 3.5: (ボート, 船, kapal) のアンケートの解答例

	(日本語話者) 回答	(中国語話者) 回答
回答者 1	A,B,G,H,I	G,H,I
回答者 2	C,G	H,I
回答者 3	なし	I
回答者 4	D,F	C,D,G,H,I
回答者 5	G,H,I	G,H

ドネシア語間で 471 件が人手で文化差あり判定となった。

3.5 評価

提案手法を正解率, 適合率, 再現率, F 値の 4 つの指標を用いて評価を行う。比較対象として, 吉野らの Wikipedia を用いた文化差検出手法 (以下, 既存手法) をベースラインとして用いる [39]。吉野らの手法は Wikipedia の記事内の内容やカテゴリから文化差を判定するもので, 日本か中国どちらか一方にしかない文化を第一種の文化差, 一方で, 両方の文化に存在するがそれぞれ別のものを指している場合を第二種の文化差と定義し, カテゴリや記述内容によってどちらの文化差かを判定している。表 3.3 の全ての単語三つ組 11,571 件からランダムで選ばれた 100 件を評価データセットとして用いる。評価データセットには人手で文化差判定した結果を正解ラベルとして付与する。

表 3.9 が提案手法と既存手法の比較結果である。日中間では, 提案手法が正解率・適合率・再現率・F 値の全ての指標で既存手法を上回った。この時, 評価データ (文化差なし:85 件, 文化差あり:15 件) に対して, 提案手法では 100 件中 30 件, 既存手法では 100 件中 18 件のデータが文化差ありと判定された。

日尼間でも提案手法が全ての指標で既存手法を上回った。評価データ (文化差なし:75 件, 文化差あり:25 件) に対して, 提案手法では 100 件中 45 件, 既存手法では 100 件中 20 件のデータが文化差ありと判定された。

表 3.6: 表 3.5 の解答例から作成される 「“craft” -“vessel ship” -“boat”」の集計表

	キーワードと一致する画像あり	キーワードと一致する画像なし
日本語話者	0	5
中国語話者	1	4

表 3.7: 表 3.5 の解答例から作成される 「“vessel ship” -“boat”」 の集計表

	キーワードと一致する画像あり	キーワードと一致する画像なし
日本語話者	1	4
中国語話者	1	4

表 3.8: 表 3.5 の解答例から作成される 「“boat”」 の集計表

	キーワードと一致する画像あり	キーワードと一致する画像なし
日本語話者	3	2
中国語話者	5	0

中尼間では, 提案手法が適合率・再現率・F 値の指標で既存手法を上回ったが, 正解率では既存手法を下回った. この時, 評価データ (文化差なし:83 件, 文化差あり:17 件) に対して, 提案手法では 100 件中 38 件, 既存手法では 100 件中 18 件のデータを文化差ありと判定した. いずれの言語間でも提案手法は既存手法と比べて, 適合率, 再現率, F 値が高かった. 特に, 正解率から, 提案手法は 69~81%の確率で人と同程度の文化差判定ができた. 一方で, 適合率は 32~44%と低かったが, 再現率は 71~87%と高水準であった. このことから, 齟齬の原因となる文化差を見逃す可能性が低いため, 本研究の目的とするコミュニケーションの齟齬を解消するために, 提案手法は有効であると言える.

表 3.9: 既存手法との比較結果

言語	手法	正解率	適合率	再現率	F 値
ja - zh	提案手法	0.81	0.43	0.87	0.58
	既存手法	0.67	0.0	0.0	0.0
ja - id	提案手法	0.70	0.44	0.80	0.57
	既存手法	0.63	0.20	0.16	0.18
zh - id	提案手法	0.69	0.32	0.71	0.44
	既存手法	0.77	0.13	0.06	0.08

※ja: 日本語, zh: 中国語, id: インドネシア語

3.6 分析

3.6.1 既存手法との比較

分析として、文化差があるデータに対して、提案手法は判定に成功したが、既存手法では判定に失敗したケースを挙げる。(ボート, 船, kapal) の日中では“ボート”の概念範囲は小舟を表現する synset のみであったが、“船”は小舟の他に船や乗り物全般を指す synset も概念範囲に含むため、提案手法では文化差あり判定であった。一方で、既存手法が判定失敗した理由として、(ボート, 船) のような文化的に特徴があまりないものが対象だったことが考えられる。既存手法では判定のプロセスに Wikipedia のカテゴリや記事などにどちらかの国名が含まれているかといった項目があり、より文化や国独自のものかを考慮して文化差を検出している。そのため、既存手法では(ボート, 船)は国独自のものではないため、文化差ありと判定できなかったのに対して、本研究は単語の概念範囲の違いを対象としていることにより、提案手法は有効だったと考えられる。

表 3.10: (日中) 文化差の分類とデータの内訳

文化差の分類	ペア数	データの内訳	
		文化差なし	文化差あり
水平 (部分共通)	43	12 (-)	31 (+)
水平 (包含)	637	410	226
垂直 (下位包含)	280	169	111
垂直 (上位包含)	634	414	220
垂直 (部分共通)	26	19	7
垂直 (包含)	27	10 (-)	17 (+)
垂直 (上位共通)	62	30 (-)	32 (+)
相等	24	16	8

3.6.2 概念レベルの文化差の分類と精度の関連性

3.3.3 項で説明した概念レベルの文化差のパターンに基づいて分析用データを分類し、種類ごとに検出精度を分析する。表 3.10 は分析用データセットの日中間での分類ごとのデータの内訳である。文化差の分類の“相等”とは日中間では概念範囲に差がなかったが、他の言語間では概念範囲に差があったデータである。一方、他の 7 タイプは提案手法で文化差ありと判定したものである。分類と検出精度に関連があるかを χ^2 検定を用いて独立性を調べた。この時、7 つの分類ごとデータの内訳で集計表を作成し検定を行なった。その結果、 p 値 < 有意水準 0.05 であったため文化差の分類によって検出精度に有意な差があった。さらに、残差分析を行いどの分類に有意な差があったかを明らかにする。調整済み残差の絶対値が 1.96 を上回っていた部分を表 3.10 内に (+), (-) をつけて表記している。残差分析の結果、水平 (部分共通) タイプ、垂直 (包含) タイプ、垂直 (上位共通) タイプでは文化差の検出精度が高いと言える。

表 3.11 は分析用データセットの日尼間で分類したデータの内訳である。同様に、 χ^2 検定を用いて分類と検出精度に関連があるかを確認する。その結果、 p 値 < 有意水準 0.05 であったため、検出精度に有意な差があった。さらに残差分析を行った結果、水平 (部分共通) タイプと垂直 (上位共

表 3.11: (日尼) 文化差の分類とデータの内訳

文化差の分類	ペア数	データの内訳	
		文化差なし	文化差あり
水平 (部分共通)	105	53(-)	52(+)
水平 (包含)	758	497	261
垂直 (下位包含)	248	150	98
垂直 (上位包含)	575	387	188
垂直 (部分共通)	62	38	24
垂直 (包含)	52	39	13
垂直 (上位共通)	111	61(-)	50(+)
相等	13	10	3

通) タイプは文化差の検出精度が高いと言える。

表 3.12 は分析用データセットの中尼間で分類したデータの内訳である。日中, 日尼と同様に χ^2 検定を用いて分類と検出精度に関連があるかを確認した結果, p 値 < 有意水準 0.05 であったため分類によって検出精度に有意な差があった。残差分析の結果, 水平 (部分共通) タイプでは検出精度が高かったが, 一方で, 垂直 (包含) タイプでは検出精度が低かったと言える。

χ^2 検定の結果から, 日中, 日尼, 中尼の全言語間で分類と検出精度に関連があることがわかった。また, どの言語間においても水平 (部分共通) タイプは他のタイプに比べて検出精度が高いことがわかった。水平 (部分共通) タイプの例として, (容器, 容器, bekas) の単語三つ組がある。日中間 (容器, 容器) では, 日本語話者と中国語話者のどちらもペットボトルや瓶などの画像をアンケートで選んでいた。しかし, 日本語話者はさらにお椀やゴミ箱など身の回りにある容れ物の画像も選んでいた。一方, 中国語話者は瓶などに加えてコンテナ船の大型コンテナの画像を選択していた。そのため, 日本語話者と中国語話者の回答に大きく差があった。このように, 水平 (部分共通) タイプでは, それぞれの概念範囲に一方しか表現しない概念があり, それぞれの言語話者で回答に差が生まれたため, 人手で文化差ありと判定されたデータが多かったと考えられる。

表 3.12: (中尼) 文化差の分類とデータの内訳

文化差の分類	ペア数	データの内訳	
		文化差なし	文化差あり
水平 (部分共通)	81	35 (-)	46 (+)
水平 (包含)	758	506	252
垂直 (下位包含)	278	179	99
垂直 (上位包含)	621	402	219
垂直 (部分共通)	59	33	26
垂直 (包含)	89	79 (+)	10 (-)
垂直 (上位共通)	109	77	32
相等	17	12	5

しかし、同じく部分共通の垂直 (部分共通) タイプでは検出精度に有意な差が見られなかった。例として、(石ころ, 礫石, kelitir) という三つ組では、(石ころ, 礫石) がこのタイプに該当する。“石ころ”は小石や石のかけらに加えて、上位クラスの鉱物などを表現する概念を持っている。一方で、“礫石”は小石や石のかけらとその下位クラスの砂利を表現する概念を持っていた。アンケートでは中国語話者が小石や砂利の画像を選択するのに対し、日本語話者は岩や小石に加えて砂利など鉱物の画像を幅広く選択していた。これは“礫石”のみが表現するはずの砂利の概念が“石ころ”の小石や鉱物の概念の下位クラスになっているため、先ほどの(容器, 容器)と比べて回答に差が生まれづらかったと考えられる。同様に他の概念差のタイプでもどちらかの概念範囲がもう一方の概念範囲を包含しているため、回答に差が生まれづらかったと考えられる。これらによって、水平 (部分共通) タイプは他の概念差タイプに比べて、文化差の検出精度が高かったと考えられる。これらのことから、概念が包含されているタイプがあり、今回の実験ではそれらが文化差として人に判定されなかったため、提案手法は適合率が低かったと考えられる。

表 3.13: 出現頻度上位 10,000 件の名詞への文化差検出結果

言語	単語数	文化差検出できた単語数	割合
ja - zh	10,000	3,911	39.11%
ja - ind		3,991	39.91%

※ja: 日本語, zh: 中国語, id: インドネシア語

3.6.3 コーパスを対象とした概念レベルの文化差検出の適用

大規模コーパスを用いて、概念レベルの文化差検出手法がどの程度適用できるかを確認した。コーパスは Wikipedia から取得した 17,478,096 文の日本語の文章をもとに作成した。本研究では、コーパスの文章を MeCab によって形態素で分割し、出現頻度が上位 10,000 に入る名詞の単語を対象とする。対象とした単語から、概念レベルの文化差検出手法で検出された単語を分析する。コーパス内の出現頻度上位 10,000 の名詞に対して、文化差検出を行った結果が表 3.13 である。このように日本語と中国語間で 10,000 単語中、3,911 個の単語で概念レベルの文化差があった。

次に、検出された単語を分析していく。“寺”は日本語と中国語間で概念レベルの文化差がある単語である。寺を表現する概念は、一般的な寺を表す synset に対応しており、synset の説明は“神聖な礼拝のための建築物からなる崇拝の場所”である。一方で、中国語では一般的な寺の他にも、中国固有の寺の synset を含めて表す。この synset の定義文は“偶像をあげめるための中国の寺院”である。

“ボート”は日本語とインドネシア語間で概念レベルの文化差がある単語である。ボートを表現する概念は、“水上に行くための小さい船”である。一方で、インドネシア語では、全ての乗り物を表現する広義な単語を指している。そのため、日本語とインドネシア語で文化差が存在する。

3.7 まとめ

対訳関係にある単語間でも言語ごとに表現する概念が異なる場合があるため、多言語コミュニケーションに齟齬を生む可能性がある。そのため、本研究では WordNet と Open Multilingual WordNet を用いて単語が表現する概念を定量化し、対訳ペア間で概念範囲を比較することで概念レベルの文化差を検出した。WordNet 上の日中尼の3言語間で、同一 synset に対応づけられた異言語の単語を単語ペアとし、概念範囲の異なる単語ペアを取得した。その結果、236,590(日中)、398,048(日尼)、111,450(中尼)個の単語ペアのうち、27,005(日中)、60,581(日尼)、14,175(中尼)個の単語ペアで文化差が検出された。また、ランダムで取り出した100件の単語三つ組を用いて、既存手法と提案手法の精度比較を行った。その結果、適合率、再現率、F値で提案手法が既存手法を上回った。さらに、提案手法の検出精度を分析するために、概念の範囲の差をトポロジーで分類し、分類の種類ごとに検出精度を統計検定を用いて比較した。その結果、日中、日尼、中尼の言語間で分類の種類によって検出精度に有意な差があった。

今後の改善として、さらなる精度向上のために、WordNet の synset 間の距離を考慮することが考えられる。WordNet では、英単語をベースに英語の概念体系を表しているため、synset 間の距離が定義されていない。そのため、hypernym で対応づけられた synset 間でも、synset によって概念の粒度が異なる。それによって、概念レベルの文化差がある場合でも、人の評価では同じものを表現していると捉えられた可能性があるため、概念間の距離を定義することでより人の判断に近い文化差検出ができると考えられる。

第4章 想起イメージレベルの文化差検出

4.1 単語によって想起されるイメージ

人が想起するイメージは、その人の文化背景や生活環境に影響を受ける。コミュニケーションでは、話者同士の共通認識や共有できるものによって、想起するイメージが形成される。例えば、学生時代の同級生同士の会話で、学生時代の昼食の話題になれば、お互いが持つ学生時代の昼食の記憶から想起イメージは形成される。同様に、同じ文化背景の話者同士では想起するイメージは似たものになる。“ごぼう”と“burdock”の例では、それぞれの単語から想起するイメージは異なる。これは、日本人は、この植物の根っこの部分を食べる文化があり、普段目にするものが根っこの見た目の食材である。日本人はごぼうから burdock で想起される花のイメージを想起する人は少なく、日本人同士ならごぼうから想起するイメージは似たようなものになる。一方で、アメリカ人はこの植物の根を食べない。そのため、根っこの見た目よりも花の方が印象が強いため、花のイメージになる。このように、同一の文化背景を持つ人同士の会話では、同じ単語から想起するイメージは似たようなものになるが、異なる文化背景を持つ人同士の会話において、想起するイメージが違いくる。

このように、多言語コミュニケーションにおいて、翻訳関係にある単語においても話者によって想起するイメージが異なる。このようなイメージレベルの文化差を検出するために、以下のような課題がある。

想起イメージの取得

単語から想起するイメージは、抽象的なものである。それらを比較

するために、各話者が想起するイメージを定量化する必要がある。

文化差の判断基準となる閾値の設定

想起するイメージを定量化できた場合でも、そこから文化差を判定するための基準が必要である。

私はこのような課題を解決するために、単語から画像検索で取得できる Web 上の画像を想起イメージと捉え、画像の類似度に基づいて検出手法を提案する。

4.2 画像の類似性の算出

柳生らは、単語の表現する概念が、どれだけ視覚的特徴と関連しているかを調査している [106]. この視覚的特徴が関連しているかを、視覚性と定義し、150 個の形容詞に対して視覚性を求めた. Rogowitz らは人は何を基準に画像を類似していると判定しているかを分析している [107]. 画像の類似性の評価において、人の認識ではまず色の類似性が重要であることを述べている. Wang らは画像類似性モデルを作成するための、学習方法を提案している [108]. 画像データベースでは、画像のデータ構造が小さいコンパクトであるため、画像の類似性を図ることが難しい. Lv らは、このようなデータ構造の画像の類似性を計算する手法を提案している [109]. Hull らは文章が載った画像の類似性を計算する手法を提案している [110].

これらの研究では、画像の類似性を図る手法や基準の提案している. 本研究では、画像から取り出した特徴ベクトルのコサイン類似度に基づいて、画像の類似性を測定する。

4.3 想起イメージレベルの文化差検出手法

本研究では、想起イメージレベルの文化差を検出するために Web 上の画像の類似性に基づいてイメージの違いを判定する。

4.3.1 画像認識技術

最初に, 画像の類似性を検出するために用いる画像認識技術について述べる. Simonyan らが発表した論文では, 大規模画像認識モデルを提案している [111]. ここでは, 画像を認識して 1000 クラスのラベルに分類する VGG16, VGG19 を提案している.

He らは, ResNet という残差を用いた画像認識を提案している [112]. Theckedath らの研究では, 怒り, 軽蔑, 嫌悪, 恐怖, 幸福, 悲しみなどの感情検出を行っている [113]. VGG16 と ResNet50, SE-ResNet50 の 3 つの手法で, 精度を判定したところ, それぞれ VGG16:96.8%, ResNet50:99.47%, SE-ResNet50:97.34% であった. Mascarenhas らの研究では, VGG16, VGG19, ResNet50 の 3 つのモデルで, 画像認識の精度比較を行っている [114]. 比較の結果, VGG16:96.67%, VGG19:97.07%, ResNet50:97.33% で, ResNet50 が最も高かったと述べている.

これらの研究では, 画像認識のモデルを提案したり, その精度比較を行っている. 本研究は, 先ほど述べた VGG16 を用いて, 画像から特徴ベクトルを抽出し, その差を元にイメージの文化差を検出する手法を提案している.

4.3.2 特徴ベクトルに基づく類似度算出

文化差を検出するために, WordNet と Multilingual WordNet を用いる.

WordNet 上で各概念は同義語の集合である synset で定義され, 一つの synset に複数の単語が割り当てられている. 特に, 英語以外の言語では, 英語の単語と比較して, 対応づいている単語の数が多い傾向にあり, これは英語の概念体系に単語を半自動的に割り当てているためである.

まず, synset の単語群を用いて, 画像検索を行い, 取得された上位 10 件の画像を取得する. この取得された画像群をその単語から想起されるイメージとする. 例えば, ごぼうを表現する Synset(“burdock”) では, 日本語と英語の情報として, “ゴボウ, 牛蒡” と “burdock, clotbur” の単語群とその synset を表現する説明が対応づけられている (表 4.1). これらの単語群から, 画像検索のクエリを作成し, 画像を取得する. 表 4.2 は, 画像検索のクエリ生成の例である. 表 4.1 の synset から取得された単語群“

表 4.1: Synset(“burdock”) の日本語と英語の情報

言語	単語群	synset の説明
日本語	ゴボウ, 牛蒡	ユーラシアの温帯地域の直立性の2年生草本の総称で, しっかりした主根を持ち, いがをつける
英語	burdock, clotbur	any of several erect biennial herbs of temperate Eurasia having stout taproots and producing burs

表 4.2: Synset(“burdock”) の単語群とその画像検索のクエリ

言語	対応づいている単語群	生成されたクエリ
英語	burdock, clotbur	“burdock” “clotbur”
日本語	ゴボウ, 牛蒡	“ゴボウ” “牛蒡”

ゴボウ”と“牛蒡”のように複数の単語が対応づいている場合, これらの単語を AND 検索とすることで対象概念を表す画像群を収集する. この場合, 画像検索のクエリは「“ゴボウ” “牛蒡”」となる. このように, 各言語でクエリを作成し, 上位 10 件ずつの画像を取得する.

次に, 2 言語間の画像の類似性を比較するために, Keras の VGG16[‡]を用いて各画像から画像の特徴ベクトルを抽出する. VGG16 は, 畳み込み層 13 層と全結合層 3 層からなる画像分類モデルで, 畳み込み層で特徴ベクトルを抽出し, 全結合層でそれらの特徴ベクトルを重ね合わせ, 1000 次元の分類ラベルのいずれかに分類している. そのため, 画像の特徴ベクトルのみを抽出するために, 畳み込み層 13 層のみを用いて, 画像の特徴ベクトルを抽出する. さらに, 抽出されたベクトルの形は $4 \times 4 \times 512$ 次元になっているため, それを後の計算のために 1×8192 次元のベクトルに整形する. そして, 各言語ごとに画像 10 枚から生成された特徴ベクトルを単純平均を用いて, 一つの平均特徴ベクトルを生成する. これを対象概念から想起されるイメージの平均特徴ベクトルとする.

[‡]<https://arxiv.org/abs/1409.1556>

そして、平均特徴ベクトル同士でベクトル間の類似度をコサイン類似度によって計算する。得られた類似度が一定の基準を下回った場合は文化差ありと判定、その基準を上回った場合を文化差なしと判定する。

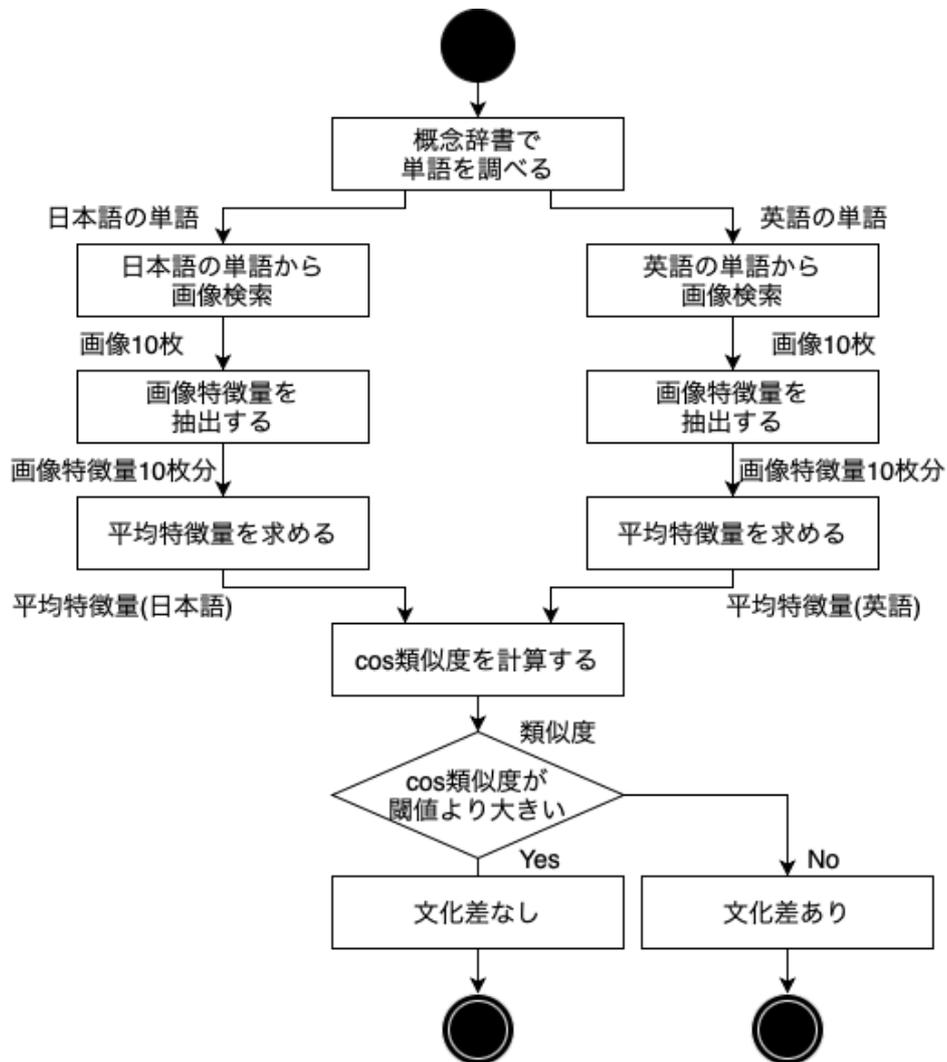


図 4.1: 想起イメージレベルの文化差検出手法

表 4.3: 人と提案手法の一致

		人手	
		文化差あり	文化差なし
提案手法	文化差あり	True	False
	文化差なし	False	True

4.3.3 閾値を用いた文化差判定

ここでは、先ほど算出した類似度から文化差を判定するための基準となる閾値について説明する。類似度はあくまで連続値であるために文化差判定のために、閾値を適切に設定する必要がある。

まず、文化差の判定基準となる閾値を最適化するために、文化差の検出精度の指標を定義する。検出精度の指標には文化差判定の正確さ (Accuracy) を用いる。Accuracy とは、人手によって文化差判定結果に対して、提案手法の文化差判定結果がどの程度一致したかの割合である。表 4.3 の True になるところが人手の判定結果と提案手法の判定結果が一致したときである。次に、Accuracy を算出する方法を式 4.1 に示す。

$$Accuracy = \frac{\text{True に該当した synset の数}}{\text{True と False に該当した synset の数}} \quad (4.1)$$

これらの指標をもとに閾値を最適化するための予備評価を行う。予備評価では、500 個の概念を用いて、提案手法で文化差判定を行う。この時、用いる 500 個の概念は、人手により文化差ありと判定された概念と文化差なしと判定された概念がそれぞれ 250 個ずつ含まれている。最初に、閾値を 0.0 と設定し、500 個の概念の文化差判定を行い、Accuracy を算出する。これを閾値 1.0 まで 0.05 ごとに繰り返し行い、最も Accuracy が高くなる閾値を文化差検出に最適な閾値とする。

4.4 評価実験

4.4.1 実験手順

人手による文化差の判定

アンケートを用いて人手で文化差の有無をラベリングする。人によって、文化差を感じる度合いは異なるため、1つの概念につき、5人のアノテーターによる文化差判定を行う。その後、アノテーターの判定結果を多数決によって、一つに定める。今回、アノテーターは20代の男女である。

用いたデータは、WordNet から 1000 個の synset をランダムで抽出したものを使用した。図 4.2 は、今回の文化差を判定するとき用いたアンケートである。1タスクにつき、1つの synset とし、各アノテーターへ、以下のようにタスクの指示をした。

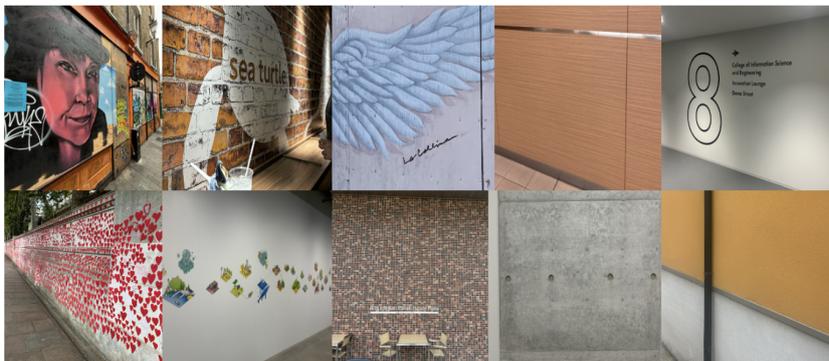
キーワードを見て、“A” と “B” のどちらの画像群がより連想できるか選んでください。
もし、両方とも連想できた場合は“両方”を選び、どちらも連想できない場合は“該当なし”を選んでください。※ () 内の数字はタスク番号です

今回のアンケートでは、日本語のキーワード、画像群 A、画像群 B の 3つをアノテーターに提示している。日本語のキーワードは、その synset を表現する日本語の単語群を用いている。画像群はその synset の日本語と英語のそれぞれの単語群から画像検索によって取得された画像である。選択肢は図 4.2 のように、“A”、“B”、“両方”、“該当なし”の 4つを用意し、キーワードから連想できる画像を選ぶように指示をした。また、キーワードに複数の単語がある場合、それらのすべての単語に共通するイメージを連想できる画像群を選ぶようにした。

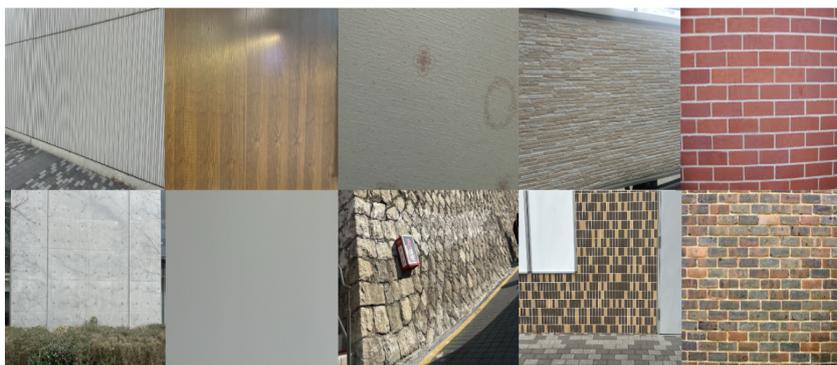
文化差の判定として、アノテーターの評価が“両方”が多数であった場合、日英の単語のイメージに差がないとして文化差なしとする。一方で、“A”もしくは“B”のどちらかが多数であった場合、どちらかの単語で検索された画像しか想起できなかったということなので、文化差ありとする。もし、“該当なし”の評価が多数であった場合は、画像検索の結果が

悪い可能性があるため、評価データから除外する。

A (1) *



B (1)



キーワード：壁

- | | | |
|----------------------------|---|---|
| <input type="radio"/> A |  | × |
| <input type="radio"/> B | | × |
| <input type="radio"/> 両方 | | × |
| <input type="radio"/> 該当なし | | × |

図 4.2: 文化差判定のアンケート

表 4.4: 交差検証での各プロセスのテスト結果

	最適な閾値	Accuracy(%)		
		全体	文化差なし	文化差あり
1 回目	0.55	76.0(76/100)	78.0(39/50)	74.0(37/50)
2 回目	0.55	79.0(79/100)	80.0(40/50)	78.0(39/50)
3 回目	0.55	74.0(74/100)	72.0(36/50)	76.0(38/50)
4 回目	0.55	75.0(75/100)	78.0(39/50)	72.0(36/50)
5 回目	0.55	76.0(76/100)	76.0(38/50)	76.0(38/50)
平均	N/A	76.0(76/100)	76.8(38.4/50)	75.2(37.6/50)

4.4.2 評価

最適な閾値の精度検証

人手で文化差の有無をラベリングしたデータを用いて、閾値の最適化とその精度の検証を行う。ラベリングの結果, synset1000 個のうち, “文化差なし” の synset 数が 713 個, “文化差あり” の synset が 255 個であった。また, 人手での判定の結果, “該当なし” が多数のため, 対象外となった synset が 32 個あった。これらの 32 個を除いた 968 個の synset を用いて評価を行う。なお, この 968 個のデータは文化差の有無に偏りがある不均衡データであるため, サンプルングによって 250:250 の均衡データとする。さらに, 評価の信頼性を高めるために, 交差検証を用いてデータの偏りを防ぐ。今回は, データ 500 件を 1 グループ 100 件の合計 5 グループに分割した 5 分割交差検証を行う。5 分割交差検証では, バリデーションデータ: テストデータ=4:1 になるように分割し, それぞれのデータを用いてバリレーションとテストを行う。このプロセスを 5 回繰り返す, それぞれのデータグループが 1 回ずつテストデータになるように選ぶ。5 回分のテストの結果を平均したものを提案手法の精度とする。それぞれのバリレーションの結果から最適な閾値を見つけ, テストでは見つけた最適な閾値をパラメータとして設定し, 提案手法の Accuracy を求める。

5 分割交差検証の各プロセスのテストの結果とその平均を示したもの

が表 4.4 である。表 4.4 の閾値は各バリデーションによって求められた最適な閾値である。Accuracy は最適な閾値を用いて、テストデータを文化差判定した結果である。全体での Accuracy とは、テストデータ全体での Accuracy を表現している。一方で、表 4.4 内の“文化差なし”はテストデータのうち文化差なしのデータに対する Accuracy, “文化差あり”はテストデータのうち文化差があるデータに対する Accuracy である。5 回分の Accuracy を平均した結果、Accuracy は 76.0%, 文化差がないデータに対する Accuracy は 76.8%, 文化差があるデータに対する Accuracy は 75.2%であった。

また、今回のデータは WordNet からランダムに取得したデータに対して、Accuracy が 76.0%であったため、文化差検出において効果的であると言える。さらに、文化差があるデータに対する Accuracy は、提案手法の再現率と同義である。本研究の目的は、コミュニケーション支援のために、齟齬の原因となる文化差を検出することである。そのため、概念レベルの文化差を 75.2%の割合で検出できることはコミュニケーション支援において有効であると考えられる。

概念の種類ごとの文化差検出精度

データを synset ごとに分類し、提案手法がどのような synset に有効なのかを分析する。WordNet では各 synset 間の is-a 関係を表した hypernym を用いて、そのデータの synset の上位 synset に基づいてデータを分類する。そして、その上位 synset の種類ごとに文化差の検出精度を調べる。対象外のデータを除いた 968 個のデータの上位 synset を調査した結果、949 個のデータの最上位 synset は Synset(“entity”)であった。残りの 15 件のデータが固有名詞を表し、4 個が動詞であったため、hypernym 関係を持たず、独立した synset であった。そのため、分析では大多数の Synset(“entity”)に属する 949 件のデータに限定して、synset の種類による検出精度の差を分析する。図 4.3 に示すように、Synset(“entity”)の子 synset に Synset(“physical entity”)と Synset(“abstraction”)がある。それぞれの synset の説明が Synset(“entity”): 表 4.5, Synset(“physical entity”): 表 4.6, Synset(“abstraction”): 表 4.7 である。

949 件のデータのうち、831 個は Synset(“physical entity”)の下位に位

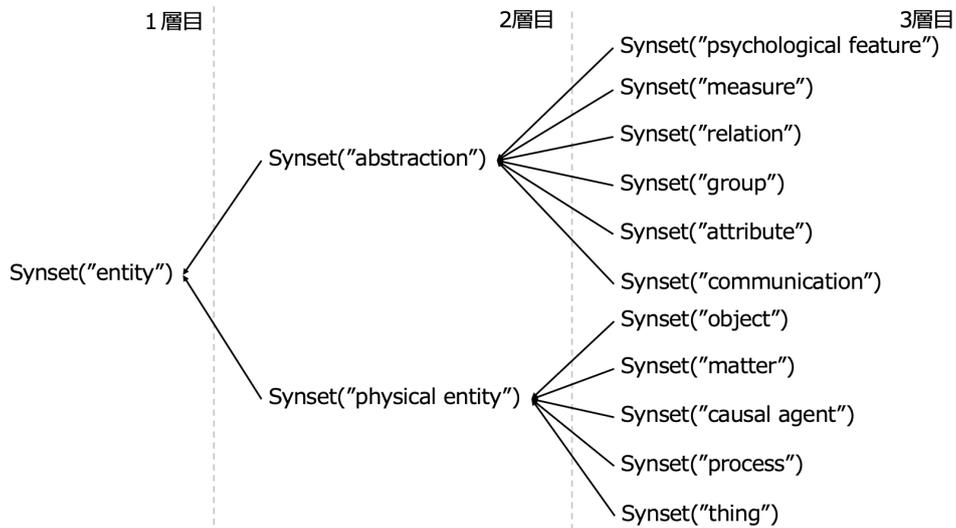


図 4.3: WordNet の上位 3 層

置する synset であった. これを “physical entity” グループとする. 一方で, 残りの 118 件のデータは Synset(“abstraction”) の下位に位置する synset である. これを “abstraction” グループとする.

“physical entity” グループには 831 件のデータのうち, “文化差あり” のデータは 195 件含まれていた. 一方, “abstraction” グループには 118 件のデータのうち, 47 件の “文化差あり” データが含まれていた. そのため, “physical entity” グループと “abstraction” グループで, それぞれ “文化差あり” と “文化差なし” が均等になるようにアンダーサンプリングを行う. サンプリングした均等データを用いて, パラメータを閾値 0.55 とした提案手法で Accuracy を算出した. その結果が表 4.8 である. “abstraction” グループに比べて, “physical entity” グループのデータは提案手法の Accuracy がわずかに高かった. 提案手法では画像から特徴ベクトルを抽出しているため, “physical entity” のような物理的に実体を持ったものでは, 画像にターゲットとなるものが含まれると考えられる. このことから, 同一のものを同じキーワードで取得された画像からは類似した特徴ベクトルを生成できることが増え, 文化差の検出精度が高まったと考えられる. 一方で, “abstraction” グループでは, 抽象的な概念を表

表 4.5: Synset(“entity”) の日本語と英語の情報

	単語群	説明
日本語	実体	(生命がある, あるいは生命がないに関わらず) それ自身の明確な存在を持つと感知される, 知られている, あるいは推定される何か
英語	entity	that which is perceived or known or inferred to have its own distinct existence(living or nonliving)

表 4.6: Synset(“physical entity”) の日本語と英語の情報

	単語群	説明
日本語	-	物理的な存在がある実体
英語	physical entity	an entity that has physical existence

現している synset のため, 具体的な形が定まっていない. そのため, 画像検索に取得された画像に固定のオブジェクトが映ることが少なく, 多様な特徴ベクトルが抽出されるため, 提案手法の検出精度が低くなったと考えられる.

次に, Synset(“physical entity”) と Synset(“abstraction”) の子 synset ごとに分類する. 図 4.3 の 3 層目の synset ごとに分類する. Synset(“object”) の下位に位置するデータを “object” グループ, Synset(“matter”) の下位に位置するデータを “matter” グループのように, それぞれのグループを定義する. synset の各グループごとに分類した時のデータ数が, 表 4.9 の “データ数” の “元” にある数字である. 先ほどと同様に, 各グループごとの “文化差あり” と “文化差なし” のデータ数が均等になるようにアンダーサンプリングし, Accuracy を算出する. 表 4.9 の “データ数” の “サンプリング” の欄はサンプリング後のデータ数である. また, “relation” グループはデータ数が 1 つであったため, 1 つのデータの判定結果を参考値として載せている. 比較した結果, “physical entity” グループの下位に位置するグループでは, “process” グループや “thing” グループの

表 4.7: Synset(“abstraction”) の日本語と英語の情報

	単語群	説明
日本語	抽象的実体	特定の例から共通点を抽出することによって形成された一般概念
英語	abstraction, abstract entity	a general concept formed by extracting common features from specific examples

表 4.8: 2 層目の synset による分類の Accuracy

synset グループ	データ数		判定結果		Accuracy(%)
	元	サンプリング	正解	誤り	
physical entity	831	390	304	86	77.9
abstraction	118	94	71	23	75.5

Accuracy が低かった。一方で、“abstraction” グループの下位に位置するグループでは、“attribute” グループ、“group” グループでは Accuracy が高かったが、“psychological feature” グループや“communication” グループの Accuracy は低かった。上位に位置する synset ごとに分類したとき、2 層目と 3 層目でそれぞれのグループによって、Accuracy の差が見受けられた。そのため、グループの種類によって、提案手法の検出精度の差があるかを確認するために、 χ^2 検定を行った。検定結果をまとめたものが、表 4.10 である。

2 層目では、表 4.8 の“physical entity” と“abstraction” の 2 群で判定の正解と誤りの件数で集計表を作成し、検定を行った結果、 $p > 0.05$ であった。同様に、3 層目では、表 4.9 の“object” から“relation” までのグループで、判定結果の正解と誤りで集計表を作成した。 χ^2 検定結果、 $p > 0.05$ であり、2 層目と 3 層目のどちらでも統計的に有意な差はみられなかった。このことから、今回のデータでは提案手法は概念の種類による大きな性能差は生じないと言える。

表 4.9: 3 層目の synset ごと分類した時の Accuracy

synset グループ	データ数		判定結果		Accuracy(%)
	元	サンプリング	正解	誤り	
object	639	296	229	67	77.4
matter	160	68	48	20	70.6
process	5	2	1	1	50
causal agent	16	4	3	1	75
thing	11	8	5	3	62.5
psychological feature	49	48	33	15	68.8
attribute	23	12	10	2	83.3
group	28	18	16	2	88.9
communication	12	8	5	3	62.5
measure	5	4	3	1	75
relation	1	1	0	1	0

4.4.3 既知の文化差への適用

Pituxcoosvarn, 柳生, 吉野らの研究において, 文化差があると報告されている言葉 (“リフォーム”, “忍者”, “noodle”, “waterfall”) に, 提案手法を適用し, 文化差の判定を行う [34, 74, 115]. 表 4.11 に示すように “リフォーム” と “忍者” では提案手法は文化差検出に成功した. 一方で, “noodle”, “waterfall” では文化差の検出に失敗した. “リフォーム” では, 日本語の単語から家の内装の画像が取得されていたが, 英語の画像検索結果では, “makeover” から, 人の見た目の変化を表す画像が取得されていた. Pituxcoosvarn らの研究で報告されているように, “リフォーム” は家の改築や模様替えの時に使用されるが, “makeover” は化粧品やダイエット後の変化を意味する. そのため, このような画像検索の結果に差が生まれたことで, 文化差の判定が成功したと考えられる. “忍者” という言葉は, 既存研究でどのような違いがあるのか具体的な説明がないため, 提案手法での画像検索の結果をもとに考察をする. “忍者” という日本語

表 4.10: 各層の検出精度の検定結果

層数	自由度	確率 (p)	χ^2 値
2 層目	1	0.297965247	1.083278689
3 層目	10	0.445922655	9.938304464

表 4.11: 既知の文化差に対する提案手法の判定結果

日本語の単語	英語の単語	類似度	判定結果
“リフォーム”	“makeover”	0.195	文化差あり
“忍者, 忍の者, 忍びの者”	“ninja”	0.351	文化差あり
“ヌードル, 麺”	“noodle”	0.585	文化差なし
“瀧, 落水, 垂水, 飛瀑, 飛泉, 滝, 瀑布, 水簾”	“waterfall, falls”	0.779	文化差なし

のキーワードからは伝統的な忍者を描写したイラストの画像が得られたのに対し、“ninja”という英語のキーワードからは実際の方が忍者のコスチュームを着た写真が取得されるという差異が生じた。このため、イラストと実物の人で画像の類似度が低くなり、文化差判定に成功したと考えられる。柳井らの研究では文化差のある単語として“noodle”や“waterfall”が述べられている。日本では、“noodle”からラーメンが最も多く想起されるが、ヨーロッパでは、“noodle”からはスパゲッティが最も想起されると報告している。提案手法で文化差の判定を行ったところ、ラーメンとスパゲッティの画像が取得された。しかし、どちらも麺の部分に焦点を当てた画像が多く、似たような特徴ベクトルが抽出されたことで、類似度が高くなり、文化差の判定に失敗したと考えられる。次に、“waterfall”という言葉はアジアでは美しい滝を想起するが、南アメリカでは力強い滝を想起するため、文化差があると柳生らは述べられている。このように言葉に対する印象は異なるが、どちらも同じ滝が想起されている。画像検索の結果はどちらも滝の画像が取得されているが、印象や

表 4.12: 文化差判定成功 - 事例: フットボール

言語	単語	画像	類似度
日本語	フットボール		0.502
英語	football		

感じ方の違いを特徴ベクトルに捉えることは難しかったため、文化差なし判定になったと考えられる。

4.5 考察

4.5.1 提案手法で判定に成功したケース

次に、文化差判定に成功したパターンは以下の2通りあり、それぞれ事例を踏まえて紹介する。

- 提案手法が文化差あり, 人手の評価が文化差あり
- 提案手法が文化差なし, 人での評価が文化差なし

提案手法が文化差あり, 人手の評価が文化差あり

文化差ありのデータを正確に判定できたケースが表 4.12 の“フットボール”である。日本語から取得された画像群ではサッカーボールが多く取得されている。一方、英語から取得された画像ではアメリカンフットボールのボールの画像がより多く取得されている。これらの画像から算出された類似度は 0.502 で、閾値を下回ったため、文化差あり判定

表 4.13: 文化差判定成功 - 事例: ワイン

言語	単語	画像	類似度
日本語	ワイン, ぶどう酒, ブドウ酒, 葡萄酒		0.702
英語	wine, vino		

となった。日本ではフットボールからはサッカーを連想し、アメリカンフットボールは“アメフト”と呼ばれることが多い。一方で、アメリカでは football からはアメリカンフットボールを連想する。そのため、画像に違いが生じ、それらのボールの形状の違いから類似度が低くなったため、文化差ありと判定できたと考えられる。

提案手法が文化差なし, 人手の評価が文化差なし

文化差のないデータを判定成功した例として、ワインがある。画像検索に用いた単語群は、日本語が「“ワイン”, “ぶどう酒”, “ブドウ酒”, “葡萄酒”」で、英語が「“wine”, “vino”」である。このケースでは、一番上の“ワイン”のケースでは、文化差なしのデータを文化差なしと判定できたケースである。この時、実際に取得された画像が表 4.13 の図である。これはアンケートによって文化差なしと判定されたデータであり、提案手法で算出された類似度が 0.702 であった。実際に取得された画像を見てみると、両方ともボトルに入ったワインの画像が取得されている。そのため、類似度が高く算出された結果、文化差なしと判定できたと考えられる。

表 4.14: 文化差判定に失敗した事例: 水

言語	単語	画像	類似度
日本語	ウォーター, ウォーター, 水		0.27
英語	water		

4.5.2 提案手法で判定に失敗したケース

提案手法で文化差を検出できなかった4つのパターンを紹介する.

- (パターン1) 提案手法で文化差あり, 人手の評価が文化差なし
- (パターン2) 提案手法で文化差なし, 人手の評価が文化差あり
- (パターン3) アンケートで“B”が多数であった場合
- (パターン4) アンケートで“該当無し”が多数であった場合

提案手法が文化差あり, 人手の評価が文化差なし

提案手法で文化差検出に失敗したデータの一つに, 日本語の単語群が“ウォーター, ウォーター, 水”で英語の単語群が“water”というものがある. それらの画像検索の結果が表 4.14 である. “water”から画像検索で取得された画像は水そのものを写している画像が多い. 一方で, “ウォーター ウォーター 水”で画像検索した結果は, ミネラルウォーターなどペットボトルに入った飲料水の画像が多かった. 取得された画像群が異なる理由に借用語が考えられる. 借用語とは, 異なる言語から取り入れられた単語のことを指し, カタカナで表記される単語の多くが借用語で

表 4.15: 文化差判定に失敗した事例: ザクロ

言語	単語	画像	類似度
日本語	ザクロ		0.322
英語	pomegranate, pomegranate tree, Punica granatum		

あるが、借用語の特徴として、元言語とは異なる意味で使われている場合がある。日本語話者が物質としての水を指す時には、“水”を使うことが一般的であり、このケースでは“ウォーター”というカタカナの言葉に影響を受けて、ミネラルウォーターの画像が取得されたと考えられる。

同様に、表 4.15 の“ザクロ”でも判定に失敗した。提案手法で算出した類似度が 0.322 であり、文化差の判定に失敗した。取得された画像群では“ザクロ”で取得された画像がザクロの実の画像が多数である。一方で、“pomegranate, pomegranate tree, Punica granatum”ではザクロの木が映った画像が多かった。英語のキーワードの中に、“pomegranate tree”があったため、tree の部分に影響されて、ザクロの木の画像が多くなった。これにより類似度が低くなり、文化差ありと判定されたと考えられる。

パターン 2: 提案手法が文化差なし、人手の評価が文化差あり

提案手法では文化差がないと判定したが、人手では文化差があったケースに日本語の単語が“タピオカ”で英語の単語が“tapioca”というデータがある。それらの画像検索の結果が表 4.16 である。“タピオカ”から画像検索で取得された画像は、飲み物の中に黒いタピオカを入れたものが多かった。一方で、“tapioca”から画像検索で取得された画像には、ゼリーやヨーグルトのような、半固形状のものに入れられている白く小さいタピオカであった。アンケートでは、5 人のアノテーターのうち 4 人

表 4.16: 文化差判定に失敗した事例: タピオカ

言語	単語	画像	類似度
日本語	タピオカ		0.69
英語	tapioca		

表 4.17: 日本語話者が英語で取得された画像のみを選択したケース

言語	単語	画像	類似度
日本語	舞踏会 ボール		0.37
英語	ball, formal		

が“タピオカ”から取得された画像の方が、より想起されると評価したため、文化差があるとラベリングされている。日本では、黒いタピオカをいれた飲み物が流行したことから、このような結果になったと考えられる。一方で、提案手法では同じような画像が取得されているため、類似度が高くなり、文化差がないと判定したと考えられる。

パターン3: アンケートで“B”が多数であった場合

今回、人手で文化差のラベリングを行ったアノテーターは全て日本人であった。しかし、アンケートで日本語から取得された画像ではなく、英

表 4.18: パターン 4 の事例

言語	単語	画像	類似度
日本語	進化, 進歩, プロGRESS, 発展, 成長, 発達, 発育		0.25
英語	growth		

語の単語から取得された画像を選んだパターンについて考察していく。それがこのアンケートにおいて、“B”が多数であったデータである。アノテーターらは、どちらの画像群が日本語から取得されたものかは伝えていないが、“B”を選ばれた理由として、キーワードに含まれている単語がよくなかったと考えられる。このようなケースになったデータの事例として、表 4.17 の“舞踏会 ボール”と“ball formal”が挙げられる。英語の“ball formal”からは、たくさんの人がダンスをする舞踏会を表現する画像が取得されている。一方で、“舞踏会 ボール”からは、ボールガウンという舞踏会用のドレスの画像が多く取得された。日本語では、舞踏会のことを、ボールとは一般的には言わない。しかし、借用語のように英語の知識をもとに、単語は存在し、ボールと舞踏会の AND 検索からボールガウンという特定のものの画像が取得されたと考えられる。そのため、アノテーターらは普段馴染みのないボールガウンよりも、たくさんの人が踊る画像群の“B”を選択したと考えられる。

パターン 4: アンケートで“該当無し”が多数であった場合

“該当無し”は対象外とし、評価データから除外している。表 4.18 のように過程や感情などの概念を表す単語であった。そのため、被験者はどちらも想起できず“該当無し”と回答した人が多かった。

4.6 まとめ

多言語コミュニケーションにおいて、話者間の文化背景の違いから起こるイメージの文化差を検出するために、本研究では画像から得られる特徴ベクトルに基づく検出手法を提案した。WordNetで同一のsynsetに紐づけられている日本語と英語の単語群から画像検索を行い、検索結果の上位10件の画像の特徴ベクトルを抽出する。これらのベクトルを言語ごとに平均し、その単語を表す平均特徴ベクトルとする。このベクトル間のコサイン類似度に基づいてイメージの文化差を検出した。本研究の貢献として、算出したベクトルの類似度から、イメージの文化差を自動的に判定するための閾値を設定した。具体的には、人手で文化差の有無をラベリングした1000個のデータ(synset)から、文化差ありと文化差なしが250:250の均等データを用いて、閾値の最適化を行った。最適化では、閾値0.0から1.0まで0.05ごとに閾値を変動させ、最も提案手法の検出精度が最大化する閾値を見つけた。この結果、閾値0.55がイメージの文化差検出に最適な閾値であるとわかった。また、その閾値をパラメータとして提案手法の評価を行ったところ、Accuracyが76.0%であった。

第5章 文化差検出によるコミュニケーション支援

5.1 多言語コミュニケーションにおける文化差

多言語コミュニケーションでは、機械翻訳の精度向上により言語面での支援環境が構築されつつある。しかしながら、機械翻訳を用いたコミュニケーションは会話がうまくいっているように感じていても、実際には、概念体系や話者の文化背景の違いから、意思疎通のための共通認識を構築することが難しいとされている [73, 116]。例えば、NPO Pangaea の KISSY ではさまざまな言語・文化背景を持つ子どもたちが機械翻訳を用いて協働タスクが行っている。その時に、ある日本語話者が“あんこ”を英訳結果の“red beans paste”という言葉を用いて、クメール語話者に説明した時に、クメール語話者が理解できずに会話に問題が起こった。この時、機械翻訳を使ったことでクメール語話者は内容を理解したと思い、日本語話者はクメール語話者の反応や返答から相手に伝わっていると思っていた。しかしながら、クメール語話者は“red beans paste”から豆をペースト状にしたものを固めた固形物を想起していたため、途中から会話が噛み合わなくなり、クメール語話者が会話を理解できなくなった。その後、英語を話すことができる別の話者が両話者の間に入り、画像などを使ってあんこを詳細に説明したため、齟齬は解消された。

このように、概念体系や文化背景の違いから、機械翻訳を介しても両話者が捉えている意味や認識は対称ではない。このような多言語コミュニケーションの非対称性を解消するために、以下のような課題がある。

文化差の認識

多言語コミュニケーションに齟齬が起こる場合、話者らはどのよ

うな単語に問題があり、認識のずれが生まれているのかを認識できない。そのため、文化差のある単語を検出し、話者らに注意を促す必要がある。

話者間の認識のずれの修正

コミュニケーションに齟齬が起こった場合、違和感に気づいた話者が相手の話者に問いかけ、両話者の認識のずれを確認する。それによって、話者同士で話し合い、認識のずれを修正する。しかしながら、多言語コミュニケーションでは、機械翻訳を用いるため、修正を図る発話にも文化差があり、話者間で認識のずれを解消することが容易ではない。

私はこれらの課題を解決するために、概念レベルと想起イメージレベルの文化差検出手法によって、文化差を自動的に検出する。これによって、話者に文化差の注意を与え、話者自身が文化差を考慮した発話を促す支援を行う。

5.2 多言語コミュニケーションの非対称性

5.2.1 予備実験

Pituxcoosuvam らの研究 [2] を参考に、多言語コミュニケーションによる予備実験を行った。実験タスクでは、先行研究と同様にサバイバルタスクによる協調作業とした。サバイバルタスクのアイテムには、本研究の提案手法で検出された文化差のある単語を用いて、多言語コミュニケーションにおいて、文化差のある単語がどのような影響を与えるかを確認した。

表 5.1 は予備実験での会話ログである。予備実験の参加者は日本語話者とインドネシア語話者で、話者同士はそれぞれの言語を使用してチャット上でタスクを行っている。表 5.1 の会話ログは日本語を表示したもので、インドネシア語話者の発話内容は翻訳を介した結果を表示している。

表 5.1: 予備実験で文化差により会話に齟齬が起こったケース

ID	Sender	Message
8	Japanese	1位「鏡」 理由「採取や護身用に使えるナイフになるから」でいいですか？
9	Indonesian	ということは2枚目はキャンドルということでしょうか？
10	Japanese	では2位を決めましょう。 私は「毛布」を選びました
11	Indonesian	え、毛布？
12	Japanese	え!?
13	Indonesian	毛皮のことですか？
14	Japanese	はい
15	Japanese	(Indonesian) さんは何を選びました？
16	Indonesian	羽根が何枚あるのか分からないのでほんの数枚だと思ってました(笑)
17	Indonesian	2本目はキャンドルを選びました
18	Japanese	キャンドル??
19	Japanese	ワックスのことですか
20	Indonesian	????
21	Indonesian	え？
22	Indonesian	それも一つの選択肢ですよ？
23	Japanese	アイテムの中に「キャンドル」はありません
24	Indonesian	さまざまなオプションがあるようです

表 5.2: 予備実験で支援を入れたケース

ID	Sender	Message
13	Indonesian	ボート自体に関しては、 外出する場合には非常に 便利だと思いますが、運転方法が わからない場合はどうすればよい ですか。さらに、道に迷って しまうと、周りの海しか 見えなくなります。 かなり内陸になる可能性が高い
14	Indonesian	もちろん！
15	Agent	“ボート” < “kapal” 日本語よりインドネシア語の方が 意味が広いから注意
16	Japanese	なるほど、 確かに初心者には運転が難しいと思います 4位「ボート」 理由：運転が難しいからで いいですか？
17	Indonesian	インドネシアのボートは 小型船の一種ですか？ 漕いで使用できるもの。
18	Indonesian	船は船長によって運転されます (船全般)
19	Japanese	漕いで使用できる物です
20	Indonesian	まだ運転が難しいと感じています
21	Japanese	免許がなくても使用できますが、 運転は難しいと思います
22	Indonesian	波が大きければ揺れるかもしれない
23	Indonesian	そうだよ、私もそう思うよ
24	Japanese	そうですね、4位ボートで

サバイバルタスクによって、複数のアイテムを生存に必要な順にランキングするように指示した。そのアイテムの一つに日本語で“ワックス”，インドネシア語で“lilin”を入れた。これは概念レベルの文化差検出手法で検出された文化差のある単語で、この文化差によって、予備実験ではコミュニケーションに齟齬が起こった。表 5.1 内の ID:17 の発話では、インドネシア語話者が“キャンドル (lilin)”を取り上げている。これは、日本語のワックスよりも lilin は広義な意味を持っており、lilin を翻訳した結果、日本語話者にはキャンドルと表示されている。そのため、日本語話者は突然キャンドルの話題になったため、ID:18 と 19 の発話のように日本語話者からインドネシア語話者に聞き返している。一方で、キャンドルやワックスは lilin に翻訳されるため、インドネシア語話者にとっては (ID:20 の発話のように) 何が問題になっているかわからないという状況になっている。このように、日本語話者は文化差によって、会話に違和感を感じているが、インドネシア語話者に聞き返した発話にも文化差があるためお互いの認識のギャップが埋まっていない。

次に、先行研究のように、各話者に文化差をアラートによって知らせた。その時の会話ログが表 5.2 である。

この場合では、表 5.2 内の ID:15 の発話のように文化差のある単語が使われた際に、エージェントが各話者にアラートを出す支援を行っている。先ほどの支援なしの場合では、表 5.1 の ID:11 や ID:17~19 の発話のように、単語のみを述べる発話を行なっている。一方で、支援を行った場合、表 5.2 の ID:16~18 の発話のように、話者がアイテムの追加情報を提供し、各単語の概念に対して、認識のすり合わせを行う様子が見られた。

5.2.2 仮説

前節の予備実験のように、アラートを出すことによって話者らは文化差の認識が向上すると考えられる。予備実験では、アラートを出していないときには単語のみの発話やランキングについての発話が多くみられた。このように単語の認識が共通のものであると各話者が考えていると考えられるため、会話に齟齬が起こったと考えられる。一方で、アラートを出しているときには、単語がどのようなものであるかを説明する発

話が見受けられるため、各話者の文化差の認識が上がることで話者の振る舞いが変わると考えられる。具体的には、お互いの話者が単語の認識を一致させるために情報を出し合い、確認を行うことが起こる。そのことから、より単語の説明の詳細度が上がると考えられる。そして、認識を一致させる行為が生まれるため、議論がより活性化すると考えられる。加えて、説明の詳細度が上がることによって、会話の理解度が向上することが考えられる。以上のことをまとめて、以下の仮説を立てる。

仮説1 支援をする場合は支援を入れない場合に比べて、各話者の文化差の認識が向上する

仮説2 支援をする場合は話者同士で認識のすり合わせが行われるため、支援を入れない場合に比べて、対話の情報量が増える

仮説3 支援をする場合は支援を入れない場合に比べて、各話者の会話の理解度が向上する

5.3 実験

実際の多言語コミュニケーションで文化差検出がどの程度効果的かを評価するために、対照実験を行う。実際に機械翻訳を通じた多言語コミュニケーションにおいて、齟齬が起こる会話をデザインし、文化差検出によって各話者にアラートを出す支援を行うグループと支援を行わないグループの2群間で会話を比較する。

5.3.1 参加者

実験の参加者には、日本語話者8人、インドネシア語話者8人の計16人に協力してもらった。被験者の基準として、使用する言語を母語とする大学生以上を対象とし、実験を行った。

参加者は日本語話者1人、インドネシア語話者1人を1グループとし、8グループで会話実験を行う。それぞれのグループの話者同士は面識がなく、チャットツールを使って、オンライン上でコミュニケーションをとる。

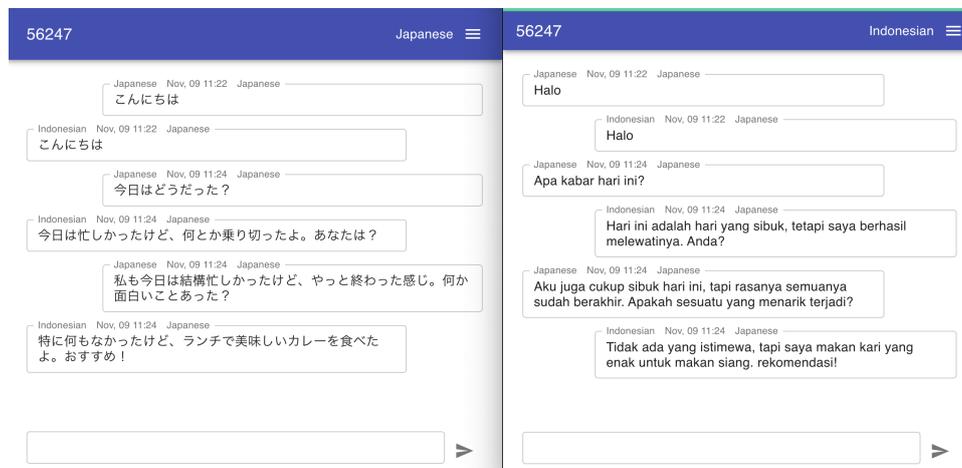


図 5.1: LangridChat

5.3.2 実験で使用するコミュニケーションツール

言語グリッドを活用した LangridChat という多言語コミュニケーションツールを用いる。図 5.1 は LangridChat の画像である。LangridChat では各話者の母語を用いてチャットツールにメッセージの記述ができる。図 5.1 の左の画面では日本語を指定しており、右の画面では言語にインドネシア語を指定している。このように、ユーザは各自の言語でチャットを行うことができ、各話者の言語に翻訳された状態で画面上に表示される。これによって、各言語話者は自分の母語を用いたコミュニケーションを行う。

LangridChat のシステム構成は図 5.2 のようになっており、サーバ側は Django で、クライアント側は React で構成されている。システムでは、ユーザーから送信されたメッセージを言語グリッド用いて、8 言語 (英語、日本語、タイ語、ベトナム語、インドネシア語、ネパール語、中国語-簡体字、中国語-繁体字) に翻訳し、各話者の言語で表示している。

また、文化差が含まれている単語がメッセージ内に含まれている場合、支援エージェントから文化差があることをアラートで知らせる。このとき、アラートには検出された文化差のある単語とどちらの言語がより広義かであることを示した内容載せている。

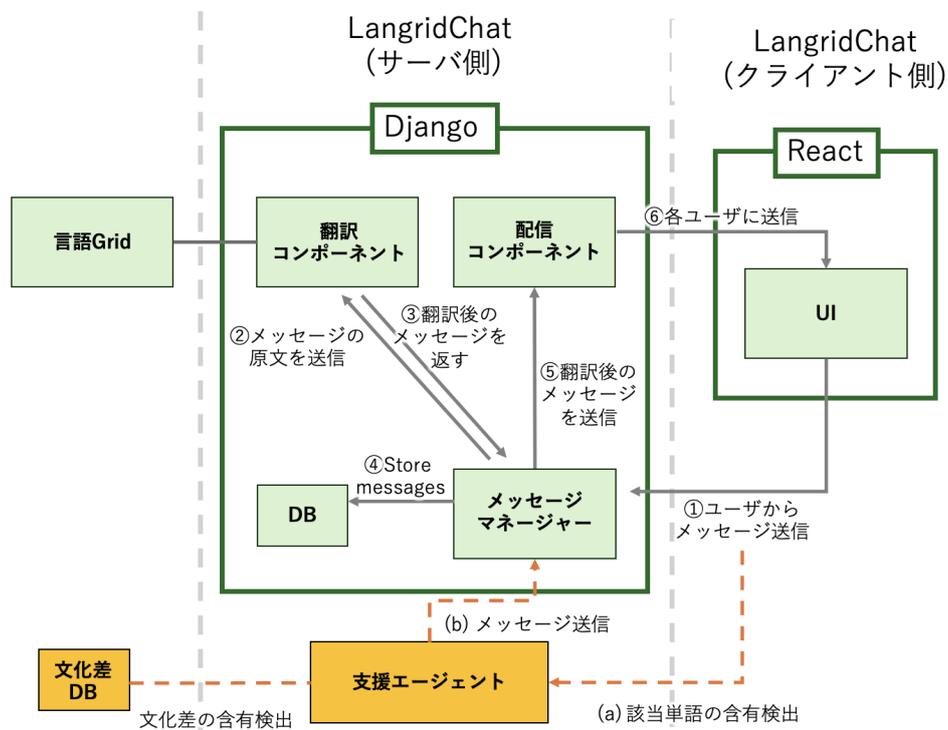


図 5.2: LangridChat のシステム構成

今回の実験ではアラートは、ウィザード・オブ・オズ法を用いて提示した。以下の内容が各話者に提示したアラートの例である。

“ワックス” < “lilin”

日本語よりもインドネシア語の単語のほうが概念が広いから注意

ワックスと lilin という概念レベルの文化差がある単語に対してのアラートである。ワックスと lilin は lilin の方が広義な意味を指し、概念差の分類は水平(包含)タイプであったため、アラートではどちらの単語が概念が広義であるかを知らせた。これによって、広義な概念を表す単語の話者がより詳細に単語を説明するように振る舞うと考えられる。

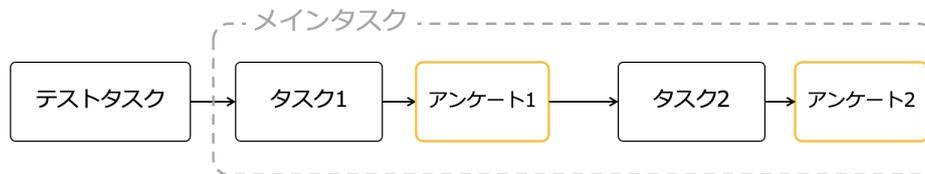


図 5.3: 実験の流れ

5.3.3 実験手順

各グループの参加者には、最初に実験の手順とチャットツールについて説明を行った。日本語話者には指示者から日本語による説明を行い、インドネシア語話者には同様の内容の説明を機械翻訳を介して、インドネシア語と英語の2言語で伝えた。

次に、チャットの練習とお互いの心理的な壁を軽減することを目的に、それぞれの自己紹介をする。その後、サバイバルタスクの練習として、10分ほどのテストタスクを行った。図 5.3 は実験の流れを示したものである。

テストタスクを行った後、メインタスクを行う。メインタスクでは、2回のサバイバルゲームを行い、それぞれタスク1とタスク2では、サバイバルゲーム内のアイテムを変更した別タスクとなっている。各タスクが終わり次第、参加者にはそのタスクに関するアンケートに回答を行ってもらい、実験の会話についてのフィードバックを得る。各メインタスクでは、2つのステップを行った。ステップ1では、5分間時間を取り、各自でアイテムのランキングについて自分の意見をまとめてもらう。ステップ2では、15分程度を目安にグループで議論を行ってもらい、アイテムのランキングを行う。ステップ1を挟むことで、前もって自分のランキングとその理由をまとめることで、その後のステップ2の議論をより活性化させる目的がある。

メインタスクのどちらかのタスクで文化差検出による支援を行う。また、サバイバルタスクは2種類のパターンを用意し、2種類のパターンと支援の有無を組み合わせに偏りがないようにした。図 5.4 はその例であ

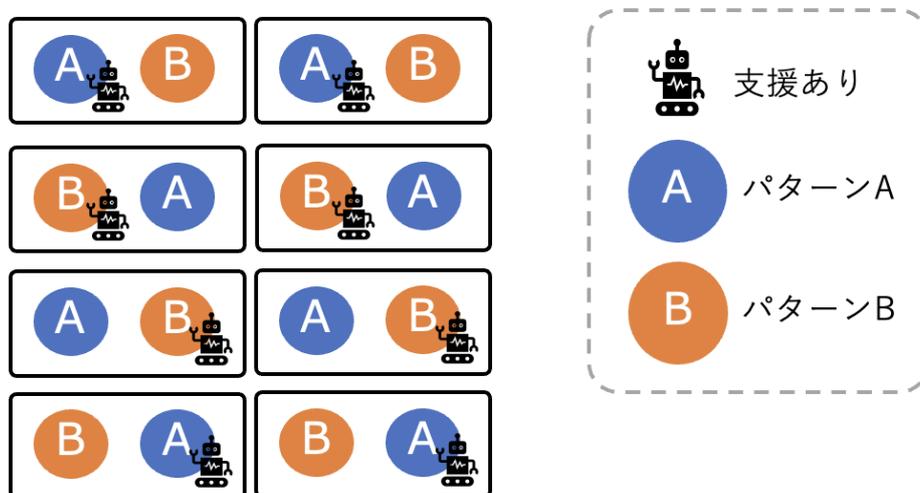


図 5.4: アイテムセットと支援の組み合わせ

る. アイテムのパターンが2種類とどちらのパターンにエージェントで支援を行うかを組み合わせた. また, 比較実験ではエージェントによる支援の効果を検証するため, エージェントなしとエージェントありをどちらを先に行うかを均一にする. そのため, 図5.4のように4種類の組み合わせになる. 本実験の参加者のグループ数は8グループのため, それぞれ2グループずつ行う.

5.3.4 実験タスク

文化差検出の効果検証のために, 文化差により問題が起こる会話実験をデザインする.

協働タスクには, Pituxcoosuvarn らの研究でも多言語コミュニケーションの実験で用いられている“サバイバルタスク [117]”を利用する. サバイバルタスクとは, 被験者らが遭難した状況を想定して行われるコミュニケーションタスクで, 被験者らにはいくつかのアイテムが与えられ, それらを必要な順に選択するものである.

以下が実際にテストタスクで使用したサバイバルタスクの内容である.

無人島サバイバル

あなた達は熱帯地域の無人島に流れ着きました。
乗ってきた船は沈没してしまっています。
島から見渡す限り海しか見えず、救助がいつになるかわかりません。
幸いなことに島には以下の3つのアイテムが落ちていました。
あなた達はこれらの3つのアイテムを生存に必要な順にランキング
して下さい。

3つのアイテム

- ナイフ
- 食料
- 水

文化差によりコミュニケーションに齟齬が起こる状況を作り出すために、文化差のある単語が会話に登場する必要がある。そこで、提案手法により文化差があると判定された単語からサバイバルタスクにあう単語を選び、それをタスクのアイテムとした。これによって、文化差のある単語を組み込んだ会話が行われるように実験を設計した。

各アイテムセットは4種類のアイテムを2パターン用意し、タスク1とタスク2でアイテムを入れ替えている。

パターンAでは表5.3の4つの単語を、パターンBでは表5.4の4つの単語をアイテムとしている。これらは、全ての単語が概念レベルの文化差検出によって、概念範囲が異なるとして抽出された単語ペアを用いている。

また、抽出された単語ペアから表5.3、表5.4の単語を選んだ基準として、以下の3つを設けた。

- サバイバルタスクに適した単語か
- 対訳関係にある単語か
- アイテム間のタスクへの難易度が同じか

表 5.3: パターン A のアイテム

	日本語	インドネシア語	概念差の分類
1	ワックス	lilin	垂直 (上位包含)
2	テープ	pita	水平 (包含)
3	毛皮	bulu	水平 (包含)
4	鏡	kaca muka	垂直 (下位包含)

表 5.4: パターン B のアイテム

	日本語	インドネシア語	概念差の分類
1	ワイン	arak	垂直 (下位包含)
2	ボート	kapal	垂直 (下位包含)
3	ランプ	lampu	水平 (包含)
4	カッター	pemotong	垂直 (包含)

1つ目の基準として、サバイバルタスクに適した単語かを判断した。抽出された単語ペアには、場所や人の職業などの単語が多く含まれている。そのため、サバイバルで使用可能な物を選んだ。

2つ目の基準として、対訳関係にある単語ペアであることを確認した。対訳関係にない場合、チャット上では異なった翻訳結果が出力される。このような誤訳で会話が破綻することを避けるため、抽出された単語ペアが、LangridChatの翻訳で対応するかを確認した。日本語の単語を w^{ja} 、インドネシア語の単語を w^{id} とした時、 w^{ja} を翻訳した結果が、 w^{id} になる。もしくは、 w^{id} を翻訳した結果が、 w^{ja} になる。このいずれかを満たした単語ペアを対象とした。

3つ目の基準は、アイテム間のタスクへの難易度が同程度であるかを考える。アイテムのランキングを決定するタスクのため、話者同士の個人ランキングが一致している場合、議論が活発化しない。そのため、人によってランキングが変わるように、アイテムの難易度を同程度にすることで、議論の活発化を図る。

ここで、先ほどのアイテムはどのような概念レベルの文化差があるかを確認する。

“ワックス” - “lilin” ワックスと lilin は概念レベルの文化差のタイプは垂直 (上位包含) タイプである. 具体的には, ワックスは英語で “wax” という単語が対応づけられた synset に対応づいている. この synset を Synset(“wax”) としたとき, Synset(“wax”) の下位概念に英単語 “beeswax” が対応づけられた synset を Synset(“beeswax”) とする. lilin は Synset(“wax”) と Synset(“beeswax”) に跨った単語で二つの概念差の関係は垂直 (上位包含) となっている. lilin は蝋燭などを意味する単語で, LangridChat で lilin を日本語に翻訳すると, “キャンドル” という結果になる. 一方で, ワックスをインドネシア語に翻訳した結果は lilin になる.

“テープ” - “pita” テープと pita は概念レベルの文化差のタイプは水平 (包含) タイプである. テープが対応づいている synset は英単語の tape が対応づいている. これを Synset(“tape”) とした時, Synset(“tape”) と兄弟関係にある Synset(“band₁”), Synset(“band₂”), Synset(“band₃”), Synset(“ribbon”) がある. Synset(“band₁”), Synset(“band₂”) と Synset(“band₃”) はそれぞれ別の概念を表現した synset である. この4つの synset と Synset(“tape”) に pita は跨って対応づけられており, テープよりも広義である.

テープと pita は LangridChat 上で, 単語のみを翻訳した場合, 両方向に翻訳関係が成り立つが, pita はロープや蝶ネクタイなども表現できるため, 文章によっては pita はテープ以外の翻訳結果になることがある.

“毛皮” - “bulu” 毛皮と bulu は概念レベルの文化差のタイプは水平 (包含) タイプである. 具体的には, 毛皮は fur という英単語が対応づけられた synset と hide という英単語が対応づけられた synset に対応づいている. これらを Synset(“fur”), Synset(“hide”) とする. 一方, bulu は Synset(“fur”) のみに対応づいた単語である. このケースでは, 日本語の方がインドネシア語の単語よりも広義な意味を持っている.

LangridChat で bulu を日本語に翻訳すると, “髪” や “毛” など体毛を表す単語に翻訳される. 一方で, 毛皮をインドネシア語に翻訳し

た結果は bulu になる。

“鏡” - “kaca muka” 鏡と kaca muka は概念レベルの文化差のタイプは垂直 (下位包含) タイプである。kaca muka は looking glass という英単語が対応づけられた Synset(“looking glass”) に対応づいている。Synset(“looking glass”) の親概念に mirror が対応づけられた Synset(“mirror”) がある。鏡は Synset(“mirror”) と Synset(“looking glass”) の二つにまたがる広義な単語である。実際に, kaca muka はインドネシア人によると顔を見る用の鏡を示すと述べており, 鏡よりも小さい概念を表していることがわかる。

LangridChat で kaca muka を日本語に翻訳すると, 鏡に翻訳されるが, 鏡をインドネシア語に翻訳した結果は cermin になる。

“ワイン” - “arak” ワインと arak は概念レベルの文化差のタイプは垂直 (下位包含) タイプである。アルコール飲料全体を表現する synset(Synset(“alcohol”)) がある。この Synset(“alcohol”) と子関係の synset に Synset(“liquor”), Synset(“gin”), Synset(“arrack”), Synset(“wine”), Synset(“neutral spilitis”) などがある。arak はこれら 7 つの synset にまたがり, アルコール飲料を表す単語である。一方で, ワインは Synset(“wine”) のみに対応づけられている。

LangridChat では arak の日本語翻訳結果は “ワイン” であるが, ワインのインドネシア語翻訳結果は “anggur” になる。

“ボート” - “kapal” ボートと kapal は概念レベルの文化差のタイプは垂直 (下位包含) タイプである。kapal は WordNet の概念階層において, 3 階層にまたがる単語である。まず, 乗り物を表現する Synset(“craft”) と Synset(“craft”) の子 synset の Synset(“vessel”) と Synset(“ship”) がある。さらに, Synset(“vessel”) の子 synset に Synset(“boat”), Synset(“ship”) の子 synset の Synset(“warship”) の 5 つの synset に kapal は対応づけられている。一方, ボートは Synset(“boat”) のみとなっている。インドネシア語話者によると kapal は船を全体的に表現する単語であるが, 一般的には大きめの船をイメージし, 船長など技術を持っている人が運転するものの

イメージを持っている。一方で、日本語話者はボートは公園などで乗れる小さめの船を指す単語と認識している。

LangridChat では両方向で翻訳関係が成り立ち、ボートを翻訳した結果は kapal, kapal の翻訳結果はボートになる。

“ランプ” - “lampu” ランプと lampu は概念レベルの文化差のタイプは水平 (包含) タイプである。source of illumination という単語が対応づけられた synset がある。この Synset(“source of illumination”) と子関係にある synset に Synset(“lamp”) と Synset(“light”) があり、日本語のランプは Synset(“lamp”) のみに対応づけられている。一方で、lampu は Synset(“lamp”) と Synset(“light”) の両方に対応づけられている。

LangridChat ではランプから lampu の方向では翻訳関係が成り立つが、lampu の翻訳結果はライトと出力される。

“カッター” - “pemotong” カッターと pemotong は概念レベルの文化差のタイプは垂直 (包含) タイプである。切断するための道具全般を指し示す Synset(“cutting implement”) がある。この Synset(“cutting implement”) の子 synset に Synset(“cutter”) がある。日本語のカッターは Synset(“cutter”) に対応づけられている。Synset(“cutter”) の子 synset に尖ったものを指す Synset(“edge tool”) がある。pemotong は Synset(“cutting implement”) から Synset(“edge tool”) までに跨っている。

LangridChat では両方向で翻訳関係が成り立ち、カッターを翻訳した結果は pemotong, pemotong の翻訳結果にカッターになる。

5.3.5 アンケート

アンケートでは、ユーザの文化差の有無への認知度がエージェントの支援によって上がるかを検証するために、以下の質問を参加者にした。

Q1: 相手の言っていることがよくわかった (1: 0%, 2: 25%, 3: 50%, 4: 75%, 5: 100%).

Q2: 相手はあなたの言っていることを理解していました? (1: 0%, 2: 25%, 3: 50%, 4: 75%, 5: 100%).

Q3: アイテムに対して、概念が違うと感じたアイテムの数 (0 個, 1 個, 2 個, 3 個, 4 個).

これら7つをアンケートで質問した。これらの Q1, Q2, Q3 は5段階評価で質問した。

これらの質問では、仮説1: 支援をする場合は支援を入れない場合に比べて、各話者の文化差の認知度が向上したことを検証するためのものである。Q3では、各話者がいくつのアイテムに対して、概念の違いを感じたかを質問した。お互いの話者が単語の認識に違いがあることに気づかずに会話を行ったときに、会話に齟齬が起こる。今回のサバイバルタスクでは、与えられる4つのアイテムは全て概念レベルの文化差がある単語にした。支援を入れない場合と支援をした場合で概念の違いがあることを各話者が認識できているかの差をみる。

参加者の理解度が向上したかを図るためには、相手話者の言っている内容が理解できる必要がある。そのため、Q1では参加者にどの程度相手話者の発話の内容が理解できたかを質問している。同時に、相手の言っている内容が理解できるには、会話の流れが断絶することなく進む必要があり、相手話者にも会話の意図や流れを汲んだ発話が求められる。そのため、Q2では相手が自分の言っていることを理解しているかを質問した。もし、会話の意図や流れを汲んだ発話がされているなら、相手の話者が自分の言っていることを理解していると感じると考えられる。

5.4 定量的分析

5.4.1 文化差の認識

仮説1を検証するために、アンケートのデータを分析する。エージェントによる支援がある場合とない場合の2群に分けて、アンケートの結果を統計検定する。具体的には、エージェントの支援がある場合のデータと支援がない場合でのデータを集計し、統計検定によってその差を比較する。対象とする質問はQ3である。

Q3の“アイテムに対して、概念が違うと感じたアイテムの数(0個, 1個, 2個, 3個, 4個)”のアンケートの集計結果が表5.5である。この表に対して、有意水準5%でウィルコクソンの符号付き順位検定を行ったところ、 $W=9.0$, $p=.045$ であったため、統計的に有意な差がある。

Q3の結果から、エージェントによる支援を入れることで、仮説1の文化差の認識が上がったことがわかった。

表 5.5: Q3(アイテムに対して、概念が違うと感じたアイテムの数)の集計表

	参加者															
	日本語話者								インドネシア語話者							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
支援なし	2	1	1	1	2	2	0	1	2	2	1	2	1	1	0	1
支援あり	3	2	1	1	1	1	3	3	2	2	2	3	2	1	1	1

5.4.2 対話の情報量

仮説2を検証するために、各実験の参加者の発話数と単語数を分析する。単語数が増えることは情報量が増えることと考えられるため、より詳細に話そうと振る舞うことによって単語数が増えると考えられる。同様に、発話数が増えることで話者同士のやり取りが増えるため、情報量が増える。よって、単語数と発話数を分析する。

まず、支援ありの場合と支援なしの場合での対話の総単語数を比較する。今回の実験では、日本語話者とインドネシア語話者の発話が含まれているため、言語の特性から日本語の場合、各話者の発話を取得し、発話ごとに形態素解析を行って、各形態素を抽出する。この形態素数を単語数として数える。一方で、インドネシア語の場合、各話者の発話からそのまま単語数を数える。また、アイテムが二パターンあり、アイテムの違いを考慮するために、各タスクの対話の単語数を集計し、タスクの平均単語数からの偏差で集計表を作成した。図5.5は対話の総単語数を集計した箱ひげ図である。それぞれ、図5.5内の“グループ”の箱ひげ図はグ

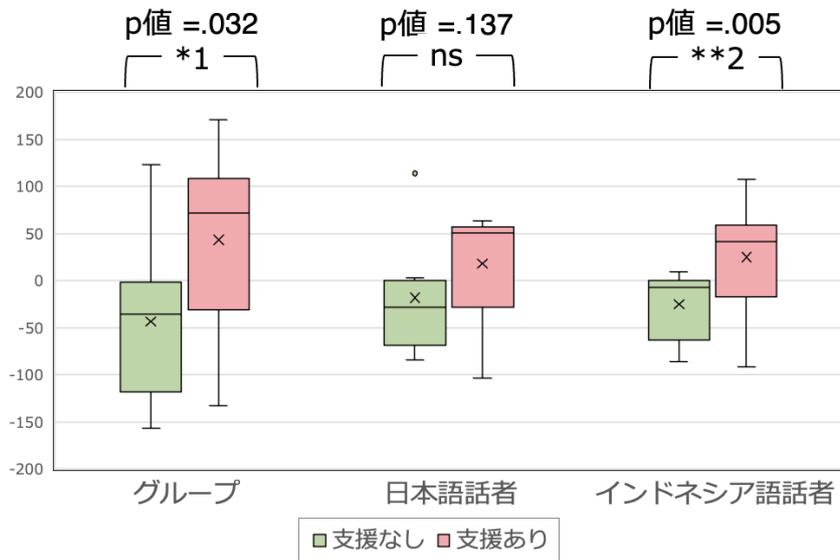


図 5.5: 総単語数

ループの日本語話者とインドネシア話者の単語数を足した総単語数である。加えて、グループの各言語話者のみの総単語数を集計したものが、図 5.5 内の“日本語話者”と“インドネシア話者”の箱ひげ図である。緑色の四分位範囲が支援なしのタスクでの総単語数を表しており、ピンク色の四分位範囲は支援ありのタスクでの総単語数を表している。

支援なしと支援ありの2群に対して、対話の総単語数が増加するかを調べるために有意水準 5%で片側検定の対応あり t 検定を行った。その結果、グループでは $t(7)=2.181$, $p=.032$ であったため、統計的に有意な差があると言える。同様に、各言語話者の総単語数についても、支援なしに比べて、支援ありの方が対話の総単語数が増加することを調べるために有意水準 5%で片側検定の対応あり t 検定を行った。その結果、日本語話者では、 $t(7)=1.184$, $p=.137$ であったため、統計的に有意な差が認められなかった。一方で、インドネシア話者では、 $t(7)=3.441$, $p=.005$ であったため、統計的に有意な差があると言える。このことから、全体的に対話の情報量が増加しており、特に、インドネシア話者の情報量が増加している。インドネシア話者の総単語数が増えた理由として、参加者に与

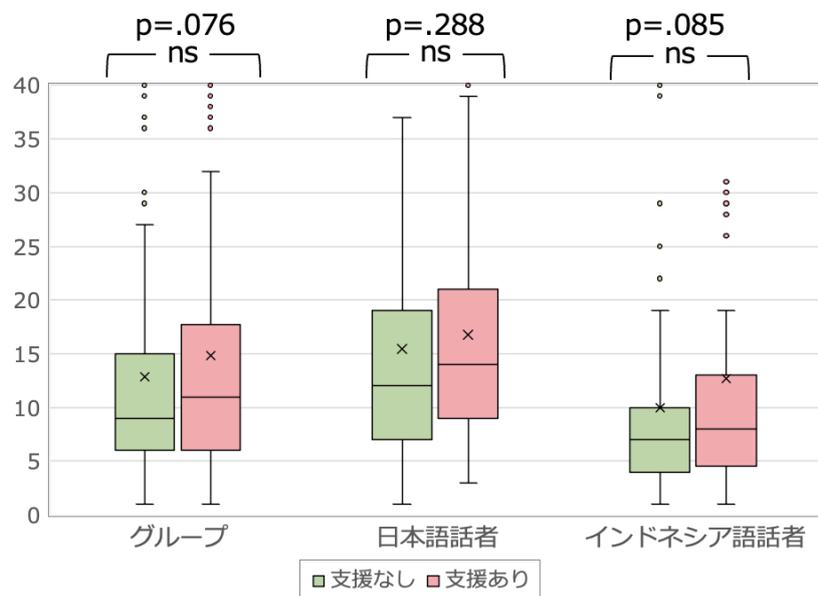


図 5.6: 1 発話の単語数

えた 8 個中 6 個のアイテムがインドネシア語の単語の方が広義な概念を表現している。そのため、日本語話者がより概念の差に気づきやすいと考えられる。それによって、日本語話者からインドネシア語話者にどういった内容かを問いかけ、それに応じてインドネシア語話者が単語の説明をしたため、インドネシア語話者の情報量が増えたと考えられる。

次に、1 発話の単語数 (発話長) と発話数に分け、情報量が増加した要因を分析する。まず、支援ありと支援なしの場合での参加者の 1 発話の単語数を比較する。

図 5.6 は会話ログから各発話ごとの単語数で作成した箱ひげ図である。支援なしでの単語数と支援ありの 2 群に対して、1 発話の単語数が増加するかを調べるために有意水準 5% で片側検定の対応なし t 検定を行った。t 検定の結果、 $t(405)=1.42$, $p=.076$ であったため、発話長に有意な差はなかった (図 5.6 の ns(No Significant))。しかしながら、統計的に有意な差は認められなかったが、箱ひげ図から支援ありの方が 1 発話の単語数は増加傾向にある。

次に、各言語話者ごとに1発話の単語数が増加するかを分析する。実験の会話ログを日本語話者とインドネシア語話者の発話で分類し、支援ありのタスクでの発話と支援なしのタスクでの発話で分析した。図5.6内の“日本語話者”と“インドネシア語話者”は各言語話者ごとに発話を分けて、1発話の単語数を箱ひげ図にしたものである。それぞれの言語話者ごとに先ほどと同様に支援ありのタスクでの発話と支援なしでの発話を分けている。同様に、対応なしt検定で単語数の増加に有意な差があるかを確認した。その結果、日本語話者は $t(219)=0.558$, $p=.288$ で、インドネシア語話者は $t(188)=1.539$, $p=.085$ であったため、支援による1発話の単語数に有意な差は見られなかった。しかしながら、インドネシア語話者の単語数においては、p値が10%未満であった。さらに、図5.6のインドネシア語話者の箱ひげ図から、単語数の平均値や中央値、第三四分位数の値が支援ありの方が支援なしに比べて高い。このことから、支援を入れることによって、インドネシア語話者は1発話の単語数が増加傾向にある。また、日本語話者でも統計的に有意な差は認められなかったが、箱ひげ図の平均値、第一四分位数や第三四分位数が支援なしでの発話よりも支援ありでの発話の方が高いことから、全体的に1発話の単語数は多い傾向にある。

次に、対話における参加者の発話回数を分析する。具体的には、各タスクにおける話者の発話回数をもとに支援を入れる場合と入れない場合の2群間で比較する。図5.7は各タスクの発話回数をもとにタスクの平均発話数からの偏差を集計した値で作成した箱ひげ図である。図5.7内の“グループ”はグループの日本語話者とインドネシア語話者の両方の発話を対象としている。これに対し、支援なしの場合に比べて、支援ありの場合は発話回数が上昇したかを見るために、有意水準5%で片側検定の対応ありt検定を行った。その結果、 $t(7)=0.995$, $p=.178$ であったため、統計的に有意な差は認められなかった。次に、図5.7内の“日本語話者”と“インドネシア語話者”は各言語話者のみを対象とした箱ひげ図である。先ほどと同様に、支援なしの場合に比べて、支援ありの場合は発話回数が上昇したかを見るために、有意水準5%で片側検定の対応ありt検定を行った。その結果、日本語話者は $t(7)=0.982$, $p=.179$ で、インドネシア語話者は $t(7)=0.861$, $p=.208$ であったため、統計的に有意な差は認められなかった。

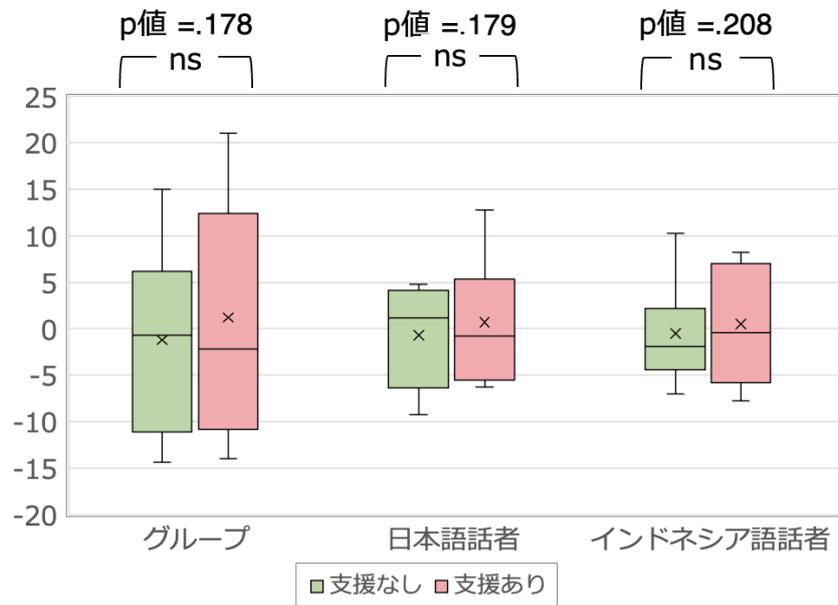


図 5.7: 対話の発話数

表 5.6: アイテムについて言及した発話数

	グループ							
	1	2	3	4	5	6	7	8
支援なし	5	13	23	18	9	19	2	20
支援あり	11	18	25	13	14	26	11	44

サバイバルタスクでは、ランキングを一致させるために、相手の意見に同意する内容の発話や、ランキングについて確認する発話が含まれる。そのため、よりアイテムについて言及した発話のみを取り出し、アイテムについての議論が増えているのかを確認した。表 5.6 はアイテムについて言及した発話の回数を集計したものである。支援をいれることによって、アイテムについての情報量が増えるかを検証するために、この集計表に対して有意水準 5% で片側検定の対応あり t 検定を行った。その結果、 $t(7)=2.785$, $p=.014$ であったため、統計的に有意な差があると言

える。

表 5.7: 1 アイテムあたりの発話数

	グループ							
	1	2	3	4	5	6	7	8
支援なし	1.25	3.25	5.75	4.5	2.25	4.75	0.5	5
支援あり	3.67	4.5	6.25	3.25	3.5	6.5	2.75	11

また、グループによっては個人のランキングが一致している場合、一致していたアイテムについて言及されないことがある。そのため、言及されたアイテム数をカウントし、アイテムについて言及した発話数を割った。表 5.7 はその集計表である。支援ありの場合は支援なしの場合に比べて、1 アイテムあたりの発話回数が上昇したかを見るために、この集計表に対して有意水準 5% で片側検定の対応あり t 検定を行った。その結果、 $t(7)=3.11$, $p=.008$ であったため、統計的に有意な差があると言える。

以上の結果から、総単語数が増加していたことから、対話の情報量は増えており、その理由として、1 発話の単語数や発話回数に統計的に有意な差はないものの箱ひげ図から支援ありの方が情報量が増えたと考えられる。

5.4.3 ユーザの会話の理解度

ユーザの会話の理解度を測るために、アンケートの結果を分析する。具体的には、Q1 と Q2 の質問の結果を集計し、統計検定によって支援エージェントが介入したときの効果をみる。Q1 では、自分が会話をどれだけ理解できたかを確認しているため、支援を入れた場合と入れない場合でその差を比較する。Q1 の“相手の言っていることがよくわかった (1: 0%, 2: 25%, 3: 50%, 4: 75%, 5: 100%)” のアンケートの集計結果が表 5.8 である。この表に対して、有意水準 5% でウィルコクソンの符号付き順位検定を行ったところ、 $W=15.5$, $p=.719$ であったため、統計的に有意な差は認められなかった。

表 5.8: Q1(相手の言っていることがよくわかったか?)の集計表

	参加者															
	日本語話者								インドネシア語話者							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
支援なし	4	4	4	4	4	2	4	4	5	4	5	5	5	3	5	4
支援あり	4	4	2	5	4	4	3	2	5	4	4	5	5	5	5	3

次に, Q2 の相手が自分の言っていることを理解していると感じたかについて分析をする. Q2 の“相手はあなたの言っていることを理解していました?(1: 0%, 2: 25%, 3: 50%, 4: 75%, 5: 100%)”のアンケートの集計結果が表 5.9 である. この表に対して, 有意水準 5% でウィルコクソンの符号付き順位検定を行ったところ, $W=2.5$, $p=.027$ であったため, 統計的に有意な差がある. 集計表の数値から, 支援を入れることによって, 話者は相手が自分の言っていることを理解していないと感じている. 加えて, Q1 の分析結果のように, 参加者はどちらの場合でも, 自分は理解できていると感じているが, Q2 の分析結果のように, 相手から客観的に理解しているかを確認した時, 支援ありと支援なしで差がある. また, 5.4.1 項の Q3 の分析結果で支援を入れることで各話者は文化差の認識が向上するとわかる.

表 5.9: Q2(相手はあなたの言っていることを理解していましたか?)の集計表

	参加者															
	日本語話者								インドネシア語話者							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
支援なし	5	3	4	4	4	5	5	5	5	4	4	5	5	5	5	4
支援あり	2	3	2	5	4	4	3	2	4	4	3	5	5	5	5	4

Q1, Q2, Q3 の結果を踏まえて, 支援なしでは各話者は自身も相手も理解できていると感じているが, 文化差による誤解に気づいていない. 一

方で、支援をいれることで、文化差の認識が高まり、客観的に相手が認識していないことに気づくことができたとわかる。

5.5 定性的分析

5.5.1 認識の乖離

表 5.10 はあるグループの支援なしのタスクの会話ログである。このタスクでは、アイテムセットが表 5.4 の (1) ワイン - arak (2) ポート - kapal (3) ランプ - lampu (4) カッター - pemotong の 4 つのアイテムであった。ID:1~2 の発話で日本語話者は個人ランキングとその理由を説明している。このグループでは、ポートやワインよりもランプとカッターがより必要であると考えていた。そして、ランプとカッターのどちらがより必要かという話題になった際に、日本語話者は自身の考えをもとに一位にランプ、二位にカッターの順番を推している (ID:17 の発話)。一方で、インドネシア語話者はランプは夜間照明専用であるが、カッターはたくさん用途があることを ID:18 と ID:19 の発話で述べている。その意見を聞いて、日本語話者は一位をカッター、二位をランプにすることを承諾している。この会話では、齟齬が起こらずに進んでいるように見られるが、文化差による問題が表面化していないだけである。実際に、日本語話者が考えるランプは、火を使って光を生み出す照明器具で、ガラスのケースの中に蝋燭やオイルに火が灯されたものである。一方で、インドネシア語の lampu はさまざまな照明器具を指しており、今回の文脈では、lampu の翻訳結果はランプになっているが、文脈によってはライトと翻訳される場合もある。そのため、天井照明など光源となるものをインドネシア語話者は考えていた。次に、日本語話者が考えるカッターは、カッターナイフであり、紙などを切るために使用される小型の刃物である。一方で、pemotong は刃物全般を表現する単語であるため、包丁や剣やサバイバルナイフを考えている場合がある。そのため、インドネシア語話者は身を守ることや食料調達や避難所の作成など多目的に使えることを述べている。このように、実際には日本語話者とインドネシア語話者が考えていることは異なるが、主な使用用途や翻訳結果が一致している

ことから、お互いの食い違いが表面化していない。

表 5.10: 話者の認識が異なっているケース

ID	Sender	Message
1	Japanese	私はランプ, カッター, ポート, ワインの順番になりました.
2	Japanese	ランプは灯りを点せませし, カッターは食料の調達, 加工に使えます. ポートは沖に出れます. ワインはいらないと思います.
16	Indonesian	一時的に滞在しなければならぬ場合は, カッターとライトが必要です.
17	Japanese	では, 一位をランプ, 二位をカッターでいいですか
18	Indonesian	カッターは非常に多用途で, 避難所を作り, 食料を集め, 信号や保護のために火を起こすのに役立ちます. 鋭利な道具がなければ, 特に過酷な環境で材料を準備する場合, 生き残ることがより困難になる可能性があります
19	Indonesian	ランプは夜間照明専用です. カッターがあれば薪から火を起こして光を作ることができます.
20	Japanese	一位をカッター, 二位をランプにしますか?
21	Indonesian	はい, 同意します.

5.5.2 文化差によって会話に齟齬が起こったケース

5.5.1 項では, 食い違いが起こっていたが, それが表面化していなかった. 一方で, それが表面化したケースを述べる. 前項で取り上げたグループとは, 別のグループの支援なしのタスクの会話ログを見ていく. このグループでは, 文化差によって会話に齟齬が生じた.

表 5.11: 文化差によって齟齬が起こったケース

ID	Sender	Message
9	Japanese	2位に何を選びましたか?
10	Indonesian	キャンドルを選びました.
11	Indonesian	夜になるとキャンドルの明かりが灯り 捜索中に救助者が私たちを 見つけやすくなるからです
12	Japanese	4つの落ちているアイテムは 何でしたか? 私は, ワックス, テープ, 毛皮, 鏡でした.
13	Indonesian	私にとっては鏡, キャンドル, リボン, 羽
14	Japanese	同じアイテムだと思ってました.
15	Indonesian	そうのようです.
22	Japanese	1位. 鏡, 2位. キャンドルは 決定でいいと思います
23	Indonesian	ランク3と4はどうでしょうか
24	Japanese	順当に進めて3と4は リボンかテープでいいと思います.
25	Indonesian	どうやって, 3. リボン, 4. 羽?
30	Japanese	二人のランキングは 1. 鏡, 2. キャンドル, 3. <u>テープ</u> , 4. <u>リボン</u> でいいですか?
31	Indonesian	ごめんなさい, 3と4は同じです
32	Indonesian	1. 鏡, 2. リボン, 3. キャンドル, 4. 羽根
33	Japanese	ワックスとキャンドルは同じですか?

表 5.11 は, このグループの会話のログである. このタスクでは, アイテムセットが表 5.3 の (1) ワックス - lilin (2) テープ - pita (3) 毛皮 - bulu (4) 鏡 - kaca muka の 4 つのアイテムであった. ID:10 の発話では, インドネシア語話者が “lilin” を 2 位にランキングしたことを述べているが, lilin は蠟燭なども含む広義な言葉である. そのため, lilin を翻訳した結

果がキャンドルとして、日本語話者には表示されている。日本語話者はアイテムにあるワックスではなく、キャンドルを説明されたため、違和感を持ち、ID:12の発話のようにアイテムが何であったかを聞いている。それに対して、インドネシア語話者はID:13の発話のように、4つのアイテムを答えている。この時、各言語話者に表示されている内容が表5.12である。ID:12の“ワックス”、“テープ”、“毛皮”、“鏡”の4つのアイテムに対して、ID:13のインドネシア語話者の発話では、“鏡”、“キャンドル”、“リボン”、“羽”と表示されている。このことから、表5.11のID:14の発話のように、日本語話者は各話者には異なるアイテムが渡されたと思っている。一方で、表5.12のID:12とID:13の発話のインドネシア語の表示では、“cermin”、“bulu”、“pita”、“lilin”と同じになっている。そのため、インドネシア語話者は同じアイテムが渡されたと思っている。日本語話者はお互いのアイテムが違うと思っているため、表5.11のID:30の発話のように、ランキングにテープとリボンを入れている。そして、インドネシア語話者から間違いを指摘されたことにより、誤解に気づいている。

表 5.12: 表 5.11 の ID12, 13 の発話の各言語の表示結果

ID	Japanese	Indonesian
12	4つの落ちているアイテムは何でしたか？ 私は、 <u>ワックス</u> 、 <u>テープ</u> 、 <u>毛皮</u> 、 <u>鏡</u> でした。	Apa saja empat benda yang jatuh itu? Saya memiliki <u>cermin</u> <u>bulu</u> dari <u>pita</u> <u>lilin</u> .
13	私にとっては <u>鏡</u> 、 <u>キャンドル</u> 、 <u>リボン</u> 、 <u>羽</u>	untuk saya <u>cermin</u> , <u>lilin</u> , <u>pita</u> dan <u>bulu</u>

このように、表5.11では、文化差によって誤解があったときに、どちらか一方の話者が違和感を感じているが、相手話者に訊ねている。しかし、この発話にも文化差によって、齟齬が生まれているため、誤解に気づかないまま会話が進んでいる。

5.5.3 支援後の認識のすり合わせ

支援後に、お互いの話者がアイテムの認識をすり合わせたケースを述べる。ここでは、5.5.2項で紹介したグループと異なるグループの会話ログを見る。先ほどのグループと同じアイテムセットであるが、このグループでは支援を行った。表 5.13 はその会話のログの一部である。

表 5.13: 支援後に認識のすり合わせが見られる会話ログ

ID	Sender	Message
1	Indonesian	アイテムの順序はフェイスグラス, キャンドル, リボン, 羽根だと思います。フェイスミラー: 太陽光を反射し, 救助者の注意を引くため, キャンドル: 火をおこすのに役立ちます リボン: 結んだり, 目印にするのに便利です 羽毛: カイロとして使用可能
2	Japanese	私は毛皮, 鏡, テープ, ワックスの順です。夜の体温低下を防ぐために毛皮は重要であり, 鏡は割ればナイフとしても使用できます。テープは水や湿気に弱く, ワックスは使用用途が思いつきませんでした。
5	Agent	“毛皮” > “bulu” インドネシア語より日本語の方が意味が広いから注意
7	Japanese	(Indonesian) さん, 毛皮は動物の毛皮, という意味で考えていたのですが合っていますか?
8	Indonesian	はい, 毛皮は動物の毛皮, という意味で考えています。

ID:1 と ID:2 の発話のように、インドネシア語話者と日本語話者は各自の考えるランキングを伝えている。それに対し、ID:5 の発話ではエージェントによる支援がされている。支援後の発話 (ID:7) では、日本語話

者がアイテムの認識をすり合わせるために、インドネシア語話者に“毛皮は動物の毛皮, という意味で考えていたのですが合っていますか”と問いかけている。それに対し、インドネシア語話者は“はい, 毛皮は動物の毛皮, という意味で考えています。”と自身の認識を相手に伝えている。このように、支援後に各話者がアイテムの認識をすり合わせる行為が見られた。

5.5.4 支援効果の継続性

次に、支援の効果が継続して他のタスクでも続いたケースを取り上げる。表 5.14 は 5.5.3 項で取り上げたグループの別のタスクを行った時の会話ログである。このグループは、支援ありでのタスクを行った後、支援なしのタスクを行った。表 5.14 はそのうちの支援なしのタスクの会話ログである。このタスクでは、エージェントによる支援を行っていないが、参加者が表 5.13 のタスクと同様にアイテムの認識が違う可能性を考慮した振る舞いが見られた。ID:2 の発話では、日本語話者が開始直後にアイテムの認識をお互いの話者ですり合わせようとしていることが見られる。

このように、エージェントの支援を受けた参加者は、次のタスクでも認識に違いがある可能性を考慮している。それによって、このタスクでも各発話でより詳細な発話を心がけていると考えられる。そのため、文化差検出によるアラートを行うことで各話者の発話の詳細度を上げる効果があると考えられる。加えて、最初のタスクで支援ありを行った直後のタスクでは、支援なしの会話でも各話者が文化差を考慮した振る舞うといった効果の継続性を見せている。

表 5.14: 支援の効果が次のタスクでも続いている例

ID	Sender	Message
1	Indonesian	私の意見では、船、カッター、ライト、ワイン、の順番が正しいと思います。(船が損傷しているか、自分を救うために運転できない場合は、船が3番目になります)
2	Japanese	まず、4つのアイテムについて、 ワイン=葡萄酒、ボート=木製の小型船 ランプ=手持ちの明かり カッター=小さくて薄い刃のナイフ なのですがありますか？
3	Japanese	私はボート、ランプ、カッター、 ワインの順で考えています。
4	Japanese	ランプとカッターについて、ランプは夜に 動物を避けたり目立つために利用でき カッターは小さく薄いため木を切ることが できないと考えました
5	Indonesian	私も (<i>Japanese</i>) さんの意見に同意しますが、 私が理解しているカッターとは、 木を切り倒すのに十分な強度を持つナタ、斧、 または切断機のような鋭利なものです

5.6 まとめ

提案した概念レベルの文化差検出と想起イメージレベルの文化差検出の効果を検証するために、日本語話者とインドネシア語話者で多言語コミュニケーションを行う実験を行った。実験では、文化差検出によって、文化差のある単語が使用された場合に、各話者にアラートを出すエージェントを用意し、エージェントの有無によって参加者の文化差の認識や対話の情報量、会話の理解度が変わるかを検証した。

まず、アラートによって参加者の文化差の認識を検証するために、アンケート(Q3: アイテムに対して、概念が違うと感じたアイテムの数)の結果を元に、統計的に有意な差があったかを確認した。その結果、アラートを提示することで参加者の文化差の認識が上がるということがわかった。また、文化差の認識が上がることで、話者がお互いの認識をすり合わせるため、対話の情報量が増えることを検証した。このとき、対話の総単語数をもとに統計検定を用いて分析した結果、対話の総単語数が増加することがわかった。このことから、アラートを提示することで対話の情報量が増加したと言える。特に、インドネシア語話者の情報量が増えており、その理由として、今回の実験で使用した単語はインドネシア語の方が概念が広がった。そのため、日本語話者が先に気づき、インドネシア語話者に問いかけ、インドネシア語話者が説明を行ったことが考えられる。次に、各話者の会話の理解度が向上することを確かめるために、アンケート(Q1: 相手の言っていることがよくわかったか?, Q2: 相手はあなたの言っていることを理解していましたか?)の二つの結果をもとに、統計検定を行った。その結果、Q1の参加者自身が思う理解度に変わりはなかったが、Q2の相手の理解度は下がることがわかった。このことから、アラートを出すことによって、話者はより慎重に発話を行うようになった。

今回の支援では会話のテンポが損なわれることがないように、文化差のある単語とどちらが広義かを表示する内容のみを示し、話者に注意を与えた。一方で、今後の展望として、WordNetのsynsetの説明や画像を用いて、両話者の認識の違いを特定し、その溝を埋めることができる支援が考えられる。

第6章 結論

6.1 研究の貢献

技術の発展により多言語でのコミュニケーションの支援環境が整いつつある。これによって、異なる言語・文化の話者とのコラボレーションが期待でき、社会や技術の発展が促進される。特に、機械翻訳による言語面の支援は効果的である。

しかしながら、翻訳が正しい場合でも言語や文化の違いによってコミュニケーションに齟齬が起こる。このような非対称性による問題を解決するために、本研究では、2つの文化差検出を構築した。それらを用いて文化差を自動的に検出することで、多言語コミュニケーション上で文化差のある単語が使われたときに警告を行う支援をした。

本研究の貢献を以下に示す。

- 概念レベルの文化差検出の構築

各言語の概念体系の違いから、翻訳結果として出力される単語間においても表現する概念が異なる場合がある。このような概念レベルの文化差を自動的に検出するために、WordNet と Multilingual WordNet を用いて概念の文化差を検出する手法を提案した。

WordNet 上での概念の最小単位である synset を基準に各単語が表現する概念の範囲を定量化した。これによって、単語の概念を比較することが可能となり、異言語間で単語の概念範囲を比較することで概念差のある単語を検出した。

この概念レベルの文化差の検出精度を評価するために、人手で概念差があるデータを作成した。人手での概念差データの作成には、画像と単語を用いたアンケートによって各単語に概念差があるか

を判定し、ラベリングを行った。作成した評価データをもとに概念レベルの文化差検出手法の評価を行った結果、既存研究よりも適合率、再現率、F値が高かった。また、どのようなデータに対して提案手法が効果的かを分析した。概念範囲のトポロジーをもとに文化差を分類し、タイプごとに分析を行った結果、タイプによって検出精度に差があることがわかった。特に、水平(部分共通)タイプがもっとも検出精度が高かった。

- 想起イメージレベルの文化差検出の構築

概念レベルの文化差がない翻訳ペアでも、話者間の文化背景の違いから想起するイメージが異なる場合がある。このような想起イメージレベルの文化差を自動的に検出するために、Web上の画像から抽出される特徴ベクトルに基づいて検出する手法を提案した。

単語のイメージの違いを検出するために、単語から画像検索で取得される画像の類似性を調べる。各言語の単語をキーワードに画像検索を行い、画像を取得する。取得された画像の類似性を調べるために、VGG16を用いて画像の特徴ベクトルを生成し、2言語間の特徴ベクトルの類似性を算出した。ベクトル間の類似度から自動的に文化差を判定するために基準となる閾値の最適化を行い、自動的な文化差検出を構築する。提案手法の評価の結果、最適な閾値は0.55でその時の検出精度は76%であった。

提案手法の分析として、WordNetの概念の種類に基づいて、検出精度を確認したところ、概念によって検出精度が異なった。特に、抽象的な概念よりも実態を持つ概念を表現する“physical entity”という概念での検出精度が高いことがわかった。

- 多言語コミュニケーションにおける文化差検出手法の効果検証

多言語コミュニケーションを支援するために、コミュニケーションの齟齬の原因となる文化差を2つのパターンに場合分けし、それぞれの検出手法を提案した。これらの検出手法が実際の多言語コミュニケーション上でどの程度効果的かを検証するために、多言語チャットツール上でのコミュニケーションで対照実験を行っ

た. 実験では, 日本語話者とインドネシア語話者を対象にサバイバルタスクという協調作業を行った.

予備実験の LangridChat を介してそれぞれの母語でコミュニケーションを行った時に文化差がコミュニケーションにどのような影響をあたえるか分析した. その結果, 文化差があることによって, 対話に齟齬が生まれることがわかった. 一方で, 文化差を検出し, 各話者にアラートを出すことで話者の振る舞いが変わるケースが見られた.

文化差がある単語がある場合に, アラートを出すエージェントを用意し, エージェントがある場合とない場合でのコミュニケーションの比較実験を行う. エージェントはウィザードオブオズを用いて, 比較実験を行ったところ, エージェントがある場合とない場合で統計的な差は見られなかった. しかし, 全体的にエージェントありでの会話の方が話者の振る舞いが変わった.

6.2 今後の展望

概念レベルの文化差の検出では, synset の密度に基づいて概念の範囲の定量化を行うことが求められる. WordNet では, synset は階層関係によって, synset 間の対応づけが行われている. しかし, 同じ階層にある synset でも同じ粒度の概念を表現しているわけではない. そのため, ある synset では, その子 synset との差が少ないが, 別の synset では, その子 synset との概念的な密度が遠いことがある. そのため, WordNet の synset を木構造にした際に, 木構造の重心や密度から synset 間の概念的な距離を計算することで, より正確な単語の概念範囲を定義することができる. これによって, 人の認識レベルで感じる概念差のみを判定できるようになり, 検出精度が向上すると考えられる.

想起イメージレベルの文化差の検出では, 正確な画像特徴ベクトルの生成が求められる. 提案手法の特性上, 取得された画像をその単語が持つイメージと捉えているが, 画像検索ではノイズとなる画像も取得される可能性がある. そのため, 画像検索のクエリの最適化やノイズ画像の

除去をおこなうことによって、より最適な画像特徴ベクトルを取得することができる。これによって、検出精度の改善が見込まれる。

今後はこれらを用いた支援エージェントの構築が将来的に考えられる。今回の実験では、文化差検出による支援が有効かを検証した。しかし、文化差のある単語に対してウィザードオブオズでアラートを出すものにとどまっている。今後は、これを改善し文化差検出を組み込んだエージェントを構築し、非対称性がある場合には、アラートを出すだけでなく、話者間のギャップがなくなるような支援方法を組み込むことが求められる。また、文化差検出の技術を用いて、翻訳技術の改善が考えられる。現在の機械翻訳では、文化差を考慮した翻訳が乏しいため、文化差を考慮した翻訳結果を出力するように改善することを提案する。

謝辞

本研究を行うにあたり、指導教員として7年間ご指導いただきました、立命館大学情報理工学部教授 村上 陽平先生に深く感謝申し上げます。村上先生には、就職活動や進路など、研究以外の相談にも乗っていただき、温かいご支援をいただきましたことを感謝申し上げます。

また、立命館大学情報理工学部教授 西原 陽子先生、並びに桑原 和宏先生には、本論文の作成にあたり、副査としての的確なご助言をいただきました。深く感謝申し上げます。

加えて、Mondheera Pituxcoosuvarn 助教には、研究や実験において、貴重なご意見・ご協力いただきました。深く感謝申し上げます。

本論文を執筆するにあたり、ごぼうの花の画像(図 1.2 の右)をご提供いただきました公益社団法人東京生薬協会 山上 勉様に深く感謝申し上げます。

研究を行うにあたり、実験にご協力いただきました方々に感謝申し上げます。特に、リアウ・イスラム大学准教授 Arbi Haza Nasution 先生、並びに立命館大学社会知能研究室のOBOG の Abia Herlianto, Shella Fitri, 張 宇軒, 李 闖には参加者を募っていただきました。深く感謝申し上げます。

立命館大学社会知能研究室の皆様には感謝いたします。皆様には、貴重なご意見をいただきました。感謝申し上げます。特に、研究室の同期メンバーとして、7年間一緒に過ごした元澤 海月さんには、研究についての貴重なご意見、サポートをいただきました。深く感謝申し上げます。

そして、9年間の大学生活において、私生活を支えてくださいました私の家族、西村 精一・西村 陽子・西村 優希に深く感謝申し上げます。

本研究は、日本学術振興会科学研究費基盤研究 (B)(21H03561, 令和3年度～令和5年度, 23K21730, 令和6年度, 21H03556, 令和3年度～令和

5年度) 及び日本学術振興会科学研究費若手研究 (21K17794, 令和3年度～令和6年度) の支援を受けました.

参考文献

- [1] Mondheera Pituxcoosuvann, Toru Ishida, Naomi Yamashita, Toshiyuki Takasaki, and Yumiko Mori. Machine translation usage in a children's workshop. In *Collaboration Technologies and Social Computing: 10th International Conference, CollabTech 2018, Costa de Caparica, Portugal, September 5-7, 2018, Proceedings 10*, pp. 59–73. Springer, 2018.
- [2] Mondheera Pituxcoosuvann, Yohei Murakami, Donghui Lin, and Toru Ishida. Effect of cultural misunderstanding warning in mt-mediated communication. In *Proceedings of the International Conference on Collaboration Technologies and Social Computing*, pp. 112–127. Springer, 2020.
- [3] Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan Cernocky, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Interspeech 2010*, pp. 1045–1048, 2010.
- [4] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pp. 746–751, 2013.
- [5] Nal Kalchbrenner and Phil Blunsom. Recurrent continuous translation models. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1700–1709, 2013.

- [6] Nal Kalchbrenner and Phil Blunsom. Recurrent convolutional neural networks for discourse compositionality. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, pp. 119–126, 2013.
- [7] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.
- [8] Ilya Sutskever, Oriol Vinyals, and Quoc Le. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, Vol. 4, , 09 2014.
- [9] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder–decoder approaches. In Dekai Wu, Marine Carpuat, Xavier Carreras, and Eva Maria Vecchi, editors, *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pp. 103–111, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [10] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [11] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations*, 2015.

- [12] Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li. Modeling coverage for neural machine translation. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 76–85, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [13] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [14] Toru Ishida. Language grid: An infrastructure for intercultural collaboration. In *International Symposium on Applications and the Internet (SAINT'06)*, pp. 5–pp. IEEE, 2006.
- [15] Toru Ishida, Yohei Murakami, and Donghui Lin. The language grid: Service-oriented approach to sharing language resources. In *The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability*, pp. 3–17. Springer, 2011.
- [16] Toru Ishida, Yohei Murakami, Donghui Lin, Takao Nakaguchi, and Masayuki Otani. Open language grid—towards a global language service infrastructure. In *The Third ASE International Conference on Social Informatics (SocialInformatics 2014), Cambridge, Massachusetts, USA*. Citeseer, 2014.
- [17] Rieko Inaba, Yohei Murakami, Akiyo Nadamoto, and Toru Ishida. Multilingual communication support using the language grid. In *Intercultural Collaboration: First International Workshop, IWIC 2007 Kyoto, Japan, January 25-26, 2007 Invited and Selected Papers*, pp. 118–132. Springer, 2007.
- [18] Georg Rehm, Maria Berger, Ela Elsholz, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Stelios Piperidis, Miltos Deligiannis, Dimitrios Galanis, Katerina Gkirtzou, et al. European language grid: An overview. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 3366–3380, 2020.

- [19] Georg Rehm. *European Language Grid: A Language Technology Platform for Multilingual Europe*. Springer Nature, 2023.
- [20] 宮部真衣, 吉野孝, 重信智宏. 折返し翻訳を用いた翻訳リペアの効果. 電子情報通信学会論文誌 D, Vol. 90, No. 12, pp. 3141–3150, 2007.
- [21] 宮部真衣, 吉野孝. 折返し翻訳文と対象言語翻訳文の精度不一致要因. 電子情報通信学会論文誌 D, Vol. 95, No. 1, pp. 11–18, 2012.
- [22] 宮部真衣, 吉野孝ほか. 機械翻訳を介したチャットコミュニケーションにおける精度判定に基づく送信拒否の適用可能性. 情報処理学会論文誌, Vol. 51, No. 3, pp. 784–795, 2010.
- [23] 宮部真衣, 吉野孝ほか. 折返し翻訳を用いた高精度なコミュニケーションのための複数翻訳機による精度不一致検出サービスの提案. 情報処理学会論文誌, Vol. 53, No. 8, pp. 2036–2043, 2012.
- [24] 藤井薫和, 重信智宏, 吉野孝ほか. 機械翻訳を用いた異文化間チャットコミュニケーションにおけるアノテーションの評価. 情報処理学会論文誌, Vol. 48, No. 1, pp. 63–71, 2007.
- [25] 宮部真衣, 吉野孝, 重野亜久里. 外国人患者のための用例対訳を用いた多言語医療受付支援システムの構築. 電子情報通信学会論文誌 D, Vol. 92, No. 6, pp. 708–718, 2009.
- [26] 福島拓, 吉野孝, 重野亜久里ほか. 用例対訳と機械翻訳を併用した多言語問診票入力手法の提案と評価. 情報処理学会論文誌, Vol. 54, No. 1, pp. 256–265, 2013.
- [27] 福島拓, 吉野孝, 重野亜久里ほか. 正確な情報共有のための多言語用例対訳共有システム. 情報処理学会論文誌コンシューマ・デバイス & システム (CDS), Vol. 2, No. 3, pp. 23–33, 2012.
- [28] 山本里美, 福島拓, 吉野孝ほか. クラウドソーシングにおける会話文を用いた応答用例対作成手法の提案. 情報処理学会論文誌, Vol. 56, No. 3, pp. 1080–1089, 2015.

- [29] Mahesh Vanjani, Milam Aiken, and Mina Park. Chatbots for multi-lingual conversations. *Journal of Management Science and Business Intelligence*, Vol. 4, No. 1, pp. 19–24, 2019.
- [30] 福島拓, 吉野孝, 喜多千草. 共通言語を用いた対面型会議における非母語話者支援システム panelive の構築. 電子情報通信学会論文誌 D, Vol. 92, No. 6, pp. 719–728, 2009.
- [31] 岡本健吾, 吉野孝ほか. 会話中の名詞の関連情報を用いた対面型異文化間コミュニケーション支援システムの構築と評価. 情報処理学会論文誌, Vol. 52, No. 3, pp. 1213–1223, 2011.
- [32] 中口孝雄, 大谷雅之, 高崎俊之, 石田亨ほか. 国際シンポジウムのための多言語支援システム. 情報処理学会論文誌, Vol. 58, No. 1, pp. 197–204, 2017.
- [33] 笹島宗彦, 井本和範, 下森大志, 山中紀子, 矢島真人, 福永幸弘, 正井康之ほか. 発話意図理解と回答誘導による異言語間会話支援ツールの試作. 情報処理学会論文誌, Vol. 48, No. 3, pp. 1262–1273, 2007.
- [34] Keiji Yanai, Keita Yaegashi, and Bingyu Qiu. Detecting cultural differences using consumer-generated geotagged photos. In *Proceedings of the 2nd International Workshop on Location and the Web*, pp. 1–4, 2009.
- [35] Tomoko Koda. Cross-cultural comparison of interpretation of avatars’ facial expressions. In *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, 2006.
- [36] Tomoko Koda, Toru Ishida, Matthias Rehm, and Elisabeth André. Avatar culture: cross-cultural evaluations of avatar facial expressions. *AI & society*, Vol. 24, pp. 237–250, 2009.
- [37] Xuhai Xu, Nan Qie, and Pei-Luen Patrick Rau. Multimodal features as a novel method for cross-cultural studies. In *Cross-Cultural De-*

sign. Experience and Product Design Across Cultures: 13th International Conference, CCD 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I 23, pp. 527–546. Springer, 2021.

- [38] Takashi Yoshino, Mai Miyabe, and Tomohiro Suwa. A proposed cultural difference detection method using data from japanese and chinese wikipedia. In *2015 International Conference on Culture and Computing (Culture Computing)*, pp. 159–166. IEEE, 2015.
- [39] 諏訪智大, 宮部真衣, 吉野孝ほか. 日本語版・中国語版 wikipedia を用いた文化差検出手法の提案. *情報処理学会論文誌*, Vol. 55, No. 1, pp. 257–266, 2014.
- [40] Mondheera Pituxcoosuvann, Yohei Murakami, and Kaede Miwa. Detection for cultural difference in impression using masked language model. In *International Conference on Human-Computer Interaction*, pp. 569–579. Springer, 2023.
- [41] Danny Livshits, Newton Howard, and Yair Neuman. Can computers help us to better understand different cultures? toward a computer-based culint. In *2012 European Intelligence and Security Informatics Conference*, pp. 172–179. IEEE, 2012.
- [42] Daniel Moritz Marutschke, Sasha Krysanova, and Hitoshi Ogawa. Clustering word co-occurrences with color keywords based on twitter feeds in japanese and german culture. In *2015 International Conference on Culture and Computing (Culture Computing)*, pp. 191–192. IEEE, 2015.
- [43] Cho Heeryon, 石田亨, 山下直美, 稲葉利江子, 高崎俊之, 神田智子. 絵文字解釈における人間の文化差判定. *ヒューマンインタフェース学会論文誌*, Vol. 10, No. 4, pp. 427–434, 2008.
- [44] Heeryon Cho, Toru Ishida, Naomi Yamashita, Rieko Inaba, Yumiko Mori, and Tomoko Koda. Culturally-situated pictogram retrieval.

In *Intercultural Collaboration: First International Workshop, IWIC 2007 Kyoto, Japan, January 25-26, 2007 Invited and Selected Papers*, pp. 221–235. Springer, 2007.

- [45] Heeryon Cho, Toru Ishida, Toshiyuki Takasaki, and Satoshi Oyama. Assisting pictogram selection with semantic interpretation. In *The Semantic Web: Research and Applications: 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, 2008 Proceedings 5*, pp. 65–79. Springer, 2008.
- [46] 西田ひろ子ほか. 異文化コラボレーション: 4. オフショア開発現場における異文化間コミュニケーション摩擦. *情報処理*, Vol. 47, No. 3, pp. 290–294, 2006.
- [47] 照井賢治, 菱山玲子. 多言語コミュニケーション環境における異文化分析. *ヒューマンインタフェース学会論文誌*, Vol. 16, No. 1, pp. 63–76, 2014.
- [48] Ruth Garcia-Gavilanes, Daniele Quercia, and Alejandro Jaimes. Cultural dimensions in twitter: Time, individualism and power. In *Proceedings of the international AAAI conference on web and social media*, Vol. 7, pp. 195–204, 2013.
- [49] Hyemin Chung. *GlobalMind: bridging the gap between different cultures and languages with common-sense computing*. PhD thesis, Massachusetts Institute of Technology, 2006.
- [50] Dhavalkumar Thakker, Stan Karanasios, Emmanuel Blanchard, Lydia Lau, and Vania Dimitrova. Ontology for cultural variations in interpersonal communication: Building on theoretical models and crowdsourced knowledge. *Journal of the Association for Information Science and Technology*, Vol. 68, No. 6, pp. 1411–1428, 2017.
- [51] Paul Laufer, Claudia Wagner, Fabian Flöck, and Markus Strohmaier. Mining cross-cultural relations from wikipedia: a study of 31 euro-

- pean food cultures. In *Proceedings of the ACM web science conference*, pp. 1–10, 2015.
- [52] Ulrike Pfeil, Panayiotis Zaphiris, and Chee Siang Ang. Cultural differences in collaborative authoring of wikipedia. *Journal of Computer-Mediated Communication*, Vol. 12, No. 1, pp. 88–113, 2006.
- [53] Geert H Hofstede, Gert Jan Hofstede, and Michael Minkov. *Cultures and organizations: Software of the mind*, Vol. 2. Mcgraw-hill New York, 2005.
- [54] Thiago Silva, Pedro Vaz De Melo, Jussara Almeida, Mirco Musolesi, and Antonio Loureiro. You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In *Proceedings of the international aaai conference on web and social media*, Vol. 8, pp. 466–475, 2014.
- [55] 元澤海月, 村上陽平, Mondheera Pituxcoosuvann. 児童の多言語コミュニケーションにおけるファシリテータの対話行為分析. ヒューマンインタフェース学会論文誌, Vol. 26, No. 1, pp. 77–86, 2024.
- [56] Mondheera Pituxcoosuvann, Mizuki Motozawa, Yohei Murakami, and Shiyumi Yokote. Facilitator agent to support low-resource language speakers in mt-mediated communication. In *International Conference on Collaboration Technologies and Social Computing*, pp. 231–244. Springer, 2022.
- [57] 宮部真衣, 吉野孝ほか. リアルタイム遠隔テキストコミュニケーションにおける対人許容応答時間の評価. 情報処理学会論文誌, Vol. 50, No. 3, pp. 1214–1223, 2009.
- [58] 船越要, 藤代祥之, 野村早恵子, 石田亨ほか. 機械翻訳を用いた協調作業支援ツールへの要求条件-機-日中韓馬異文化コラボレーション実験からの知見. 情報処理学会論文誌, Vol. 45, No. 1, pp. 112–120, 2004.

- [59] 鈴木宏, 菱山玲子ほか. 多言語環境における専門知識伝達サービスの開発および知識伝達レシピの考案. *情報処理学会論文誌*, Vol. 58, No. 5, pp. 955–966, 2017.
- [60] Kaori Kita, Toshiyuki Takasaki, Donghui Lin, Yuu Nakajima, and Toru Ishida. Case study on analyzing multi-language knowledge communication. In *The International Conference on Culture and Computing (ICCC 2012) was Organized with a Symposium on Digital Media and Digital Heritage to Show the Latest Research and Development Results in the State of the Art on Cultural Computing Technologies And Traditional Culture*, pp. 35–42. Citeseer, 2012.
- [61] Toshiyuki Takasaki and Yumiko Mori. Design and development of a pictogram communication system for children around the world. In Toru Ishida, Susan R. Fussell, and Piek T. J. M. Vossen, editors, *Intercultural Collaboration*, pp. 193–206, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.
- [62] Toshiyuki Takasaki. Pictnet: Semantic infrastructure for pictogram communication. In *The 3rd International WordNet Conference (GWC-06)*, pp. 279–284. Citeseer, 2006.
- [63] 宗森純, 大野純佳, 吉野孝ほか. 絵文字チャットによるコミュニケーションの提案と評価. *情報処理学会論文誌*, Vol. 47, No. 7, pp. 2071–2080, 2006.
- [64] Victoria Abou-Khalil, Toru Ishida, Masayuki Otani, Brendan Flanagan, Hiroaki Ogata, and Donghui Lin. Learning culturally situated dialogue strategies to support language learners. *Research and Practice in Technology Enhanced Learning*, Vol. 13, pp. 1–20, 2018.
- [65] Xiaoyan Li, Naomi Yamashita, Wen Duan, Yoshinari Shirai, and Susan R Fussell. Improving non-native speakers’ participation with an automatic agent in multilingual groups. *Proceedings of the ACM on Human-Computer Interaction*, Vol. 7, No. GROUP, pp. 1–28, 2023.

- [66] Keisuke Tsunoda and Reiko Hishiyama. Design of multilingual participatory gaming simulations with a communication support agent. In *Proceedings of the 28th ACM International Conference on Design of Communication*, pp. 17–25, 2010.
- [67] 仁科喜久子. 理工系留学生のセミナーでの対話理解過程の分析–理工系学生のシラバス作成に向けて. *日本語教育= Journal of Japanese language teaching/日本語教育学会学会誌委員会 編*, No. 84, pp. p40–52, 1994.
- [68] 徳井厚子. 誤解はどこから生まれるか: 留学生と日本人学生のコミュニケーション・ブレイクダウンへの対処をめぐって. PhD thesis, Shinshu University Library, 1995.
- [69] 仁科浩美. 理工系留学生の発表場面における質疑応答の課題 コミュニケーション・ブレイクダウンの観点から. *専門日本語教育研究*, Vol. 16, pp. 37–44, 2014.
- [70] Barry R Chiswick and Paul W Miller. Linguistic distance: A quantitative measure of the distance between english and other languages. *Journal of multilingual and multicultural development*, Vol. 26, No. 1, pp. 1–11, 2005.
- [71] Christiane Fellbaum and Piek Vossen. Challenges for a multilingual wordnet. *Language Resources and Evaluation*, Vol. 46, pp. 313–326, 2012.
- [72] Naomi Yamashita and Toru Ishida. Automatic prediction of misconceptions in multilingual computer-mediated communication. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pp. 62–69, 2006.
- [73] 山下直美, 石田亨ほか. 翻訳機を用いた対話における参照方法に関する分析. *情報処理学会論文誌*, Vol. 48, No. 2, pp. 939–948, 2007.

- [74] Mondheera Pituxcoosuvann, Donghui Lin, and Toru Ishida. A method for automated detection of cultural difference based on image similarity. In *Collaboration Technologies and Social Computing: 25th International Conference, CRIWG+ CollabTech 2019, Kyoto, Japan, September 4–6, 2019, Proceedings 25*, pp. 129–143. Springer, 2019.
- [75] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, Vol. 26, , 2013.
- [76] Xinxiong Chen, Zhiyuan Liu, and Maosong Sun. A unified model for word sense representation and disambiguation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1025–1035, 2014.
- [77] Siddharth Patwardhan and Ted Pedersen. Using wordnet-based context vectors to estimate the semantic relatedness of concepts. In *Proceedings of the Workshop on Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together*, 2006.
- [78] Marianna Apidianaki. Data-driven semantic analysis for multilingual wsd and lexical selection in translation. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pp. 77–85, 2009.
- [79] Daiki Ko and Koichi Takeuchi. Evaluation of embedded vectors for lexemes and synsets toward expansion of japanese wordnet. In *Computational Linguistics: 16th International Conference of the Pacific Association for Computational Linguistics, PACLING 2019, Hanoi, Vietnam, October 11–13, 2019, Revised Selected Papers 16*, pp. 79–87. Springer, 2020.
- [80] Andrés Montoyo and Manuel Palomar. Wsd algorithm applied to a nlp system. In *International Conference on Application of Natural Language to Information Systems*, pp. 54–65. Springer, 2000.

- [81] Eneko Agirre, Oier Lopez De Lacalle, Aitor Soroa, and Informatika Fakultatea. Knowledge-based wsd and specific domains: Performing better than generic supervised wsd. In *IJCAI*, pp. 1501–1506, 2009.
- [82] Els Lefever, Veronique Hoste, and Martine De Cock. Five languages are better than one: an attempt to bypass the data acquisition bottleneck for wsd. In *Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part I 14*, pp. 343–354. Springer, 2013.
- [83] George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, Vol. 3, No. 4, pp. 235–244, 1990.
- [84] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, Vol. 38, No. 11, pp. 39–41, 1995.
- [85] Christiane Fellbaum. Wordnet. In *Theory and applications of ontology: computer applications*, pp. 231–243. Springer, 2010.
- [86] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. Development of the japanese wordnet. In *LREC*, 2008.
- [87] Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. Enhancing the japanese wordnet. In *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*, pp. 1–8, 2009.
- [88] H Chu-Ren, H Shu-Kai, H Jia-Fei, C Yun-Zhu, S I-LI, C Yong-Xiang, and H Sheng-Wei. Chinese wordnet: Design and implementation of a cross-lingual knowledge processing infrastructure. *Journal of Chinese Information Processing*, No. 24, p. 2, 2010.

- [89] Chu-Ren Huang, Ru-Yng Chang, and Hshiang-Pin Lee. Sinica bow (bilingual ontological wordnet): Integration of bilingual wordnet and sumo. In *LREC*, 2004.
- [90] Renjie Xu, Zhiqiang Gao, Yingji Pan, Yuzhong Qu, and Zhisheng Huang. An integrated approach for automatic construction of bilingual chinese-english wordnet. In *The Semantic Web: 3rd Asian Semantic Web Conference, ASWC 2008, Bangkok, Thailand, December 8-11, 2008. Proceedings. 3*, pp. 302–314. Springer, 2008.
- [91] Shan Wang and Francis Bond. Building the chinese open wordnet (cow): Starting from core synsets. In *Proceedings of the 11th Workshop on Asian Language Resources*, pp. 10–18, 2013.
- [92] Shan Wang and Francis Bond. Theoretical and practical issues in creating chinese open wordnet (cow). 2013.
- [93] Desmond Darma Putra, Abdul Arfan, and Ruli Manurung. Building an indonesian wordnet. In *Proceedings of the 2nd International MALINDO Workshop*, pp. 12–13, 2008.
- [94] L Lian Tze and Nur Hussein. Fast prototyping of a malay wordnet system. In *Proceedings of the Language, Artificial Intelligence and Computer Science for Natural Language Processing Applications (LAICS-NLP) Summer School Workshop*, pp. 13–16, 2006.
- [95] Nuril Hirfana Bte Mohamed Noor, Suerya Sapuan, and Francis Bond. Creating the open wordnet bahasa. In *Proceedings of the 25th Pacific Asia conference on language, information and computation*, pp. 255–264. Waseda University, 2011.
- [96] Sareewan Thoongsup, Thatsanee Charoenporn, Kergrit Robkop, Tan Sinthurahat, Chumpol Mokarat, Virach Sornlertlamvanich, and Hitoshi Isahara. Thai wordnet construction. In *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*, pp. 139–144, 2009.

- [97] Ae-Sun Yoon, Soon-Hee Hwang, Eun-Ryoung Lee, and Hyuk-Chul Kwon. Construction of korean wordnet. *Journal of KIISE: Software and Applications*, Vol. 36, No. 1, pp. 92–108, 2009.
- [98] Piek Vossen. Eurowordnet: a multilingual database of autonomous and language-specific wordnets connected via an inter-lingualindex. *international journal of Lexicography*, Vol. 17, No. 2, pp. 161–173, 2004.
- [99] Marten Postma, Emiel Van Miltenburg, Roxane Segers, Anneleen Schoen, and Piek Vossen. Open dutch wordnet. In *Proceedings of the 8th Global WordNet Conference (GWC)*, pp. 302–310, 2016.
- [100] Benoît Sagot and Darja Fišer. Building a free french wordnet from multilingual resources. In *OntoLex*, 2008.
- [101] Magdalena Derwojedowa, Maciej Piasecki, Stanisław Szpakowicz, Magdalena Zawisławska, and Bartosz Broda. Words, concepts and relations in the construction of polish wordnet. In *Proceedings of GWC*, pp. 162–177, 2008.
- [102] Ewa Rudnicka, Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. A strategy of mapping polish wordnet onto princeton wordnet. In *Proceedings of COLING 2012: Posters*, pp. 1039–1048, 2012.
- [103] Ewa Rudnicka, Lukasz Grabowski, Maciej Piasecki, and Tomasz Naskret. In search of gaps between languages and wordnets: the case of polish-english wordnet. *International Journal of Lexicography*, Vol. 36, No. 1, pp. 68–92, 2023.
- [104] Francis Bond and Ryan Foster. Linking and extending an open multilingual wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1352–1362, 2013.

- [105] Francis Bond, Luis Morgado Da Costa, Michael Wayne Goodman, John Philip McCrae, and Ahti Lohk. Some issues with building a multilingual wordnet. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 3189–3197, 2020.
- [106] 柳井啓司ほか. 一般物体認識のための単語概念の視覚性の分析. 情報処理学会論文誌コンピュータビジョンとイメージメディア (CVIM), Vol. 48, No. SIG1 (CVIM17), pp. 88–97, 2007.
- [107] Bernice E Rogowitz, Thomas Frese, John R Smith, Charles A Bouman, and Edward B Kalin. Perceptual image similarity experiments. In *Human Vision and Electronic Imaging III*, Vol. 3299, pp. 576–590. SPIE, 1998.
- [108] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. Learning fine-grained image similarity with deep ranking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1386–1393, 2014.
- [109] Qin Lv, Moses Charikar, and Kai Li. Image similarity search with compact data structures. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pp. 208–217, 2004.
- [110] Jonathan J Hull. Document image similarity and equivalence detection. *International Journal on Document Analysis and Recognition*, Vol. 1, No. 1, pp. 37–42, 1998.
- [111] K Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations (ICLR 2015)*. Computational and Biological Learning Society, 2015.
- [112] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE*

conference on computer vision and pattern recognition, pp. 770–778, 2016.

- [113] Dhananjay Theckedath and RR Sedamkar. Detecting affect states using vgg16, resnet50 and se-resnet50 networks. *SN Computer Science*, Vol. 1, No. 2, p. 79, 2020.
- [114] Sheldon Mascarenhas and Mukul Agarwal. A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification. In *2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)*, Vol. 1, pp. 96–99. IEEE, 2021.
- [115] 吉野孝, 宮部真衣ほか. 文化差データの収集サービスの提案. 第75回全国大会講演論文集, Vol. 2013, No. 1, pp. 515–516, 2013.
- [116] Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka, and Toru Ishida. Difficulties in establishing common ground in multiparty groups using machine translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 679–688, 2009.
- [117] J Clayton Lafferty and Alonzo William Pond. *The Desert Survival Situation: Manual: a Group Decision Making Experience for Examining and Increasing Individual and Team Effectiveness*. Human Synergistics, 1974.

研究業績

主要論文

論文誌

1. 西村 一球, 村上 陽平, Mondheera Pituxcoosuvam, “画像特徴量に基づく文化差検出,” *ヒューマンインタフェース学会論文誌*, Vol. 23, No. 2, pp. 145–152, 2021.

国際会議

1. Ikkyu Nishimura, Yohei Murakami, Mondheera Pituxcoosuvam, “Image-Based Detection Criteria for Cultural Differences in Translation,” in *Collaboration Technologies and Social Computing : 26th International Conference (CollabTech2020)*, pp. 81-95, 2020.
2. Ikkyu Nishimura, Yohei Murakami, Mondheera Pituxcoosuvam, “Evaluation Dataset for Cultural Difference Detection Task,” in *International Conference on Human-Computer Interaction (HCI2022)*, pp. 357-369, 2022.
3. Ikkyu Nishimura, Yohei Murakami, Mondheera Pituxcoosuvam, “Extracting Conceptual Differences between Translation Pairs Using Multilingual WordNet,” in *Global WordNet Conference (GWC2025)*, 2025, To be published.

その他

国内会議

1. 西村 一球, 村上 陽平, Mondheera Pituxcoosuvann, “異常値検出を用いた単語の想起イメージの文化差検出,” 第 85 回情報処理学会全国大会講演論文集, pp. 3–4, 2023.
2. 西村 一球, 村上 陽平, Mondheera Pituxcoosuvann, “概念差検出による多言語コミュニケーション支援,” 電子情報通信学会ヒューマンコミュニケーション基礎 (HCS) 研究会, 2025.