

Language Management in a Japanese Multinational Company: A Data-Driven Approach

Ziyuan Zhang

NHH



Department of Professional and Intercultural Communication

Norwegian School of Economics

Helleveien 30, 5045 Bergen

Norway

(+47) 46342826

ziyuan.zhang@nhh.no

Preface

The motivation for this PhD thesis originally emanated from my working experience in Rakuten. I worked as a software engineer in Rakuten from 2014 to 2017. I also have a Master's degree in Software Engineering from Peking University, and hands-on experience executing data-driven solutions to increase efficiency, accuracy, and utility of data processing. I am experienced at comparing generative statistical models, using dynamic topic modeling, and analyzing data mining algorithms to deliver insights and implement action-oriented solutions to complex business problems.

Abstract

Globalization poses a challenge for businesses with linguistically diverse staff, prompting the choice of English as the default corporate language. In Japan, research on the use of English in business contexts from both corporate and employees' perspectives is very limited, let alone studies adopting a data-driven approach. This study focuses on Rakuten, a Japanese multinational corporation (MNC), with the aim of illustrating the key challenges the company faces when it adopts English as its official language. The research is interdisciplinary and is positioned at the intersection of business communication, computational sociolinguistics, and language management. The first article, "Content analysis of language-sensitive recruitment influenced by corporate language policy using topic modeling", explores the match (or mismatch) between language-sensitive recruitment (English, Japanese, or bilingual) and corporate language policy. The second article, "It is all about TOEIC: discovering topics and trends in employee perceptions of corporate language policy", examines the barriers in multinational companies that have adopted a foreign language and analyzes employees' attitudes. The third and final article, "Analyzing cultural expatriates' attitude toward 'Englishnization' using dynamic topic modeling", investigates changes in employee' perceptions of Japanese work practices and values over time. The results of my study have implications for the implementation of language-sensitive recruitment in a multilingual corporate context. Furthermore, the thesis also highlights the evolutionary nature of corporate language policy topics by exploring and categorizing large amounts of text. Overall, the results presented in the three articles expand the understanding of the challenges associated with the

use of English in a Japanese business.

Keywords

language management; language-sensitive recruitment; employee perspective;
bilingual job advertisements; computer-assisted text analysis; topic modeling

Acknowledgements

My PhD thesis is the result of an exciting adventure. Many people have contributed to this thesis in important ways.

First and foremost, I would like to thank my supervisor, Gisle Andersen, for motivating and supporting me all along the way. This thesis is only possible because of his skilled guidance and insightful comments on my work. I would also like to express my gratitude to my co-supervisor, Kristin Rygg, for her detailed comments throughout the writing of this thesis. Further, the thesis also benefited greatly from feedback by Yao Liu, my external supervisor.

I want to thank my colleagues at the department of Professional and Intercultural Communication at NHH for making the four years so memorable. Beate Sandvei, Margrete Cardona, Trine Dahl, Claudia Förster Hegrenæs, Bjørnulf Hinderaker, Christian Langerfeld, Kaisa Pietikäinen, Kari Johanne Oma, Klara Sjo, Ingrid Simonnæs, Maja Dame, Silje Grønner Stang, Rikke Salbu and Fitia Andria are just some of those who deserve my gratitude. Thank you Anne Kari Bjørge and Annelise Ly for reading and commenting my article and providing me with constructive feedback.

My fellow PhD scholars at the department of Professional and Intercultural Communication, Victoria S. Nydegger Schrøder, José Luis Rojas Díaz, Ina Celise Sortland and Agnes Bamford also deserve my gratitude for the mutual support and the friendship.

I would like to express my deepest gratitude to my parents for caring about me from afar.

I would also like to thank the evaluation committee members for their

thoughtful suggestions and insights. The thesis has benefited from these insightful suggestions.

Last but not the least, special thanks go to Mario and Melody for their emotional support and endless encouragement.

“乘風破浪会有時, 直掛雲帆濟滄海。”

Oslo, June 2022

Ziyuan Zhang

Contents

List of Figures	10
List of Tables	11
1 Introduction	12
1.1 Language management in Japanese multinational corporations	12
1.2 Research questions	18
1.3 Research setting: Rakuten	19
1.4 The structure of the thesis	20
2 Outline of the articles	22
2.1 Article 1: Content analysis of language-sensitive recruitment influenced by corporate language policy using topic modeling	22
2.2 Article 2: It is all about TOEIC: Discovering topics and trends in employee perceptions of corporate language policy	23
2.3 Article 3: Analyzing cultural expatriates’ attitudes toward “Englishnization” using dynamic topic modeling	25
3 Theoretical framework	27
3.1 Language issues in international business	27
3.1.1 The concept of language in international business	27
3.1.2 The concept of business language and communication	28
3.1.3 The concept of English	29
3.1.4 The concept of lingua franca and English as lingua franca .	30
3.1.5 The concept of BELF	30

3.1.6	The concept of English as corporate language	31
3.2	Language management	33
3.2.1	Language management at the corporate level	35
3.2.2	Language management at the employee level	37
3.2.3	The concept of corporate language policy	40
3.2.4	The concept of linguistic instrumentalism	42
3.3	The concept of cultural expatriates	45
3.4	Computational sociolinguistics	47
4	Methodology	50
4.1	Data presentation	50
4.2	Data analysis	57
4.3	Computer-assisted text analysis and topic modeling	60
4.4	Topic popularity	62
5	Overall contributions and implications	65
5.1	Key findings and theoretical contributions	65
5.2	Methodological contributions and applied perspectives	67
5.3	Managerial implications	68
5.4	Limitations and suggestions for future research	71
6	References	73
7	Article 1	93
8	Article 2	109

List of Figures

1	Disciplines used in this thesis and topics of the thesis	17
2	Data-driven analysis	60

List of Tables

1	Data sources	56
---	------------------------	----

1 Introduction

1.1 Language management in Japanese multinational corporations

This study examines language management in the context of a Japanese multinational corporation (MNC). Over the years, language management has been defined differently by various scholars such as Jernudd and Neustupný (1987), Spolsky (2004), and Marschan et al. (1997). The European Commission (EC) defines it based on strategic considerations as “the planned adoption of various techniques to facilitate effective communication with clients and suppliers abroad” (EC 2007: 5). Recently, language management has received considerable attention in research on international business and management (Marschan et al. 1997; Piekkari and Tietze 2011; Tenzer et al. 2014; Sanden 2015), and many studies on corporate language policies focus on the role of English as a common corporate language (Piekkari et al. 2005; Luo and Shenkar 2006; Kankaanranta and Louhiala-Salminen 2007; Tietze 2010; Lønsmann 2011). The introduction of English as a corporate language has proven helpful in internal and external communication because it provides a common language for all members of the organization and offers easy access to official information channels (Tange and Luring 2009). Neeley (2011) finds that English as a *lingua franca* is used for instrumental goals, such as changing organizational culture through cross-border communication. Kankaanranta and Louhiala-Salminen (2010) have covered perceptions of individual English as a business lingua franca (BELF) users. However, more micro-analysis of BELF users’ attitudes and costs of using English

as a corporate language are needed (Takino 2016).

Japanese MNCs offer a useful context to study the choice of English as a corporate language (or “Englishnization”; Sekiguchi et al. 2016). Today, many Japanese MNCs are oriented toward the international market and pay much attention to globalization. Switching to English has served as the foundation for the desired global expansion, as fast and direct communication is the only way to integrate business across multiple nations. Despite this, Englishnization has roused much debate and some strong opposition in the fields of academia and business (Tsuda 2011; Neeley 2011). Japan is often considered to be a highly monolingual nation (Hashimoto 2011; Kobayashi 2013). The Japanese are considered to possess low English proficiency (Takino 2016) and are consistently at the bottom of international rankings in terms of English competence. The Internet-based Test (IBT) scores for the Test of English as a Foreign Language, which indicate a country’s average English skill level, show that Japan ranks third from the bottom among all Asian countries with 72 points (out of 120 points) (ETS 2020). Furthermore, Japan is among the low-performing countries in the EF English proficiency index published by EF Education First, an international education company (EPI 2020).

Recent studies on language management in the Japanese context focus on the role of corporate language policy in Japanese MNCs to promote international skills among their Japanese employees and hire more highly skilled non-Japanese employees inside and outside Japan (Sekiguchi et al. 2016). This policy includes guidelines and recommendations for appropriate language use. According to the perspective of language cost, individual employees may face implicit cost of using English as lingua franca (Takino 2016). Many companies have begun to use scores

on the Test of English for International Communication (TOEIC) as a criterion for HR decisions related to hiring, promotion, training, and on-the-job performance (Nakamura 2016). In the past few years, Sumida (electronic components), Rakuten (e-commerce hosting), Fast Retailing (fashion retailing), Nippon Sheet Glass (industrial glass), Nissan (the automotive industry) and others have adopted English as the official corporate language (Neeley et al. 2012; Yamao and Sekiguchi 2015). Despite the increase in demographic and cultural diversity in Japanese headquarters (HQs), Japanese MNCs tend to hire and socialize with non-Japanese employees so that employees can acclimate to traditional Japanese work culture (Maki et al. 2015). To the best of my knowledge, however, no prior study has been done to explore this tension between conformity and diversity. Thus, there is a need for research to better understand of the influence of English-only corporate language policy on the practice of language-sensitive recruitment in the Japanese business context. Language-sensitive recruitment refers to the employment of language-competent personnel (Sanden 2015). There is also a more general need for analyses of employees' perceptions of such policies, changes in language use, and the transformation of corporate culture that follows from language policy decisions in business organizations. Moreover, the abundance of social media data makes it possible to study employee satisfaction and corporate value, provides new insight into employee perceptions (Huang et al. 2015; Moniz 2016), and becomes an important source for the reproduction of discursive power in society (Törnberg and Törnberg 2016). However, few studies use anonymous media platforms as data sources to examine employees' perceptions of corporate language policy. It also remains unclear how these perceptions change over time.

The present study aims to fill this gap by focusing on a single case, namely Rakuten, a Japanese MNC that operates internationally and that made a conscious decision to adopt English as its official corporate language in 2010. According to Yoshihara's (2001) study, Sumida Corporation was the first Japanese MNC to adopt English as their corporate language worldwide, and the next is Rakuten after ten years. Neeley (2011, 2013, 2017) conducted a series of studies on this case. The CEO of Rakuten allowed her full access to his organization, stating he "has always wanted others to learn from his experience, be they positive or negative, in order to empower their journey". Although Neeley (2017: 146) claims that her work is not a piece of commissioned research, her work takes a celebratory approach; she is even featured on Rakuten's designated website showcasing successful "Englishnization"¹.

Against this background, it seems clear that a more critical analysis of the case is needed. Owing to the complexity of the phenomenon, this study could not be tackled solely using theories on language management. My thesis therefore draws from computational sociolinguistic analysis of language use and the study of business communication relating to MNCs. More specifically, I used a computational linguistics approach (details in 3.3). Thus, the research is interdisciplinary and is positioned at the intersection of business communication, computational sociolinguistics, and language management (see Fig. 1). As shown in Fig. 1, I study business English communication by Japanese employees from both a macro and micro perspective. The macro perspective is taken in a study

¹"Separation of language and culture key to global success: HBS' Tsedal Neeley", Rakuten, last modified November 10, 2017, <https://rakuten.today/blog/global-success-depends-separating-language-culture.html>

of the influence of corporate language policy on the practice of language-sensitive recruitment from a corporate perspective, whilst the micro perspective is taken in order to study how employee perceptions of corporate language policy change over time. Unlike traditional survey-based studies, I use data-driven content analysis of social media, which lets the data tell the story (Schwartz and Ungar 2015). In such a bottom-up approach, the exploration is not based on previous categories (details in section 4.3).

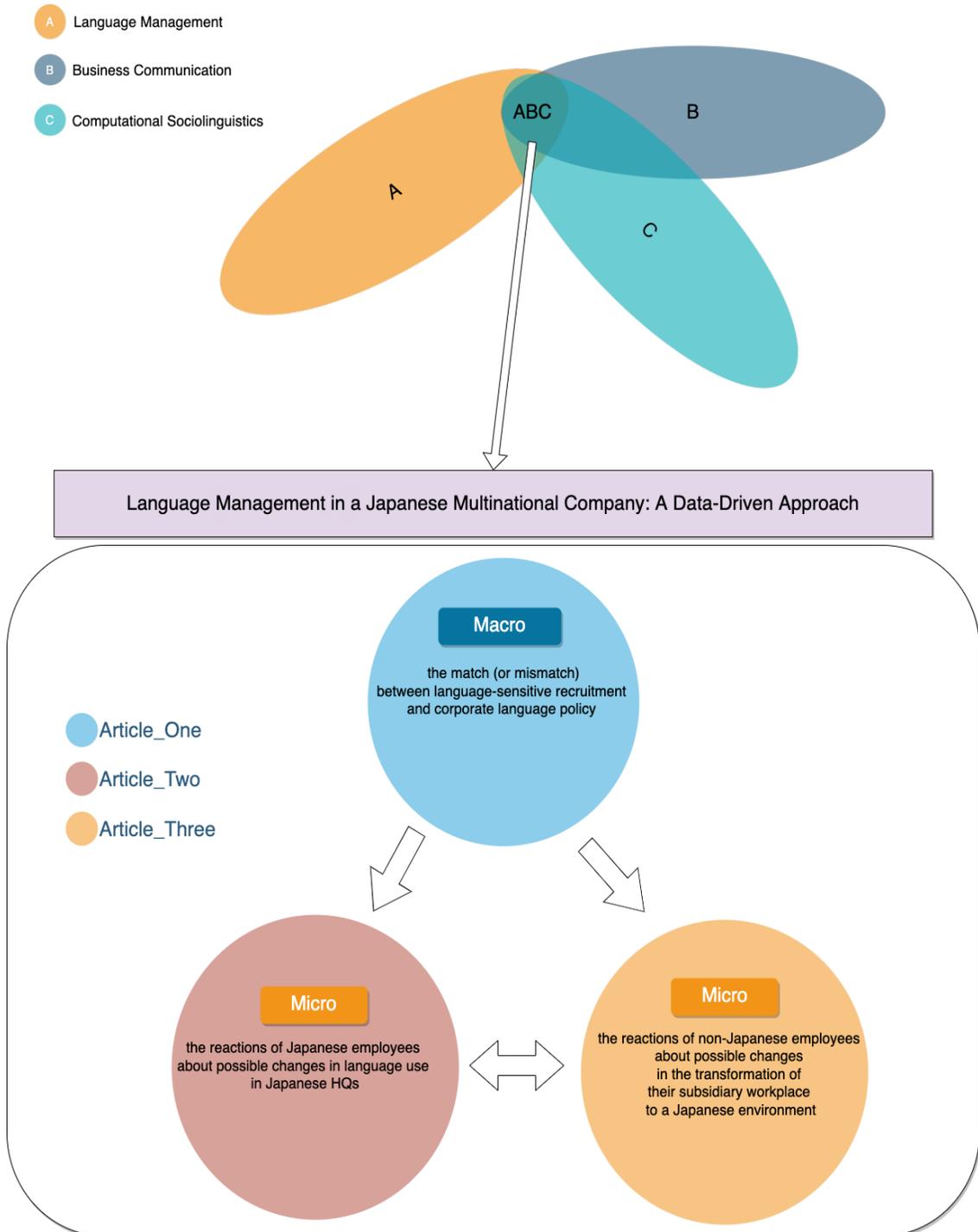


Figure 1: Disciplines used in this thesis and topics of the thesis

1.2 Research questions

Given the existing studies on language management in Japanese MNCs, the Japanese business community increasingly practices the discourse of “English only for global business” (Takino 2016). Japanese government has also made a proposal for needs of developing Global Human Resources (METI 2010). Despite the reluctance of Japanese employees to use English at work, Japanese companies need to adopt English as an official language more widely (OKABE 2009; Takino 2016). In this thesis, official is defined as mandated by the corporation. Neeley (2011) argues that adopting English as the corporate language is the only way to integrate the business across multiple nations and operate effectively worldwide. However, Rakuten and its employees must have faced enormous challenges resulting from Englishnization. Thus, in this doctoral thesis, I specifically investigate the present overarching question:

- RQ1: What are the key challenges facing a Japanese MNC when it adopts English as the official language?

More specifically, the study aims to address the following research questions and is exemplified through the case of Rakuten:

- RQ2: How is corporate language policy implemented in language-sensitive recruitment?
- RQ3: What are the relevant recruitment needs for bilinguals in job advertisements?

- RQ4: What are Japanese employees' attitudes toward the implementation of English as an official corporate language in Japanese HQs?
- RQ5: How do non-Japanese employees perceive changes in corporate culture following from change in language policy?
- RQ6: How could computer-assisted text analysis be exploited to help researchers and practitioners better understand issues in language management on social media?

1.3 Research setting: Rakuten

A number of studies have drawn attention to the fact that implementing an English language policy in Japanese MNCs could be problematic and therefore deserves further investigation (Tsuda 2011; Neeley 2011). As mentioned, this study focuses on Rakuten, a leading Japanese e-commerce MNC, one of the world's largest internet companies with a presence in 30 countries and regions. The company's HQs in Tokyo employs over 10,000 employees, 80% of whom are Japanese. On March 1, 2010, Hiroshi Mikitani, the CEO of Rakuten introduced an English-only policy known as "Englishnization" to the 7,000 Japanese staff gathered for his weekly address and the 3,000 employees watching online abroad. "From now on," he said, "everything at Rakuten, from meetings to the canteen menus, would be in English" (Skapinker 2017). His objective was to improve employees' English language proficiency, prepare them to better deal with global competition, and enable them to meet the English language proficiency requirements of the workplace. The CEO advocated the use of the TOEIC listening and reading test to meet this objective, by requiring all employees to achieve a satisfactory score. From Neeley's perspective,

Englishnization in Rakuten's case is a strict corporate language policy; it was required for every employee irrespective of rank, role, or tenure at the company (Neeley 2017). Current Rakuten employees face severe consequences if they are not able to obtain an ideal TOEIC score (Neeley 2017). For instance, if an employee cannot pass the test within two years, then they might be demoted. However, the policy is not based on a formal language policy document that includes guidelines and recommendations for appropriate language use. Instead, it is based on a mandate from the CEO's regulatory documents (Neeley 2011). The implementation of Englishnization is divided into three phases. First, after the program was launched in 2010, Rakuten employees had two years to obtain a score above 650 in the TOEIC or face demotion (Neeley 2017). Second, the proficiency bar rose gradually to reach a score of 800 and the points became a mandatory requirement for all new hires (Neeley 2017). The 800-point phase is still underway and emphasizes speaking. In the third phase, the policy documents stress global understanding as a goal, but they do not go into details about what they mean by that.

1.4 The structure of the thesis

This doctoral thesis is constituted by three papers preceded by this introductory section, which summarizes and compares the research questions and provides conclusions and suggestions for future research. The introductory section also aims to demonstrate the coherence of the thesis and summarizes the contribution to the research field. The articles present studies using a data-driven approach and an interdisciplinary attempt to answer the overarching research question RQ1. Article 1 tackles RQ2 and RQ3, exploring the match (or mismatch) between

language-sensitive recruitment (English, Japanese, or bilingual) and corporate language policy. Article 2 addresses RQ4, examining the barriers experienced in multinational companies that have adopted a language other than that of their country of origin. Moreover, the study analyzes the perceived change in employees' attitudes toward corporate language policy. Finally, the thesis answers RQ5 and RQ6, mainly in article 3. In the next section, I offer an outline of the three articles that compose this thesis.

2 Outline of the articles

The thesis comprises three articles with Ziyuan Zhang as sole author. Article 1 was published in July 2021: Zhang, Ziyuan. 2021a. “Content Analysis of Language-Sensitive Recruitment Influenced by Corporate Language Policy Using Topic Modeling” *HERMES - Journal of Language and Communication in Business*, no. 61: 77–91. Article 2 was published in May 2021: Zhang, Ziyuan. 2021b. “It is All about TOEIC: Discovering Topics and Trends in Employee Perceptions of Corporate Language Policy” *Journal of Multilingual and Multicultural Development*. 1–19. Article 3 was published in December 2021: Zhang, Ziyuan. 2021c. “Analyzing cultural expatriates’ attitudes toward ‘Englishnization’ using dynamic topic modeling” *Journal of Computer-Assisted Linguistic Research*. 1-26.

2.1 Article 1: Content analysis of language-sensitive recruitment influenced by corporate language policy using topic modeling

Zhang, Ziyuan. 2021a. “Content Analysis of Language-Sensitive Recruitment Influenced by Corporate Language Policy Using Topic Modeling” *HERMES - Journal of Language and Communication in Business*, no. 61: 77–91.

The first article analyzes the complex relationship between language-sensitive recruitment and corporate language policy and explores the role of language management in a large international corporation. Language-sensitive recruitment is

a language management tool frequently used by corporate organizations. However, its relationship with corporate policy is lacking; hence, this study aims to consider language-sensitive job advertisements from a computational text analysis perspective and explore the match (or mismatch) between language-sensitive recruitment (English, Japanese or bilingual) and corporate language policy. This study uses corpus methods combined with topic modeling and text analysis to investigate the influence of corporate language policy on the textual practice of language-sensitive recruitment in Rakuten (see Chapter 3, where these methods are described in more detail). The findings suggest that there exists a considerable discrepancy between recruitment needs and corporate language policy. It also finds that bilinguals still play a key role in crossing language boundaries ten years after the English language policy in this Japanese MNC was mandated. The study contributes to business language by exploring an additional scenario for linking language competency with actual recruitment needs. Thus, the results shed light on the implementation of language-sensitive recruitment in a multilingual corporate context, affecting the communication patterns and recruitment tactics.

2.2 Article 2: It is all about TOEIC: Discovering topics and trends in employee perceptions of corporate language policy

Zhang, Ziyuan. 2021b. "It is All about TOEIC: Discovering Topics and Trends in Employee Perceptions of Corporate Language Policy" *Journal of Multilingual and Multicultural Development*. 1–19.

The second article examines the barriers in multinational companies that have adopted a foreign language and analyzes employees' attitudes. Nowadays, many companies intend to enter the international market and pay considerable attention to global trends, and some corporations adopt English as their corporate language. Although many studies extensively describe the role of corporate language policy in large corporations, employees' perceptions of such policy have not been explored adequately. Few studies investigate the trends in such perceptions in social media texts. This study uses computer-assisted text analysis to examine employees' perceptions of the mentioned English-only policy over time in Rakuten. It is based on the analysis of a corpus consisting of 704 posts (about 66,141 words) from the social media platform OpenWork for the period from 2010 to 2018. The study finds some general trends in employees' attitudes, such as topics on the inconsistency between expectation and reality for the language policy. The findings suggest that Rakuten is not global in the true sense; the value of the TOEIC score is overestimated and the neoliberal employment system affects employees in complex ways. This study contributes to international business language with a bottom-up, employee-centered, and diachronic perspective on language management. The results shed light on the implementation of corporate language policy to effectively respond to global expansion and business needs.

2.3 Article 3: Analyzing cultural expatriates’ attitudes toward “Englishnization” using dynamic topic modeling

Zhang, Ziyuan. 2021c. “Analyzing cultural expatriates’ attitudes toward ‘Englishnization’ using dynamic topic modeling” *Journal of Computer-Assisted Linguistic Research*. 1-26.

The final article uses computer-assisted text analysis to investigate changes in cultural expatriates’ perceptions of Japanese work practices and values over time. Cultural expatriates are “people who are temporarily or permanently detached from their home culture while still operating in their own country” (Neeley 2017). They are a significant but underexplored outcome of globalization. Despite the recent proliferation of studies on the internationalization of Japanese MNCs, few studies focus on cultural expatriates’ perceptions of corporate language policy over time in social media texts. This study analyzes a corpus of 208 posts on Glassdoor from 2009 to 2020 (see details in section 4.4). The findings suggest that these posts can be divided into three content groups: the threat of a foreign corporate culture, embracing the Rakuten way, and perception of leadership and marginalized status. Further, the posts reveal how corporate language policy, as an instrument of internal internationalization, impacts the external internationalization of Japanese MNCs. With regard to method, the study demonstrates that dynamic topic modeling can enhance our understanding of cultural expatriates and English-only policy affecting the internationalization of Japanese MNCs. It contributes to interdisciplinary research by examining cultural expatriates’ perceptions of Japanese work practices

and values from a short-term diachronic perspective.

3 Theoretical framework

This chapter presents the theoretical framework applied in this thesis and contains a survey of the most relevant literature, which describes and discusses previous research on the topic of language management in business organizations. By drawing on a combination of computational sociolinguistics, language management, and business communication, this chapter seeks to explain key concepts of this thesis, including language issues in international business (3.1), language management (3.2), the concept of cultural expatriates (3.3), and computational sociolinguistics (3.4). Thus, the thesis contributes to the literature by including an interdisciplinary study of language management in the Japanese business context. In each section, I give a brief outline of previous research and an account of how the current thesis contributes to and complements these fields of knowledge.

3.1 Language issues in international business

This section discusses the most important concepts related to language and communication in international business.

3.1.1 The concept of language in international business

A growing number of researchers in international business have started to recognize the importance of language in MNCs and have approached language issues from different angles. Language and communication are important for all businesses (Clausen 2007). The globalized labor market also brings international candidates to the local, domestic markets. Language has therefore become a hot topic for large

multinational corporations. The concept of language itself is central to the study of language issues, and yet, there is a lack of consensus in international business as to how it should be conceptualized.

Three facets of language feature most prominently in the conceptualization of language in international business: English as the language of international business, national languages spoken in MNCs, and mandated corporate languages by the top management of an organization (Brannen 2004). English, as the first facet of language, is typically chosen by default as the preferred language for corporate use (Marschan-Piekkari et al. 1999*b*). Many scholars focus on the notion of a common corporate language, which is defined as a management tool in MNCs (Marschan-Piekkari et al. 1999*b*; Feely and Harzing 2003; Piekkari et al. 2005; Harzing et al. 2011; Latukha et al. 2016). Others explore the national languages of corporate HQs and globally dispersed subsidiaries (Steyaert et al. 2011; Angouri 2014; Janssens and Steyaert 2014).

Following a few scholars (Berthoud et al. 2015; Neeley 2011), I adopt the current definition of language: a specific national code (in this particular case English), which is mandated by the top management of an organization. I take a critical approach to English as an officially mandated corporate language with a point of view that it is reconceptualized as a commodifiable skill.

3.1.2 The concept of business language and communication

The HQ of a Japanese MNC offers a good context to study the language issue. For example, I explore the reactions of Japanese employees to possible changes in language use and the transformation of their HQ workplace into a

multilingual environment in article 2. The terms “business language” and “business communication” are used to signify two slightly different perspectives to the same study object. I use the terms “business language” (article 1) and “international business language” (article 2) to mean a language choice of code (in this particular case English) which is mandated by top management of an organization. I use “business communication” to represent the quality, mechanisms, and consequences of communication in business contexts (Takino 2016). Finally, I focus on close textual analysis from bilingual viewpoints, covering not only English but also Japanese viewpoints.

3.1.3 The concept of English

Another core concept plays a fundamental role in my thesis is English. International business scholars view English as a bounded linguistic system. This view suggests that English is one of the multiple language systems in a multilingual workplace (Luo and Shenkar 2006). In my thesis, I hold that English is not “a resource which emerges and is mobilized in a bottom-up manner in interaction with other languages as part of a social, multilingual practice” but a bounded entity chosen by the top management in a multilingual organization (Kankaanranta et al. 2018: 338). Following Peltokorpi and Vaara (2014), who argue that employees who master English as corporate language are likely to have advantages in recruitment, promotion, or performance rewards, I conceptualize English in my thesis as a commodity that can be measured and assessed (Heller 2010; Piller and Lising 2014). By this I mean that English is a commodifiable skill.

3.1.4 The concept of lingua franca and English as lingua franca

Researchers in sociolinguistics also discuss English as the lingua franca in business contexts. With this research traditions English is viewed as a social practice (Louhiala-Salminen et al. 2005; Jenkins et al. 2011; Pullin 2010; Ehrenreich 2015). Rather than using this definition, I define English as a commodity. I conceptualize lingua franca as “a common language different from the parties’ native language, very often English” (RP Cuypers et al. 2015: 430). Some researchers have aimed to understand what characterizes English as a Lingua Franca (ELF), that is, “the resource used among speakers of diverse linguistic backgrounds in their interpersonal communication” (Jenkins et al. 2011: 337). Jenkins (2015) strongly emphasizes the use of ELF as a multilingual practice.

3.1.5 The concept of BELF

This concept is not the choice of my theoretical fundamental. However, this is so close to my thesis. I will briefly explain it below. Researchers in sociolinguistics conceptualize English as a more neutral daily communicative tool in the form of English as Business Lingua Franca (BELF) (Kankaanranta and Planken 2010). The term BELF was introduced by Louhiala-Salminen et al. (2005) to apply the concept of ELF into a business context. BELF scholars emphasize the business domain and is used in multilingual contexts (Louhiala-Salminen et al. 2005; Ehrenreich 2010; Kankaanranta and Louhiala-Salminen 2010). The term “BELF” has been criticized for preserving the concept of English in the name of BELF as it refers to a bounded linguistic system (Makoni and Pennycook 2012; Kankaanranta et al.

2018). The notion of BELF competence emphasizes the need to be able to communicate in a particular situation related to work, rather than relying on various formats of language tests (Heller 2010; Kankaanranta et al. 2018). There are many empirical studies on the perceptions of BELF by employees (Nickerson 2005; Kankaanranta and Louhiala-Salminen 2010; Louhiala-Salminen and Kankaanranta 2012; Kankaanranta and Lu 2013). However, few studies look at the perceptions of BELF by employees who must achieve a TOEIC score above a certain level.

3.1.6 The concept of English as corporate language

The concept of corporate language is also a central notion in my thesis. I adopt the definition of corporate language as the official language, representing a mandate announced by the corporation and “taking standard English (ENL) for granted” in article 1 (Kankaanranta et al. 2018; Cornelissen 2020). I also adopt the definition of corporate language as a working language used in interactions among individual MNC employees in articles 2 and 3 (Kankaanranta et al. 2018). The concept of (English as) corporate language was explicitly defined by Kankaanranta et al. (2018) for the first time. Kankaanranta et al. (2018) argue that neither “corporate language” in general nor “English as corporate language” in particular are explicitly defined in the research fields of international management and sociolinguistics. They advance the notion of English as corporate language by distinguishing its frontstage and backstage representations: official, standard English, and working language BELF (Kankaanranta et al. 2018). English is typically chosen as a corporate language in MNCs owing to its dominance and hegemonic position as the language of international business (Tietze and Dick 2013). The corporate language

has a position of superiority compared with the local languages in the MNC as a multilingual organization (Marschan-Piekkari et al. 1999*b*; Luo and Shenkar 2006). Multilingualism in the MNC is conceptualized by Tange and Luring (2009) as the co-existence of more than one language within the same organization. The meta-language of English as corporate language is used to override rival speech communities based on employees' native languages (Tange and Luring 2009). The implementation of English as corporate language has also been linked to power relations (Marschan-Piekkari et al. 1999*a,b*; Piekkari et al. 2005; Vaara et al. 2005), discrimination and ingroup-outgroup division (Andersen and Rasmussen 2004; Lønsmann and Mortensen 2018). Over the last few years, however, an increasing number of voices have challenged the traditional view on the superior status of English as corporate language in MNCs (Barner-Rasmussen et al. 2014; Lønsmann 2014; Gaibrois 2016). These voices advocate for an interdisciplinary approach that combines sociolinguistics with business management research, including the interplay of different languages in MNCs (Janssens and Steyaert 2014). Within sociolinguistics, a group of researchers has used the term BELF to discuss English as corporate language in business contexts. Language-sensitive international business and management research typically views corporate language as an organizational artifact that shows how thoughts are formulated and how they are communicated and discussed (Brannen et al. 2017). BELF research is often interested in how corporate language is conceptualized, how it is used, or what the discourse is about. English as corporate language, the only shared language among MNC employees, is typically the medium of everyday communication from employees' perspectives (Ehrenreich 2010; Louhiala-Salminen and Kankaanranta 2012; Lønsmann 2015). In my thesis,

article 1 focuses on the role of English as corporate language that is instrumental for the corporate image. Article 2 and 3 focus on the attitudes of Japanese employees in the HQs of a MNC and non-Japanese employees in a subsidiary towards English as corporate language. To the best of my knowledge, no prior study has examined the difference between corporate language as official language in a public genre (the bilingual job advertisements, see article 1) and corporate language as a working language in a different genre (employees' anonymous comments, see articles 2 and 3).

3.2 Language management

This section starts by discussing one fundamental concept in the thesis: language management. Two sections will follow: language management at the corporate level and the employee level. Recently, the issue of language management has received considerable attention in international business studies. I use the term “language management” to mean the management of language users, as a form of human resource management. I also adopt the present definition of language management: “the strategic use of language from the managerial point of view” (Sanden 2015).

Language management theory (LMT), which is introduced by Jernudd and Neustupný (1987), is concerned with the management of discourse in corporate settings (Nekvapil and Sherman 2009). In contrast, Spolsky (2004, 2009, 2012) defines language management as a wider concept of language policy. However, his theory does not consider important factors of the corporate context such as enterprise business strategies and costs (Sanden 2015; Van den Born and Peltokorpi 2010).

In contrast to general sociolinguistic approaches mentioned above, Marschan et al. (1997) established language management research in international business and management, advocating for the importance of language in multinational management (Volk et al. 2014). Language management is a practice-oriented approach to language, focusing on the management of language users and their competencies (Sanden 2015). Scholars such as Piekkari (2006), Marschan-Piekkari et al. (1999*a*), and Tenzer et al. (2014) discuss the strategic use of language from a managerial perspective. “Business strategy tool” and “language-sensitive approach” are expressions used in this field to refer to the strategic use of language by management with the purpose of controlling the linguistic-communicative arena of a firm or business unit based on the assumption that it will benefit the company (Sanden 2015). I provide an overview of these in the theoretical framework of article 2.

The unit of analysis in language management studies is often the language needs of the company, that is, the company’s need for employees with sufficient knowledge of foreign languages to initiate and consolidate business relations with speakers of these languages (Vandermeeren 2003). In language management studies, there is less focus in individual segments of discourse than in the sociolinguistics approaches such as LMT and BELF. Marschan et al. (1997: 596) argue that the final target of language management is to “develop a fit between a company’s language profile and its strategies. An important first step is to include language aspects at the highest level of strategic planning and implementation.” Following Marschan et al. (1997), I argue that language must be considered a key element in strategically managing a multinational company. As language

management has been referred to as “a course of action where language regulation is designed and implemented on the basis of strategic considerations,” that is, strategic language management or simply language strategies (EC 2007; Piekkari et al. 2014), I adopt the definition of language strategies mentioned in the introduction (see Chapter 1). However, academic researchers have divergent conceptions of strategies, either strategy as a plan or strategy as a practice. Strategy as a plan is defined by Chandler Jr (1969: 13) as “the determination of the basic long-term goals and objectives of an enterprise, and the adoption of courses of action and the allocation of resources necessary for carrying out these goals”, and by Mintzberg (1998: 13) as a : “consciously intended course of action, a guideline (or set of guidelines) to deal with a situation.” In contrast, the conception of strategy as a practice is represented by Golsorkhi et al. (2010) and Whittington (1996), who focus on “how managers and consultants practice strategies, rather than developing strategies through corporate policies” (Sanden 2015). To my knowledge, very few studies look at the perception of language strategies as practice in the Japanese business context, let alone focus on employees’ perceptions. This thesis, therefore, aims to how language strategies as practiced in a Japanese MNC are perceived by employees in articles 2 and 3.

3.2.1 Language management at the corporate level

I define “language management at the corporate level” as management initiatives to address the language needs of the corporation. Sanden (2015) refers to the same phenomenon by using the expression “corporate language measures.” These measures are controlled at the management level to modify the language

practice of the employees. Sanden's framework is composed of six measures: namely (1) linguistic auditing, (2) translation and interpretation, (3) technological solutions, (4) language training, (5) language-sensitive recruitment, and (6) language intermediates. I refer to Sanden's work for a detailed explanation of language management at the corporate level. Language-sensitive recruitment is the measure that is particularly relevant for the present study. The concept of language-sensitive recruitment has been widely used as a language management tool (EC 2007). Language-sensitive recruitment is defined by Peltokorpi and Vaara (2014: 601) as "practices in which a certain proficiency in the corporate language is used as a precondition for employment".

I adopt the definition of language-sensitive recruitment, referring to the employment of language-competent personnel in the context of this study. A job advertisement is one way to attract the desired applicants to an organization (Blackman 2006). Organizations create job advertisements as a routine activity in the recruitment process (Rafaeli and Oliver 1998). Job advertisements are potential sites where ideology is naturalized (Fairclough 2013). That is to say, it is almost impossible to separate the messages in an advertisement from the set of values, attitudes, and shared knowledge that comprise a culture. Walters and Fage-Butler (2014) illustrate how the genre of the job advertisement has developed over time, as well as indicating the features of the genre that characterize the richer job advertisements written today. Backhaus (2004) assesses the importance of word choice in producing recruitment materials. The author advocates that the task of business communicators is to find the appropriate words to pique the curiosity of the desired potential workers and encourage them to continue through the application

process (Backhaus 2004). Backhaus (2004) proposes that the easiest and cheapest way to approach the language problem is to hire people with sufficient language skills. In the business world, a foreign language capability is strictly an ancillary skill (Inman 1980). MNCs can remove language barriers by recruiting employees with the required language proficiency (Reeves and Wright 1996). However, language-sensitive recruitment is not always an appropriate choice and may lead to many challenges. One challenge is finding qualified candidates because there is a shortage of potential employees with sufficient English skills in the local labor market, including Japan (Peltokorpi 2010) and South Korea (Park et al. 1996).

To my knowledge, no prior study has specifically looked at language-sensitive recruitment's relationship with corporate language policy. In article 1, I investigate the influence of corporate language policy on the textual practice of language-sensitive recruitment in a Japanese MNC. I further explore the match (or mismatch) between language-sensitive recruitment (English, Japanese, or bilingual) and corporate language policy.

3.2.2 Language management at the employee level

Language management at the employee level refers to the management of language practices emanating from the front-line level rather than the management of discourse at the corporate level. Front-line employees may decide what their language practices should be, for example, which language to use in what situation (Sanden 2015). They have to deal with language and communication issues every day (Andersen and Rasmussen 2004). For instance, employees often switch to the local language, which is different from the corporate language (Harzing et al. 2011).

The concept of language management at the employee level has been further applied to the choice of communication medium in organizations by Charles and Marschan-Piekkari (2002), Shachaf (2008), and Harzing et al. (2011). These scholars discuss that the use of a written communication medium, such as email, may reduce miscommunication as compared with an oral communication medium. Instead of communicating via phone, Harzing et al. (2011) argue that the use of email is a common way of managing language at the employee level.

However, as argued by Björkman et al. (2012), there is a great need to engage in more in-depth analysis at the micro-level. In the present study, I focus on employees' motivation to learn the corporate language and their perceptions of language management for analyzing front-line practices. Companies with international aspirations attempt to adopt the English language to increase their probability of success (Sanden 2015). If English is not the native language of employees, the officials' decision may be undermined by employees' preferences for the local language (Feely 2004). In language learning, motivation is also an important factor, and those who score low in language tests may be discouraged (Newfields et al. 2005). Motivation in self-determination theory can be divided into two types: intrinsic and extrinsic (Ryan and Deci 2000). Extrinsic motivation is based on external outcomes such as promotion or punishment, while intrinsic motivation encourages employees to learn without expectation of reward (Newfields et al. 2005). Nantakaew (2012) argues that the most important factor affecting employees' motivation to learn English is the perceived need for self-improvement in language ability. Lønsmann (2017) focuses on employees' reactions to the implementation of a new corporate policy from the top and finds three stances

among employees: resisting English, embracing English, or a middle position. Many employees must invest heavily in learning the corporate language and ultimately accept built-in inequality, seek to escape the corporate language through internal mobility, or leave the company (Vaara et al. 2005). Managers must consider the significance of language and communication skills for staff motivation and job satisfaction in multilingual organizations (Charles 2007). Another factor is Lønsmann's (2011) observation of a discrepancy between employees' definition of language policy and the company's reality. Danish in her study was in fact the default language rather than English, which was used primarily with non-Danish speakers. Similarly, for a quantitative analysis of CEO leadership and employee confidence, Neeley (2017: 155) collected data from 4,000 eligible employees in Rakuten to examine employee perceptions of how the impact of the corporate language policy on their productivity is influenced by leadership. She conducted two surveys, at 14 months (Time 1) and 20 months (Time 2) after Rakuten implemented the Englishnization policy (Neeley 2017: 107) and then compared the responses in the two time periods to observe the changes in leadership, beliefs, and attitudes from Time 1 to Time 2. She then performed a quantitative analysis of CEO leadership and employee confidence. However, the survey was unable to determine how employees' beliefs and job attitudes toward language management change over time because of only two measurement points. I address some trends in employees' attitudes toward language management in article 2. Moreover, as only a few studies look at employee data in terms of designation, age, gender, and tenure (Neeley 2017), the present thesis intends to contribute with comparative micro-level analysis, which is conducted at the level of business units and technical

units, gender, and varying tenures.

3.2.3 The concept of corporate language policy

In the present study, I emphasize corporate language policy, which is among the main measures by which language strategies are implemented in practice. I define corporate language policy as a type of language policy that is developed specifically for a company or business organization. Academic researchers in sociolinguistics and business management have divergent conceptions of language policy (Sanden 2015). Within the sociolinguistic research tradition, a language policy is a plan for language use (Baldauf Jr 2012). The term of language policy is often used together with “language planning” as “language policy and planning”. As explained by Baldauf Jr (2012), language planning refers to implementation of the language policy across the particular regions and organizations in which the policy is addressing. The conception of language policy, represented by Kangasharju et al. (2010) in the business and management domain, is described as everyday practices, abstract principles, and political statements about language use. Language policy therefore is defined as “both the general guidelines and the practical procedures and instructions for improving and harmonizing internal and external communications, which the management of an organization or another responsible party has presented in a written form” (Kangasharju et al. 2010). As discussed above, the term corporate is central to the study of corporate language policy, which is a policy about language use in a corporation. However, the term “corporate” is not only used to denote the corporation setting in which the language policies are implemented but also the hierarchical order within the corporation (Goodijk 2002). Thus, I argue that

a corporate language policy is developed by the company's leadership and applies to the entire corporation. Following Sanden (2015), I adopt a definition of corporate language policy: a type of language management for an organization.

Many studies have explored the role of English as a corporate language policy (Marschan-Piekkari et al. 1999*b*; Thomas 2008; Tietze 2010; Lønsmann 2011; Neeley 2011; Piekkari et al. 2014). The choice of ELF is motivated by its extensive use in global communication (Nickerson 2005; Kankaanranta and Louhiala-Salminen 2007; Ammon 2010). The rationale for this choice is that it could benefit companies to exchange information in a shared linguistic framework (Thomas 2008), to ease access to professional and technical literature (Marschan-Piekkari et al. 1999*b*), and to prevent the loss of information due to translation (Sanden 2015). Moreover, Lønsmann (2011) claims that the lack of an official language policy document has led to different definitions and expectations among employees who experience a discrepancy between their interpretations of the language policy and the company's reality. Furthermore, English-only policies may lead to "language clustering", where employees with similar language backgrounds combine into groups to resort to their non-official common language within the workplace (Lønsmann and Mortensen 2018). When using a second or foreign language, employees find that they do not achieve the same level of communication or information flow, which has implications for their ability to participate in and benefit from peer-based networks and knowledge-sharing (Tange and Lauring 2009). Moreover, if English is not the native language, the official decision may be undermined by employees' preference for the local language (Feely 2004). Previous research has focused more closely on the unstructured and informal group of people

who engage in interaction more because of their common linguistic background (Ahmad and Widén 2015; Takino 2016). Comparative research across languages is rare but might yield interesting insights in the future (Takino 2016).

Studies seeking to capture existing employees' perceptions of corporate language policy are limited in drawing inferences and generalizing results (Luo et al. 2016). A few studies examine employee perceptions of corporate language policy using qualitative methods such as interviews and focus groups (Lønsmann 2017; Lønsmann and Mortensen 2018). However, few studies have illustrated temporal trends in employees' opinions regarding corporate language policy. In the present thesis, I use anonymous media platforms as data sources to examine employees' perceptions of corporate language policy. In article 2, I further investigate employees' opinions regarding the implementation of the corporate language policy and its impact in the context of the policy in Rakuten.

3.2.4 The concept of linguistic instrumentalism

To examine the implementation of language-sensitive employment in a Japanese MNC, I analyze how linguistic instrumentalism has influenced corporate language policy and the language test score. According to Piller and Cho (2013), neoliberalism, an economic ideology, serves as a covert language policy mechanism that pushes the global spread of English. Following Phelan (2014), I conceptualize neoliberalism as a discourse. The concept of linguistic instrumentalism was initially introduced by Wee (2003) and further developed by Kubota (2011). Briefly put, linguistic instrumentalism underlines the importance of English skills for work and for achieving individual economic success. This perspective has influenced

the notion of human capital and the unstable employment conditions of neoliberal society (Kubota 2011). In neoliberal discourse, human capital development relies on acquiring communication skills that become fetishized as economically viable (Kubota 2011). Communication skills in English—a language with worldwide dominance—are required for global competitiveness in non-English-speaking countries (Phillipson 2008). One may draw a parallel to the concept of the discourse of linguistic instrumentalism carried out in nations and individuals, as a way of promising economic return from English language competency (Kubota 2011). Several studies, such as those of Piekkari (2006), Lønsmann (2011), and Neeley (2017), posit that using English as a corporate language medium may attract global talent. Lønsmann (2011) argues that ELF is often used for instrumental goals.

In relation to linguistic instrumentalism, I raise three points that are relevant to my thesis. First, a plethora of outsourcing and part-time employment are demonstrated to reduce labor costs in the pursuit of economic productivity. In the Japanese business context, the previous life-long regular employment has been replaced by more flexible contract employment to reduce labor costs (Kubota 2011). To stay employable, as Urciuoli (2008) argues, communication skills have become a commodity. If native citizens with insufficient skills are employed while there is a particular language requirement, they may not perform effectively (Peltokorpi and Vaara 2014).

Second, as stated above, linguistic instrumentalism assumes a connection between English competence and individual economic benefits (Kubota 2011). The emphasis on communicative competence in English is meant to enhance the value of human capital and national competitiveness (Park and Wee 2013). Human capital, is

defined by Keeley (2007: 29) as “the importance of people—their abilities, their knowledge, and their competences—to economic growth.” More attention needs to be devoted to the role of English competence for assessing human capitals.

Third, TOEIC scores are typically used internationally to make HR decisions related to hiring, promotion, training, and on-the-job performance (Anthony 2003; Stahl et al. 2007; Peltokorpi and Vaara 2014; Oliveri and Tannenbaum 2017). A regular employee should achieve a TOEIC score of above 800 in Rakuten’s case (see article 1). However, if TOEIC scores are used as a criterion for recruitment and promotion, the stakes are high, meaning the test scores will significantly affect the test-taker’s life (Chapman 2006). The TOEIC test has been criticized because some test preparation strategies are unrelated to English proficiency. Test-takers may focus less on language skills unrelated to TOEIC (Newfields et al. 2005). The TOEIC test may misjudge the test-taker’s communication competency based on a specific context of language usage (Pilcher and Richards 2017). Thus, there is a danger that people may take a high score as a guarantee of effective communication in BELF (Takino 2016). These issues prompted me to examine whether English language skills actually generate economic returns in reality.

In the present thesis, I elaborate on the power of English and the neoliberal employment system and their implications for employees in articles 1 and 2. I examine how corporate language policy is implemented in language-sensitive recruitment (RQ2) in article 1 and what the temporal trends in employees’ opinions regarding corporate language policy are (RQ4) in article 2. I also investigate how English as a corporate language (re)produces social hierarchies and who wins and who loses in those scenarios, from a critical analytical lens in article 2. Moreover,

this particular case may also be applicable to other organizations.

3.3 The concept of cultural expatriates

The concept of cultural expatriates was coined by Neeley (2017). Cultural expatriates are defined as “people who are temporarily or permanently detached from their home culture while still operating in their own country” (Neeley 2017: 5). For example, American native-English speakers were detached from their American culture as they worked in a subsidiary of Japanese MNC (see article 3). Neeley repurposed the word “expatriate” as an alternative approach to the conventional ways of looking at biculturals. Following the work of Brannen and Thomas (2010), I define biculturals as people who have internalized more than one cultural profile. Culture is defined by Varner and Beamer (2005: 5) as “the coherent, learned, shared view of a group of people about life’s concerns that ranks what is important, furnishes attitudes about what things are appropriate, and dictates behavior.” As such, biculturals have a simultaneous awareness of being a member of two cultures without having to choose one culture over the other (LaFromboise et al. 1993; Brannen et al. 2009). I refer to Brannen’s work on biculturals for a detailed explanation of cultural expatriates. I provide an overview of four types of biculturals in the theoretical framework of article 3. I adopt the definition of cultural expatriates as people who have been socialized between two distinct cultures as proposed by Brannen et al. (2009). A cultural expatriate can have knowledge of another culture without identifying with it. For example, American employees may be able to acquire knowledge about the Japanese culture and apply this knowledge to guide their behavior, without actively identifying with that culture.

Cultural expatriates portray self-identity, which is defined as “the self as reflexively understood by the individual in terms of his or her biography” (Giddens 1991: 244). Following Yagi and Kleinberg (2011), I define cultural identity as a dimension of self-identity and one’s self-concept that concerns perceptions of “who I am” as a cultural being. Jameson (2007: 207) makes a similar distinction by describing cultural expatriates as people with “an individual’s sense of self derived from formal or informal membership in groups that transmit and inculcate knowledge, beliefs, values, attitudes, traditions, and ways of life.” They consist of six components: vocation, class, geography, philosophy, language, and biological traits with cultural aspects. There are three aspects of cultural identity that are particularly relevant to the present study. First, employees who work for a company for a length of time often buy into its values and cultural practices (Jameson 2007). Sometimes, divisions or departments of corporations create strong cultural affiliations (Kuhn and Nelson 2002). Second, a corporate strategy often needs to account for cultural differences related to regional distinctions (Jameson 2007). Third, language defines cultural groups, often interacts with other cultural components, and can unite or divide people in the organization (Jameson 2007). For instance, some native English-speaking informants who exhibit linguistic ethnocentrism do not understand that English is the lingua franca of international business, and assume that “it was their country, their culture, and their market that were of interest” (Neeley 2017: 59).

Recent studies have looked at individuals’ cultural identities in the organizational context. Keeley (2001) acknowledges, interestingly, that non-Japanese employees are considered an out-group, even when they work in

subsidiaries in their home country. Kankaanranta and Lu (2013) argue that the use of BELF—or a common language—in international business can have a homogenizing effect on the cultural identities of the speakers. Jameson (2007) argues that cultural identity is also influenced by close relationships, and that it changes over time. The close relationships could include family members, friends, and colleagues, with whom interaction occurs frequently (Kankaanranta and Lu 2013). To the best of my knowledge however, no prior study has looked at how cultural identity changes over time in a Japanese MNC .

In the present thesis, I discuss Rakuten’s English-only language policy, and examine the consequences of this policy. I examine the use of English in Rakuten’s HQs and its relation to non-Japanese employees’ opposition to Japanese work practices and values transmitted by the English-only policy. I also analyze the relationship between language, culture, and local identity in article 3.

3.4 Computational sociolinguistics

The present thesis also shows how large-scale data-driven methods can complement existing sociolinguistics studies. Computational sociolinguistics is an emerging research field that integrates sociolinguistics and computer science to study the relationship between language and society from a computational perspective (Nguyen et al. 2016). An increasing number of computational linguists show interest in studying language in a social context. Although computational linguistics has significantly contributed to our understanding of the informational dimension of language, few studies have explored modeling of the social dimension (Elangovan and Eisenstein 2015).

Traditionally, sociolinguists collect data using surveys and ethnographic research as the main methods (Milroy and Gordon 2008; Tagliamonte 2006). Formal texts, such as large corpora and recorded conversation, were used in early computational studies on social identity and language use (Argamon et al. 2003; Van Durme 2012; Nguyen et al. 2016). Social media is a rich and easy-to-access source of abundant informal information, providing an opportunity to investigate language in social contexts (Johannsen et al. 2015). The recent increase in literature on computational linguistics studying language in social contexts is partly driven by the ever-increasing availability of social media data, which help investigate language variation more broadly (Nguyen et al. 2016). Some studies automatically infer social variables, such as gender and age, from texts (Nguyen et al. 2014).

The content generated on social media platforms provides contextual information that can be used for large-scale studies of social variables. Among the platforms that have been studied are microblogs (Eisenstein et al. 2014), web forums (Garley and Hockenmaier 2012) and online review sites (Hovy et al. 2015). However, no study has yet been conducted on the trend analysis of employees' perceptions of corporate language policy and possible changes in the corporate culture using anonymous media platforms as data sources. Therefore, I employ natural language processing tools in order to gain insights into language management issues from modeling and analyzing data in a bottom-up manner. The data is drawn from different genres (see Chapter 4). In article 2, I analyze the evolutionary nature of topics on the reactions of Japanese employees about possible changes in language use before and after Englishnization. In article 3, I investigate non-Japanese employees and possible changes in the transformation of

their subsidiary workplaces.

4 Methodology

In this section, I describe the data and explain my methodology and research design, by accounting for factors such as data presentation (4.1), data analysis (4.2), computer-assisted text analysis and topic modeling (4.3), and topic popularity (4.4).

4.1 Data presentation

Lønsmann (2015: 13) emphasizes that “what we see –or what we are allowed to see –depends on where we stand and who we are at the moment”. In present thesis, my data presentation is influenced by my positionality and identities that I detail below.

As a software engineer in the company, I was considered an employee, and therefore an insider with knowledge of the corporate language policy. As an insider, I also had knowledge of the problems and difficulties the employees encountered. One of the potential criticisms of the insider position is that it may be biased toward interpretation (Brannick and Coghlan 2007).

To offset this built-in bias, I also tried to position myself as a researcher. I am an outsider and do not have an interest that conflicts with Rakuten while doing my PhD. I am seen as an objective observer (Bonner and Tolhurst 2002). I also have adopted the approach of computer-assisted text analysis and topic modeling to decrease subjective interpretation.

Finally, I also positioned myself as a non-local. I am a native Chinese speaker. Many texts in the dataset are in Japanese. I have lived in Japan for three years while working for Rakuten, and I have passed Japanese-Language Proficiency Test (JLPT) level N2. The JLPT has been offered by the Japan Foundation and

Japan Educational Exchanges and Services (formerly Association of International Education, Japan) since 1984 as a reliable means of evaluating and certifying the Japanese proficiency of non-native speakers (Foundation 2022*b*). Regarding linguistic competence required for level N2, I can read materials written clearly on a variety of topics, such as articles and commentaries in newspapers and magazines as well as simple critiques, and comprehend their contents (Foundation 2022*a*). I can also read written materials on general topics, follow their narratives, and understand the intent of the writers (Foundation 2022*a*). It means I feel relatively comfortable with Japanese language. Even if I am not a native Japanese speaker, I have a fair understanding of the Japanese culture and texts.

For the first article I collected online bilingual job advertisements in 2019 from Rakuten in the tenth year of the corporate language policy mandate. A web crawler² was used to obtain 1,109 advertisements (1,053,178 words) posted to Rakuten's talent section³ on November 12, 2019. In total, 683 (61.5%) of these advertisements were bilingual, 129 (11.6%) were in Japanese, and 297 (26.9%) were only in English. The study extracted all the sentences, including English and Japanese requirements keywords, from 150 bilingual advertisements, which accounted for 25% of 683 bilingual advertisements.

Employee review websites contain self-initiated and anonymous information that stands to mitigate many of the biases introduced in survey studies (Das Swain et al. 2020). Glassdoor⁴ and OpenWork⁵ are examples of career community websites where current and former employees anonymously review

²The library was installed with pip: pip install scrapy.

³available at <https://japan-job-en.rakuten.careers/search-job>

⁴<https://www.glassdoor.com/>

⁵<https://www.vorkers.com/>

companies and their management (Luo et al. 2016). OpenWork (formerly known as Vorkers) mainly focuses on Japan. It contains objective information such as pay, benefits and free-form text that represents organizational culture. There are about 9,800,000 reviews and evaluated scores of 36,000 companies posted by current and former employees at the time of writing. The website has 2,900,000 registered users and approximately 3,000 new registrations every day. Glassdoor is one of the largest review websites for employers worldwide. The website provides extensive details about the latest jobs, company ratings and reviews, CEO approval ratings, salary reports, interview reviews and questions, benefits reviews, and office photos. At the time of the data collection (December 2019), there were approximately 70,000,000 reviews and evaluated scores of 1,300,000 employers posted by current and former employees. The website has approximately 50,000,000 unique visitors per month.

For article 2, the contents of the OpenWork website between 2008 and 2018 were collected by means of a customized web crawler. The website comprised 9,324 Japanese posts (1,337,652 words). As the study relied only on data written in Japanese, it does not account for employees who speak languages other than Japanese at the HQ or employees located in Rakuten offices outside Japan.

For article 3, I used a customized web crawler to download data on Glassdoor from February 2009 to November 2020, resulting in 208 reviews and a corpus of 49,497 words.

The challenge that I have experienced is data collection bias. I identify three ways in which the data can be biased. First, users on employee review websites can introduce demographic bias into datasets. The bias is often unintentionally caused

by the user base of the sites. Second, malicious users can also introduce bias into the site. Many individuals coordinate with each other to leave fake reviews and spread negative information (Morstatter and Liu 2017). Third, online bots, software programs that perform automated tasks, can run to change the statistics of the site. They can make a topic to trend or misrepresent already trending topics (Cook et al. 2014).

In the present thesis, I adopt a variety of approaches to deal with bias, such as including different data sources (job advertisements, Glassdoor, and OpenWork), adjusting my study focus, and accepting respective limitations. Regarding representativeness, my data collection has appropriate sampling to ensure generalizability of the study findings. For example, in article two, for the Japanese employees' perspective, 9,324 posts were collected over a period of eight years. There are 12,053 employees in Japan region in 2018. It means I got the voices of 77.4% of the entire population of the company if I assume that each of these posts came from a different employee. A corpus of 709 posts containing terms 英語 "English", toEIC and 公用 "official" was obtained from 9,324 posts and analyzed owing to my focus on language management. This sub-sampling method was used in three articles. Unlike interviews, employees in the websites were not asked about language issues in the company but they freely shared their thoughts about the company. It provides more valuable insights into employees' attitudes toward the corporate language policy: employees in those 709 posts (7.6%) voluntarily commented on the corporate language in the company. The other 8615 posts (92.4%) were not related to these language topics. Moreover, the data sample is also diverse and tells us about employees' function, gender, tenures and employee

status (see article 2). For example, (no.673、2018、トラベル事業、営業、在籍5～10年、女性) (Translated from Japanese, no. 673, 2018, female, BU, employee, 5–10 years). Article 2 investigated the temporal topic variation for the different profiles of employees. The study found that there is a link between their profile and the views they have (e.g., male vs. female, business unit vs. development unit, and different tenures at Rakuten). It would have been interesting if I had explored more profiles in article 3. However, since the anonymous comments miss a lot of profile information such as employment status and staff levels, it is still unknown whether the trend differs at these levels.

Further, the reviews are not uniformly distributed over time as half of the reviews were written after 2014, which biases our findings toward the recent periods—for example, in article 2, the early years vs. more recent years: 2010 (7), 2011 (25), 2012 (42), 2013 (37), 2014 (64) …2018 (157). The overall trend of the sample size is increasing comparatively fast. There are three reasons why the data sample are different across time. First, as mentioned before, these data are extracted from the larger dataset by using key words. It means at the time point the number of those commenting on language issues, while the rest of the comments are pertinent to other topics that this thesis does not focus on. Second, the number of employees is growing over time. In 2010 the company had 5,810 employees, compared to 17,214 employees in 2018. The number of employees has nearly tripled. Third, the number of employees using the employee review website is also growing. According to Software Advice’s survey, 48 percent of respondents had used an employee review website at some point in their job search in 2017.

Finally, since the anonymous comments cover only data from employees

who are active on the employee review websites, the generalizability of our findings is limited to the perceptions of this group. However, the purpose of my application examples in article 2 and 3 is to illustrate the general use of dynamic topic modeling for language management research and provide insights into the factors that influence employees' perception of corporate language policy. Thus, despite these limitations, I decided to proceed.

The details of the procedures and the data collected in the three papers are shown in the table below (see Table 1).

Article	Website	URL	Number of Posts	Words	Language	Time of data collection
1	Rakuten's talent section	https://japan-job-en.rakuten.careers/	1,109	1,053,178	Japanese, English	November 12, 2019
2	OpenWork	https://www.vorkers.com/	9,324	1,337,652	Japanese	from 2008 to 2018
3	Glassdoor	https://www.glassdoor.com/	208	49,497	English	from 2009 to 2020

Table 1: Data sources

4.2 Data analysis

On the basis of the data collected and described in section 4.1, the data analysis proceeded in three steps, as shown in Fig. 2 below: corpus construction, pattern detection using human-centered computational exploratory analysis, and pattern refinement using guided deep reading.

In Step 1, data were collected from the websites. As I explain in article 3, the data included upper- and lower-case letters, different parts of speech, and commonly used stop words. Therefore, the following preprocessing steps were taken: converting the text data to lower-case, eliminating punctuation, removing certain stop words, and lemmatizing the words. Modules in the *gensim* library⁶ written in Python⁷ language such as *gensim.parsing.preprocessing* and *gensim.utils* were used to perform these tasks.

In Step 2, the examination of lists of topic words formed the foundational part of the analysis. After data collection and preprocessing, the *ldaseqmodel* function in the *gensim* package was used to implement DTM analysis and generate topics. To decide the number of topic models generated objectively from the corpus, this study adopted the selection method proposed by Tanimoto (1958), which checks the difference between the coherence and stability per number of topics. A general rule of the method is to create LDA models across different topic numbers, and then check the Tanimoto coefficient and coherence for each topic. The Tanimoto coefficient, also known as the Tanimoto index, is a statistic used for measuring the similarity and diversity of sample sets (Chung et al. 2019). The coherence measure

⁶The library was installed with pip: `pip install gensim`.

⁷Downloaded from <https://www.python.org/downloads>.

defines a single topic by the degree of semantic similarity between high-scoring words which co-occur across the text corpus. The ideal number of topics will maximize coherence and minimize the topic overlap based on Tanimoto coefficient. I detail the procedure in article 3.

In Step 3, a computationally guided in-depth reading was also applied to the top five texts in each time slice that showed the highest proportion of that topic, as evaluated by the topic modeling algorithm. After topics are discovered by topic modeling, these topics are conventionally represented by their word lists. It is sometimes not easy for users to understand each topic based on the words. In order to make topic representations more interpretable and make the topics easier to understand, I adopt a close reading approach and focus on the representative texts in each topic. The texts in each time slice of that topic are ranked based on the conditional probability judged by the topic modeling algorithm. This close reading analysis is qualitative and topics are presented with manual labeling. Following Labov and Waletzky (1997: 370), I am sensitive to “the possibility for the texts in our data to contain one or more of the following narrative components: abstract (what the comment is about), orientation (who,when, what, where?), complicating action (then what happened?), evaluation (so what?), result or resolution (what finally happened?) and coda (how do the events relate to the present time and place?).” My awareness of these narrative mechanisms and their potential discourse functions aided my interpretation of the topics and evaluations of the data (Brookes and McEnery 2019). The topic explanations are based on the top five texts in each time slice rather than a single word or sentence.

To better illustrate the importance of the words used to discuss each topic,

I use the visualization tool *pyLDAvis*⁸ to explore the relations between the topics and their most relevant terms. In article 1 and 3, the visualization tool helps me interpret the topics in a topic model that has been fit to a corpus of text data. The package extracts information from a fitted LDA topic model to inform an interactive web-based visualization. The visualization can also be saved to a stand-alone HTML file for easy sharing. This analysis was a blended approach using topic modeling as a reference for a close reading of sub-sample documents. This framework is similar to what Gkotsis et al. (2017) applied to online health-related texts. The process and framework are detailed in articles 2 and 3.

⁸The library was installed with pip: `pip install pyLDAvis`.

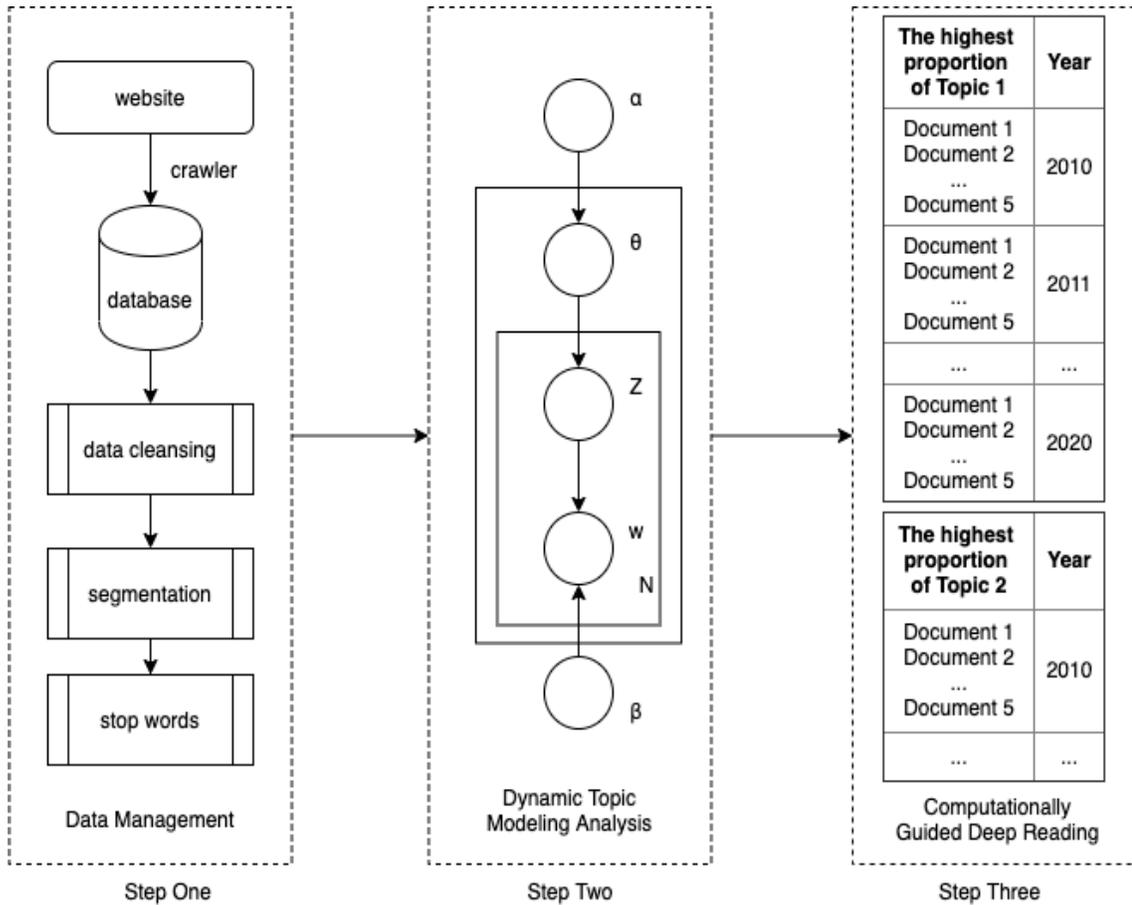


Figure 2: Data-driven analysis

4.3 Computer-assisted text analysis and topic modeling

Computer-assisted text analysis techniques belong to three main categories: lexical-based (Monroe et al. 2008; Lee and Martin 2015), text classification (Chuang et al. 2012; DiMaggio 2015), and natural language processing (Nelson 2020). Lexical-based techniques are done at the word level and aim to identify the important content words in a corpus (Monroe et al. 2008). Computer-assisted text classification can be divided into supervised machine learning and unsupervised

machine learning algorithms. In text analysis, supervised machine learning categorizes text into predetermined categories. Unsupervised text classification is a more sophisticated way to identify patterns across texts than simple word frequencies, enabling researchers to reproduce the classification results every time (Lancichinetti et al. 2015). Natural language processing incorporates language structures, such as grammar or syntax, into text analysis (Nelson 2020). Computational techniques mitigate the shortcomings of purely qualitative research by realistically and credibly incorporating large-scale data into theory-generating research (Nelson 2020). The combination of interpretive and computational approaches allows the interpretation of the output from computational methods, delivering quality and quantity, depth, and breadth (Franzosi 2010). It also allows researchers to better measure meaning by weighing the choice of close and distant reading (Moretti 2013).

As an unsupervised text classification method, topic modeling is used to identify patterns in text and is an effective tool for understanding and managing a sequential collection of documents (Blei and Lafferty 2006). It provides an overview of the corpus, complements content analysis (Nelson 2020), reveals hidden themes, and helps model the evolution of topics over time. The most commonly used algorithm for topic modeling (Blei et al. 2003) is Latent Dirichlet Allocation (LDA). The output from the algorithm is a list of the words which are highest-weighted per document. This list of words forms the basis of a human researcher's interpretation of the estimated categories. It is also common to integrate visualization as a means to summarize the corpus. Topic modeling is now increasingly applied to many social science fields to automatically categorize

large datasets and capture theoretical concepts (DiMaggio et al. 2013; DiMaggio 2015; Grimmer and Stewart 2013; Törnberg and Törnberg 2016). Because of its features, topic modeling can serve as a bridge in the social sciences, as it lies at the interface between case studies and big data, unstructured and structured analysis, and induction and deduction (DiMaggio et al. 2013; Grimmer and Stewart 2013; Mützel 2015; Hannigan et al. 2019).

However, a word of caution is warranted. Brookes and McEnery (2019: 18) question “the thematic coherence of the topic generated using topic modeling methods” and “the topics intended to exhibit limited thematic consistency.” They demonstrate that “the usefulness of topic word lists alone as basis is unlikely to be sufficient for accurately inferring the themes present in the texts grouped into a given topic” (Brookes and McEnery 2019: 18). They conclude that topic modeling methods lack an adequate theoretical basis for defining the topic (Brookes and McEnery 2019). Törnberg and Törnberg (2016) also claim that people must be careful about over-interpreting the information gleaned by the method because its theoretical implications are still not fully understood. In this thesis, I attempt to remedy this by analyzing the topics extensively, and employ content analysis that provides elaborate analytical techniques and theoretical depths for interpretation. This method was used successfully in articles 1, 2 and 3.

4.4 Topic popularity

There are three approaches by which topic models consider the time dimension: jointly modeling word occurrence and time, pre-discretized analysis, and post-discretized analysis (Wang and McCallum 2006). Post-discretized analysis

involves fitting a time-unaware topic model, slicing time-ordering documents into discrete subsets, and reviewing the topic distributions in each time-slice (Wang and McCallum 2006). An example of this type is identifying hot and cold topics based on LDA model analysis (Griffiths and Steyvers 2004). The joint modeling of word occurrence and time introduces time as an observed continuous variable and relies on the topics as a constant, as topic co-occurrence patterns change over time. The topic over time model is a typical example of joint modeling and analyzes documents with timestamps to discover topics. However, this approach fails to identify evolving topics.

In the DTM as a pre-discretized analysis, time is discretized into several periods, and time within each LDA is used to analyze documents (Blei and Lafferty 2006). The continuous time DTM (cDTM) uses the Brownian motion to model latent topics (Wang et al. 2012). DTM leverages the knowledge of different documents belonging to different time-slices to map how words in a topic change over time (Blei and Lafferty 2006). In the DTM method, documents are divided into sequential groups, assumed to be exchangeable within a group, and topics of each slice evolve from the topics in the previous slice (Wang et al. 2012). In contrast to DTM, the cDTM method is more suitable for analyzing text with dense time stamps and high sparsity (Wang et al. 2012). Topic evolution has been extended and applied to email (Wang and McCallum 2006), news (Wang et al. 2012), bulletin board systems (Lina and Xijin 2014), scientific literature (Alsumait et al. 2008), and movie reviews (Meng et al. 2012).

Topic popularity is calculated using the occurrence of topics in the documents (Griffiths and Steyvers 2004; Sun and Yin 2017). Combining DTM

with content analysis allows consideration of the context for a better understanding of the dynamics of social representations in a comparative and diachronic view (Chandelier et al. 2018). However, few studies have analyzed topic evolution to examine changes in employee perceptions over time. In the present thesis, I investigate the representation of corporate language policy in social media by identifying topics at a particular point and examining their evolutionary nature. I further analyze shifts in the opinions of employees on each topic. Article 2 considers the posterior document-topic distribution and shows how information in each post can be used to explicitly quantify the dynamics of the inferred topics that evolve over time at the levels of gender, function and tenure with Rakuten.

5 Overall contributions and implications

5.1 Key findings and theoretical contributions

This section highlights key findings and theoretical contributions in the present thesis.

First, the present study contributes to filling a gap in language issues in the Japanese business context by focusing on a micro level of analysis that has largely been overlooked in the existing literature on international business and management. This thesis has demonstrated the manner in which actual recruitment needs, the English-only policy, and employees' attitudes toward corporate language policy affect the internationalization of a Japanese MNC. The main contribution of this work therefore lies in generating knowledge on the interplay between Englishnization and organizational culture in the study of language management.

Second, one of the most profound findings is related to corporate language policy. While previous studies have emphasized the role of corporate language policy in large corporations, the empirical study in article 2 explored the trends in employees' perception of such policy in social media texts. In my thesis, I have identified inconsistency between expectations and reality in terms of Englishnization. The results not only reflect employees' emotions and dissatisfaction regarding the corporate language policy but also reveal that employees' motivation to learn English gradually declines. The findings raise questions regarding one argument in the previous research on whether Rakuten would achieve the intended outcomes associated with Englishnization (Neeley 2017: 171).

Third, the thesis contributes to language management in the field of language-sensitive recruitment implementation under corporate language policy. The findings reveal a discrepancy between recruitment needs and corporate language policy. There is also a misalignment between corporate language policy and functional skill requirements in job advertisements. This is in line with the existing literature on language-sensitive recruitment (Grin et al. 2011; Brandt and Chancellor 2011; Peltokorpi and Vaara 2014). The thesis also contributes to the ongoing discussion on the implementation of language-sensitive recruitment by examining a shortage of potential employees with sufficient English skills in the Japanese labor market (Peltokorpi 2010).

Fourth, the thesis nuances previous theories in linguistic instrumentalism (Kubota 2011). The study demonstrates that the TOEIC score, as used by Japanese MNCs, can be misleading if not considered carefully. Clearly, the TOEIC score satisfies a commercial need for a practical, objective, and fair test of receptive language skills. However, an overreliance on TOEIC scores has negative implications for viewing language as a social practice rather than a neutral, mechanical instrument. At the same time, this study suggests that test results cannot be used as a simple measure of the English skills employees must develop, because the excessive use of the multiple-choice format in the test cannot simulate real-world language use for business communication purposes. Given this result, the findings are in line with existing research on the significant impact of the TOEIC test as a high-stakes test on the lives of the test takers and other stakeholders (Chapman 2006).

Fifth, another contribution of this thesis lies in reconceptualizing English as

a commodifiable skill. The findings of articles 1 and 2 allow for participation in the ongoing theoretical discussion on the neoliberal employment system. Piller and Cho (2013) argue that internationalization is tied to English as an index of global competitiveness. The spread of English is a result not of the free linguistic market but of a “systematic, organized, and orchestrated policy” (Piller and Cho 2013: 38). The results described in article 2 demonstrate that the power of English and the neoliberal employment system affect employees in complex ways, with gender, units and types of employment playing significant roles. Further, the data (Article 1) reveals that bilinguals are still in high demand. Consequently, the findings of article 2 are in line with Urciuoli (2008), who shows that language is a commodifiable skill in the neoliberal employment system.

Finally, the present thesis contributes to the literature on cultural expatriates in MNCs, both in terms of the wide use of Japanese, counter to official organizational language policy, as a symbol of discontent and ethnic loyalty (article 2), and by examining cultural expatriates’ linguistic ethnocentrism and perceptions of Japanese work practices and values transmitted by the English-only policy over time (article 3).

5.2 Methodological contributions and applied perspectives

The present thesis exemplifies how computer-assisted text analysis can be carried out and how it may contribute to an interdisciplinary study of language and communication practices in international business and management.

First, while there has been extensive study of social media including Twitter, Facebook, blogs and discussion boards, the data sources in this thesis have not

been as thoroughly investigated: online bilingual job advertisements, as represented by OpenWork, and Glassdoor. I have found that analysis of these professionally oriented data sources offered complementary insights into employees' thoughts and concerns about corporate language policy and in other respects.

Second, the thesis has developed a computer-assisted analytical framework. Through anonymous posts, employees freely post their activities, feelings, and thoughts about their company. This framework could be used by researchers or human resources departments to leverage these data to monitor and promote employees' motivation. It could also help human resources departments design and plan English training programs to improve corporate language policy in the long term.

Finally, the thesis discussed, developed, and implemented data-driven content analysis to gain understanding of trends in the reactions of employees as suggested by language patterns, thus gathering insight for further investigation. This method of using language as a feature is attractive and often easily interpreted. For example, word clouds in articles 2 and 3 provide a useful method for visually communicating how users orientation to different topics evolves over time.

5.3 Managerial implications

Two different concepts of English as corporate language in job advertisements and social media have repercussions for a MNC's language policy regarding recruitment needs and language instrumentalism.

First, to address the disparity between recruitment needs and the official language, a language policy must align with the language requirements in Japanese

and English job advertisements. “Business level English” in job advertisements has been defined by using TOEIC scores. However, the required scores of TOEIC in job advertisements can differ significantly depending on what role has to be fulfilled in the organization. For example, most of the advertisements require an 800 TOEIC score or more for new hires, whereas some have a lower-score requirement. Moreover, Japanese-only advertisements present hegemony in professional skills and defy the corporate language policy. In my view, increasing recruitment and compensation budgets for employing candidates who possess the required language and professional skills could solve the mismatch between corporate policy and recruitment practice.

Second, to provide employees with an incentive to increase their “commodity value” in communication, I call for adopting the idea of competence in working language other than using the TOEIC score. I would like to emphasize the limitations and weaknesses of using the TOEIC to measure competence in “working” language, as discussed in the previous chapter. Although I understand that the TOEIC as a standard test is widely used in Japan, the current trend of over-emphasizing the TOEIC scores may mislead MNCs into thinking that employees with high scores can be an automatic guarantee of successful communication in a particular job-related situation. The present study illustrates clearly that the policy should encourage employees to communicate at work instead of depending on various formats of language tests.

Third, to bypass some of these challenges on using the TOEIC score in the recruitment, I argue that a job simulation assessment in speaking and listening skills should be implemented as part of the recruitment process. Job simulations, or work

simulations, can be defined as a type of assessment used during the hiring process. When designing this practically oriented job simulation assessment, one should make sure to present job applicants with a realistic scenario similar to what they will experience while they are in their actual work. For instance, for a MNC is looking to hire a software engineer, a comprehension and speaking task might involve playing the candidate a snippet of a conversation, in which the applicant then has to explain his previous work to a non-technical employee. An alternative assessment might involve the candidate having to confirm what they heard in lecture snippets from international tech conferences where the dominant language is English. I strongly suggest that such job simulation tasks must represent typical on-the-job communication situations in a realistic way. For example, if a candidate need to communicate with customers with specific vocabularies, that should be included in the simulation.

Last but not the least, this study and previous research have shown that managers need to consider the significance of language and communication skills for staff motivation and job satisfaction in a multilingual organization (Charles 2007). Anonymous feedback encourages employees to bring up tough topics in a safe environment. Once MNCs collect data from anonymous employee review websites, one of the most effective ways to enhance employees' job satisfaction is implementing a feedback loop, a management mechanism that facilitates continuous improvement. For example, Glassdoor supports "Adding an employer response" function. An organization can give continuous feedback to any company review posted on its company profile. Within an organization, encouraging regular, two-way communication between management and employees, such feedback loops

could considerably help to address potential language policy challenges at an early stage.

5.4 Limitations and suggestions for future research

Undoubtedly, the present thesis has some limitations. In order to increase the validity of these results, I have used various data (bilingual advertisements, Japanese and non-Japanese employees' posts), which review the Japanese MNC's adoption of English into its organization from both a macro and micro perspective. However, it is challenging to generalize this study's results because all the data were collected from the same company. More case studies are needed to support these results.

One may claim that the posts may not be representative of language management in other Japanese MNCs. However, the computer-assisted text analysis framework that I have adopted is valuable and may be applied to other companies that make the choice to use English as a corporate language. Additionally, the method looks at longitudinal data, which is suitable for investigating before-and-after scenarios (e.g., how do Japanese employees' attitudes change before and after Englishnization). It may help language policy makers to make more informed decisions in implementing HR practices that promote language learning.

One may also argue that the data-driven analysis reveals correlation but not causality. Sampling from a single social media platform may be considered biased. Moreover, although the data collection period is long, the dataset is not particularly large. A larger sample size is one way to mitigate errors and bias in future research. Social media platforms are gradually being used by broader demographic segments

(Schwartz and Ungar 2015). Articles 2 and 3 clearly demonstrate that career community websites have gained in popularity and trend toward increased usage by employees.

Finally, future research could combine DTM with other techniques such as sentiment analysis, which could demonstrate the changes in employee sentiment related to Japanese work practices and values. For a broader view, there is a need for future studies to investigate the questions posed in this thesis from different data sources and companies in different locations.

6 References

- Ahmad, F. and Widén, G. (2015), 'Language clustering and knowledge sharing in multilingual organizations: A social perspective on language', *Journal of Information Science* **41**(4), 430–443.
- Alsumait, L., Barbará, D. and Domeniconi, C. (2008), On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking, in '2008 Eighth IEEE International Conference on Data Mining', IEEE, pp. 3–12.
- Ammon, U. (2010), 'World languages: Trends and futures', *The handbook of language and globalization* pp. 101–122. Hoboken: John Wiley & Sons.
- Andersen, H. and Rasmussen, E. S. (2004), 'The role of language skills in corporate communication', *Corporate Communications: An International Journal* .
- Angouri, J. (2014), 'Multilingualism in the workplace: Language practices in multilingual contexts', *Multilingua* **33**(1-2), 1–9.
- Anthony, L. (2003), Integrating the TOEIC into the university English curriculum, in 'Proceedings of JACET 42th Annual Convention', pp. 149–149.
- Argamon, S., Koppel, M., Fine, J. and Shimoni, A. R. (2003), 'Gender, genre, and writing style in formal written texts', *Text & Talk* **23**(3), 321–346.
- Backhaus, K. B. (2004), 'An exploration of corporate recruitment descriptions on monster.com', *The Journal of Business Communication (1973)* **41**(2), 115–136.
- Baldauf Jr, R. B. (2012), 'Introduction-language planning: Where have we been? where might we be going?', *Revista Brasileira de Linguística Aplicada* **12**(2), 233–248.

- Barner-Rasmussen, W., Ehrnrooth, M., Koveshnikov, A. and Mäkelä, K. (2014), 'Cultural and language skills as resources for boundary spanning within the mnc', *Journal of International Business Studies* **45**(7), 886–905.
- Berthoud, A.-C., Grin, F. and Lüdi, G. (2015), 'The dylan project: "language dynamics and management of diversity" ', *European Journal of Applied Linguistics* **3**(1), 145–153.
- Björkman, I., Barner-Rasmussen, W. and Vaara, E. (2012), 'Towards crossvergence, micro-level approaches and critical perspectives: Introduction to the special issue on new perspectives on international management research'.
- Blackman, A. (2006), 'Graduating students' responses to recruitment advertisements', *The Journal of Business Communication (1973)* **43**(4), 367–388.
- Blei, D. M. and Lafferty, J. D. (2006), Dynamic topic models, in 'Proceedings of the 23rd International Conference on Machine Learning', ACM, pp. 113–120.
- Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003), 'Latent dirichlet allocation', *Journal of Machine Learning Research* **3**(Jan), 993–1022.
- Bonner, A. and Tolhurst, G. (2002), 'Insider-outsider perspectives of participant observation', *Nurse Researcher (through 2013)* **9**(4), 7.
- Brandt, W. and Chancellor, F. G. (2011), 'Report on language management strategies and best practice in European SMEs: The PIMLICO project'. <https://ec.europa.eu/assets/eac/languages/policy/>.
- Brannen, M. Y. (2004), 'When mickey loses face: Recontextualization, semantic fit, and the semiotics of foreignness', *Academy of Management Review* **29**(4), 593–616.
- Brannen, M. Y., Garcia, D. and Thomas, D. C. (2009), Biculturals as natural bridges

- for intercultural communication and collaboration, *in* 'Proceedings of the 2009 international workshop on Intercultural collaboration', pp. 207–210.
- Brannen, M. Y., Piekkari, R. and Tietze, S. (2017), The multifaceted role of language in international business: Unpacking the forms, functions and features of a critical challenge to mnc theory and performance, *in* 'Language in international business', Springer, pp. 139–162.
- Brannen, M. Y. and Thomas, D. C. (2010), 'Bicultural individuals in organizations: Implications and opportunity'.
- Brannick, T. and Coghlan, D. (2007), 'In defense of being "native" : The case for insider academic research', *Organizational research methods* **10**(1), 59–74.
- Brookes, G. and McEnery, T. (2019), 'The utility of topic modelling for discourse studies: A critical evaluation', *Discourse Studies* **21**(1), 3–21.
- Chandelier, M., Steuckardt, A., Mathevet, R., Diwersy, S. and Gimenez, O. (2018), 'Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling', *Biological Conservation* **220**, 254–261.
- Chandler Jr, A. D. (1969), *Strategy and structure: Chapters in the history of the American industrial enterprise*, Vol. 120, Cambridge: MIT press.
- Chapman, M. (2006), An over-reliance on discrete item testing in the Japanese business context, *in* 'International Conference on English Instruction and Assessment', Taiwan.
- Charles, M. (2007), 'Language matters in global communication: Article based on ORA lecture, October 2006', *The Journal of Business Communication* (1973) **44**(3), 260–282.
- Charles, M. and Marschan-Piekkari, R. (2002), 'Language training for enhanced

- horizontal communication: A challenge for mncs', *Business Communication Quarterly* **65**(2), 9–29.
- Chuang, J., Ramage, D., Manning, C. and Heer, J. (2012), Interpretation and trust: Designing model-driven visualizations for text analysis, in 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', pp. 443–452.
- Chung, N. C., Miasojedow, B., Startek, M. and Gambin, A. (2019), 'Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data', *BMC Bioinformatics* **20**(15), 1–11.
- Clausen, L. (2007), 'Corporate communication challenges: A 'negotiated' culture perspective', *International Journal of Cross Cultural Management* **7**(3), 317–332.
- Cook, D. M., Waugh, B., Abdipanah, M., Hashemi, O. and Rahman, S. A. (2014), 'Twitter deception and influence: Issues of identity, slacktivism, and puppetry', *Journal of Information Warfare* **13**(1), 58–71.
- Cornelissen, J. (2020), *Corporate communication: A guide to theory and practice*, London: Sage.
- Das Swain, V., Saha, K., Reddy, M. D., Rajvanshy, H., Abowd, G. D. and De Choudhury, M. (2020), Modeling organizational culture with workplace experiences shared on glassdoor, in 'Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems', pp. 1–15.
- DiMaggio, P. (2015), 'Adapting computational text analysis to social science (and vice versa)', *Big Data & Society* **2**(2), 2053951715602908.
- DiMaggio, P., Nag, M. and Blei, D. (2013), 'Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper

- coverage of us government arts funding’, *Poetics* **41**(6), 570–606.
- EC (2007), ‘Effects on the European Union economy of shortages of foreign language skills in enterprise (ELAN)’, *London: CILT, the National Centre for Languages* .
- Ehrenreich, S. (2010), ‘English as a business lingua franca in a german multinational corporation: Meeting the challenge’, *The Journal of Business Communication* (1973) **47**(4), 408–431.
- Ehrenreich, S. (2015), English as a lingua franca (elf) in international business contexts: key issues and future perspectives, in ‘Exploring ELF in Japanese academic and business contexts’, Routledge, pp. 147–167.
- Eisenstein, J., O’Connor, B., Smith, N. A. and Xing, E. P. (2014), ‘Diffusion of lexical change in social media’, *PloS one* **9**(11), e113114.
- Elangovan, V. K. and Eisenstein, J. (2015), ” you’re mr. lebowski, i’m the dude”: Inducing address term formality in signed social networks, in ‘HLT-NAACL’.
- EPI, E. (2020), ‘EF English proficiency index a ranking of 100 countries and regions by English skills’, *Jakarta: EF* .
- ETS (2020), *TOEFL iBT Test and Score Data Summary 2019*, Princeton, NJ: Educational Testing Service.
- Fairclough, N. (2013), *Critical discourse analysis: The critical study of language*, Abingdon: Routledge.
- Feely, A. J. (2004), The impact of the language barrier on the management of multinational companies, PhD thesis, Aston University, Birmingham.
- Feely, A. J. and Harzing, A.-W. (2003), ‘Language management in multinational companies’, *Cross Cultural Management: An International Journal*

- 10(2), 37–52.
- Foundation, J. (2022a), ‘N1-n5: Summary of linguistic competence required for each level’.
- URL:** <https://www.jlpt.jp/e/about/levelsummary.html>
- Foundation, J. (2022b), ‘What is the japanese-language proficiency test?’.
- URL:** <https://www.jlpt.jp/e/about/purpose.html>
- Franzosi, R. (2010), *Quantitative narrative analysis*, number 162, Ca: Sage.
- Gaibrois, C. (2016), Turning away from national languages: How hybrid forms of language use affect communication in multilingual organizations., in ‘10th International GEM&L (Groupe d’Etudes Management et Langage) Conference’.
- URL:** <http://www.alexandria.unisg.ch/250321/>
- Garley, M. and Hockenmaier, J. (2012), Beefmoves: dissemination, diversity, and dynamics of English borrowings in a German hip hop forum, in ‘Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)’, pp. 135–139.
- Giddens, A. (1991), *Modernity and self-identity: Self and society in the late modern age*, Stanford university press.
- Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J., Dobson, R. J. and Dutta, R. (2017), ‘Characterisation of mental health conditions in social media using informed deep learning’, *Scientific Reports* 7, 45141.
- Golsorkhi, D., Rouleau, L., Seidl, D. and Vaara, E. (2010), *Cambridge handbook of strategy as practice*, Cambridge: Cambridge University Press.
- Goodijk, R. (2002), ‘Partnership at corporate level: The meaning of the stakeholder model’, *Journal of Change Management* 3(3), 225–241.

- Griffiths, T. L. and Steyvers, M. (2004), 'Finding scientific topics', *Proceedings of the National Academy of Sciences* **101**(suppl 1), 5228–5235.
- Grimmer, J. and Stewart, B. M. (2013), 'Text as data: The promise and pitfalls of automatic content analysis methods for political texts', *Political Analysis* **21**(3), 267–297.
- Grin, F., Sfreddo, C. and Vaillancourt, F. (2011), *The economics of the multilingual workplace*, Abingdon: Routledge.
- Hannigan, T. R., Haans, R. F., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S. and Jennings, P. D. (2019), 'Topic modeling in management research: Rendering new theory from textual data', *Academy of Management Annals* **13**(2), 586–632.
- Harzing, A.-W., Köster, K. and Magner, U. (2011), 'Babel in business: The language barrier and its solutions in the hq-subsiidiary relationship', *Journal of World Business* **46**(3), 279–287.
- Hashimoto, K. (2011), 'Compulsory 'foreign language activities' in Japanese primary schools', *Current Issues in Language Planning* **12**(2), 167–184.
- Heller, M. (2010), 'The commodification of language', *Annual review of Anthropology* **39**, 101–114.
- Hovy, D., Johannsen, A. and Søgaaard, A. (2015), User review sites as a resource for large-scale sociolinguistic studies, in 'Proceedings of the 24th international conference on World Wide Web', pp. 452–461.
- Huang, M., Li, P., Meschke, F. and Guthrie, J. P. (2015), 'Family firms, employee satisfaction, and corporate performance', *Journal of Corporate Finance* **34**, 108–127.

- Inman, M. (1980), 'Foreign languages and the US multinational corporation', *The Modern Language Journal* **64**(1), 64–74.
- Jameson, D. A. (2007), 'Reconceptualizing cultural identity and its role in intercultural business communication', *The Journal of Business Communication* (1973) **44**(3), 199–235.
- Janssens, M. and Steyaert, C. (2014), 'Re-considering language within a cosmopolitan understanding: Toward a multilingual franca approach in international business studies', *Journal of International Business Studies* **45**(5), 623–639.
- Jenkins, J. (2015), 'Repositioning english and multilingualism in english as a lingua franca', *Englishes in Practice* **2**(3), 49–85.
- Jenkins, J., Cogo, A. and Dewey, M. (2011), 'Review of developments in research into english as a lingua franca', *Language teaching* **44**(3), 281–315.
- Jernudd, B. H. and Neustupný, J. V. (1987), Language planning: for whom, in 'Proceedings of the international colloquium on language planning', Les Presses de l' Université Laval Québec, pp. 69–84.
- Johannsen, A., Hovy, D. and Søgaard, A. (2015), Cross-lingual syntactic variation over age and gender, in 'Proceedings of the Nineteenth Conference on Computational Natural Language Learning', pp. 103–112.
- Kangasharju, H., Piekkari, R. and Sääntti, R. (2010), The language policy in international firms: Where is it hiding?, in '26th European Group of Organizational Studies Colloquium (EGOS), Lisbon, Portugal, July 1-3, 2010', European Group of Organizational Studies (EGOS).
- Kankaanranta, A., Karhunen, P. and Louhiala-Salminen, L. (2018), ' "english

- as corporate language” in the multilingual reality of multinational companies’, *Multilingua* **37**(4), 331–351.
- Kankaanranta, A. and Louhiala-Salminen, L. (2007), ‘Business communication in BELF’, *Business Communication Quarterly* **70**(1), 55–59.
- Kankaanranta, A. and Louhiala-Salminen, L. (2010), “‘English?–Oh, it’s just work!’: A study of BELF users’ perceptions’, *English for Specific Purposes* **29**(3), 204–209.
- Kankaanranta, A. and Lu, W. (2013), ‘The evolution of english as the business lingua franca: Signs of convergence in chinese and finnish professional communication’, *Journal of Business and Technical Communication* **27**(3), 288–307.
- Kankaanranta, A. and Planken, B. (2010), ‘Belf competence as business knowledge of internationally operating business professionals’, *The Journal of Business Communication (1973)* **47**(4), 380–407.
- Keeley, B. (2007), ‘Human capital: A revolution?’, *OECD Observer* (261), 25–27.
- Keeley, T. (2001), *International human resource management in Japanese firms: Their greatest challenge*, Springer.
- Kobayashi, Y. (2013), ‘Global English capital and the domestic economy: The case of Japan from the 1970s to early 2012’, *Journal of Multilingual and Multicultural Development* **34**(1), 1–13.
- Kubota, R. (2011), ‘Questioning linguistic instrumentalism: English, neoliberalism, and language tests in Japan’, *Linguistics and Education* **22**(3), 248–260.
- Kuhn, T. and Nelson, N. (2002), ‘Reengineering identity: A case study of multiplicity and duality in organizational identification’, *Management*

- communication quarterly* **16**(1), 5–38.
- Labov, W. and Waletzky, J. (1997), ‘Narrative analysis: Oral versions of personal experience.’
- LaFromboise, T., Coleman, H. L. and Gerton, J. (1993), ‘Psychological impact of biculturalism: evidence and theory.’, *Psychological bulletin* **114**(3), 395.
- Lancichinetti, A., Sirer, M. I., Wang, J. X., Acuna, D., K rding, K. and Amaral, L. A. N. (2015), ‘High-reproducibility and high-accuracy method for automated topic classification’, *Physical Review X* **5**(1), 011007.
- Latukha, M., Doleeva, A., J rlstr m, M., Jokinen, T. and Piekkari, R. (2016), ‘Does corporate language influence career mobility? evidence from mncs in russia’, *European Management Journal* **34**(4), 363–373.
- Lee, M. and Martin, J. L. (2015), ‘Coding, counting and cultural cartography’, *American Journal of Cultural Sociology* **3**(1), 1–33.
- Lina, C. and Xijin, T. (2014), ‘Trends of bbs topics based on dynamic topic model’, *Journal of Management Sciences in China* **17**(11).
- L nsmann, D. (2011), English as a corporate language: language choice and language ideologies in an international company in Denmark, PhD thesis, Roskilde Universitet.
- L nsmann, D. (2014), ‘Linguistic diversity in the international workplace: Language ideologies and processes of exclusion’, *Multilingua* **33**(1-2), 89–116.
- L nsmann, D. (2015), ‘Language ideologies in a danish company with english as a corporate language: ‘it has to be english’ ’, *Journal of Multilingual and Multicultural Development* **36**(4), 339–356.
- L nsmann, D. (2017), ‘Embrace it or resist it? Employees’ reception of corporate

- language policies’, *International Journal of Cross Cultural Management* **17**(1), 101–123.
- Lønsmann, D. and Mortensen, J. (2018), ‘Language policy and social change: A critical examination of the implementation of an English-only language policy in a Danish company’, *Language in Society* **47**(3), 435–456.
- Louhiala-Salminen, L., Charles, M. and Kankaanranta, A. (2005), ‘English as a lingua franca in nordic corporate mergers: Two case companies’, *English for Specific purposes* **24**(4), 401–421.
- Louhiala-Salminen, L. and Kankaanranta, A. (2012), ‘Language as an issue in international internal communication: English or local language? if english, what english?’, *Public Relations Review* **38**(2), 262–269.
- Luo, N., Zhou, Y. and Shon, J. (2016), Employee satisfaction and corporate performance: Mining employee reviews on glassdoor. com, in ‘Proceedings of the 2016 International Conference on Information Systems’.
- Luo, Y. and Shenkar, O. (2006), ‘The multinational corporation as a multilingual community: Language and organization in a global context’, *Journal of International Business Studies* **37**(3), 321–339.
- Maki, M., Azusa, E. and Tomoki, S. (2015), ‘Nihon kigyō honsha ni okeru jinji kokusaika no genjō to kadai’, *The Internationalization of Human Resource Practices in Japanese Headquarters*. *Multinational Enterprises* **8**(1), 93–113.
- Makoni, S. and Pennycook, A. (2012), Disinventing multilingualism: From monological multilingualism to multilingua francas, in ‘The Routledge handbook of multilingualism’, Routledge, pp. 451–465.
- Marschan-Piekkari, R., Welch, D. and Welch, L. (1999a), ‘Adopting a common

- corporate language: IHRM implications', *International Journal of Human Resource Management* **10**(3), 377–390.
- Marschan-Piekkari, R., Welch, D. and Welch, L. (1999b), 'In the shadow: The impact of language on structure, power and communication in the multinational', *International Business Review* **8**(4), 421–440.
- Marschan, R., Welch, D. and Welch, L. (1997), 'Language: The forgotten factor in multinational management', *European Management Journal* **15**(5), 591–598.
- Meng, C., Zhang, M. and Guo, W. (2012), 'Evolution of Movie Topics Over Time', *UR L: <http://cs229.stanford.edu/proj2012/MengZhangGuo-EvolutionofMovieTopicsOverTime.pdf>*. Cited April 1, 2014.
- METI (2010), 'Develop global human resources through industry-academia-government collaboration'.
- Milroy, L. and Gordon, M. (2008), *Sociolinguistics: Method and interpretation*, Vol. 13, John Wiley & Sons.
- Mintzberg, H. (1998), 'The structuring of organizations' in H. Mintzberg, J. B. Quinn & S. Ghoshal', *The Strategy Process* pp. 332–353.
- Moniz, A. (2016), 'Inferring employees' social media perceptions of goal-setting corporate cultures and the link to firm value', *Unpublished Working Paper*.
- Monroe, B. L., Colaresi, M. P. and Quinn, K. M. (2008), 'Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict', *Political Analysis* **16**(4), 372–403.
- Moretti, F. (2013), *Distant reading*, London: Verso Books.
- Morstatter, F. and Liu, H. (2017), 'Discovering, assessing, and mitigating data bias in social media', *Online Social Networks and Media* **1**, 1–13.

- Mützel, S. (2015), 'Facing big data: Making sociology relevant', *Big Data & Society* **2**(2), 2053951715599179.
- Nakamura, S. (2016), 'Nihon kigyo no kokusai jinteki shigen kanri', *International human resource management of Japanese firms. In T. Sekiguchi, N. Takeuchi, & C. Iguchi, Kokusai Jinteki Shigen Kanri, [International Human Resource Management]* pp. 240–249.
- Nantakaew, A. (2012), *A survey study on motivation in attending English classes among employees of an engineering services company*, Language Institute, Thammasat University.
- Neeley, T. (2011), 'Language and Globalization: 'Englishization' at Rakuten', *Harvard Business School organizational behavior unit case* (412–002).
- Neeley, T. (2017), *The language of global success: How a common tongue transforms multinational organizations*, Princeton: Princeton University Press.
- Neeley, T. B. (2013), 'Language matters: Status loss and achieved status distinctions in global organizations', *Organization Science* **24**(2), 476–497.
- Neeley, T. B., Hinds, P. J. and Cramton, C. D. (2012), 'The (un) hidden turmoil of language in global collaboration', *Organizational Dynamics* **41**(3), 236–244.
- Nekvapil, J. and Sherman, T. (2009), 'Pre-interaction management in multinational companies in Central Europe', *Current Issues in Language Planning* **10**(2), 181–198.
- Nelson, L. K. (2020), 'Computational grounded theory: A methodological framework', *Sociological Methods & Research* **49**(1), 3–42.
- Newfields, T. et al. (2005), 'TOEIC washback effects on teachers: A pilot study at one university faculty', *The Economic Review of Toyo University* **31**(1), 83–105.

- Nguyen, D., Dođruöz, A. S., Rosé, C. P. and De Jong, F. (2016), 'Computational sociolinguistics: A survey', *Computational Linguistics* **42**(3), 537–593.
- Nguyen, D., Trieschnigg, D. and Meder, T. (2014), Tweetgenie: Development, evaluation, and lessons learned, in 'Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations', pp. 62–66.
- Nickerson, C. (2005), 'English as a lingua franca in international business contexts', *English for Specific Purposes* **4**(24), 367–380.
- OKABE, Y. (2009), 'Language Benefits and Language Costs in Japanese-style Management', pp. 1–25. Kyoto Sangyo University Library.
- Oliveri, M. E. and Tannenbaum, R. J. (2017), 'Insights into using TOEIC test scores to inform human resource management decisions', *ETS Research Report Series* **2017**(1), 1–7.
- Park, H., Dai Hwang, S. and Harrison, J. K. (1996), 'Sources and consequences of communication problems in foreign subsidiaries: The case of United States firms in South Korea', *International Business Review* **5**(1), 79–98.
- Park, J. S.-Y. and Wee, L. (2013), *Markets of English: Linguistic capital and language policy in a globalizing world*, Abingdon: Routledge.
- Peltokorpi, V. (2010), 'Intercultural communication in foreign subsidiaries: The influence of expatriates' language and cultural competencies', *Scandinavian Journal of Management* **26**(2), 176–188.
- Peltokorpi, V. and Vaara, E. (2014), 'Knowledge transfer in multinational corporations: Productive and counterproductive effects of language-sensitive recruitment', *Journal of International Business Studies* **45**(5), 600–622.

- Phelan, S. (2014), Neoliberal discourse: Theory, history and trajectories, in 'Neoliberalism, Media and the Political', Springer, pp. 34–55.
- Phillipson, R. (2008), 'The linguistic imperialism of neoliberal empire', *Critical inquiry in language studies* 5(1), 1–43.
- Piekkari, R. (2006), 'Language effects in multinational corporations: A review from an international human resource management perspective', *Handbook of research in international human resource management* pp. 536–550.
- Piekkari, R. and Tietze, S. (2011), 'A world of languages: Implications for international management research and practice', *Journal of World Business* 46(3), 267–269.
- Piekkari, R., Vaara, E., Tienari, J. and Sääntti, R. (2005), 'Integration or disintegration? Human resource implications of a common corporate language decision in a cross-border merger', *The International Journal of Human Resource Management* 16(3), 330–344.
- Piekkari, R., Welch, D. and Welch, L. S. (2014), *Language in international business: The multilingual reality of global business expansion*, Cheltenham: Edward Elgar Publishing.
- Pilcher, N. and Richards, K. (2017), 'Challenging the power invested in the International English Language Testing System (IELTS): Why determining 'English' preparedness needs to be undertaken within the subject context', *Power and Education* 9(1), 3–17.
- Piller, I. and Cho, J. (2013), 'Neoliberalism as language policy', *Language in Society* 42(1), 23–44.

- Piller, I. and Lising, L. (2014), 'Language, employment, and settlement: Temporary meat workers in australia', *Multilingua* **33**(1-2), 35–59.
- Pullin, P. (2010), 'Small talk, rapport, and international communicative competence: Lessons to learn from belf', *The Journal of Business Communication (1973)* **47**(4), 455–476.
- Rafaeli, A. and Oliver, A. L. (1998), 'Employment ads: A configurational research agenda', *Journal of Management Inquiry* **7**(4), 342–358.
- Reeves, N. and Wright, C. (1996), *Linguistic auditing: A guide to identifying foreign language communication needs in corporations*, Vol. 9, Bristol: Multilingual Matters.
- RP Cuypers, I., Ertug, G. and Hennart, J.-F. (2015), 'The effects of linguistic distance and lingua franca proficiency on the stake taken by acquirers in cross-border acquisitions', *Journal of International Business Studies* **46**(4), 429–442.
- Ryan, R. M. and Deci, E. L. (2000), 'Intrinsic and extrinsic motivations: Classic definitions and new directions', *Contemporary educational psychology* **25**(1), 54–67.
- Sanden, G. R. (2015), *Language strategies in multinational corporations: A cross-sector study of financial service companies and manufacturing companies*, Frederiksberg: Copenhagen Business School (CBS).
- Schwartz, H. A. and Ungar, L. H. (2015), 'Data-driven content analysis of social media: a systematic overview of automated methods', *The ANNALS of the American Academy of Political and Social Science* **659**(1), 78–94.
- Sekiguchi, T., Froese, F. J. and Iguchi, C. (2016), 'International human resource

- management of Japanese multinational corporations: Challenges and future directions’, *Asian Business & Management* **15**(2), 83–109.
- Shachaf, P. (2008), ‘Cultural diversity and information and communication technology impacts on global virtual teams: An exploratory study’, *Information & Management* **45**(2), 131–142.
- Skapinker, M. (2017), ‘The Japanese company that ordered staff to speak English’.
URL: <https://www.ft.com/content/2fdd6626-ba3b-11e7-8c12-5661783e5589>
- Spolsky, B. (2004), *Language policy*, Cambridge: Cambridge University Press.
- Spolsky, B. (2009), *Language management*, Cambridge: Cambridge University Press.
- Spolsky, B. (2012), *The Cambridge handbook of language policy*, Cambridge: Cambridge University Press.
- Stahl, G., Björkman, I., Farndale, E., Morris, S., Paauwe, J., Stiles, P., Trevor, J. and Wright, P. (2007), ‘Global talent management: How leading multinationals build and sustain their talent pipeline. INSEAD Faculty and Research Working Papers’, *Fontainebleau: INSEAD* .
- Steyaert, C., Ostendorp, A. and Gaibrois, C. (2011), ‘Multilingual organizations as ‘linguascapes’ : Negotiating the position of english through discursive practices’, *Journal of World Business* **46**(3), 270–278.
- Sun, L. and Yin, Y. (2017), ‘Discovering themes and trends in transportation research using topic modeling’, *Transportation Research Part C: Emerging Technologies* **77**, 49–66.
- Tagliamonte, S. A. (2006), *Analysing sociolinguistic variation*, Cambridge: Cambridge University Press.

- Takino, M. (2016), Negotiating the challenges of using English in business communication: listening narratives of Japanese BELF users, PhD thesis, University of Southampton.
- Tange, H. and Lauring, J. (2009), 'Language management and social interaction within the multilingual workplace', *Journal of Communication Management* **13**(3), 218–232.
- Tanimoto, T. T. (1958), 'Elementary mathematical theory of classification and prediction'.
- Tenzer, H., Pudelko, M. and Harzing, A.-W. (2014), 'The impact of language barriers on trust formation in multinational teams', *Journal of International Business Studies* **45**(5), 508–535.
- Thomas, C. A. (2008), 'Bridging the gap between theory and practice: Language policy in multilingual organisations', *Language Awareness* **17**(4), 307–325.
- Tietze, S. (2010), 'International managers as translators', *European Journal of International Management* **4**(1-2), 184–199.
- Tietze, S. and Dick, P. (2013), 'The victorious english language: Hegemonic practices in the management academy', *Journal of Management Inquiry* **22**(1), 122–134.
- Törnberg, A. and Törnberg, P. (2016), 'Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum', *Discourse & Society* **27**(4), 401–422.
- Tsuda, Y. (2011), 'Eigo o shanai koyogo ni shitewa ikenai mittsu no riyu (three reasons why English should not be made the official language of a company)', *Tokyo: Hanshin Communications* .

- Urciuoli, B. (2008), 'Skills and selves in the new workplace', *American ethnologist* **35**(2), 211–228.
- Vaara, E., Tienari, J., Piekkari, R. and Sääntti, R. (2005), 'Language and the circuits of power in a merging multinational corporation', *Journal of Management Studies* **42**(3), 595–623.
- Van den Born, F. and Peltokorpi, V. (2010), 'Language policies and communication in multinational companies: Alignment with strategic orientation and human resource management practices', *The Journal of Business Communication* (1973) **47**(2), 97–118.
- Van Durme, B. (2012), Streaming analysis of discourse participants, in 'Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning', pp. 48–58.
- Vandermeeren, S. (2003), 'German language needs in danish companies', *HERMES-Journal of Language and Communication in Business* (31), 13–29.
- Varner, I. and Beamer, L. (2005), 'Nonverbal language in intercultural communication', *Intercultural Communication in the Global Workplace, McGraw-Hill* pp. 175–203.
- Volk, S., Köhler, T. and Pudelko, M. (2014), 'Brain drain: The cognitive neuroscience of foreign language processing in multinational corporations', *Journal of International Business Studies* **45**(7), 862–885.
- Walters, N. and Fage-Butler, A. (2014), 'Danish job advertisements: Increasing in complexity', *Communication & Language at Work* **3**(3), 38–52.
- Wang, C., Blei, D. and Heckerman, D. (2012), 'Continuous Time Dynamic Topic Models', *arXiv* pp. arXiv–1206.

- Wang, X. and McCallum, A. (2006), Topics over time: a non-markov continuous-time model of topical trends, *in* 'Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', ACM, pp. 424–433.
- Wee, L. (2003), 'Linguistic instrumentalism in singapore', *Journal of Multilingual and Multicultural Development* **24**(3), 211–224.
- Whittington, R. (1996), 'Strategy as practice', *Long range planning* **29**(5), 731–735.
- Yagi, N. and Kleinberg, J. (2011), 'Boundary work: An interpretive ethnographic perspective on negotiating and leveraging cross-cultural identity', *Journal of international business studies* **42**(5), 629–653.
- Yamao, S. and Sekiguchi, T. (2015), 'Employee commitment to corporate globalization: The role of English language proficiency and human resource practices', *Journal of World Business* **50**(1), 168–179.

7 Article 1

Zhang, Ziyuan. 2021a. “Content Analysis of Language-Sensitive Recruitment Influenced by Corporate Language Policy Using Topic Modeling” *HERMES - Journal of Language and Communication in Business*, no. 61: 77–91.

Ziyuan Zhang*

Content Analysis of Language-Sensitive Recruitment Influenced by Corporate Language Policy Using Topic Modeling

Abstract

Language-sensitive recruitment is a language management tool frequently used by corporate organizations. However, its relationship with corporate policy is lacking; hence, this study aims to consider language-sensitive job advertisements from a computational text analysis perspective and explore the match (or mismatch) between language-sensitive recruitment (English, Japanese, or bilingual) and corporate language policy. This study uses corpus methods combined with topic modeling and text analysis to investigate the influence of corporate language policy on the textual practice of language-sensitive recruitment in a Japanese multinational corporation (MNC). This study finds a considerable discrepancy between recruitment needs and corporate language policy. It also finds that bilinguals still play a key role in crossing language boundaries 10 years post-mandate of the English language policy in this Japanese MNC. The study contributes to business language by exploring an additional scenario for linking language competency with actual recruitment needs. Thus, this study sheds light on the implementation of language-sensitive recruitment in a multilingual corporate context, affecting the communication patterns and recruitment tactics.

Keywords

language management; language-sensitive recruitment; computer-assisted text analysis; topic modeling; bilingual job advertisements

1. Introduction

Language management is a practice-oriented approach to language, focusing on the management of language users (Sanden 2015). Recent research within language management finds that English-only policies may lead to ‘selective recruitment’, also known as language-sensitive recruitment, which refers to ‘practices in which a certain proficiency in the corporate language is used as a precondition for employment’ (Peltokorpi/Vaara 2014). Language-sensitive recruitment appears to be a frequently used language management tool, which may appear relatively easy to implement in practice (Lester 1994; Brandt/Chancellor 2011). However, it is a time-consuming solution with higher recruiting and compensation costs (Grin et al. 2011; Peltokorpi/Vaara 2014). Despite this significant research, an analysis of the complex relationship between language-sensitive recruitment and corporate language policy is lacking. Research is also needed to better understand the role of language management in large international corporations. This study aims to fill the research gap by examining the influence of corporate language policy on the practice of language-sensitive recruitment. More specifically, the study aims to address the following research questions:

1. How is corporate language policy implemented in language-sensitive recruitment?
2. What are the recruitment needs for bilinguals in job advertisements?

This study focuses on Rakuten, one of the world’s largest internet companies with businesses based in 30 countries and regions. The company’s headquarters in Tokyo employs over 10,000 employees, 80% of whom are Japanese. This multinational corporation (MNC) decided to

* Ziyuan Zhang
Department of Professional and Intercultural Communication
Norwegian School of Economics
E-mail: ziyuan.zhang@nhh.no

change its corporate language from Japanese to English in 2010. This policy, referred to as ‘Englishnization’, was seen by CEO and Chairman Hiroshi Mikitani as a means to dismantle cultural and linguistic barriers (Neeley 2011). The CEO proposed the Test of English for International Communication (TOEIC) listening and reading test to improve employees’ English language proficiency, prepare them for global competition, and enable them to meet the English language proficiency requirements of the workplace.

This study uses quantitative and qualitative methods to investigate this case of language policy at both macro and micro levels of language management. In particular, the study is based on a corpus of online bilingual job advertisements from 2019. It uses corpus methods combined with topic modeling and text analysis to investigate the implementation of a corporate language policy ten years after its mandate. This study considers language-sensitive job advertisements from a computational text analysis perspective and explores the match (or mismatch) between language-sensitive recruitment (English, Japanese, or bilingual) and corporate language policy.

This article continues with related literature; it considers language management and language-sensitivity recruitment. Subsequently, it presents methodological and data procedures and findings, and then presents the discussion and conclusion in the last section.

2. Theoretical considerations

2.1 Language Management

Recently, language management has received considerable attention in international business studies. A business strategy tool or language-sensitive approach implies that language management relates to the strategic use of language by management to deliberately control the linguistic-communicative arena of a firm or business unit based on the assumption that it will benefit the company (Sanden 2015). Language needs to be considered a key element in strategically managing a multinational company (Sanden 2015; Marschan-Piekkari et al. 1999b).

Corporate language policy is a type of language management for a company or business organization (Sanden 2015). Many studies, such as that of Lønsmann (2011), Neeley (2017), and Piekkari (2006), posit that using English as a corporate language medium may attract global talent. Lønsmann (2011) claims that the lack of a language policy document has led to different definitions and expectations among employees who experience a discrepancy between their interpretations and the company’s reality. In the case of Rakuten, Englishnization is a strict corporate language policy from Neeley’s perspective; it was required for every employee irrespective of rank, job role, or tenure at the company (Neeley 2017). Current Rakuten employees would face the consequences if they failed to obtain an ideal score on the TOEIC (Neeley 2017). For instance, if an employee fails the test within 2 years, then they might be demoted.

However, the policy is not based on a formal language policy document that explains the Englishnization policy that includes guidelines and recommendations for appropriate language use. Instead, it is based on a mandate from the CEO’s regulatory documents (Neeley 2011). Based on the previous literature, the implementation of Englishnization is divided into three phases. First, after the program was launched in 2010, Rakuten employees had two years to obtain a score above 650 in the TOEIC or face demotion (Neeley 2017). Second, the proficiency bar rose gradually to reach a score of 800 and the points became the mandatory requirement for all new hires (Neeley 2017). The 800-point phase is still underway and emphasizes speaking. Finally, the implementation will focus on global understanding.

2.2 Language-Sensitive Recruitment

A job advertisement is one way to attract the desired applicants to an organization (Blackman 2006). Organizations create job advertisements as a routine activity in the recruitment process (Rafaeli/Oliver 1998). Job advertisements are potential sites where ideology is naturalized (Fairclough 2013). Walters/Fage-Butler (2014) illustrate how the genre of the job advertisement

has developed over time, as well as indicate the features of the genre that characterize the richer job advertisements written today. Backhaus (2004) assesses the importance of word choice in producing recruitment materials. The author advocates that ‘the task of business communicators is to find the appropriate words to pique the curiosity of the desired potential workers and encourage them to continue through the application process’ (Backhaus 2004: 116). Lester (1994: 43) has proposed that ‘the easiest and cheapest way to approach the language problem’ is to hire people with sufficient language skills.

In the business world, a foreign language capability is strictly an ancillary skill (Inman 1980). MNCs can remove language barriers by recruiting employees with the required language proficiency (Reeves/Wright 1996). However, language-sensitive recruitment is not always an appropriate choice and may face many challenges. One challenge is to find qualified candidates due to a shortage of potential employees with sufficient English skills in the local labor market, for example, in Japan (Peltokorpi 2010) and South Korea (Park et al. 1996). If native citizens with insufficient skills are employed while there is a language requirement, they may not perform effectively (Peltokorpi/Vaara 2014). Another challenge is the higher recruitment and compensation costs (Grin et al. 2011; Peltokorpi/Vaara 2014).

Bilinguals in the context of this study, also known as ‘language intermediates’, ‘language brokers’, or ‘bridge individuals’, refer to the employees who become communication channels for people who have difficulty understanding each other (Harzing et al. 2011). Brannen (2004) defines language skills as an individual’s ability to manipulate the linguistic and semantic signals, that together constitute a particular language, to transfer meaning in that language. Barner-Rasmussen et al. (2014) argue that language skills can, in principle, refer to the language of a country, a corporate language, a professional or functional language, a regional lingua franca, and/or ‘company speak’ (Vecchi 2014). Language intermediates are obliged to assist other employees who have insufficient language competency as ‘language nodes’ (Harzing et al. 2011; Marschan-Piekkari et al. 1999a), ‘parallel information networks’ (Harzing et al. 2011) or an emergent language management tool (Sanden, 2015). Takino (2016) states that Japanese Business English as a lingua franca (BELF) users use Japanese as a system in their business community or as an effective strategy, to deal with the challenges in the use of English as a lingua franca in global business. After the lingua franca mandate emerged and people’s language fluency evolved, the demand for bilinguals lessened and then became nearly nonexistent because they were no longer unique (Neeley 2017). However, Neeley (2017) fails to capture sufficient data to focus on this group because she believes demands for bilinguals reduced dramatically, and they represent a miniscule portion of the total employee population.

3. Methodology: Computer-Assisted Text Analysis

Computer-assisted text analysis techniques fall into three main categories: lexical-based (Monroe et al. 2008; Lee/Martin 2015), text classification (Chuang et al. 2012; DiMaggio 2015), and natural language processing (Nelson 2020). Content analysis uses variables to represent the counts (or proportions) of keywords encountered within or across the records of text (Sodhi/Son 2010), and it can be effectively complemented with topic modeling techniques (Nelson 2020). Topic modeling is a generative model for a document: a topic is modeled as a probability distribution over words, and a document as a distribution over topics (Blei et al. 2003; Griffiths/Steinert 2004; Hofmann 1999; Concannon et al. 2018). The latent Dirichlet allocation (LDA) is the most common form of topic modeling that automatically discovers topics within a large collection of text. The model seeks to automatically discover thematically coherent topics, extracted from a large collection of texts based on word co-occurrence (Brookes/McEnery 2019). Topic modeling has been applied to study various themes in different text types, such as political science (Grimmer/Stewart 2013), academic English papers (Murakami et al. 2017), and online forums (Törnberg/Törnberg 2016). Isoaho et al. (2019) argue that content, thematic, and vocabulary analyses can be combined with topic modeling through embedded design, as a predominant method that guides the analysis and

an auxiliary methods, to enhance this process (Creswell and Creswell, 2017). Törnberg/Törnberg (2016) combine critical discourse analysis with topic modeling to automatically and inductively structure the text to find recurrent patterns and then analyze the resulting topics to discern what they are about, by focusing both on the terms most closely associated with each topic and the related documents. The objective of the combination approach is not to merge text analysis and topic modeling into a single synthesis but to use them as complementing steps in the analysis. Content analysis can utilize topic modeling at the exploratory stage, deep reading at the pattern refinement stage, and supervised machine learning approaches at the pattern confirmation stage to provide a robust, rigorous, and reproducible methodological framework (Nelson 2020).

This study's analytical procedure broadly mirrors the approach, including corpus construction, pattern detection using human-centered computational exploratory analysis, and pattern refinement using guided deep reading. Applying the framework to advertisements, the study can identify the important content words in the corpus, reduce complicated text to informative groups of words, and interpret text without the burden of reading the full text.

4. Data and Procedure

This study collected online bilingual job advertisements in 2019 from a Japanese multinational company in its tenth year of the mandate. The study used web crawler to obtain 1,109 advertisements (1,053,178 words) posted to Rakuten's talent section available at (<https://japan-job-en.rakuten.careers/search-jobs>) on November 12, 2019. In total, 683 (61.5%) of these advertisements are bilingual, 129 (11.6%) are in Japanese, and 297 (26.9%) are only in English. The bilingual advertisements can be displayed in the English or Japanese versions, and users can change the display language by clicking the upper right button (Fig. 1).



Figure 1: English and Japanese versions of a bilingual advertisement.

The analysis consisted of three steps. First, the study assessed frequency and occurrence by using the eight most relevant language requirement terms: TOEIC, English, Japanese, English and Japanese, 英語(English), 日本語(Japanese), 英語(English) and 日本語(Japanese), and Japanese Language Proficiency Test (JLPT). This study compared specific language requirement words used for the English and Japanese versions of bilingual advertisements to further explore the differences in the language requirement for bilingual positions. Second, the study extracted all the

sentences, including English and Japanese requirements, from 150 bilingual advertisements. Five main topics were extracted from LDA using the subjective analysis of iterative results.

This study also analyzed 60 job advertisements only in Japanese without any language requirement keywords, which accounts for 50% of the total Japanese-only advertisements. It aimed to dissect texts about misalignment between corporate language and functional skill requirement.

The study used pyLDAvis as the visualization tool (refer to Figs. 3 and 4). The circle area in the left panel represents the importance of each topic over the entire corpus. The distance between the centers of the circles indicates the similarity between topics. For each topic, one should click a circle in the left panel to select a topic; the bar chart on the right side displays the top 30 most relevant terms for the selected topic. When a given term is highlighted, the default areas of circles change to the circles for the term. The red bars indicate the frequency of a term in a given topic, and the blue bars indicate the term's frequency in the entire corpus. Sievert/Shirley (2014) define the relevance for ranking terms within topics. The relevance metric λ determines the weight given to the probability of term w under topic k relative to its lift (Sievert/Shirley 2014). One can change the value of λ to modify the term rankings – small values of λ highlight potentially rare but exclusive terms for the selected topic, and large values of λ (near 1) highlight frequent but not necessarily exclusive terms for the selected topic (Sievert/Shirley, 2014). By default, the terms related to a topic are ranked in decreasing order, according to their topic-specific probability ($\lambda = 1$) and the suggested 'optimal' value of λ is 0.6 (Sievert/Shirley 2014). Changing λ from 1 to 0 allows adjustment of the ranking of the terms based on more relevant terms for the specific topic.

5. Findings

In the following analysis, this article provides insights into how language-sensitive recruitment is implemented in the Rakuten corporation; it explores the construction of candidates in the job advertisements.

5.1 Notion of Corporate Language Policy

The use of the 'official language', 'English', and 'TOEIC' in both Japanese and English in job advertisements reveals that the language-sensitive recruitment of Rakuten considers language policy from three overarching perspectives. First, only 14 job advertisements in Japanese mention language policy as '社内英語公用語化 (the introduction of the company's official language as English)', reflected in the following job description as a supplementary condition of employment, which relates to the employment type: regular employee or contract employee.

The two types of texts below reveal that the TOEIC score is mentioned as a recommendation rather than a mandatory requirement for regular employees. Potential candidates could achieve it after entering the company or taking the TOEIC test during the selection process, and there are no TOEIC requirements for contract employees.

社内英語公用語化に伴い、TOEIC800 点が正社員としての採用推奨要件ですが、TOEIC800 点がなくとも正社員で採用することもございます。正社員採用の場合、入社後にTOEIC800 点を目標して頂きます。契約社員採用の場合、TOEIC のスコアは必要ございません。応募時点で 800 点以上の TOEIC スコアを証明するエビデンスをお持ちでない方については、選考過程で弊社主催 の TOEIC IP テスト（無料）を受験頂きます。

After the introduction of English as the company's official language, the company required TOEIC score of 800 points from its regular employees. Nonetheless, employees would still be hired even if their TOEIC of 800 points are not available. Regarding the employment of regular employees, you will be expected to aim for a TOEIC score of 800 after joining the company. TOEIC scores are not required for contract employees. If you do not have evidence to prove a TOEIC score of 800 or higher at the time of application, you will be required to take the TOEIC IP test sponsored by the company during the selection process. (Translated from Japanese)

【雇用形態】正社員／契約社員【雇用形態の補足（TOEIC スコアについて）】社内英語公用語化に伴い、TOEIC800 点が正社員としての採用要件となります。スコアが 800 点以上の場合は正社員としての採用が可能となります。TOEIC800 未満の場合は契約社員として採用となります。応募時点で 800 点以上の TOEIC スコアを証明するエビデンスをお持ちでない方については、選考過程で弊社主催のTOEIC IP テスト（無料）を受験頂きます。

Employment type: Regular employees / Contract employees [Supplementary information on employment types (about TOEIC score)] With the introduction of the company's official language as English, the TOEIC score of 800 is a requirement for regular employees. If the score is over 800, you can be hired as a regular employee. If your TOEIC is less than 800, you will be hired as a contract employee. If you do not have evidence of a TOEIC score of 800 or higher at the time of application, please take the TOEIC IP test (free) sponsored by our company during the selection process. (Translated from Japanese)

Second, TOEIC is a key index for the lingua franca mandate in Rakuten and represents an 'entity' that implements corporate language policy – this is a dominant perspective. However, the requirement for the TOEIC score is not uniform in the 337 advertisements with the term 'TOEIC'. Most of the advertisements require an 800 score or more for new hires, whereas some have lower requirements for TOEIC, such as 'TOEIC over 800 (or same English level)', 'TOEIC score over 650', 'TOEIC over 600 (or related skill)', and 'TOEIC score of 800+ is required after entering', '(TOEIC 800+) or a desire to reach a similar level', 'TOEIC higher than 800 (acceptable to mark 800+ during job-interview process)', '(better) over TOEIC score 800', 'over 800 TOEIC score at present, or candidate who can score more than 800 within 2 years', and 'TOEIC score is not required'.

5.2 Prevalent Terminology

Table 1 presents the results of the analysis. The proficiency TOEIC score bar has become 800 for all new hires in Rakuten. However, only 18.81% of the 1,792 advertisements mentioned the term 'TOEIC'. The term 'English' was used in 292 advertisements, and this figure rose to 317 for the term '英語 (English)', an increase of 25. By contrast, the term 'Japanese' is mentioned in 317 advertisements, almost half of this figure (167) used the term '日本語 (Japanese)', as shown in the table.

Keywords	Frequency of Occurrence	Number of Advertisements with Words	Occurrence by % of 1792 Advertisements
TOEIC	511	337	18.81
English	387	292	16.29
Japanese	423	317	17.69
English and Japanese	810	207	11.55
英語 (English)	498	317	17.69
日本語 (Japanese)	210	167	9.32
英語 (English) and 日本語 (Japanese)	708	146	8.15
JLPT	6	6	0.33

Table 1: Language requirement keywords and their frequency of occurrence.

Table 2 presents the statistics for the three types of advertisements: the English version of bilingual advertisements, the Japanese version of bilingual advertisements, and advertisements only in Japanese. Interestingly, for bilingual jobs, the average number of words used in the Japanese version (403.70) is much greater than the English version's average number (265.71); linguistic differences may have played a role. However, the average number of words used in the Japanese-only advertisements (455.03) is more than the average number in the Japanese version of bilingual

jobs (403.70). This might reflect more information about Japanese-only advertisements compared to the Japanese version of bilingual advertisements.

Language	Num. of Ads	Statistics
English version of bilingual advertisements	683	Average number of words per advertisement: 265.70
		Number of advertisements with no language requirement keywords/phrases: 357
		Number of advertisements with language requirement keywords/phrases: 326
Japanese version of bilingual advertisements	683	Number of advertisements in Japanese: 20
		Average number of words per advertisement: 403.71
		Number of advertisements with no language requirement keywords/phrases: 393
Only Japanese	129	Number of advertisements with language requirement keywords/phrases: 290
		Average number of words per advertisement: 455.03
		Number of advertisements with no language requirement keywords/phrases: 60
		Number of advertisements with language requirement keywords/phrases: 69

Table 2: Overview of the breakdown for each language.

English Subsample

Language requirement keywords	Frequency of Occurrence	Number of Advertisements with Words	Occurrence by % of 683 Advertisements
TOEIC	173	119	17.42
English	271	205	30.01
Japanese	322	231	33.82
English and Japanese	593	149	21.82
英語 (English)	52	20	2.93
日本語 (Japanese)	19	7	1.02
英語 (English) and 日本語 (Japanese)	71	7	1.02
JLPT	2	2	0.29

Japanese Subsample

Language requirement keywords	Frequency of Occurrence	Number of Advertisements with Words	Occurrence by % of 683 Advertisements
TOEIC	253	148	21.67
English	25	16	2.34
Japanese	7	6	0.88
English and Japanese	32	6	0.88
英語 (English)	363	239	34.99
日本語 (Japanese)	171	141	20.64
英語 (English) and 日本語 (Japanese)	534	120	17.57
JLPT	1	1	0.15

Table 3: Language requirement keywords for bilingual subsamples.

5.3 High Demand for Bilinguals

The study extracted 150 job advertisement texts of bilingual (English and Japanese) requirements, which accounted for 25% of 683 bilingual advertisements. In the Japanese version of bilingual job advertisements, there were further explanations for TOEIC scores in 54 such advertisements. If candidates had the same level of English as the required TOEIC score (800), they could apply

for the position. By contrast, only the TOEIC score requirement was written without further explanation in the English version. The same requirement for the TOEIC score (800) could be described as different levels of English skills by using terms such as ‘proficient’, ‘fluent’, ‘business’, ‘proactive’, and ‘conversational’. Moreover, some bilingual advertisements in the English version had the requirement for Japanese skills using terms such as ‘proficient’, ‘fluency’, ‘business’, or ‘native’. By contrast, their Japanese versions do not mention any requirement for Japanese skills. The content explicitly shows a misalignment between the Japanese version and the English version.

By further analyzing the advertisements, the study found that the job category was widely distributed in both technical positions (e.g., ‘Software Engineer’, ‘Engineering’, ‘R&D scientist’, ‘Architect’, ‘System Engineer’, and ‘QA Engineer’) and business positions (e.g., ‘Marketing’, ‘Legal’, ‘Sales & Business Development’, ‘Business Planning & Strategy’, ‘Service Planner’, and ‘HR & Admin’; Figure 2). Surprisingly, the location has a narrow distribution in Japan, ranging from the Tokyo headquarters to Japanese subsidiaries, and Tokyo headquarters accounts for more than 80% (Fig. 2).

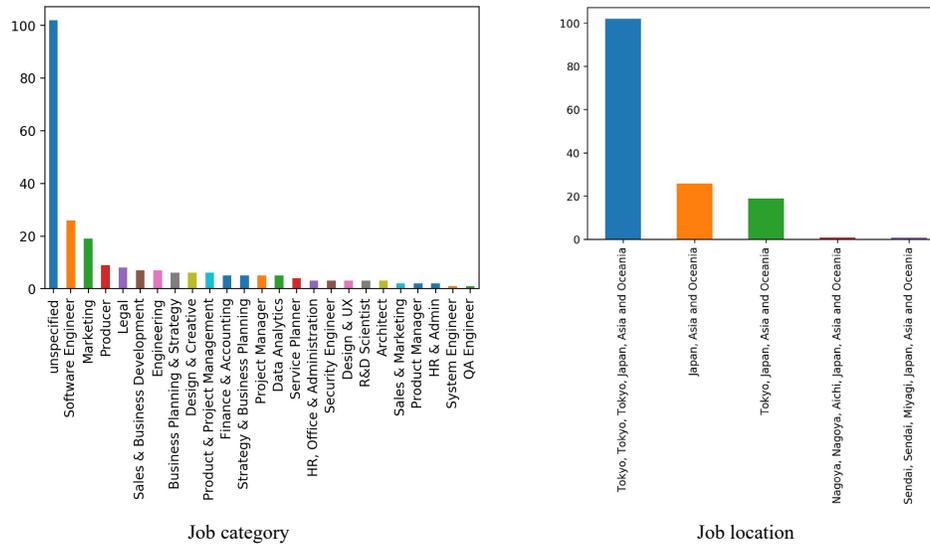


Figure 2: Bilingual advertisements.

The data show that bilingual employees who can speak English and Japanese are in high demand in different roles across different departments in the Japanese headquarters. As can be observed in Table 4, the LDA model presents five separate topics for bilingual requirements.

Num	Topic	Keywords ($\lambda = 1$)	Keywords ($\lambda = 0.2$)
1	Business level	English, Japanese, business, level, toEIC, communication, skills, native, skill, required	business, toEIC, English, Japanese, level, requirement, communication, proficient, speaking, required
2	Management skills	English, level, Japanese, ability, communication, business, native, skill, language, conversational	ability, management, intermediate, advanced, overseas, conversational, subsidiaries, web, conference, mtg
3	Speaking, reading and writing abilities	Japanese, level, business, English, written, skills, fluent, native, skill, spoken	reading, written, spoken, excellent, fluent, position, essential, welcome, subtitles, years
4	Customer-first mindset	Japanese, English, skills, communication, language, native, fluent, experience, excellent, comfortable	comfortable, looking, logic, customer, mindset, communicating, talented, Korean, recruiting, interpersonal
5	Working proficiency	Japanese, English, language, skills, proficiency, speaking, working, native, preferred, presentation	working, professional, role, proficiency, issues, convincing, presentation, discuss, language, preferred

Table 4: Five topics for bilingual requirements.

The screenshot of the pyLDAvis result is shown in Figure 3.

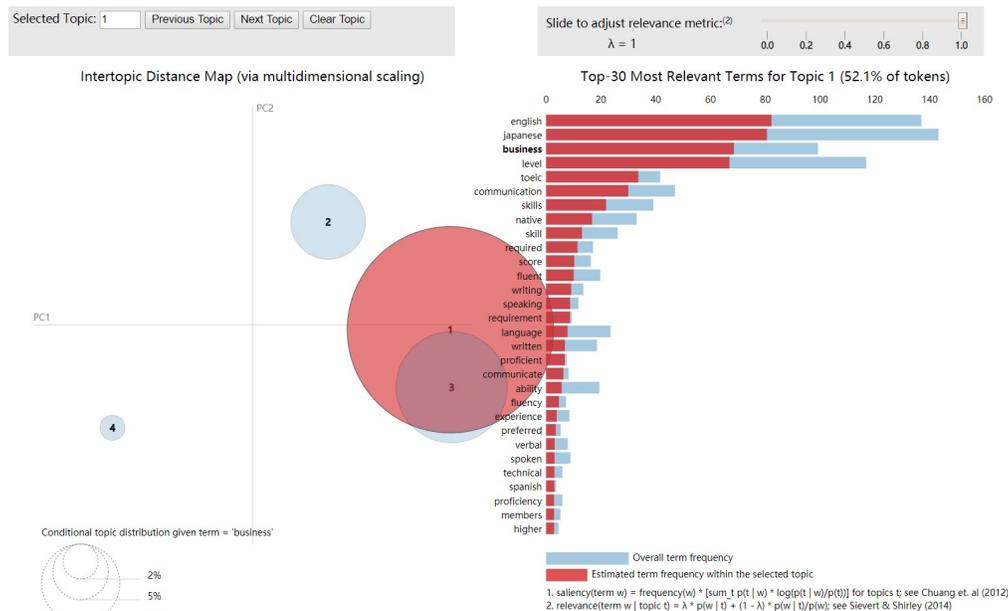


Figure 3: Visualized topic models for bilingual requirements.

For topic one (business level), the top 10 most relevant terms, given that $\lambda = 1$ (setting results in the ranking of terms in decreasing order by probability), are {'English', 'Japanese', 'business', 'level', 'toEIC', 'communication', 'skills', 'native', 'skill', 'required'}, and setting $\lambda = 0$ ranks terms solely by lift, which is the ratio of a term's probability within a topic to its marginal probability across the corpus. The top 10 most relevant terms given $\lambda = 0.2$ are {'business', 'toEIC', 'English', 'level', 'requirement', 'communication', 'proficient', 'speaking', 'required' and 'skills'}. The second set of terms is much more descriptive of the topic being discussed than the first. A set of topics (one and three) together, contains a topic category of bilingual requirements for recruitment. Topic one (business level) focuses on the business level, which is a necessary demand for English and Japanese, adding that TOEIC is required for English, but Japanese does not have a specific requirement for a standard test. Topic three (speaking, reading

and writing abilities) focuses on the description of specific important skills by using terms, such as reading, written, spoken, and verbal, which are suggestive of this requirement.

Hence, topics one and three are interpreted as being about the business level and speaking, and reading and writing abilities, respectively. The presence of these terms in the text suggests that at least a business level of English and Japanese in reading, writing, and speaking are necessary requirements for these 150 jobs.

Excellent communication skills (written, verbal and presentational) in English (Spoken Japanese is highly welcomed) with the ability to communicate issues effectively and comprehensively. (job_description 45)

Topic two (management skills) was interpreted as a bilingual ability with management experience and overseas subsidiaries. Systematically linking bilingual ability to terms such as ‘comfortable’, ‘logic’, ‘customer’, ‘mindset’, and ‘communicating’, topic four (customer-first mindset) indicates that candidates should have a customer-first mindset.

We are looking for someone with a customer-first mindset who is comfortable communicating his logic both in English and in Japanese. (job_description 58)

The keywords ‘working’, ‘email’, ‘presentation’, and ‘telephone’ that characterize topic five (working proficiency) suggest that good presentation skills and the ability to interact effectively in English and Japanese should be a capability. The discourse of job advertisements shows that the demand for these bilinguals does not reduce but still widely exists in Japan because they are still needed after ten years of the post-mandate era:

First, communication with some branch offices in Japan requires Japanese. This is well illustrated in this job advertisement:

Fluent communication using Japanese or English. Producers often need to discuss with members who work at other branch offices in Japan or in other countries. (job_description 64)

Then, one works with internal and external clients to identify their business issues and opportunities in Japan, as two job advertisements detail:

This position also requires the candidate to possess business-level or higher skills in Japanese and English, both written and spoken. (job_description 74)

No fear approach to new business development through cold calling and networking. (job_description 126)

Finally, Japanese is mainly used at work, and English is only as a regulation rule for some positions, such as consultant and producer, which is illustrated in these bilingual advertisements.

ビジネスレベルのコミュニケーション力（日本語中心、英語は当社規定に準ずる）(job_description 126, Japanese version)

Business-level communication skills (mainly Japanese, English conforms to our regulations) (job_description 126, English version)

ご入社後直ぐは英語の使用はございませんが追って発生します。全業務の2~3割 (job_description 321, Japanese version)

Immediately after joining the company, there is no use of English, but it will occur later in 20 to 30% of all operations (job_description 321, English version)

5.4 No Demand for English

The result also shows that English is not in demand, but can be used as a professional skill, such as business and Japanese language skills, for hiring contract employees in some advertisements from the Rakuten travel, e-commerce, and fintech departments. This study ran an LDA set to five topics on this subsample only in Japanese without any language requirement keywords (Table 5).

Num	Topic	Keywords ($\lambda = 1$)	Keywords ($\lambda = 0.6$)
1	Sales experience	楽天 (Rakuten), 経験 (experience), トラベル (travel), 様 (suffix), 営業 (sales), 施設 (facilities), 休暇 (holiday), 宿泊 (lodging), 提案 (proposal), web	楽天 (Rakuten), 経験 (experience), トラベル (travel), 営業 (sales), 様 (suffix), 施設 (facilities), 休暇 (holiday), 宿泊 (lodging), 提案 (proposal), web
2	Business skills	経験 (experience), 楽天 (Rakuten), 様 (suffix), 宿泊 (lodging), 事業 (business), 休暇 (holiday), トラベル (travel), 年 (year), 提案 (proposal), 業務 (work)	経験 (experience), 楽天 (Rakuten), 事業 (business), 業務 (work), 年 (year), 宿泊 (lodging), 目標 (goal), 以上 (above), 制度 (system), 提案 (proposal)
3	Travel business	経験 (experience), 楽天 (Rakuten), トラベル (travel), 施設 (facilities), 保険 (insurance), 月 (month), 休暇 (holiday), 宿泊 (lodging), web, 的 (modifier)	経験 (experience), トラベル (travel), 楽天 (Rakuten), 施設 (facilities), 保険 (insurance), 的 (modifier), 月 (month), 業務 (work), 休暇 (holiday), web
4	Experience requirement	経験 (experience), 楽天 (Rakuten), 年 (year), 様 (suffix), トラベル (travel), 保険 (insurance), 休暇 (holiday), 提案 (proposal), 営業 (sales), 以上 (above)	経験 (experience), 年 (year), 楽天 (Rakuten), サービス (service), 保険 (insurance), 様 (suffix), 提案 (proposal), 休暇 (holiday), トラベル (travel), 以上 (above)
5	Benefits	楽天 (Rakuten), 経験 (experience), トラベル (travel), 様 (suffix), 休暇 (holiday), 施設 (facilities), 集客 (attracting customers), 保険 (insurance), 宿泊 (lodging), 年 (year)	楽天 (Rakuten), 経験 (experience), トラベル (travel), 休暇 (holiday), 様 (suffix), 集客 (attracting customers), 保険 (insurance), 施設 (facilities), 年 (year), 的 (modifier)

Table 5: Japanese-only ads without keywords.

As observed in Figure 4 generated by pyLDavis, the five topics capturing Japanese-only texts present a hegemony in professional and Japanese language skills. For example, the keywords ‘経験 (experience)’ and ‘トラベル (travel)’ that characterize all topics show that professional experience is a primary demand in Rakuten Travel. An optimal value of $\lambda = 0.6$ is set to interpret the topics. Topic one (sales experience), two (business skills), three (travel business), and four (experience requirement) were interpreted as highlighting the importance of sales experience and work experience in travel business by linking to terms such as ‘営業 (sales)’, ‘事業 (business)’, ‘トラベル (travel)’, and ‘業務 (work)’. Topic five (benefits) includes content of employee benefits, bonus, paid time off, and insurance explicitly presented by terms such as ‘年 (year)’, ‘休暇 (holiday)’, and ‘保険 (insurance)’. The topics here focus on how the organization wishes to hire persons who possess ‘primary’ skills, rather than English to serve as their main occupation, which is well illustrated in this Japanese-only advertisement:

必須条件: seo 経験 (2年以上) / web 製作経験者
 歓迎条件: 広告代理店での勤務経験 / web サービス開発経験 ui/ux

Minimum qualifications: SEO experience (2 years or more) / Web production experience
 Preferred qualifications: Work experience at an advertising agency / Web service development experience ui / ux (Translated from Japanese)

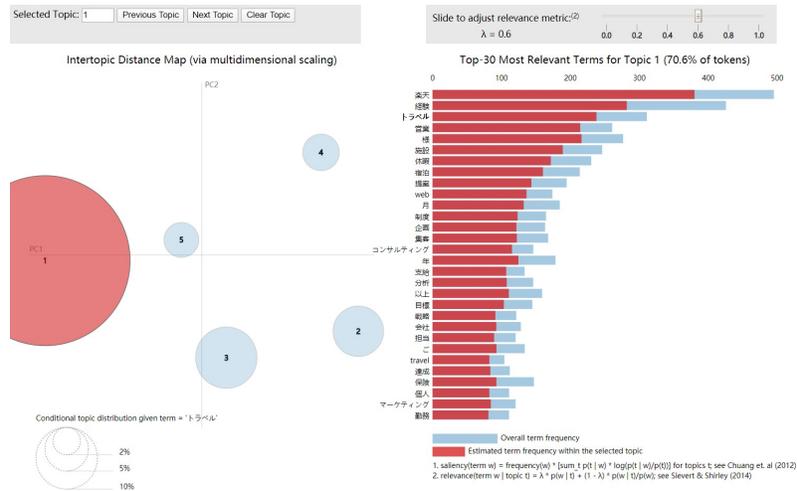


Figure 4: Visualized topic models for Japanese advertisements without language requirement keywords.

6. Discussion and Conclusion

The study aimed to consider language-sensitive job advertisements from a content analysis perspective and explore the match (or mismatch) between language-sensitive recruitment (English, Japanese, or bilingual) and corporate language policy and finally to examine the recruitment needs for bilinguals.

In response to the first research question ‘How is corporate language policy implemented in the language-sensitive recruitment?’, the study found a discrepancy between recruitment needs and corporate language policy. Rakuten’s English policy is strict: any employee contravening the policy would be demoted (Neeley 2017). To ensure the seriousness of the policy, the company requires an 800 TOEIC score for all recruits (Neeley 2017). However, only 337 (18.81%) of the 1,792 advertisements that mentioned the term ‘TOEIC’, 129 advertisements are written solely in Japanese. Japanese and English Rakuten job advertisements’ coverage of corporate language policy differs, contrary to the corporate language policy.

Furthermore, there is a misalignment between the corporate language policy and functional skill requirements in the job advertisements. As previously mentioned, Inman (1980) concludes that there is no demand for foreign language majors unless those individuals also possess another primary skill to serve as their main occupation. Conversely, Japanese-only advertisements present hegemony in professional skills and defy the corporate language policy.

Data reveals that hiring managers prefer potential candidates with a particular proficiency in Japanese and professional ability that satisfies business needs over English skills, corroborating the extant literature on language-sensitive recruitment (Grin et al. 2011; Brandt/Chancellor 2011; Peltokorpi/Vaara 2014). This study can also contribute to the ongoing discussion on the implementation of language-sensitive recruitment. Grin et al. (2011) conclude that English skills are associated with sizeable wage premiums, indicating higher recruitment costs. Another problem with language-sensitive recruitment is a shortage of potential employees with sufficient English skills in the Japanese labor market (Peltokorpi 2010). Hence, increasing recruitment and compensation budgets for employing candidates who possess the required language and professional skills could solve the mismatch between corporate policy and recruitment practice.

The present study can also contribute to the ongoing discussion on recruitment needs for bilinguals. As previously mentioned, Takino (2016) concludes that the Japanese language resource is often used to make English communication more effective in their bilingual business

communication. In contrast to Neeley's (2017) conclusion that negates the need for bilinguals because of employees' improvement in their English level, this study finds that bilinguals are still in high demand. Job advertisements lower the TOEIC score requirements for regular employees, and contract employees are not covered by the corporate language policy. Possible reasons are the shortage of potential employees with sufficient English skills in the Japanese labor market and Japanese still being a required language for business in Japan.

Moreover, the study shows that Japanese is still widely used at work, while English remains a regulation rule for certain professionals, such as consultants and producers. This study also shows that bilinguals act as communication channels to connect internal and external clients to identify their business issues and opportunities in Japan and play a key role in communicating with some branch offices, including domestic and overseas offices.

However, the limitations of the study point to areas for future research. The bilingual advertisements were collected only from a single company. Hence, one cannot generalize the findings to job advertisements in other companies and industries. Future comparative studies are needed to confirm or nuance these results. Furthermore, there needs to be continued research into bilingual advertisements to follow up on on-going changes.

In summary, this study sheds light on the implementation of language-sensitive recruitment in a multilingual corporate context, affecting communication patterns and recruitment tactics. The Rakuten context introduces an additional scenario for linking language competency with actual recruitment needs.

References

- Backhaus, Kristin B. 2004: An exploration of corporate recruitment descriptions on monster.com. In *The Journal of Business Communication* 41:2, 115–136.
- Barner-Rasmussen, Wilhelm/Ehrnrooth, Mats/Koveshnikov, Alexei/Mäkelä, Kristiina 2014: Cultural and language skills as resources for boundary spanning within the MNC. In *Journal of International Business Studies* 45(7), 886–905.
- Blackman, Anna 2006: Graduating students' response to recruitment advertisements. In *The Journal of Business Communication* 43(4), 367–388.
- Blei, David M./Ng, Andrew Y./Jordan, Michael I. 2003: Latent Dirichlet allocation. In *Journal of Machine Learning Research* 3:Jan, 993–1022.
- Brandt, Willy/Chancellor, Former German 2011: *Report on language management strategies and best practice in European SMES: The pimlico project*.
- Brannen, Mary Yoko 2004: When mickey loses face: Recontextualization, semantic fit, and the semiotics of foreignness. In *Academy of Management Review* 29(4), 593–616.
- Brookes, Gavin/McEnery, Tony 2019: The utility of topic modelling for discourse studies: A critical evaluation. In *Discourse Studies* 21(1), 3–21.
- Chuang, Jason/Ramage, Daniel/Manning, Christopher/Heer, Jeffrey 2012: Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Austin.
- Concannon, Shauna Julia/Balaam, Madeline/Simpson, Emma/Comber, Rob 2018: Applying computational analysis to textual data from the wild: A feminist perspective. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM.
- Creswell, John W./Creswell, J. David 2017: *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. New York: Sage Publications.
- DiMaggio, Paul 2015: Adapting computational text analysis to social science (and vice versa). *Big Data & Society* 2(2), 1-5.
- Fairclough, Norman 2013: *Critical Discourse Analysis: The Critical Study of Language*. New York: Routledge.
- Griffiths, Thomas L./Steyvers, Mark 2004: Finding scientific topics. In *Proceedings of the National Academy of Sciences* 101, 5228-5235.
- Grimmer, Justin/Stewart, Brandon M. 2013: Text as data: The promise and pitfalls of automatic content analysis methods for political texts. In *Political Analysis* 21(3), 267–297.

- Grin, François/Sfreddo, Claudio/Vaillancourt, François 2011: *The Economics of the Multilingual Workplace*. New York: Routledge.
- Harzing, Anne-Wil/Köster, Kathrin/Magner, Ulrike 2011: Babel in business: The language barrier and its solutions in the HQ-subsiary relationship. In *Journal of World Business* 46(3), 279–287.
- Hofmann, Thomas 1999: Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 50–57.
- Inman, Marianne 1980: Foreign languages and the US multinational corporation. In *The Modern Language Journal* 64(1), 64–74.
- Isoaho, Karoliina/Gritsenko, Daria/Mäkelä, Eetu 2019: Topic modeling and text analysis for qualitative policy research. In *Policy Studies Journal* 49(1), 300–324.
- Lee, Monica/Martin, John Levi 2015: Coding, counting and cultural cartography. In *American Journal of Cultural Sociology* 3(0), 1–33.
- Lester, Tom 1994: Pulling down the language barrier. In *International Management* 49(6), 42–45.
- Lønsmann, Dorte 2011: *English as a corporate language: language choice and language ideologies in an international company in Denmark*. PhD Thesis. Roskilde Universitet.
- Marschan-Piekkari, Rebecca/Welch, Denice/Welch, Lawrence 1999a: Adopting a common corporate language: IHRM implications. In *International Journal of Human Resource Management* 10(3), 377–390.
- Marschan-Piekkari, Rebecca/Welch, Denice/Welch, Lawrence 1999b: In the shadow: The impact of language on structure, power and communication in the multinational. In *International Business Review* 8(4), 421–440.
- Monroe, Burt L./Colaresi, Michael P./Quinn, Kevin M. 2008: Fightin’ words: Lexical feature selection and evaluation for identifying the content of political conflict. In *Political Analysis* 16(4), 372–403.
- Murakami, Akira/Thompson, Paul/Hunston, Susan/Vajn, Dominik 2017: ‘What is this corpus about?’: Using topic modelling to explore a specialised corpus. In *Corpora* 12(2), 243–277.
- Neeley, Tsedal 2011: Language and globalization: ‘Englishnization’ at Rakuten. In *Harvard Business School Organizational Behavior Unit Case*, 412–002.
- Neeley, Tsedal 2017: *The Language of Global Success: How a Common Tongue Transforms Multinational Organizations*. Princeton: Princeton University Press.
- Nelson, Laura K. 2020: Computational grounded theory: A methodological framework. In *Sociological Methods & Research* 49(1), 3–42.
- Park, Hoon/Dai Hwang, Sun/Harrison, J. Kline 1996: Sources and consequences of communication problems in foreign subsidiaries: The case of united states firms in South Korea. In *International Business Review* 5(1), 79–98.
- Peltokorpi, Vesa 2010: Intercultural communication in foreign subsidiaries: The influence of expatriates’ language and cultural competencies. In *Scandinavian Journal of Management* 26(2), 176–188.
- Peltokorpi, Vesa/Vaara, Eero 2014: Knowledge transfer in multinational corporations: Productive and counterproductive effects of language-sensitive recruitment. In *Journal of International Business Studies* 45(5), 600–622.
- Piekkari, Rebecca 2006: Language effects in multinational corporations: A review from an international human resource management perspective. In: Stahl, G. K. / Björkman, I. (Eds.) *Handbook of Research in International Human Resource Management*, 536–550.
- Rafaeli, Anat/Oliver, Amalya L. 1998: Employment ads: A configurational research agenda. In *Journal of Management Inquiry* 7(4), 342–358.
- Reeves, Nigel/Wright, Colin 1996: *Linguistic Auditing: A Guide to Identifying Foreign Language Communication Needs in Corporations* 9, Bristol: Multilingual Matters.
- Sanden, Guro Refsum 2015: *Language strategies in multinational corporations: A cross-sector study of financial service companies and manufacturing companies*. Frederiksberg: Copenhagen Business School (CBS).
- Sievert, Carson/Shirley, Kenneth 2014: Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. Stroudsburg: Association for Computational Linguistics.
- Sodhi, MS/Son, Byung-Gak 2010: Content analysis of or job advertisements to infer required skills. In *Journal of the Operational Research Society* 61(9), 1315–1327.
- Takino, Miyuki 2016: *Negotiating the challenges of using English in business communication: listening narratives of Japanese BELF users*. PhD Thesis. University of Southampton.
- Törnberg, Anton/Törnberg, Petter 2016: Combining cda and topic modeling: Analyzing discursive connections between islamophobia and anti-feminism on an online forum. In *Discourse & Society* 27(4), 401–422.

- Vecchi, Dardo De 2014: Company-speak: An inside perspective on corporate language. In *Global Business and Organizational Excellence* 33(2), 64–74.
- Walters, Nina/Fage-Butler, Antoinette 2014: Danish job advertisements: Increasing in complexity. *Communication & Language at Work* 3(3), 38–52.

8 Article 2

Zhang, Ziyuan. 2021b. "It is All about TOEIC: Discovering Topics and Trends in Employee Perceptions of Corporate Language Policy" *Journal of Multilingual and Multicultural Development*. 1-19.



It is all about TOEIC: discovering topics and trends in employee perceptions of corporate language policy

Ziyuan Zhang 

Department of Professional and Intercultural Communication, Norwegian School of Economics, Bergen, Norway

ABSTRACT

Globalisation poses a challenge for businesses with linguistically diverse staff, prompting the choice of English as the default corporate language. Although many studies extensively explored the role of corporate language policy in large corporations, employees' perceptions of such policy has not been explored adequately. Fewer studies investigate the trends in such perceptions in social media texts. This study fills this gap; it examines the barriers in multinational companies that have adopted a foreign language and analyses employees' attitudes. The study uses computer-assisted text analysis to investigate changes in employees' perceptions of an English-only policy in Rakuten, a Japanese company. It analyses a corpus of 704 social media posts on OpenWork from 2010 to 2018. The study finds some trends in employees' attitudes such as the inconsistency between expectation and reality in terms of 'Englishnization' and the popularity of the Test of English for International Communication (TOEIC). The findings suggest that Rakuten is not global in the true sense; the value of the TOEIC score is overestimated and the neoliberal employment system affects employees in complex ways. This study contributes to international business language with a bottom-up, employee-centred, and diachronic perspective on language management.

ARTICLE HISTORY

Received 6 January 2021
Accepted 26 May 2021

KEYWORDS

Corporate language policy; social media; computational text analysis; dynamic topic modelling; employee perspective; neoliberalism

Introduction

Today, many companies intend to enter the international market and hence, pay considerable attention to global trends. However, language barriers seem to be presenting challenges for many corporations in their internationalisation efforts. English has been promoted as the lingua franca of international business by political, economic, and technological power over the last three decades (Tsedal Neeley 2017, 2). Corporate language policy improves internal and external communication because it acts as a common language medium for all organisational members and offers easy access to official information channels (Tange and Lauring 2009). Language is reconceptualised as a commodifiable skill under neoliberalism (Urciuoli 2008; De Costa, Park, and Wee 2019). Adopting English as the default corporate language can serve as the foundation for corporate global expansion, since fast and direct communication ensures effective business integration globally.

Corporate language policy involves language usage in business (Sanden 2015). According to Kaplan and Baldauf (1997), language policy and planning activities comprise the linguistic ecosystem. The extant literature has investigated language policy challenges, which include the implications for employees. Relating to Japanese MNCs' efforts to increase internationalisation, their

Japanese employees are encouraged to acquire English skills and a high-stake test has been implemented in the companies (Sekiguchi, Froese, and Iguchi 2016). Piller and Cho (2013) argue that internationalisation is tied to English as an index of global competitiveness. ‘The spread of English’ is not a result of the free linguistic market but of a ‘systematic, organized, and orchestrated policy’ (Piller and Cho 2013, 38). Tsedal B. Neeley’s (2013) qualitative study of a French high-tech company showed that non-native English-speaking employees experienced status loss regardless of their English fluency level. Charles (2007) argues that language choice has implications for management, human resource (HR), and employee satisfaction. Vaara et al. (2005) illustrate the long-term implications of the language policy issue in terms of the reality of career segregation. The choice of a non-native corporate language could frustrate employees as they might feel deprived of their ability to communicate and express themselves adequately, resulting in dissatisfaction with the job and the company as well as lower performance levels (Charles 2007).

Social media has emerged as a significant platform that reflects public opinions (Alomari 2017). Compared to traditional media such as newspapers and television, social media has unique features such as anonymity and the speed of information dispersion (Anstead and O’Loughlin 2015). Thus, the abundance of social media data makes it possible to study employee satisfaction and corporate value, provides new insight into employee perceptions (Huang et al. 2015; Moniz 2016), and becomes an important source for the reproduction of discursive power in society (Törnberg and Törnberg 2016). Glassdoor (<https://www.glassdoor.com/>) and OpenWork (<https://www.vorkers.com/>) are examples of career community websites where current and former employees anonymously review companies and their management (Ning Luo, Zhou, and Shon 2016). However, few studies use such anonymous media platforms as data sources to examine employees’ perceptions of corporate language policy. It also remains unclear how these perceptions change over time. This study employs dynamic topic modelling (DTM) and social media datasets to examine these corporate language policy perceptions.

The study focuses on Rakuten, a Japanese company. Rakuten is one of the world’s largest Internet companies; it is present in 30 countries and regions. The corporation decided to change its corporate language from Japanese to English in 2010. This policy, referred to as ‘Englishnization’, was promoted by the CEO and Chairman Hiroshi Mikitani to dismantle cultural and linguistic barriers (Tsedal Neeley 2011). Rakuten’s challenges with language and cultural diversity are germane to other global organisations (Tsedal Neeley 2017, 3).

In Rakuten, a transgression of the ‘Englishnization’ policy was associated with potential demotion; any employee who failed to comply with the policy, or more specifically, who failed to pass the Test of English for International Communication (TOEIC) within two years, would be demoted. This study investigates topics relating to the ‘Englishnization’ policy and the associated trends to understand employees’ perceptions of corporate language policy through social media. The study aims to (1) better understand corporate language policy by employing a comparative and diachronic view, and (2) provide a valuable method for researchers and language policy makers to make more informed decisions in implementing HR practices that promote language learning.

Employee perceptions of corporate language policy

Corporate language policy is used specifically by a company, a unit within the company, or business organisations (Sanden 2015). The abundant literature on corporate language policies focuses on the role of English as a common corporate language (Piekkari et al. 2005; Yadong Luo and Shenkar 2006; Thomas 2008; Tietze 2010). Language standardisation offers promising avenues for sociolinguistic and international business researchers to engage in collaborative interdisciplinary theoretical and empirical research (Linn, Sanden, and Piekkari 2018). English is used as the lingua franca because of its global use (Nickerson 2005; Ammon 2010; Ferguson 2012). Emphasis on communicative competence in English is presented in the name of enhancing the value of human capital and national competitiveness (Park and Wee 2013). Companies with international aspirations try to

adopt the English language to increase their probability of success (Sanden 2015). However, if English is not the native language, the official decision may be undermined by employees' preference for the local language (Feely 2004). Lønsmann (2017) focussed on employees' reaction to the implementation of a new corporate policy from the top and found three stances among employees: resisting English, embracing English, or a middle position. Many employees must invest heavily in learning the corporate language and ultimately accept in-built inequality or seek to escape the corporate language through internal mobility or leaving the company (Vaara et al. 2005). Lønsmann (2011) concludes that employees may also experience a discrepancy between their definition of language policy and the company's reality. Therefore, managers need to consider the significance of language and communication skills for staff motivation and job satisfaction in a multilingual organisation (Charles 2007).

For a quantitative analysis of CEO leadership and employee confidence, Tsedal Neeley (2017, 155) collected data from 4,000 eligible employees in Rakuten and studied employee perceptions of how the impact of the corporate language policy on their productivity is influenced by leadership. Neeley administered two surveys, at 14 months (Time 1) and 20 months (Time 2) after Englishnization. Tsedal Neeley (2017, 107) then compared the responses in the two time periods to observe the changes in leadership, beliefs, and attitudes from Time 1 to Time 2 and performed a quantitative analysis of CEO leadership and employee confidence. However, the survey, which relied on measurements taken only in two points of time, failed to prove how employees' beliefs and job attitudes toward language management change over time. Here, language means the lingua franca (common language), and management refers to the mandate of 'Englishnization'. Although the study collected employee data in terms of designation, age, gender, and tenure, no comparison analysis was conducted at the level of business units (BUs) and technical units (DUs), gender, and varying tenures. Studies that seek to capture existing employees' perceptions of corporate language policy and rely on surveys suffer from selection bias, and have limitations with regard to drawing inferences and generalising the results (Ning Luo, Zhou, and Shon 2016). A few studies examine employee perceptions of corporate language policy using qualitative methods such as interviews and focus groups (Lønsmann 2017; Sanden and Lønsmann 2018); however, these studies fail to illustrate the temporal trends in employees' opinions regarding corporate language policy.

Context of the study: Rakuten

TOEIC was launched in Japan in 1979 and is now one of the most popular English-language proficiency tests worldwide (Chapman 2006). On March 1, 2010, Hiroshi Mikitani, the CEO of Rakuten, the leading Japanese e-commerce company, introduced an English-only policy known as 'Englishnization' to the 7000 Japanese staff gathered for his weekly address and the 3000 overseas employees watching online. 'From now on', he said, 'everything at Rakuten, from meetings to the canteen menus, would be in English' (Skapinker 2017). His objective was to improve employees' English language proficiency, prepare them to better deal with global competition, and enable them to meet the English language proficiency requirements of the workplace. The CEO advocated the use of the TOEIC listening and reading test to meet this objective, by requiring all employees to achieve a satisfactory score. The implementation of Englishnization was divided into different phases: (1) After the programme was launched in 2010, Rakuten employees had two years to obtain TOEIC scores above 650 to avoid demotion (Tsedal Neeley 2017, 15). (2) The proficiency cutoff rose gradually to 800, also becoming mandatory for all new recruits (Tsedal Neeley 2017, 53). Under neoliberalism, the previous life-long regular employment in Japan has been replaced by more flexible contract employment to reduce labour costs (Kubota 2011). In Rakuten's case, a regular employee should achieve a TOEIC score of above 800.

TOEIC scores are used internationally to make HR decisions related to hiring, promotion, training, and on-the-job performance (Anthony 2003; Stahl et al. 2007; Peltokorpi and Vaara 2014; Oliveri and Tannenbaum 2017). However, if TOEIC scores are used as a criterion for recruitment and

promotion, the stakes are high, meaning the test scores will significantly affect the test-taker's life (Chapman 2006).

TOEIC has been criticised because some of the test preparation strategies are unrelated to English proficiency. The test-takers may focus less on language skills unrelated to the TOEIC (Newfields 2005). The test may misjudge the test-taker's communication competency based on a specific context of language usage (Pilcher and Richards 2017). Then there is the danger that people may take a high score as conferring an automatic guarantee of successful communication in Business English as a Lingua Franca (BELF) (Takino 2016). In language learning, motivation is also an important factor and those who score low may be discouraged (Newfields 2005). Based on Gardner and Lambert's theory, there are two types of motivation: integrative and instrumental motivation (Gardner 1985). In addition to Gardner's motivation model, motivation in self-determination theory is divided into two types: intrinsic and extrinsic motivation (Ryan and Deci 2000). Extrinsic motivation is based on external outcomes such as promotion and punishment, while intrinsic motivation encourages the employee to learn without expectation of reward (Newfields 2005). The most important factor affecting employees' motivation to learn English was the perceived need for self-improvement regarding language ability (Nantakaew 2012). In the context of the corporate language policy in Rakuten, employees' opinions regarding the implementation of the policy and its impact warrant further investigation.

Methodology

Computer-assisted text analysis

Computer-assisted text analysis techniques belong to three main categories: lexical-based (Monroe, Colaresi, and Quinn 2008; Lee and Martin 2015), text classification (Chuang et al. 2012; DiMaggio 2015), and natural language processing (Nelson 2020). As an unsupervised text classification method, topic modelling is used to identify patterns in text, and is as an effective tool for understanding and managing a sequential collection of documents (Blei and Lafferty 2006). It provides an overview of the corpus, complements content analysis (Nelson 2020), reveals hidden themes, and helps model the evolution of topics over time. However, Brookes and McEnery (2019, 18) have questioned 'the thematic coherence of the topic generated using topic modelling methods' and 'the topics intended to exhibit limited thematic consistency'. They demonstrate that 'the usefulness of topic word-lists alone as basis is unlikely to be sufficient for accurately inferring the themes present in the texts grouped into a given topic' (Brookes and McEnery 2019, 18). They conclude that topic modelling methods lack an adequate theoretical basis for defining the topic (Brookes and McEnery 2019). There are three approaches by which topic models consider time: jointly modelling word occurrence and time, pre-discretised analysis, and post-discretised analysis. Post-discretised analysis involves fitting a time-unaware topic model, slicing time-ordering documents into discrete subsets, and reviewing the topic distributions in each time-slice (Wang and McCallum 2006). An example of this type is identifying hot and cold topics based on Latent Dirichlet Allocation (LDA) model analysis (Griffiths and Steyvers 2004).

The joint modelling of word occurrence and time introduces time as an observed continuous variable and relies on the topics as the constant, as topic co-occurrence patterns change over time. The topics over time model is a typical example of joint modelling and analyses documents with timestamps to discover topics. However, this approach fails to identify evolving topics. In the DTM as a pre-discretised analysis, time is discretised into several periods, and time within each LDA is used to analyse documents (Blei and Lafferty 2006). The continuous time dynamic topic model(cDTM) uses the Brownian motion to model latent topics (Wang, Blei, and Heckerman 2012). DTM leverages the knowledge of different documents belonging to different time-slices to map how the words in a topic change over time (Blei and Lafferty 2006). In the DTM, documents are divided into sequential groups, assumed to be exchangeable within a group, and topics of each

slice evolve from the topics in the previous slice (Wang, Blei, and Heckerman 2012). In contrast to DTM, cDTM is more suitable for analysing text with dense time stamps and high sparsity (Wang, Blei, and Heckerman 2012). Topic evolution has been extended and applied to email (Wang and McCallum 2006), news (Wang, Blei, and Heckerman 2012), bulletin board system (BBS) (Lina and Xijin 2014), scientific literature (AlSumait, Barbará, and Domeniconi 2008), movie reviews (Meng, Zhang, and Guo 2012), and so on.

Topic popularity is calculated by occurrences of topics in the documents (Griffiths and Steyvers 2004; Sun and Yin 2017). $\theta_{d,t}$ is the distribution of topics in document d at year t and $\theta_{k,t}$ is the popularity of topic k at year t that shows the proportion of topic k at year t . The change in $\theta_{k,t}$ reflects the trend of the focus of employees' opinions on the topic k . The posterior document-topic distribution and employees' information (i.e. gender, function, positions and tenures with Rakuten) of each post can explicitly quantify the dynamics about the inferred topics that evolve over time at the levels of gender, function and tenures with Rakuten.

Combining DTM with content analysis enables one to consider the context for a better understanding of the dynamics of social representations in a comparative and diachronic view (Chandelier et al. 2018). In this case, this study reveals the evolutionary nature of topics and then analyses the topics extensively. Topic modelling can be employed with other methods that provide elaborate analytical techniques and theoretical depth for interpretation (Törnberg and Törnberg 2016). Because of its features, topic modelling can serve as a bridge in the social sciences, for it sits at the interfaces between case studies and big data, unstructured and structured analysis, and induction and deduction (DiMaggio, Nag, and Blei 2013; Grimmer and Stewart 2013; Mützel 2015; Hannigan et al. 2019). However, few studies have used an analysis of topic evolution to examine changes in employee perceptions over time. Identifying topics at a particular point provides quantitative measures that may help investigate the representation of corporate language policy in social media.

Data and procedure

Employee review websites contain self-initiated and anonymous information that stands to mitigate many of the biases introduced by survey studies (Das Swain et al. 2020). OpenWork (formerly known as Vorkers) is one such review and job website in Japan. It contains objective information such as pay, benefits and the free-form text that encapsulates organisational culture. There are about 9,800,000 reviews and evaluated scores of 36,000 companies posted by current and former employees at the time of writing. The website has 2,900,000 registered users and about 3000 new registrations every day.

The analysis proceeded in three steps (see Figure 1). In Step 1, the contents of the OpenWork website between 2008 and 2018 were crawled by using a customised web crawler. The website comprised 9324 Japanese posts (1,337,652 words). Since the study only relied on data written in Japanese, it does not account for employees who speak languages other than Japanese at the HQ or employees located in Rakuten offices outside Japan. Furthermore, several stop words from the analysis were excluded by using a comprehensive collection for Japanese. To build the corpus for the analysis, the study extracted all posts containing terms 英語 'English', *toeic* and 公用 'official', which are chosen by their relevance and frequency in the database. A corpus of 704 posts (about 66,141 words) over time (from 2010 to 2018) was obtained. In this study, five topics were selected given subjective analysis of iterative results.

A significant input for DTM to work is the time slice input that contains the number of documents in each time slice (Blei and Lafferty 2006). In this case, the first year (2010) had 7 articles, the second (2011) 25, the third (2012) 42, the fourth (2013) 37, the fifth (2014) 64, the sixth (2015) 96, the seventh (2016) 141, the eighth (2017) 135 and the last year (2018) had 157 articles. In Step 2, the examination of lists of topic words formed the fundamental part of the analysis. In Step 3, a computationally guided in-depth reading was also applied to the top five texts in each time slice that

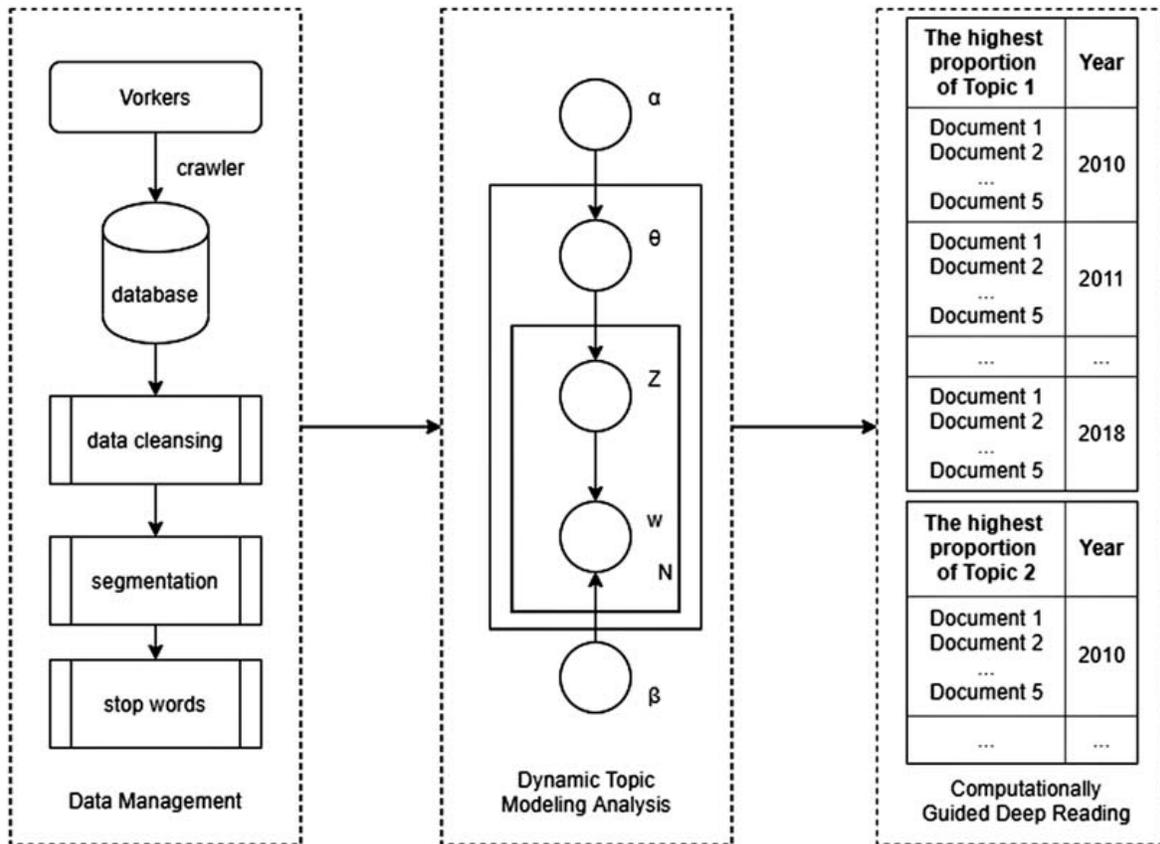


Figure 1. Data and procedure.

shows the highest proportion of that topic as evaluated by the topic modelling algorithm. This analysis was a blended approach that used topic modelling as reference to a close reading of sub-sample documents. This framework is similar to what Gkotsis et al. (2017) applied in online health-related texts. Through this analysis, the results obtained are presented in the next section.

Findings

DTM discovered five interpretable topics (Table 1). Figure 2 displays the trends in each topic, calculated by the topic popularity of all topics each year. In the following paragraphs, the temporal topics are presented separately, and changes in these topics at different levels are analysed. Figures 3–7 show wordclouds of five topics. The size of each word is based on the proportion of its probability and changes over time.

Inconsistency between expectation of the employees and reality for Englishnization

For topic one (see Figure 3), the top 20 most relevant words are {入社 'joining company', 英語 'English', 理由 'reason', べき 'should do', おく 'suffix', 性 'suffix', 事 'thing', 認識 'perceive', 妥当 'valid', 決め 'conclusive factor', たい 'would like to', 化 '-ification', グローバル 'global', 公用 'official', いい 'good', 語 'language', 思い 'think', 思っ 'think', 日本語 'Japanese', 海外 'overseas'}. The top relevant words *joining company* and *English* remain highly stable throughout the period (from 2010 to 2018). Additionally, the increasing usage of *good* and the decline of the use of *global* are noted. Topic one focuses on 'Englishnization' as a motivation for joining the company. The focus contains the content of the gap between before and after joining the company, articulating that there is a considerable discrepancy between the motivation for joining the company and the validity of the

Table 1. Five topics.

Num	Topic	Keywords (time = 2010)
1	Inconsistency between expectation of the employees and reality for Englishnization	入社 'joining company', 英語 'English', 理由 'reason', べき 'should do', おく 'suffix', 性 'suffix', 事 'thing', 認識 'perceive', 妥当 'valid', 決め 'conclusive factor', たい 'would like to', 化 '-ification', グローバル 'global', 公用 'official', いい 'good', 語 'language', 思い 'think', 思っ 'think', 日本語 'Japanese', 海外 'overseas'
2	The interplay between Englishnization and organisational leadership	英語 'English', 化 '-ification', 公用 'official', 語 'language', 社内 'office', 社員 'employee', 的 'suffix', 勉強 'study', 時間 'time', 多い 'many', 海外 'overseas', 事業 'business', 力 'skill', 仕事 'work', 感じる 'feel', 学習 'learn', 等 'and so on', 系 'system', いく 'will be ...ing', 会 'association'
3	The interplay between Englishnization and organisational culture	英語 'English', 化 '-ification', 人 'people', 外国 'foreign', グローバル 'global', 社内 'office', 企業 'company', 多い 'many', 公用 'official', 多く 'much', 語 'language', 文化 'culture', 方 'the one's side', 海外 'overseas', 環境 'environment', 機会 'opportunity', 増え 'increase', 仕事 'work', 思い 'think', 会社 'company'
4	Supremacy of TOEIC	toeic, 英語 'English', 点 'point', 給与 'salary', 制度 'system', 点数 'point', スコア 'score', 特徴 'features', 入社 'joining company', 年収 'annual income', 以上 'the above', 必要 'necessarily', 必須 'required', 評価 'assessment', 万 'ten thousand', 昇格 'promotion', なけれ 'if not', 場合 'case', 事例 'example', 年 'year'
5	Power of English skills	英語 'English', 力 'skill', 仕事 'work', スキル 'skill', 会議 'meeting', 化 '-ification', 非常 'extremely', 身 'the person', 出来 'result', 思う 'think', 業務 'business', 人 'people', 使う 'use', 語学 'language study', 環境 'environment', キャリア 'career', 今 'now', 給料 'salary', アップ 'up', 部署 'position'

reason. Five posts were each extracted from 2010 to 2018, which exhibited the highest proportion of topic one. Interestingly, the top texts exhibited similarities over time. Employees discussed the inconsistency, including their reason for joining the company: Rakuten is a global company, and English is the official language. However, they found that the frequency of using English appeared quite low after joining the company, especially in domestic BUs. The discussion here focuses on how Englishnization falls below their expectations starting from 2014 to 2018, which is well illustrated in this post:

Extract 1

入社を決めた理由: グローバルHQとしてビジネスを展開と英語が公用語が理由です。入社理由の妥当性と認識しておくべき事: 英語が使えるかどうかの方が本当に配属部署次第です。(No.583、コーポレート部門、在籍3~5年、男性、2018)

Reasons for deciding to join the company: developing a business as the global HQ and English is the official language 'Reason for joining the company' and "Things to be aware of: It is really up to the assigned department to determine whether English can be used. (Translated from Japanese, no. 583, 2018, male, BU, employee, 3-5 years)

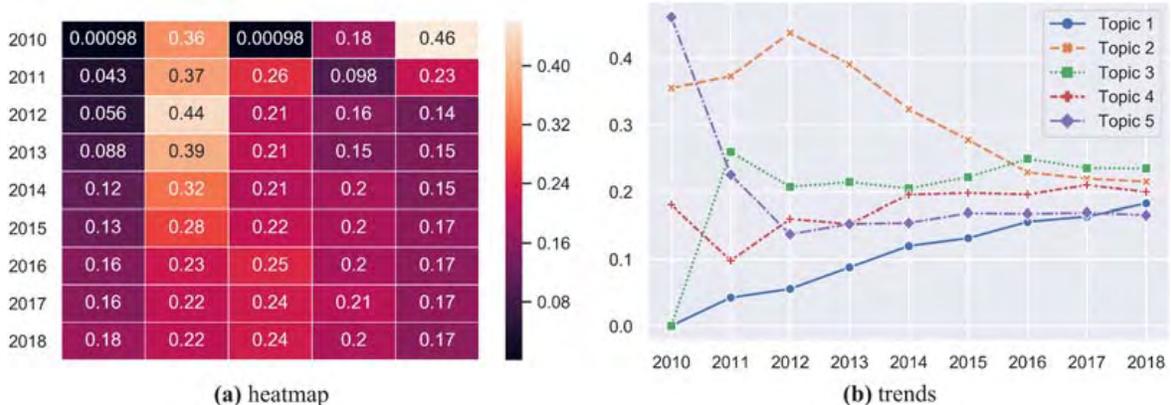


Figure 2. Topic popularity over time. (a) heatmap and (b) trends.

show that Englishnization attracts more female employees to return to work after childbirth. One of the reasons is an in-house daycare center that provides English education. Topic 3 peaks in 2011 (Figure 2), driven by discussions regarding how corporate culture changes over time.



Figure 5. Wordcloud of Topic 3 from 2010 to 2018.

Supremacy of TOEIC

This topic focuses on the supremacy of TOEIC – a requirement for promotion – as a hiring standard, an evaluation system, and a criterion for annual income (see Figure 6). The topic is central and remains internally consistent over time in Figure 2. The focus in the first years (2010–2015) lies foremost on salary deduction and the inability to receive a promotion due to TOEIC: ‘those who do not have a TOEIC score cannot raise their salary’ in 2011, ‘if the TOEIC score is below the requirement, there is also a salary reduction (10%)’ in 2014, ‘if a certain score is not achieved, you cannot get promoted’ in 2011 and ‘no matter how good your business performance is, you cannot be promoted unless you reach the TOEIC score’ in 2012 and 2015, ‘there is a trend that many Japanese who do not have web experience are recruited just because they have a TOEIC score’ in 2015 and 2017. While some opinions on this topic seem to be explicitly related to TOEIC as a hiring standard, many employees also discussed consequences of adopting this standard in 2015. Experienced employees are likely to be hired as contract workers even if their TOEIC scores are less than 800.

Moreover, HR is concerned with the TOEIC score of employees. Some employees who obtained a high score still struggle to converse in English or construct an email. There is a trend of recruiting Japanese employees based on the TOEIC score even though inexperienced in information technology. Last but not least, since a TOEIC score of over 800 is used for hiring, it is difficult to recruit qualified Japanese candidates. Some opinions in 2013 and 2016 focus on the reason for leaving the company; they are against TOEIC and Englishnization. Parallels are frequently drawn between the non-necessity for business, the obligation of learning English, a tool to switch jobs to foreign competitors, and the inability to get a promotion. One employee said the following:



Figure 7. Wordcloud of Topic 5 from 2010 to 2018.

Extract 4

英語化に関しては、大きな会議はほとんど英語で、日本人同士の会議やメールは日本語が用いられている。(no.479、2017、企画、在籍5~10年、女性)

Regarding Englishnization, large meetings are mostly in English, but meetings or emails with Japanese colleagues are in Japanese. (Translated from Japanese, no. 479, 2017, female, BU, employee, 5–10 years)

Opinions in 2018 focus on how English is associated with career development that impacts salary and promotion, as highlighted in the next post:

Extract 5

ただし、英語力が最重要なので実力で必ずしも評価されることはない。昇進には英語が必須となる。(no.673、2018、トラベル事業、営業、在籍5~10年、女性)

However, English skills are the most important, so employees are not always evaluated based on their ability. English is required for promotion. (Translated from Japanese, no. 673, 2018, female, BU, employee, 5–10 yrs)

As shown in Figure 2, this topic peaks in 2010, driven by opinions particular on language policy shocks on the Japanese employees. however, it seems to decline dramatically from 2011 to 2012, and it is stable from around 2013 to 2018.

Topic popularity of function/gender/tenures with Rakuten over time

Although Figure 2 shows each topic's popularity over time, it is still unknown whether the trend differs at the function, gender, and tenure levels. The study applied the same procedure and analysis to investigate the temporal topic variation for each level. The temporal trend for males and females is presented in the data set in Figure 8. This plot is interesting as it reveals the temporal variation of popularity among different genders. For example, the results indicate that females have a particular focus on Topic 2: The interplay between Englishnization and leadership and Topic 3: The interplay between Englishnization and organisational culture. From 2010 to 2019, the focus of females on Topic 2 exhibits a declining trend that is more apparent from 2013. In contrast, there is an increased

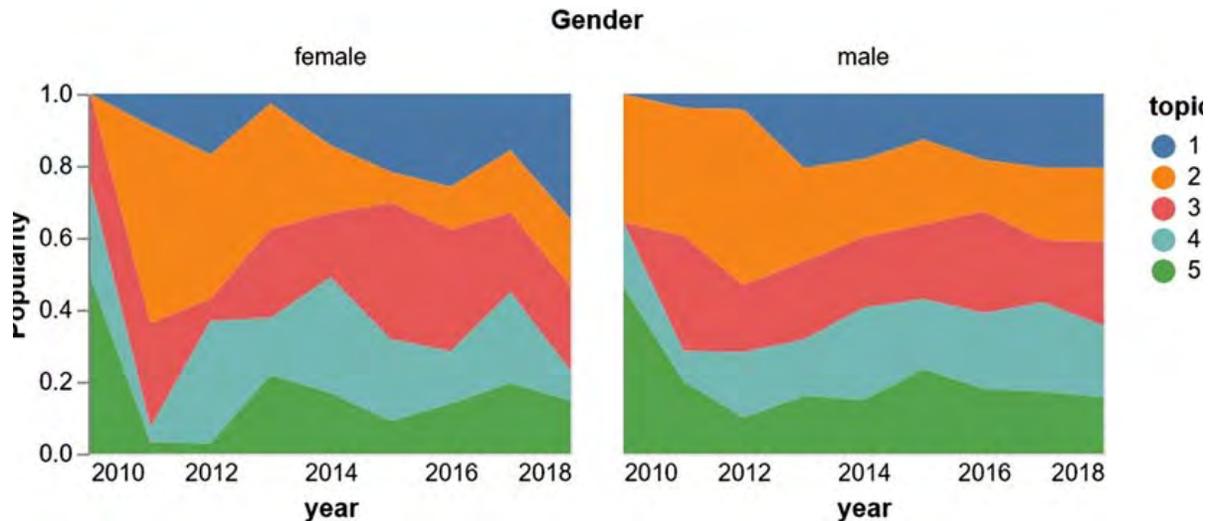


Figure 8. Topic popularity over time for gender.

particular focus within Topic 3. The focus of males has been on Topic 2, which might be somewhat declining in the corpus overall. These examples clearly show that Topic 2 has particular popularity or is more important for both female and male employees.

Subsequently, the exercises are performed at the function level (see Figure 9). For BUs, Topic 2 on the interplay between Englishnization and leadership covers a much larger proportion than the other four topics. However, the focus of DUs has been on Topic 4 on the supremacy of TOEIC, which peaks in 2013, driven by opinions particularly on leaving the company due to TOEIC.

Extract 6

業務に必要なない英語習得を義務付けられた時退職。(no.84、2013、管理部門、在籍3~5年、男性)
Left the company when required to learn English that is not relevant to work. (Translated from Japanese, no. 84, 2013, male, BU, employee, 3-5 years)

Extract 7

TOEIC最低800点以上、目標900点以上今後も海外展開を加速する会社。ついていけないと感じたら、転職をするべき。(no.110、2013、管理部門、在籍3~5年、男性)
The company will accelerate overseas expansion in the future with a minimum TOEIC score of 800 and a target of 900 or more. If you feel you can't keep up, you should change jobs. (Translated from Japanese, no. 110, 2013, male, DU, employee, 10-15 years)

Extract 8

英語化はやめた方が良い人材が集まると思う。国内営業に外人は不要。(no.88、2013、営業、在籍3~5年、男性)
I think that it would be better for people to stop speaking English. No foreigners are required for domestic sales. (Translated from Japanese, no. 88, 2013, male, BU, employee, 3-5 years)

The next analysis of topic popularity over time is performed at the tenure level at Rakuten, which is shown in Figure 10. Essentially, results indicate that topic popularity of tenures at Rakuten during 0-3 years, 3-5 years, and 5-10 years are consistent over time. From 2010, there seems to be an increased particular focus on Topic 1 in these three tenures – a tendency that is even more apparent after 2013. Tenures of 0-3 years, 3-5 years and 5-10 years have been centred on Topic 2, but they seem to decline from 2013 to 2016, and recur from 2017 to 2018. Topic 4 stands out at 10-15 years, probably because these individuals require TOEIC to get a promotion. The category 20+ years only appears around 2012, driven by a focus on Topic 2. This is rather interesting since it seems that after 2013, there is no category of 20+ years, probably because these topics are not relevant to them.

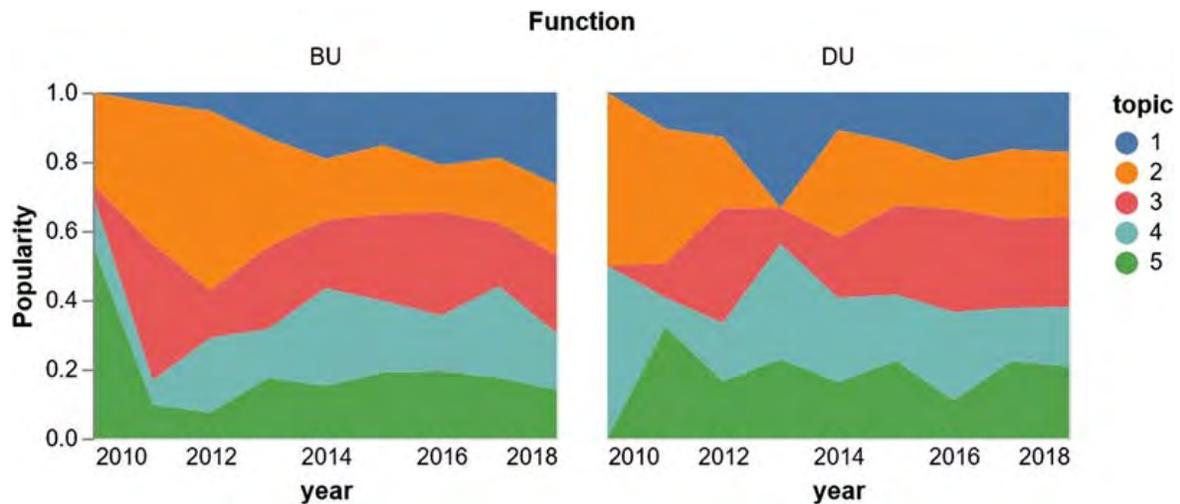


Figure 9. Topic popularity over time for function.

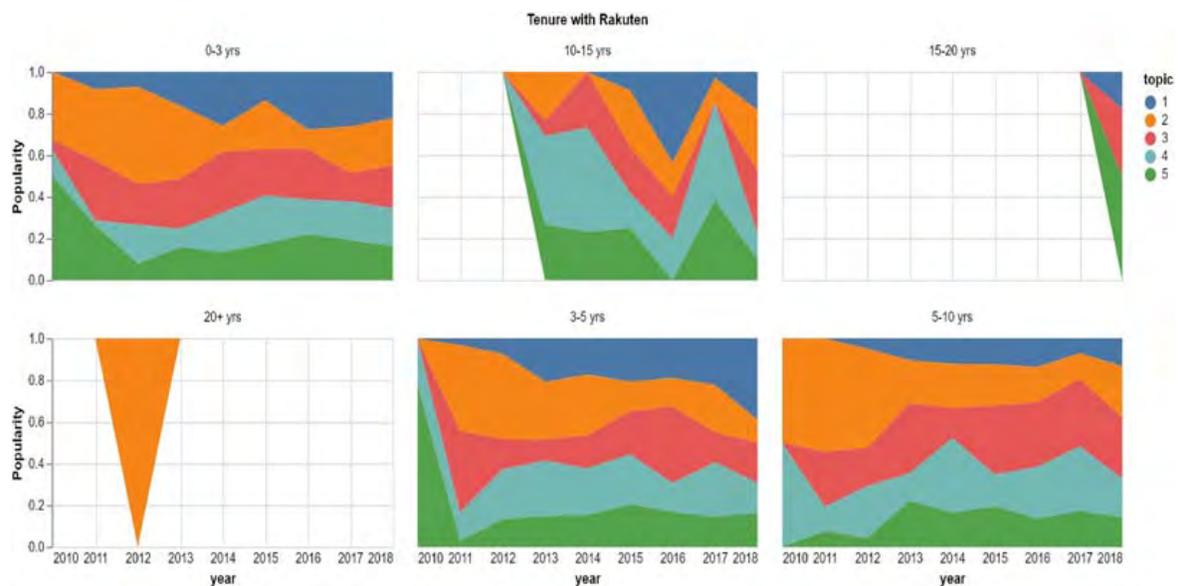


Figure 10. Topic popularity over time for tenures.

Discussion

Not a real global company: less intrinsically motivated employees

Based on the above findings, the online texts about 'Englishnization' are often associated with a discrepancy between employees' perception of corporate language policy before and after joining the company, and dissatisfaction with the Englishnization's status quo. From the analysis, the first topic seems to attract growing attention. In Rakuten's case, many employees joined the company because they wanted to improve their English and looked forward to embracing it as a corporate policy. The need for English competence may lead employees to seek challenges optimal for their capacities and to maintain English skills persistently. However, the text shows that the need for autonomy, or being the perceived origin or source of one's behaviours (Deci and Ryan 2002, 8) is not fulfilled when employees perceive that they do not have the opportunity to make personal choices (i.e. they have to use Japanese). Employees highlight the negative elements of Englishnization and

their attitudes toward corporate language policy in the above contents, which change from cynical at the beginning to dissatisfied at present. The data showed that Englishnization makes the company increasingly attractive to Japanese employees. The company has also recognised the Englishnization policy as a brand which serves as a touchpoint to attract employees. This view is confirmed by the statement of the CEO (Mikitani 2014):

Englishnization has already helped Rakuten. It has made us a more attractive workplace to the world's best and brightest employees. (Englishnization: Not Just for Rakuten Anymore, 2014)

However, employees have realised that the English-only policy is just a publicity stunt. Texts on OpenWork not only reflect employees' emotions and dissatisfaction toward the corporate language policy but also reveal that employees' motivation in learning English declines because there is no demand in their departments. Negative and aggressive language, such as *dictatorial leader* and *religion*, is utilised in the CEO leadership description. This analysis raises questions regarding one argument in the previous research on whether groups or departments in Rakuten would achieve the intended outcomes associated with Englishnization (Tsedal Neeley 2017, 171).

Furthermore, although English has been the official company language since 2010, trends in inconsistency between expectation of the employees and reality for Englishnization and the power of English skills reveal that Japanese skills are still more valuable than English skills in many Japanese departments. Global talents are mostly placed at DUs which results in mixed corporate culture (Japanese and English). These results align with existing research on the de facto vs. de jure language policies (Feely 2004; Tange and Lauring 2009; Lønsmann 2011). As previously mentioned, Lønsmann (2011) concludes that while English was used primarily with international non-Danish speakers, many employees experience a discrepancy between their definition of the language policy and the reality at the company where Danish was the default language. A deeper analysis of the results in this study leads to a similar conclusion in the Japanese business context. This study provides the company with a clear picture of the shortcomings that de-motivate employees to learn and use English in the workplace, encouraging the company to monitor and promote employees' motivation. It also helps the HR department to design and plan the English training programme to improve corporate language policy in the long term.

It is all about TOEIC: the score is everything

This study has identified a prevailing content of the negative impact of an over-reliance on the TOEIC scores by analysing these topics and their associated documents. The score has two major functions: to select employees for recruitment, punishment, and promotion; and to exert the intended washback effect on employees' learning. Although TOEIC satisfies a commercial environment for a practical, objective, and fair test of receptive language skills, 'Supremacy of TOEIC' is used as a discursive strategy to criticise the corporate language policy. The score is everything. Meanwhile, Englishnization is represented as 'TOEICnization' in the context. Moreover, the representation of Englishnization in the latest corporate report focuses on the TOEIC score, showing that employees' average TOEIC score improved from 526.2 in 2010 to 839.7 in 2019 (Rakuten Inc 2019, 38). The trend in communication and acquiring global talent is also reflected in the following statement:

With the commencement of Englishnization project in 2010 – Rakuten's initiative to make English the official company language – communication between employees regardless of nationality or place of birth has become much smoother and easier. It has also exponentially expanded our pool of potential hires from the relatively small number of people who speak Japanese to English speakers worldwide. Since the launch of the project, proportion of non-Japanese employees has grown significantly; we now have over 70 nationalities represented, making up more than 20% of our total head count at Rakuten, Inc. (Rakuten Corporate Report, 2019)

The findings match the representation and discursive construction of Englishnization in the corporate report, where the TOEIC score is used to prove the policy's success. This misuse of the

test results pressurises managers and employees to strive for high TOEIC scores, thus impeding the intended changes or washback effect. It seems like very few Rakuten employees take the English-only policy seriously, and most of them only focus on passing the test.

Moreover, the study results have implications for viewing language as a social practice rather than a neutral, mechanical instrument. When crucial decisions are made based on test results, and when one's interests are seriously affected, it is not surprising that employees learn for the test rather than for using English in the workplace, which is the original intention behind the corporate language policy. Furthermore, the long-term goals of learning English are overlooked. The excessive use of the multiple-choice format in the test fails to simulate real-world language use for communication. This study confirms that current trends of over-emphasising TOEIC scores may mislead BELF into thinking that the TOEIC skills are the main English skills they have to develop (Takino 2016). The results also corroborate established research on the significant impact of the TOEIC test as a high-stakes test on the lives of the test takers and other stakeholders (Chapman 2006).

The neoliberal employment system

The power of English and the neoliberal employment system affect employees in complex ways, with gender, units and types of employment playing significant roles. 'Markets of English' consist of linguistics, educational and employment markets, and they are too many and overlapping (Park and Wee 2013; Price 2014).

For non-regular employment, learning English helps pursue a specific goal: to become regular employment, more secure and more remunerated in a long run. This is compatible with language as a commodifiable skill in the neoliberal employment system (Urciuoli 2008; De Costa, Park, and Wee 2019). Contrary to non-regular employment, learning English in regular employment was directly related to demotion. Exceeding the score of 800 gives regular employees an economic return. However, both regular and non-regular employees working for domestic branches have very few opportunities to use English at work. The low frequency of English use at BUs also indicates that language skills are not valued. This has also been discussed in Kubota's (2011) study where 'the language tests such as TOEIC work as a convenient tool to measure the level of effort rather than proficiency itself'.

Nonetheless, Englishnization attracts more female employees to return to work after childbirth, and improves gender equality in the neoliberal socio-economic system. This corresponds with linguistic instrumentalism and contradicts the finding that women with higher English proficiency may not gain significant economic return (Kubota 2011). Overall, linguistic instrumentalism works to keep Japanese women in regular employment, and contains hierarchies with regular employment perched at the top as winners and non-regular employment at the bottom as losers under the neoliberal discourse of competition.

In general, this exploratory study shows that employees' attitudes toward corporate language policy should be examined. While previous studies have shown how employees react to the introduction of a new policy of 'English only' (Vaara et al. 2005; Lønsmann 2017), this study has focussed on employees' perceptions of an English-only policy as a mandate over time. The study shows the clear implications for a society where discrete item testing (TOEIC) is an easy 'solution' to language issues in the Japanese business context.

Conclusion

This study has demonstrated that DTM allows the user to take the context into account, to understand better the use of English in Japanese businesses, that is, the use of English as a corporate language. This approach provides a bottom-up and employee-centred perspective regarding the advantages, disadvantages, and consequences of using English as a corporate language. It also reveals the evolutionary nature of corporate language policy topics by exploring and categorising large amounts of text.

It is challenging to generalise this study's results because all posts are in Japanese and from the same company. However, the study helps understand trends in employees' attitudes toward corporate language policy as well as other aspects. As an unsupervised learning algorithm, the DTM used in this study inferred topics based purely on the statistical structure of the term data. The study used statistical analysis to investigate the context of each topic and the temporal variation in topics. It also provides insights into how TOEIC may be misused when dealing with language issues in the Japanese business context. Future directions include integrating other data sources to quantify temporal topic variation and predict employee attitude trends.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ziyuan Zhang  <http://orcid.org/0000-0002-1877-6937>

References

- Alomari, Amirah. 2017. "Distance Impact on Quality of Video Streaming Services in Cloud Environments." *International Journal of Space-Based and Situated Computing* 7 (3): 119–128.
- AlSumait, Loulwah, Daniel Barbará, and Carlotta Domeniconi. 2008. "On-Line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking." In *2008 Eighth IEEE International Conference on Data Mining*, 3–12. IEEE.
- Ammon, Ulrich. 2010. "World Languages: Trends and Futures." In *The Handbook of Language and Globalization*, 101–122. Hoboken: John Wiley & Sons.
- Anstead, Nick, and Ben O'Loughlin. 2015. "Social Media Analysis and Public Opinion: The 2010 UK General Election." *Journal of Computer-Mediated Communication* 20 (2): 204–220.
- Anthony, Laurence. 2003. "Integrating the TOEIC into the University English Curriculum." Proceedings of Jacet 42th Annual Convention, Sendai, 149–149.
- Blei, David M., and John D. Lafferty. 2006. "Dynamic Topic Models." In *Proceedings of the 23rd International Conference on Machine Learning*, 113–120. ACM.
- Brookes, Gavin, and Tony McEnery. 2019. "The Utility of Topic Modelling for Discourse Studies: A Critical Evaluation." *Discourse Studies* 21 (1): 3–21.
- Chandelier, Marie, Agnes Steuckardt, Raphael Mathevet, Sascha Diwersy, and Olivier Gimenez. 2018. "Content Analysis of Newspaper Coverage of Wolf Recolonization in France Using Structural Topic Modeling." *Biological Conservation* 220: 254–261.
- Chapman, Mark. 2006. "An Over-Reliance on Discrete Item Testing in the Japanese Business Context." International Conference on English Instruction and Assessment, Taiwan.
- Charles, Mirjaliisa. 2007. "Language Matters in Global Communication: Article Based on Ora Lecture, October 2006." *The Journal of Business Communication* (1973) 44 (3): 260–282.
- Chuang, Jason, Daniel Ramage, Christopher Manning, and Jeffrey Heer. 2012. "Interpretation and Trust: Designing Model-Driven Visualizations for Text Analysis." Proceedings of the Sigchi Conference on Human Factors in Computing Systems, Austin, 443–452.
- Das Swain, Vedant, Koustuv Saha, Manikanta D. Reddy, Hemang Rajvanshy, Gregory D. Abowd, and Munmun De Choudhury. 2020. "Modeling Organizational Culture with Workplace Experiences Shared on Glassdoor." Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems, Honolulu, 1–15.
- Deci, Edward L., and Richard M. Ryan. 2002. "Overview of Self-Determination Theory: An Organismic Dialectical Perspective." In *Handbook of Self-Determination Research*, 3–33. Rochester: University of Rochester Press.
- De Costa, Peter I., Joseph Park, and Lionel Wee. 2019. "Linguistic Entrepreneurship as Affective Regime: Organizations, Audit Culture, and Second/foreign Language Education Policy." *Language Policy* 18 (3): 387–406.
- DiMaggio, Paul. 2015. "Adapting Computational Text Analysis to Social Science (and Vice Versa)." *Big Data & Society* 2 (2). doi:10.1177/2053951715602908.
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities Between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of Us Government Arts Funding." *Poetics* 41 (6): 570–606.

- Feely, Alan J. 2004. "The Impact of the Language Barrier on the Management of Multinational Companies." Unpublished PhD dissertation, Aston University, Birmingham.
- Ferguson, Gibson. 2012. "English in Language Policy and Management." In *The Cambridge Handbook of Language Policy*, 475–498. Cambridge: Cambridge University Press.
- Gardner, Robert C. 1985. *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London: Arnold.
- Gkotsis, George, Anika Oellrich, Sumithra Velupillai, Maria Liakata, Tim J. P. Hubbard, Richard J. B. Dobson, and Rina Dutta. 2017. "Characterisation of Mental Health Conditions in Social Media Using Informed Deep Learning." *Scientific Reports* 7: Article ID 45141.
- Griffiths, Thomas L., and Mark Steyvers. 2004. "Finding Scientific Topics." *Proceedings of the National Academy of Sciences* 101 (suppl 1): 5228–5235.
- Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–297.
- Hannigan, Timothy R, Richard F. J. Haans, Keyvan Vakili, Hovig Tchalian, Vern L. Glaser, Milo Shaoqing Wang, Sarah Kaplan, and P. Devereaux Jennings. 2019. "Topic Modeling in Management Research: Rendering New Theory From Textual Data." *Academy of Management Annals* 13 (2): 586–632.
- Huang, Minjie, Pingshu Li, Felix Meschke, and James P. Guthrie. 2015. "Family Firms, Employee Satisfaction, and Corporate Performance." *Journal of Corporate Finance* 34: 108–127.
- Kaplan, Robert B., and Richard B. Baldauf. 1997. *Language Planning From Practice to Theory*. Vol. 108. Bristol: Multilingual Matters.
- Kubota, Ryuko. 2011. "Questioning Linguistic Instrumentalism: English, Neoliberalism, and Language Tests in Japan." *Linguistics and Education* 22 (3): 248–260.
- Lee, Monica, and John Levi Martin. 2015. "Coding, Counting and Cultural Cartography." *American Journal of Cultural Sociology* 3 (1): 1–33.
- Lina, Cao, and Tang Xijin. 2014. "Trends of Bbs Topics Based on Dynamic Topic Model." *Journal of Management Sciences in China* 17 (11): 109–121.
- Linn, Andrew, Guro Refsum Sanden, and Rebecca Piekkari. 2018. "Language Standardization in Sociolinguistics and International Business: Theory and Practice Across the Table." In *English in Business and Commerce: Interactions and Policies*, 19–45. Berlin: De Gruyter Mouton.
- Lønsmann, Dorte. 2011. "English as a Corporate Language." In *Language Choice and Language Ideologies in An International Company in Denmark*.
- Lønsmann, Dorte. 2017. "Embrace it Or Resist it? Employees' Reception of Corporate Language Policies." *International Journal of Cross Cultural Management* 17 (1): 101–123.
- Luo, Yadong, and Oded Shenkar. 2006. "The Multinational Corporation as a Multilingual Community: Language and Organization in a Global Context." *Journal of International Business Studies* 37 (3): 321–339.
- Luo, Ning, Yilu Zhou, and John Shon. 2016. "Employee Satisfaction and Corporate Performance: Mining Employee Reviews on Glassdoor.Com."
- Meng, Cong, Mian Zhang, and Wenqiong Guo. 2012. "Evolution of Movie Topics Over Time." Accessed April 1 2014. <http://cs229.stanford.edu/proj2012/MengZhangGuo-EvolutionofMovieTopicsOverTime.pdf>.
- Mikitani, Mickey. 2014. "Englishnization: Not Just for Rakuten Anymore." <https://www.linkedin.com/pulse/20140304144111-52782505-englishnization-not-just-for-rakuten-anymore/>.
- Moniz, Andy. 2016. *Inferring Employees' Social Media Perceptions of Goal-Setting Corporate Cultures and the Link to Firm Value*. Unpublished Working Paper.
- Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. 2008. "Fightin' words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16 (4): 372–403.
- Mützel, Sophie. 2015. "Facing Big Data: Making Sociology Relevant." *Big Data & Society* 2 (2). doi:10.1177/2053951715599179.
- Nantakaew, Atchara. 2012. *A Survey Study on Motivation in Attending English Classes Among Employees of An Engineering Services Company*. Bangkok: Language Institute, Thammasat University.
- Neeley, Tsedal. 2011. *Language and Globalization: 'englishnization' at rakuten*. Harvard Business School organizational behavior unit case, 412-002.
- Neeley, Tsedal B. 2013. "Language Matters: Status Loss and Achieved Status Distinctions in Global Organizations." *Organization Science* 24 (2): 476–497.
- Neeley, Tsedal. 2017. *The Language of Global Success: How a Common Tongue Transforms Multinational Organizations*. Princeton: Princeton University Press.
- Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49 (1): 3–42.
- Newfields, Tim. 2005. "TOEIC Washback Effects on Teachers: A Pilot Study At One University Faculty." *The Economic Review of Toyo University* 31 (1): 83–105.
- Nickerson, Catherine. 2005. "English As a Lingua Franca in International Business Contexts." *English for Specific Purposes* 24 (4): 367–380.

- Oliveri, Maria Elena, and Richard J. Tannenbaum. 2017. "Insights into Using TOEIC® Test Scores to Inform Human Resource Management Decisions." *ETS Research Report Series 2017* (1): 1–7.
- Park, Joseph Sung-Yul, and Lionel Wee. 2013. *Markets of English: Linguistic Capital and Language Policy in a Globalizing World*. Abingdon: Routledge.
- Peltokorpi, Vesa, and Eero Vaara. 2014. "Knowledge Transfer in Multinational Corporations: Productive and Counterproductive Effects of Language-sensitive Recruitment." *Journal of International Business Studies* 45 (5): 600–622.
- Piekkari, Rebecca, Eero Vaara, Janne Tienari, and Risto Sääntti. 2005. "Integration Or Disintegration? Human Resource Implications of a Common Corporate Language Decision in a Cross-Border Merger." *The International Journal of Human Resource Management* 16 (3): 330–344.
- Pilcher, Nick, and Kendall Richards. 2017. "Challenging the Power Invested in the International English Language Testing System (IELTS): Why Determining 'english' preparedness Needs to Be Undertaken Within the Subject Context." *Power and Education* 9 (1): 3–17.
- Piller, Ingrid, and Jinhyun Cho. 2013. "Neoliberalism as Language Policy." *Language in Society* 42 (1): 23–44.
- Price, Gareth. 2014. "English for All? Neoliberalism, Globalization, and Language Policy in Taiwan." *Language in Society* 43 (5): 567–589.
- Rakuten Inc. 2019. "Corporate Report 2018." [http://global.rakuten.com/corp/investors/assets/doc/documents\\$EP](http://global.rakuten.com/corp/investors/assets/doc/documents$EP).
- Ryan, Richard M., and Edward L. Deci. 2000. "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions." *Contemporary Educational Psychology* 25 (1): 54–67.
- Sanden, Guro Refsum. 2015. "Language Strategies in Multinational Corporations." *A Cross-Sector Study of Financial Service Companies and Manufacturing Companies*.
- Sanden, Guro Refsum, and Dorte Lønsmann. 2018. "Discretionary Power on the Front Line: A Bottom-Up Perspective on Corporate Language Management." *European Journal of International Management* 12 (1/2): 111–137.
- Skiguchi, Tomoki, Fabian Jintae Froese, and Chie Iguchi. 2016. "International Human Resource Management of Japanese Multinational Corporations: Challenges and Future Directions." *Asian Business & Management* 15 (2): 83–109.
- Skapinker, Michael. 2017. "The Japanese Company That Ordered Staff to Speak English." <https://www.ft.com/content/2fdd6626-ba3b-11e7-8c12-5661783e5589> .
- Stahl, G. K., I. Björkman, E. Farndale, S. S. Morris, J. Paauwe, P. Stiles, J. Trevor, and P. Wright. 2007. *Global Talent Management: How Leading Multinationals Build and Sustain Their Talent Pipeline*. Insead Faculty and Research Working Papers. Fontainebleau: INSEAD.
- Sun, Lijun, and Yafeng Yin. 2017. "Discovering Themes and Trends in Transportation Research Using Topic Modeling." *Transportation Research Part C: Emerging Technologies* 77: 49–66.
- Takino, Miyuki. 2016. "Negotiating the Challenges of Using English in Business Communication: Listening Narratives of Japanese Belf Users." University of Southampton dissertation.
- Tange, Hanne, and Jakob Lauring. 2009. "Language Management and Social Interaction Within the Multilingual Workplace." *Journal of Communication Management* 13 (3): 218–232.
- Thomas, Chris Allen. 2008. "Bridging the Gap Between Theory and Practice: Language Policy in Multilingual Organisations." *Language Awareness* 17 (4): 307–325.
- Tietze, Susanne. 2010. "International Managers As Translators." *European Journal of International Management* 4 (1–2): 184–199.
- Törnberg, Anton, and Petter Törnberg. 2016. "Combining Cda and Topic Modeling: Analyzing Discursive Connections Between Islamophobia and Anti-feminism on An Online Forum." *Discourse & Society* 27 (4): 401–422.
- Urciuoli, Bonnie. 2008. "Skills and Selves in the New Workplace." *American Ethnologist* 35 (2): 211–228.
- Vaara, Eero, Janne Tienari, Rebecca Piekkari, and Risto Sääntti. 2005. "Language and the Circuits of Power in a Merging Multinational Corporation." *Journal of Management Studies* 42 (3): 595–623.
- Wang, Chong, David Blei, and David Heckerman. 2012. "Continuous Time Dynamic Topic Models." arXiv-1206.
- Wang, Xuerui, and Andrew McCallum. 2006. "Topics Over Time: A Non-Markov Continuous-Time Model of Topical Trends." In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 424–433. ACM.

9 Article 3

Zhang, Ziyuan. 2021c. “Analyzing cultural expatriates’ attitudes toward ‘Englishnization’ using dynamic topic modeling” *Journal of Computer-Assisted Linguistic Research*. 1-26.

Analyzing cultural expatriates' attitudes toward "Englishnization" using dynamic topic modeling

Ziyuan Zhang

Norwegian School of Economics, Norway

Corresponding author: ziyuan.zhang@nhh.no

Received: 16 July 2021 / Accepted: 14 October 2021 / Published: 22 December 2021

Abstract

Several Japanese multinational corporations (MNCs) have recently adopted an English-only policy known as "Englishnization". This study examines the impact of this policy using computer-assisted text analysis to investigate changes in cultural expatriates' perceptions of Japanese work practices and values over time. Cultural expatriates are a significant but underexplored outcome of globalization. Despite the recent proliferation of studies on the internationalization of Japanese MNCs, few studies have focused on cultural expatriates' perceptions of corporate language policy in social media texts. This study analyzes a corpus of 208 posts from Rakuten, a Japanese MNC, on Glassdoor from 2009 to 2020. The findings suggest that these posts can be divided into three content groups: the threat of a foreign corporate culture, embracing the Rakuten way, and perceptions of leadership and marginalized status. Further, the posts reveal how Rakuten's corporate language policy, as an instrument of internal internationalization, impacts external internationalization. The dynamics of "Englishnization" reveal a pressing issue facing Rakuten: namely, how to balance multinational cohesion with monolingualism and multiculturalism. This paper aims to demonstrate that dynamic topic modeling could enhance our understanding of the manner in which cultural expatriates and the English-only policy affect the internationalization of Japanese MNCs. It contributes to the literature by examining cultural expatriates' perceptions of Japanese work practices and values from a diachronic perspective.

Keywords: cultural expatriates, corporate language policy, social media, computational sociolinguistics, dynamic topic modeling, Englishnization

1. INTRODUCTION

Several Japanese multinational corporations (MNCs) have adopted English as the official corporate language in the past few years. Because Japanese is the dominant business language in Japanese headquarters (HQ), language has become a major hindrance to internationalization (Conrad and Meyer-Ohle 2019). Communicating in English directly, rather than through bilingual employees, gives native English speakers more nuanced knowledge and facilitates communication between HQ and subsidiaries (Neeley 2017). Neeley (2017) showed that an organization-wide English-language mandate in a Japanese company also transferred Japanese work practices and values to American subsidiaries. Employees in subsidiaries are compelled to assume a foreign way of thinking and behaving. Such employees are conceptualized as “cultural expatriates,” representing a growing and underexplored demography (Neeley 2017).

Many employees find themselves in complex cultural settings (Brannen, Garcia, and Thomas 2009). Zhang (2021a) used computer-assisted text analysis to examine employees’ perceptions of an English-only policy over time in a Japanese company; he analyzed a corpus of 704 posts from a social media platform from 2010 to 2018. Since his study relied only on data written in Japanese, it did not account for employees who spoke languages other than Japanese at the HQ or employees located in offices outside Japan. The focus of the current study is non-Japanese employees working in foreign subsidiaries of a Japanese MNC.

Social media could provide insights into a company’s culture. Currently, social media platforms attract more interest than alternative sources of information that reflect public opinions (Alalwan 2018). In particular, the wealth of social media data makes it easier to study employee satisfaction and corporate culture and provides better insight into employee perceptions (Moniz 2016). Of interest to this study is Glassdoor (<https://www.glassdoor.com/>), an important social media platform where current and former employees anonymously review companies and their management (Luo, Zhou, and Shon 2016). However, few studies have utilized Glassdoor as a data source to examine how employees’ perceptions of corporate culture change over time.

The present study uses social media datasets combined with dynamic topic modeling (DTM), corpus methods, and text analysis to investigate these corporate cultural perceptions. Research on cultural expatriates and the role of language as a trigger of cultural frame switching is scant; therefore, this study addresses the gap and explores the change in cultural expatriates’ attitudes toward corporate culture over time. More specifically, the study investigates non-Japanese employees’ perceptions of Japanese work practices and values transmitted by English-only language policy over time. Therefore, the following research questions arise:

Q1 - How do cultural expatriates perceive changes in corporate culture?

Q2 - How could computer-assisted text analysis be exploited to help researchers and practitioners better understand temporal trends in employees’ opinions on social media?

The study focuses on Rakuten, an electronic commerce company based in Tokyo, Japan. The company has 14,826 employees worldwide and operates across 29 countries and regions. In February 2010, Hiroshi Mikitani—the founder and chief executive officer (CEO) of Rakuten—mandated English as the official language of the corporation (Nixon 2015). Rakuten employees

worldwide were required to be proficient in English (Neeley 2011). In 2005 and 2010, Rakuten bought US-based LinkShare and Buy.com and rebranded them as Rakuten Advertising and Rakuten.com, respectively.

This study contributes to the existing literature in four ways. First, the study analyzes the attitudes of cultural expatriates toward "Englishnization" over time. Second, the study uses a methodological synergy for research on cultural expatriates. Previous research has analyzed various contexts of cultural expatriates through traditional methodologies such as interviews, surveys, and ethnography (Brannen, Garcia, and Thomas 2009; Brannen and Thomas 2010; Neeley 2017). Third, the study aims to identify how Englishnization, as one instrument of internal internationalization, impacts external internationalization. Fourth, the study combines DTM with social media datasets and provides insights into cultural expatriates' perceptions of Japanese work practices and values from a diachronic perspective.

2. THEORETICAL BACKGROUND

2.1. Cultural Expatriates

Cultural expatriates are "people who are temporarily or permanently detached from their home culture while still operating in their own country" (Neeley 2017, 5). Jameson (2007) makes a similar distinction by describing cultural expatriates as people with "an individual's sense of self derived from formal or informal membership in groups that transmit and inculcate knowledge, beliefs, values, attitudes, traditions, and ways of life," and they consist of six components: vocation, class, geography, philosophy, language, and biological traits with cultural aspects. Cultural expatriates portray self-identity, defined as "the self as reflexively understood by the individual in terms of his or her biography" (Giddens 1991, 244). Yagi and Kleinberg (2011) define cultural identity as a dimension of self-identity and one's self-concept that concerns perceptions of "who I am" as a cultural being. Keeley (2001) acknowledges that non-Japanese employees are considered an out-group, even when they work in subsidiaries in their home country. Kankaanranta and Lu (2013) argue that the use of business English as a lingua franca—or a common language—in international business can have a homogenizing effect on the cultural identities of the speakers. Jameson (2007) argues that cultural identity is also influenced by close relationships, and changes over time. The close relationships could include family members, friends, and colleagues, with whom interaction occurs frequently (Kankaanranta and Lu 2013). Some native English-speaking informants exhibit linguistic ethnocentrism, do not understand that English is the lingua franca of international business, and assume that "it was their country, their culture, and their market that were of interest" (Neeley 2017, 59). This conception of language ideology, which has been defined as "shared bodies of commonsense notions about the nature of language in the world" (Rumsey 1990, 346), can lead to potentially discriminatory or stereotyped assumptions (Cogo and Yanaprasart 2018). Language plays an important role in expressing cultural norms and beliefs (Ringberg et al. 2010) and can be used as an instrument in the creation and recreation of power (Gaibrois 2015). However, the policy of monolingualism is not a guarantee of success in interpersonal interactions (Welch and Welch 2008). Brannen, Garcia, and Thomas (2009) propose a four-

category framework to understand the different ways that bicultural individuals or biculturals experience and manage their two identities (see TABLE 1). Biculturals are people who have internalized more than one cultural profile (Brannen and Thomas 2010). In Table 1 below, a person categorized as “Both/And” identifies with and integrates two cultures. A person typified as “Either/Or” also identifies with both cultures but changes orientation and behavior according to context.

Num	Bicultural Types	Definition
1	One-Home	One-home biculturals identify mostly with one of their two cultures
2	Neither/Nor	Although aware of—and perhaps even participating in—both of their cultural identities and customs, Neither/Nor biculturals feel marginalized and are conscious about not fitting in with either of their cultures
3	Either/Or	Either/Or biculturals identify with both of their cultural identities, but change their orientation and behaviors based on context
4	Both/And	Both/And biculturals identify with both of their cultures and combine them in various ways

TABLE 1. BICULTURAL TYPES

2.2 Internationalization of Japanese MNCs

To date, several Japanese MNCs have adopted traditional Japanese human resources (HR) practices, characterized by long-term employment and development, seniority-based compensation and promotion, enterprise-based unions, and teamwork orientation (Tung 1984; Morishima 1995; Jacoby 2005). They exhibit an ethnocentric management style, wherein Japanese expatriates are delegated to control overseas operations, and HQs in Japan make decisions for the subsidiaries (Conrad and Meyer-Ohle 2019). This Japanese style of human resources management (HRM), along with the ethnocentric management style, is the primary reason for the inability of Japanese MNCs to attract and retain global talent (Kopp 1994a; Keeley 2001; Froese and Kishi 2013). To counter these challenges, some Japanese companies, especially newly emerging firms, are aggressively internationalizing their HRM practices by emphasizing “internal internationalization” at HQs and “external internationalization” in foreign subsidiaries. Japanese MNCs tend to enhance Japanese employees’ language and intercultural skills to effectively manage foreign operations and hire highly skilled non-Japanese employees

(Sekiguchi, Froese, and Iguchi 2016). Extant studies have consistently reported major issues of "external internationalization." The ethnocentric staffing policies and Japanese style of HRM practices, such as seniority-based compensation, make companies unattractive to non-Japanese employees (Froese and Kishi 2013). Japanese expatriates hold key positions in foreign subsidiaries (Wong 1996, 2010; Black and Morrison 2010), while the local staff has limited opportunities for promotion (Kopp 1994b). Japanese expatriates lack the language skills and cultural background to communicate with highly skilled local employees (Shiraki 2007; Conrad and Meyer-Ohle 2019). Neeley (2017) argues that while Englishnization enables US-based native speakers and Japanese colleagues to have direct exchanges without a translator, the use of the common language enables greater dissemination of Japanese cultural values.

2.3 Computational Sociolinguistics

Computational sociolinguistics is an emerging research field that integrates sociolinguistics and computer science to study the relationship between language and society from a computational perspective (Nguyen et al. 2016). An increasing number of computational linguists have shown an interest in studying language in a social context. Although computational linguistics has significantly contributed to our understanding of the informational dimension of language, it has made little progress in modeling the social dimension (Elangovan and Eisenstein 2015). Sociolinguists traditionally collect data using surveys and ethnographic research as main methods (Milroy and Gordon 2008; Tagliamonte 2006).

Social media is a rich and easy-to-access source of abundant informal information, providing the opportunity to investigate language in social contexts (Johannsen, Hovy, and Søgaaard 2015). Language is an instrument used by people to construct their social identity (Bucholtz and Hall 2005); hence, some studies have focused on automatically inferring social variables—such as gender and age—from texts (Nguyen, Trieschnigg, and Meder 2014). However, to the best of the author's knowledge, no study has conducted a trend analysis of cultural expatriates' perceptions over time. Therefore, this study employed a natural language processing (NLP) tool based on sociolinguistic insights to model and analyze data that reveal the evolutionary nature of topics on cultural expatriates' perceptions of corporate culture.

3. METHODOLOGY

3.1 Data Collection and Preprocessing

The corpus developed for this study consisted of posts extracted from Glassdoor, one of the largest review websites for employers in the world. The website provides extensive details about the latest jobs, company ratings and reviews, CEO approval ratings, salary reports, interview reviews and questions, benefits reviews, and office photos. At the time of research, there were approximately 70,000,000 reviews and evaluated scores of 1,300,000 employers posted by current and former employees. The website has about 50,000,000 unique visitors per month. By using customized web crawlers, the author downloaded data on Rakuten from February 2009 to November 2020, resulting in 208 reviews and a corpus of 49,497 words. The posts included upper- and lower-case letters, different parts of speech, and commonly used

stop words (such as “the,” “a,” “an,” “in”). Therefore, the following preprocessing steps were adopted: converting the text data to lower-case, eliminating punctuation, removing certain stop words, and lemmatizing the words. Modules in the *gensim* library¹ written in Python² language such as *gensim.parsing.preprocessing* and *gensim.utils* were used to perform these tasks.

3.2. Computer-assisted Text Analysis

Computer-assisted text analysis is performed using NLP as a critical technique (Nelson 2020). NLP, together with topic modeling, was employed by this study to model and analyze data. Topic modeling is a collection of NLP algorithms, which use the co-occurrence of words to uncover the latent thematic structure in documents (Blei 2012). Thus, topic modeling can be employed with critical discourse analysis that provides elaborate analytical techniques and theoretical depth for interpretation (Törnberg and Törnberg 2016). Latent Dirichlet Allocation (LDA) proposed by Blei et al. (2003) is a popular topic modeling tool for discovering main themes from large datasets. LDA structures the data into words, topics, and documents, which are generated from randomized mixtures of hidden topics (Vayansky and Kumar 2020). Thus, LDA calculates the percentage of words in a document d that are currently assigned to topic t , and the percentage of times the word w was assigned to topic t over all documents—or $p(\text{topic } t \mid \text{document } d)$ and $p(\text{word } w \mid \text{topic } t)$, respectively. The approach is especially recommended for studies using social media where space is often restricted (Hong and Davison 2010). DTM, the algorithm applied in this study, is used to model the evolution of topics in a corpus over time (Blei and Lafferty 2006). In the DTM as a pre-discretized analysis, time is discretized into several periods, and time within each LDA is used to analyze documents (Blei and Lafferty 2006, Zhang 2021b). As an unsupervised text classification method, DTM can reduce complicated text into more interpretable lists of words. Other topic modeling algorithms such as the LDA do not expect time-tagged data; however, DTM leverages the knowledge of different documents belonging to a different time-slice to map how the words in a topic change over time. Combining DTM with content analysis provides a better understanding of the dynamics of social representations in a comparative and diachronic view (Nguyen et al. 2016; Zhang 2021b).

This study uses a three-step text analysis process (Zhang 2021a, 2021b): corpus construction, pattern detection using human-centered computational exploratory analysis, and pattern refinement using guided deep reading (FIGURE 1). In the first step, patterns were detected by using DTM. In this case, the first year (2009) had 2 articles, the second (2010) 3, the third (2011) 2, the fourth (2012) 11, the fifth (2013) 5, the sixth (2014) 19, the seventh (2015) 22, the eighth (2016) 30, the ninth (2017) 43, the eleventh (2018) 40, the twelfth (2019) 19, and the final year (2020) 12. In the second step, the groups of topic words produced in the quantitative step were examined qualitatively to gain insights into each topic. In the third step, the patterns identified in the first step are inspected to determine if the earlier interpretation of the computational output in the second step was valid. The patterns in the third step were identified through a computationally guided deep reading of the top five documents for each of these topics as well as representative

¹ The library was installed with pip: `pip install gensim`.

² Downloaded from <https://www.python.org/downloads>.

documents from the remaining topics. Computer-assisted text analysis played a key role in processing the corpus and reducing complicated text to informative groups of words.

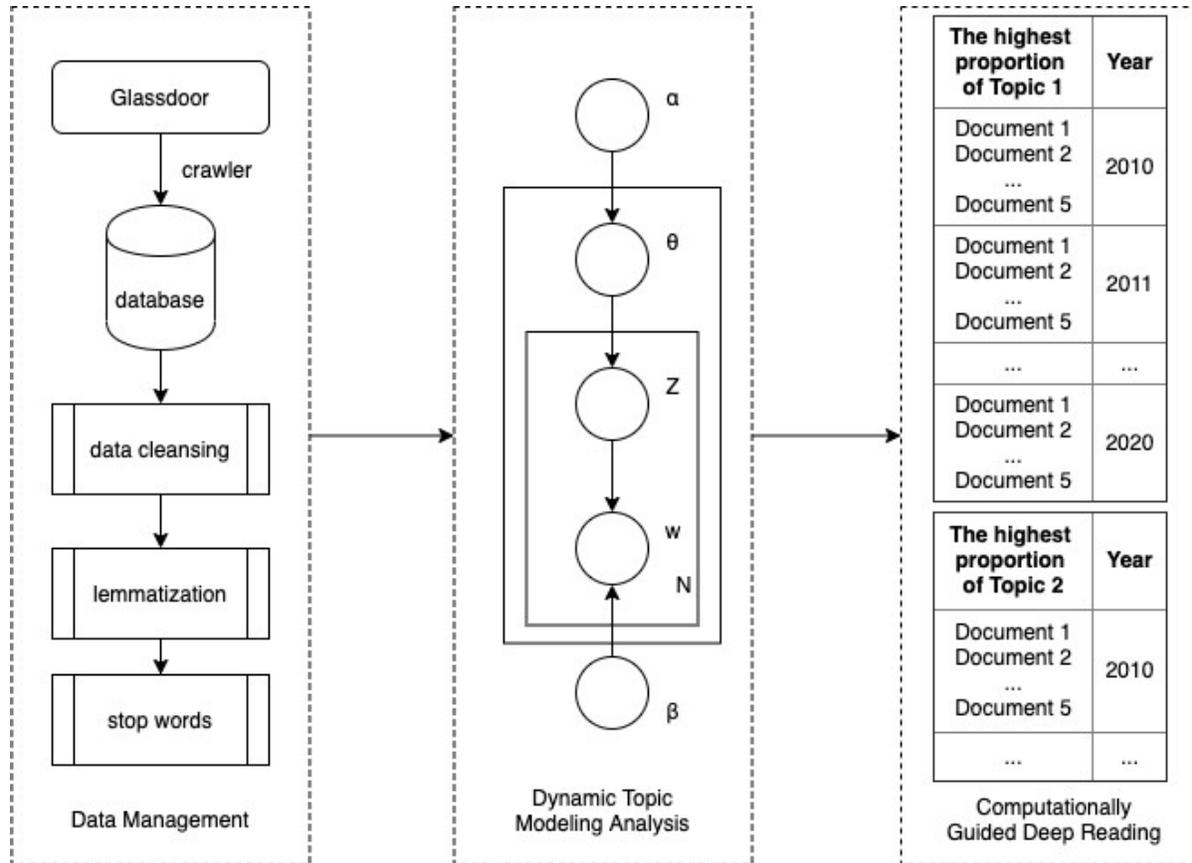


FIGURE 1: DATA AND PROCEDURE

4. FINDINGS

After data collection and preprocessing, the *ldaseqmodel* function in the *gensim* package was used to implement DTM analysis and generate topics. To decide the number of topic models generated objectively from the corpus, this study adopted the selection method proposed by Tanimoto (1958), which checks the difference between the coherence and stability per number of topics. A general rule of the method is to create LDA models across different topic numbers, and then check the Tanimoto coefficient and coherence for each topic. The Tanimoto coefficient, also known as the Tanimoto index, is a statistic used for measuring the similarity and diversity of sample sets (Chung et al. 2019). Coherence measures a single topic by the degree of semantic similarity between high-scoring words which co-occur across the text corpus. The ideal number of topics will maximize coherence and minimize the topic overlap based on the Tanimoto coefficient. FIGURE 2 shows that the ideal number of topics is three. Therefore, this study chooses three topics for the DTM model.

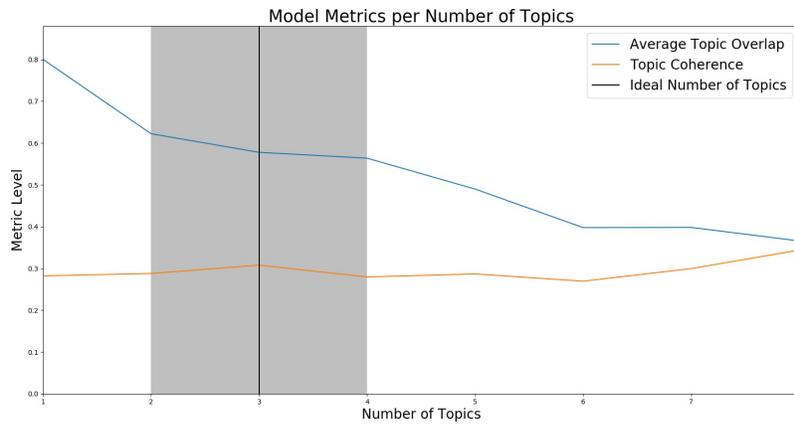
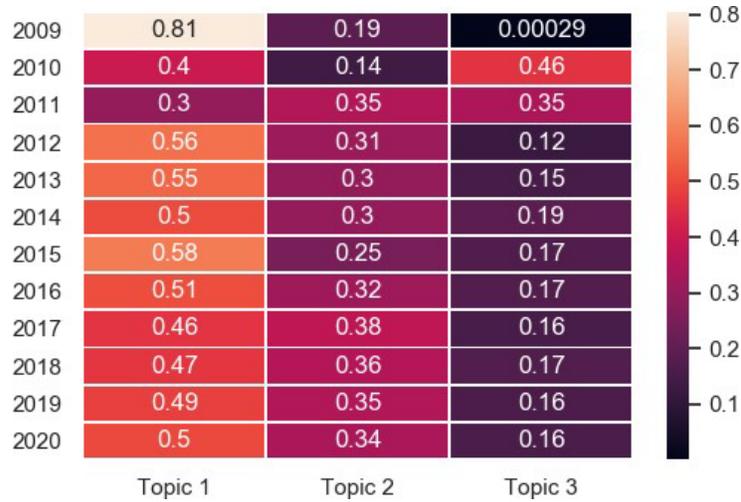


FIGURE 2: NUMBER OF TOPICS

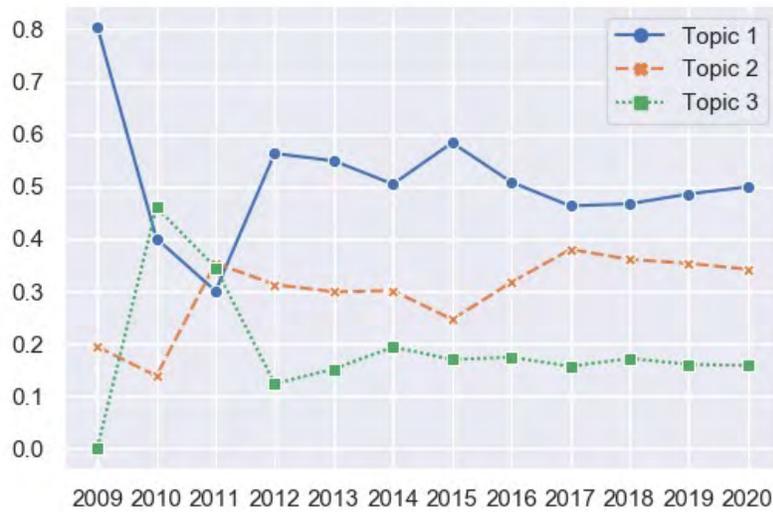
The posts were clustered into three prevalent topics using DTM. The words generated for each topic are listed in TABLE 2. FIGURE 3 displays the trends in each topic, calculated by the topic popularity of all topics for each year.

Num	Topic	Keywords (time = 2020)
1	The threat of a foreign corporate culture	"work," "company," "great," "person," "employee," "good," "management," "lot," "team," "office," "culture," "benefit," "lunch," "leadership," "time," "place," "free," "environment," "manager," "project"
2	Embracing the Rakuten way	"company," "work," "team," "employee," "rakuten," "great," "person," "office," "opportunity," "good," "place," "sale," "management," "culture," "client," "job," "environment," "business," "time," "best"
3	Employees' perceptions of leadership and marginalized status	"company," "management," "employee," "person," "rakuten," "manager," "work," "japan," "business," "problem," "lack," "leave," "hr," "good," "team," "promote," "time," "low," "high," "com"

TABLE 2: THREE TOPICS AND THEIR RELEVANT KEYWORDS



A. HEATMAP



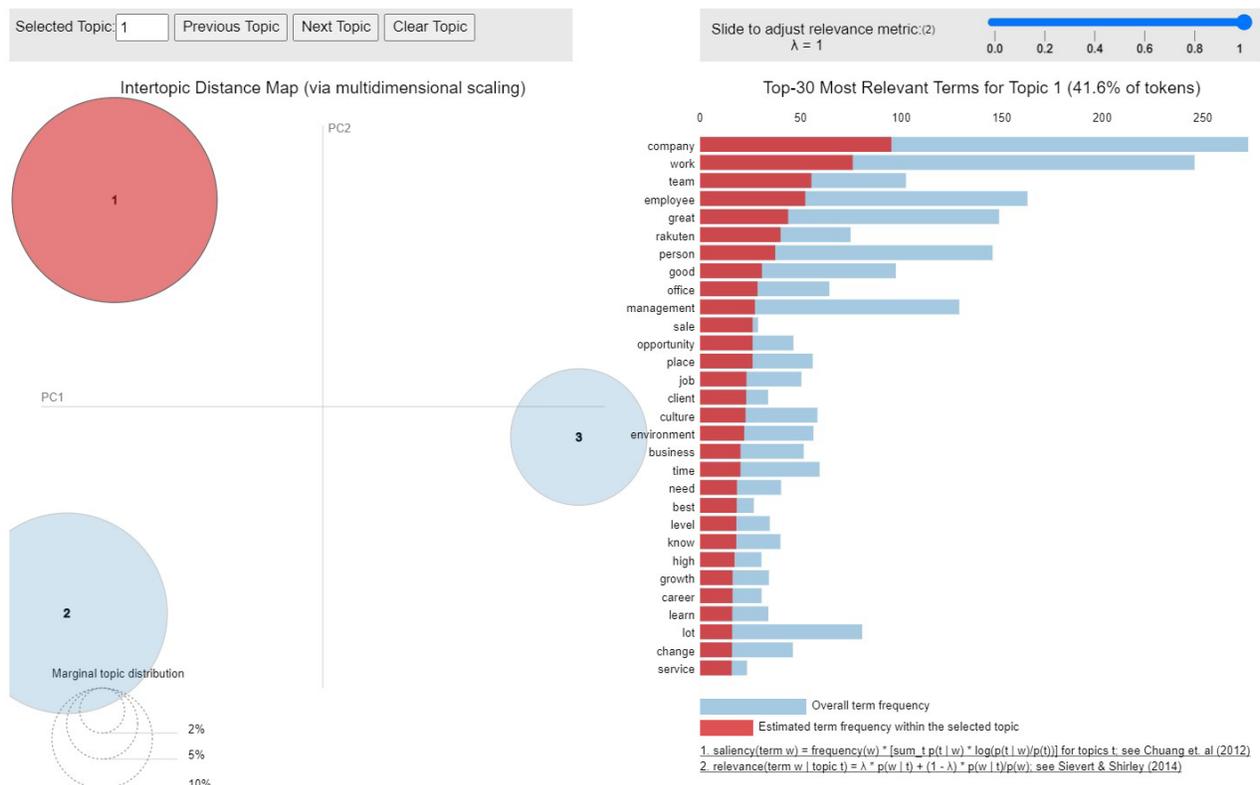
B. TRENDS

FIGURE 3: TOPIC POPULARITY OVER TIME

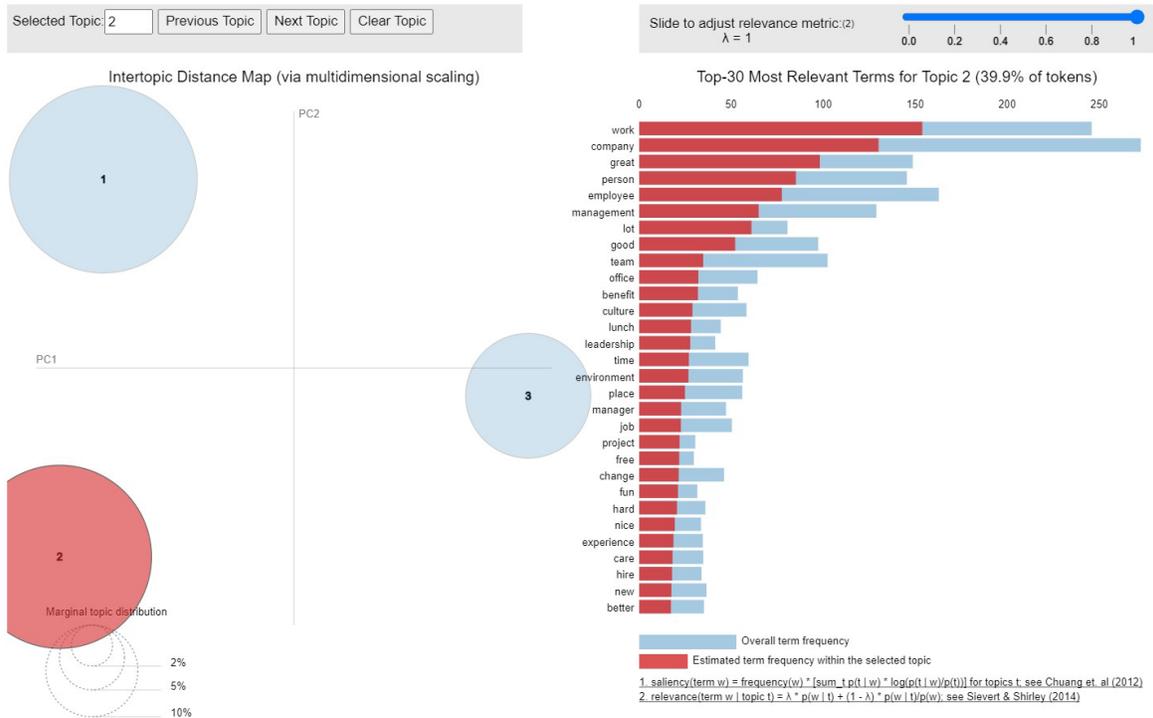
To better illustrate the importance of the words used for discussing the evolution of each topic, a visualization of these topics is presented below, with word clouds that highlight core words and phrases using size and color depth. A topic modeling exploration tool, pyLDAvis³, is used to explore the relationships between the three topics and their most relevant terms. FIGURE 4 presents the spatial distance between the interactions and the three topics among the top 30 most relevant terms for each topic. The pie charts on the left, which represent the proportions of each topic in the corpus, indicate three topics. The presence of an interaction between two topics means they are connected. If there is no interaction, then they are mutually independent. The right bar represents the 30 most relevant terms for the selected topic. FIGURE 4 shows that all

³ The library was installed with pip: pip install pyLDAvis.

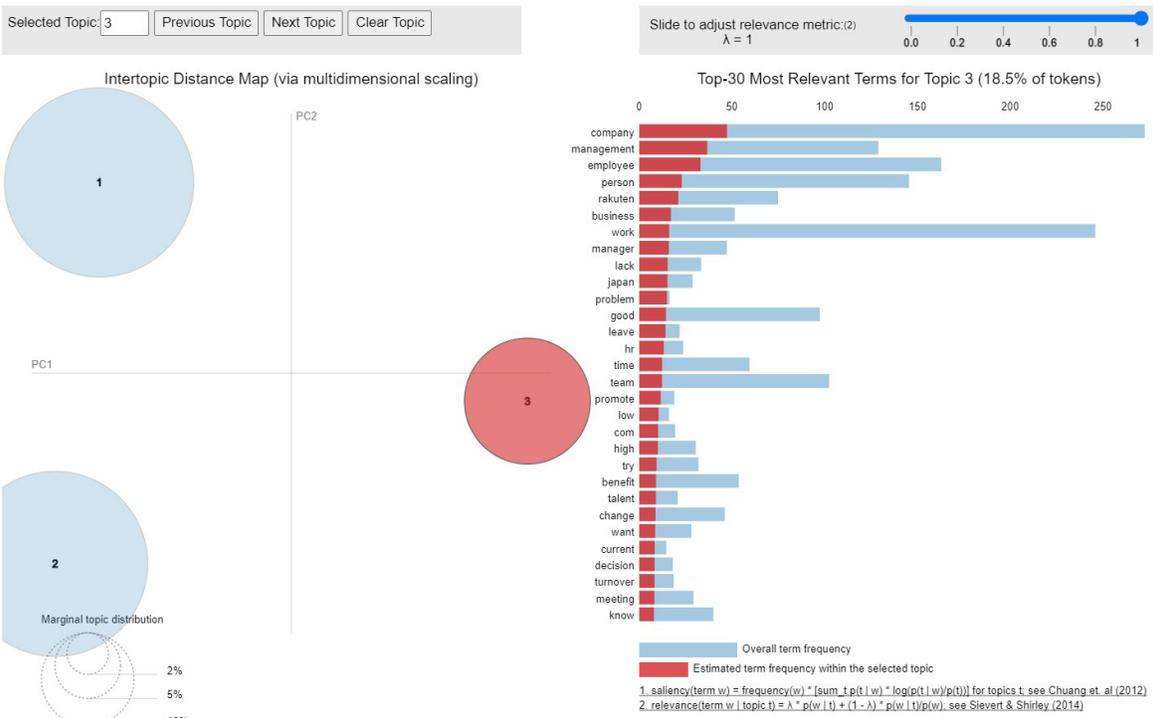
three topics are independent of each other. The right blue bars represent the overall frequency of individual words in the corpus. These proportions are significant because they indicate the level of importance of a term in the corpus. For example, the words “Englishnization,” “English,” and “policy” do not appear in the three topics in the corpus. Therefore, these words are not representative of the three topics. The identified topics are the threat of a foreign corporate culture, embracing the Rakuten way, and perception of leadership and marginalized status. FIGURES 5, 6, and 7 show word clouds of three topics. The size of each word is based on the proportion of its probability of occurrence and changes over time.



a. EMBRACING THE RAKUTEN WAY



b. THE THREAT OF A FOREIGN CORPORATE CULTURE



c. EMPLOYEES' PERCEPTIONS OF LEADERSHIP AND MARGINALIZED STATUS

FIGURE 4: INTER-TOPIC DISTANCE MAP OF TOPICS

4.1 The Threat of a Foreign Corporate Culture

The first topic—related to Rakuten’s organizational culture changes in daily work practices—is rather broad, and has remained central to the discussions over the years. Discussions generally focus on how non-Japanese employees adjusted their perceptions and their place in the company. As shown in FIGURE 5, the most frequently used words in Topic 1 were “company,” “work,” “team,” and “employee,” and these remained highly stable from 2009 to 2020. The focus in the first years of Englishnization (2010 to 2013) was primarily on the subsidiary employees’ adaption to the organizational change due to the acquisition by Rakuten: “lots of work, competition with other companies is fierce” in 2010; “fast-paced (learned more here in 1 year than at other places in 2–3 years)” in 2011; “a lot of pressure and unrealistic expectations (goals, work hours, tasks, etc.) put on employees. Hearing the same feedback every week/day does not make anyone work harder” in 2012; and “Ambitious goals, supported by the parent company Rakuten” in 2013. A close reading of the documents from 2014 revealed that some employees discerned that Rakuten had infused the American subsidiaries with Japanese culture four years after the mandate, which is illustrated in the following posts:

Extract 1

Post title: It’s an interesting company

Content: It is a Japanese company, but expanding globally, so they welcome non-Japanese employees. (No. 39, 2014, San Francisco, Former Employee, Positive Outlook)

Extract 2

Post title: Good place to learn and move on

Content: Weird culture. The company is a mix of OC culture (pretty fair on work/life balance) and Japanese culture (late hours, put in your face-time, don’t leave till your boss leaves, etc.). It partially depends on which group you’re in. (No. 35, 2014, New York, Current Employee, Positive Outlook)

Extract 3

Post title: Lack of leadership!

Content: Japanese militaristic management which means management through fear and intimidation. This is NOT a modern US workplace with a sharing flat structure. It is top-down and managed by Japanese account managers elevated to all C level positions including the CEO. Leadership is weak and only knows how to emulate their Japanese counterpart which doesn’t seem to work in the US. Leadership knows little of the US market and competition which makes it difficult for this company to compete in the US. It’s not a happy place to be especially for account managers or if you are female. (No. 56, 2015, Aliso Viejo, Former Employee, Negative Outlook)

Some clearly expressed their desire to "stay away" from the company from 2018. The superficial change in corporate language did not westernize work practices and work values. Instead, it transformed the company into a more hierarchical organizational structure, making it less attractive to non-Japanese employees. For example, employees were forced to watch the Asakai meeting, held in the HQ every Monday morning, due to Englishnization. Excerpts below provide more detailed reasons and highlight the threat:

Extract 4

Post title: Stay away

Content: * It's difficult to communicate with coworkers in other locations or departments since tools aren't standardized * Constant meaningless meetings where attendance is required and monitored working from home is treated suspiciously * Meetings are scheduled outside working hours regularly * Act on employee feedback and stop trying to force the Japanese working style on American workers * Also invest in training for employees. (No. 128, 2017, San Mateo, Former Employee, Negative Outlook)

Extract 5

Post title: Run away as fast as you can

Content: * Tons of red tape when working with the teams and process in Tokyo * Forced to sit through a 45-minute presentation at 9 a.m. every Monday morning where you're force-fed the latest company propaganda * HR is really lackluster, not enough training available * Lots of office politics and insane matrixed management structure * If you are not Japanese, prepare to hit the "bamboo ceiling" at some point (No. 144, 2018, San Mateo, Current Employee, Negative Outlook)

All the excerpts above elucidate that the source of the threat was Japanese work ethics and social customs. With the implementation of Englishnization, local employees continued to have limited opportunities for promotion and faced difficulties communicating with the staff in the Japanese HQ. It seemed these concerns were not rare. Thus, the topic highlighted that Englishnization helped Japanese HQ to promote Japanese work practices and values and turn the American subsidiaries into a cultural mix.

Before the implementation of Englishnization, it appeared that the frequency of Topic 1 might be decreasing dramatically in the corpus overall. However, one year after the mandate, there seemed to be an increase in the popularity of the topic, which is particularly central from 2012 to 2020 (see FIGURE 3). In the last seven years, the focus of this topic remained at around 0.5, and it was also consistently the most frequent of the three topics.

severely as mission-critical software. (No. 45, 2015)

Extract 7

Content: The Rakuten One strategy is not for everyone. Some employees of acquired companies wish to keep their old ways, resist change, dig in. Hence natural conflict takes a high degree of collaboration to resolve. As a company born in Japan, Rakuten has certain cultural ways that are not for everyone (or new to the Valley). Some employees make fun of our weekly all-hands meetings (conducted globally), our inspirational comments from the CEO and other founders, the respect and care for our workplace, and our willingness to give non-performing employees a chance to hang-on a little longer and try to improve rather than your too-typical Valley company cut-and-run approach. The Rakuten One strategy also means changes, reorganizations, consolidations, new training programs, new ways of recruiting, common ways of paying and measuring performance. Change is sometimes difficult for some people. And some people are afraid they will be passed by with this movement. Our coworkers from Japan are sometimes slow to adapt to Valley/US tech norms, and often break the rules with side-bar conversations in Japanese (we're an English-speaking/writing company). This makes some people uncomfortable, especially those with less experience working in culturally diverse environments. (No. 162, 2018, Salt Lake City, Current Employee, Positive Outlook).

The two excerpts above illustrate that some employees adapted to the cultural change and embraced the mixed working method. Some cultural expatriates, the most skilled in the lingua franca, perceived the "Japanese way" as a change in a culturally diverse environment rather than an "identity threat". Other comments highlight the pride more directly: "Nothing negative from me. I love this company," "Hats off to the management for making this sudden and unexpected transition for all of our team to work remotely. We have been given wonderful support and care for the wellbeing of all of our employees. Thank you, Rakuten."

As shown in FIGURE 3, the popularity of the topic seems to decline from 2009 to 2010, but increases significantly from 2010 to 2011, then declines gradually from 2011 to 2015, and increases again from 2015.



continue to provide the morale-boosting activities that used to happen. (No. 2, 2010, New York)

Extract 10

I found it to be the poorer side of how management treat[s] their staff. This includes the infiltration of the parent company Rakuten and their pawns that run the show. (No. 6, 2011, Client Development Manager, New York)

Extract 11

Inability of the management to guide and bring employees together under the parent company's (Rakuten) vision (No. 10, 2012, Aliso Viejo)

Extract 12

New Rakuten (Japanese Amazon) management and lack of job security. (No. 17, 2012, Aliso Viejo, Former Employee, Negative Outlook)

The above excerpts from 2009 to 2012 show that local employees criticized the insufficient understanding of the leadership teams about the work ethics of the US subsidiaries. The changes in daily work practices, including key performance indicators (KPIs) and Asakai meetings, decided employees' perceptions of the status quo and future. The feelings of marginalization expressed by employees implied that in Rakuten, many US-based employees were prompted to leave the company due to a lack of leadership direction: "job growth is limited; good amount of turnover" in 2013; "Lots of turnover. Plenty of people [are] moving on to greener pastures" in 2014; "Upper level management lacks direction and vision that could align this company with tech progression. High-employee turnover. Lots of employees leaving" in 2015; "Due to the high turnover of employees, people from other departments get put into marketing positions, when they clearly have no marketing experience. Even with the new CEO, CTO, COO...etc. announcements, if you don't change or at least reevaluate your current management and leadership... this company is still bound to fail from these same individuals... a big 'restructuring' is in progress" in 2016; and "Ton of layoffs and also heard [a] bunch of sales guys left" in 2017.

Employees also compared their situations with Japanese expatriates who could rent luxurious houses and will never get fired:

Extract 13

In case you don't do what they tell you to do: a) If you were assigned from Japan, probably you will be assigned back to Japan and never again assigned to [the] US again. b) If you are [a] US employee, you will be fired or invited to leave very quickly. After you leave the company, all the current problems will be assigned to you. It is how they survive and hide the problem. So, they are always hiring new employees.

P.S.: Japan Rakuten Card is [the] worst. They are the source of current management in Card US. It means the source of [the] issue may be on the Japanese side. Those problems [explain] the high turnover of very good employees and also why US employees have low [salaries]. It also explain[s] why the president thinks the US employees are lazy. But I don't know if the president does not know it. (If he knows, then something is very wrong). My advice to the president would be: I think you need to solve these problems as soon as possible. Please, fix Japan and US management. Only permitting managers from one team to another does not solve the problems only hides because the current team managers, group managers and department managers are the ones causing all these problems. This is my last review in this company. (No. 150, 2018, Software Engineer, San Mateo, Former Employee, Negative Outlook)

Extract 14

Post title: Ethnocentric in their management style

Content: No one gets fired, low performing, problematic employees are moved to other departments or shipped back to Japan. [The] overall company strategy is a shotgun approach. It literally changes on [a] quarterly basis, and it's mainly driven by the CEO and passed down to his Yes-men! (No. 186, 2019, Product, San Mateo, Current Employee, Negative Outlook)

The above fragments show strong claims: after an eight-year mandate, the company increased expatriation and adopted a more ethnocentric staffing policy. Employees expressed their perceptions of leadership and sense of dissatisfaction with the organizational hierarchy through remarks such as “Shake off the Yes-men, from IT to HR to Facilities, to all business groups. Stop bending backward to please the CEO” and “Moving to California is useless if you keep the rigid, hierarchical mindset that prevents creativity from thriving.”

The topic of employees' perceptions of leadership and feelings of being marginalized emerged in 2009 and was spurred by the announcement of Englishnization in 2010. It decreases significantly from 2010 to 2012 but increases again from 2012 to 2014 and appears to remain relatively stable from 2014 to 2020. As shown in FIGURE 7, words such as “management,” “employee,” “company,” and “work” stand out from 2009 to 2014. However, after 2015 “management” and “company” were still major words, but “lack,” “low,” “Japan,” and “leave” became more prominent.

career advancement on the part of cultural expatriates should be ascribed to global expansion rather than linguistic ethnocentrism. This study can contribute to the ongoing discussion on monolingual solutions, which constitute a conventional and simplistic style of thinking (Martin 2009).

Second, this study has proposed automatic content analysis using DTM and social media data to measure perceptions of Japanese work practices and values transmitted by the English-only policy over time. These results align with the literature on DTM, which enables one to develop a better understanding of the dynamics of social representations from a comparative and diachronic perspective within a contextual framework (Chandelier et al. 2018).

Extant studies of biculturals have provided topic categories, including the following: One-Home, Neither/Nor, Either/Or, and Both/And (Brannen, Garcia, and Thomas 2009; Brannen and Thomas 2010). Topic 1 (The threat of a foreign corporate culture) was the most popular topic between 2012 and 2020. This topic draws attention to the Either/Or type of cultural expatriates, suggesting that cultural expatriates started to identify with their American and Japanese identities around 2014; however, they perceived the forcible imposition of Japanese work practices and values at the workplace as a threat.

Neeley (2017) also observed that the Japanese way, described pejoratively as “isolationist” or “pre-global” in mindset, was a threat to the cultural expatriates’ sense of their own American identity. The results of this study are consistent with previous findings, which suggest that while cultural expatriates’ perceptions of the contrast between their two cultures are strengthened, they resist the “Japanese way” and link “stay away from the company” to “getting rid of the threat.” Japanese work practices and values are seen as posing a threat to the status of English as the corporate language and symbol of American identity.

Topic 2 (embracing the Rakuten way) was the second most popular topic, and its popularity remained stable between 2011 and 2020. The prevalence of this topic shows the popularity of Both/And cultural expatriates. The literature suggested these Both/And individuals identify with both of their cultures and leverage various cultural identities in a unique and idiosyncratic way (Brannen, Garcia, and Thomas 2009). The data show that these cultural expatriates disagree with other employees who resist the cultural change. They think “the change is not for everyone,” and even persuade people not to believe those negative reviews on Glassdoor and “come and see for yourself.”

Topic 3 (perception of leadership and marginalized status) was the least popular. The data show that the posts are often connected to cultural expatriates’ perceptions and portray them as a marginalized group. Arrival at the Neither/Nor type happens in several ways. For example, these cultural expatriates choose not to identify with the American or Japanese operating styles because of leadership headed in the wrong direction. The study results have sinister implications for the adoption of a common corporate language, which often introduces new barriers (Welch and Welch 2008). Cultural expatriates with a strong competence in the company language may still feel excluded if they are not comfortable with the Japanese work culture they must adopt.

Finally, another relevant insight drawn from the findings is related to the traditional emphasis of employing local staff to fill the positions formerly dominated by Japanese expatriates to attract and retain valuable non-Japanese talent (Sekiguchi, Froese, and Iguchi 2016). Researchers have argued that changing from an ethnocentric toward a polycentric or geocentric staffing policy could potentially help overcome obstacles for the external internationalization of Japanese MNCs (Beamish and Inkpen 1998; Schaaper et al. 2013). Conversely, this study suggests that setting English as the official corporate language could generate unintended consequences of increasing expatriation and impeding localization. When translators, "bridge individuals" or bilingual employees who facilitate communication between HQ and subsidiaries, became obsolete at Rakuten with the implementation of Englishnization, the Japanese HQ, sent more expatriates to fill senior positions in the American subsidiaries. Neeley (2017) also highlighted this in her study and explained that the superficial change in language patterns to become more direct did not necessarily translate cross-company into a less hierarchical organizational culture.

Thus, the present study contributes to the ongoing discussion on the interplay between Englishnization and organizational culture. As previously mentioned, Zhang (2021a) concluded that an English-only corporate language policy results in a mixed corporate culture at the Japanese HQ. However, the present study found that the organizational culture of overseas subsidiaries has become more hierarchical due to Englishnization. These results align with existing research on the internationalization of Japanese MNCs (Wong 1996, 2010; Black and Morrison 2010). A close analysis of the dynamics of Englishnization reveals a pressing issue facing Rakuten and other Japanese MNCs with regard to the balance of multinational cohesion with monolingualism and multiculturalism.

Based on the Glassdoor analytics data of 208 reviews from 2009 to 2020, this study reveals key insights into the analysis of cultural expatriates. The DTM and content analysis elucidate the cultural expatriates' perceptions of Japanese work practices and values from a diachronic perspective. Moreover, the study results have implications for viewing Englishnization as one instrument of internal internationalization, which could have a negative impact on the external internationalization of Japanese MNCs.

This study is a small step toward using the computational text analysis method to examine the experiences of cultural expatriates. However, the study has some limitations. First, the data collection was limited to a single company and social media platform. Second, although the data collection period was long (2009–2020), the data size was not particularly large. Taking a broader view, future studies could collect data from various companies and data sources. Last but not least, future research could combine DTM with other techniques, such as sentiment analysis, to examine changes in cultural expatriates' sentiments regarding Japanese work practices and values.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Gisle Andersen and Kristin Rygg for their thoughtful comments and constructive suggestions. I would also like to thank Carlos Perrián-Pascual, the

editor-in-chief of this journal, and the two reviewers for their detailed and insightful comments on my paper.

REFERENCES

Alalwan, Ali Abdallah. 2018. "Investigating the Impact of Social Media Advertising Features on Customer Purchase Intention." *International Journal of Information Management* 42: 65–77.

Beamish, Paul W., and Andrew C. Inkpen. 1998. "Japanese Firms and the Decline of the Japanese Expatriate." *Journal of World Business* 33 (1): 35–50.

Black, J., and A. Morrison. 2010. *Sunset in the Land of the Rising Sun: Why Japanese Multinational Corporations Will Struggle in the Global Future*. New York: Palgrave Macmillan.

Blei, David M. 2012. "Probabilistic topic models." *Communications of the ACM* 55(4): 77–84.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet allocation." *Journal of Machine Research* 3: 993–1022.

Blei, David M, and John D. Lafferty. 2006. "Dynamic Topic Models." In *Proceedings of the 23rd International Conference on Machine Learning*, 113–120.

Brannen, Mary Yoko, and David C. Thomas. 2010. "Bicultural Individuals in Organizations: Implications and Opportunity." *International Journal of Cross Cultural Management* 10(1): 5–16.

Brannen, Mary Yoko, Dominie Garcia, and David C. Thomas. 2009. "Biculturals as Natural Bridges for Intercultural Communication and Collaboration." In *Proceedings of the 2009 International Workshop on Intercultural Collaboration*, 207–210.

Bucholtz, Mary, and Kira Hall. 2005. "Identity and Interaction: A Sociocultural Linguistic Approach." *Discourse Studies* 7(4-5): 585–614.

Chandelier, Marie, Agnes Steuckardt, Raphael Mathevet, Sascha Diwersy, and Olivier Gimenez. 2018. "Content Analysis of Newspaper Coverage of Wolf Recolonization in France using Structural Topic Modeling." *Biological Conservation* 220: 254–261.

Chung, Neo Christopher, Błażej Miasojedow, Michał Startek, and Anna Gambin. 2019. "Jaccard/Tanimoto Similarity Test and Estimation Methods for Biological Presence-Absence Data." *BMC Bioinformatics* 20(15): 1-11. <https://doi.org/10.1186/s12859-019-3118-5>.

Cogo, Alessia and Yanaprasart, Patchareerat. 2018. "English is the language of business": An exploration of language ideologies in two European corporate contexts." In *English in Business and Commerce: Interactions and Policies; English in Europe, Volume 5*, edited by Tamah Sherman

and Jiri Nekvapil, 96-116. Berlin, Boston: De Gruyter Mouton.
<https://doi.org/10.1515/9781501506833-005>

Conrad, Harald, and Hendrik Meyer-Ohle. 2019. Overcoming the Ethnocentric Firm? – Foreign Fresh University Graduate Employment in Japan as a New International Human Resource Development Method. *The International Journal of Human Resource Management* 30(17): 2525–2543.

Elangovan, Vinodh Krishnan, and Jacob Eisenstein. 2015. "You're Mr. Lebowsky, I'm the Dude": Inducing Address Term Formality in Signed Social Networks. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies*. 1616-1626.

Froese, Fabian J., and Yasuyuki Kishi. 2013. "Organizational Attractiveness of Foreign Firms in Asia: Soft Power Matters." *Asian Business & Management* 12(3): 281–297.

Gaibrois, Claudine. 2015. "Power at Work: the Discursive Construction of Power Relations in Multilingual Organizations." Thesis. Bamberg: Difo-Druck.

Giddens, Anthony. 1991. *Modernity and Self-identity: Self and Society in the Late Modern age*. Stanford: Stanford University Press.

Hong, Liangjie, and Brian D. Davison. 2010. "Empirical Study of Topic Modeling in Twitter." In *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*.
<https://doi.org/10.1145/1964858.1964870>.

Jacoby, Sanford M. 2005. "Business and Society in Japan and the United States." *British Journal of Industrial Relations* 43(4): 617.

Jameson, Daphne A. 2007. "Reconceptualizing Cultural Identity and its Role in Intercultural Business Communication." *The Journal of Business Communication* (1973) 44(3): 199–235.

Johannsen, Anders, Dirk Hovy, and Anders Søgaard. 2015. "Cross-lingual Syntactic Variation over Age and Gender." In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, 103–112.

Kankaanranta, Anne, and Wei Lu. 2013. "The Evolution of English as the Business Lingua Franca: Signs of Convergence in Chinese and Finnish Professional Communication." *Journal of Business and Technical Communication* 27(3): 288–307.

Keeley, Tim. 2001. *International Human Resource Management in Japanese Firms: Their Greatest Challenge*. London: Palgrave Macmillan UK.

- Kopp, Rochelle. 1994a. "International Human Resource Policies and Practices in Japanese, European, and United States Multinationals." *Human Resource Management* 33(4): 581–599.
- Kopp, Rochelle. 1994b. *The Rice-paper Ceiling: Breaking Through Japanese Corporate Culture*. Berkeley: Stone Bridge Press.
- Kwok, Linchi, and Bei Yu. 2013. "Spreading Social Media Messages on Facebook: An Analysis of Restaurant Business-to-Consumer Communications." *Cornell Hospitality Quarterly* 54(1): 84–94.
- Luo, Ning, Yilu Zhou, and John Shon. 2016. "Employee Satisfaction and Corporate Performance: Mining Employee Reviews on Glassdoor.com". In *Proceedings of the 2016 International Conference on Information Systems*. 1-16.
- Martin, Roger L. 2009. *The Opposable Mind How Successful Leaders Win Through Integrative Thinking*. Boston: Harvard Business School Press.
- Milroy, Lesley, and Matthew Gordon. 2008. *Sociolinguistics: Method and Interpretation*, vol. 13. Chichester: John Wiley & Sons.
- Moniz, Andy. 2016. "Inferring Employees' Social Media Perceptions of Goal-setting Corporate Cultures and the Link to Firm Value." *Unpublished Working Paper*.
- Morishima, Motohiro. 1995. "Embedding HRM in a Social Context." *British Journal of Industrial Relations* 33(4): 617–640.
- Neeley, Tsedal. 2011. "Language and Globalization: 'Englishnization' at Rakuten." *Harvard Business School Organizational Behavior Unit Case* (412-002).
- Neeley, Tsedal. 2017. *The Language of Global Success: How a Common Tongue Transforms Multinational Organizations*. Princeton: Princeton University Press.
- Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49(1): 3–42.
- Nguyen, Dong, A Seza Dođruöz, Carolyn P Rosé, and Franciska De Jong. 2016. "Computational Sociolinguistics: A Survey." *Computational Linguistics* 42(3): 537–593.
- Nguyen, Dong, Dolf Trieschnigg, and Theo Meder. 2014. "Tweetgenie: Development, Evaluation, and Lessons Learned." In *Proceedings of Coling 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, 62–66.
- Nixon, Richard Mark. 2015. "Workplace English Usage in Japan." *The Journal of the Faculty of Foreign Studies* 47: 21–34.

Ringberg, Torsten V., David Luna, Markus Reihlen, and Laura A. Peracchio. 2010. "Bicultural-bilinguals: The Effect of Cultural Frame Switching on Translation Equivalence." *International Journal of Cross Cultural Management* 10(1): 77–92.

Rumsey, Alan. "Wording, Meaning, and Linguistic Ideology." 1990. *American Anthropologist* 92, no. 2: 346–61.

Schaaper, Johannes, Bruno Amann, Jacques Jaussaud, Hiroyuki Nakamura, and Shuji Mizoguchi. 2013. "Human Resource Management in Asian Subsidiaries: Comparison of French and Japanese MNCs." *The International Journal of Human Resource Management* 24 (7): 1454–70.

Sekiguchi, Tomoki, Fabian Jintae Froese, and Chie Iguchi. 2016. "International Human Resource Management of Japanese Multinational Corporations: Challenges and Future Directions." *Asian Business & Management* 15(2): 83–109.

Shiraki, Mitsuhide. 2007. *Role of Japanese Expatriates in Japanese Multinational Corporations: From the Perspective of the Multinational Internal Labor Market.* Working Paper No.42. Tokyo: School of Political Science and Economics, Waseda University.

Tagliamonte, Sali A. 2006. *Analysing Sociolinguistic Variation.* Cambridge: Cambridge University Press.

Tanimoto, Taffee T. 1958. "Elementary Mathematical Theory of Classification and Prediction." International Business Machines Corp.

Törnberg, Anton, and Petter Törnberg. 2016. "Combining CDA and Topic Modeling: Analyzing Discursive Connections Between Islamophobia and Anti-feminism on an Online Forum." *Discourse & Society* 27(4): 401–422.

Tung, Rosalie L. 1984. "Strategic Management of Human Resources in the Multinational Enterprise." *Human Resource Management* 23(2): 129–143.

Vayansky, Ike, and Sathish A.P. Kumar. 2020. "A Review of Topic Modeling Methods." *Information Systems* 94: 101582. <https://doi.org/10.1016/j.is.2020.101582>.

Welch, Denise E., and Lawrence S. Welch. 2008. "The Importance of Language in International Knowledge Transfer." *Management International Review* 48(3): 339–60. <https://doi.org/10.1007/s11575-008-0019-7>.

Wong, Heung Wah. 2010. "Why a Globalizing Corporate Culture Still Inhibits Localization of Management; the Yaohan Case." *Sangyō Keiei Kenkyū* 32(1): 31–48.

Wong, May M.L. 1996. "Shadow Management in Japanese Companies in Hong Kong." *Asia Pacific Journal of Human Resources* 34(1): 95–110.

Xiang, Zheng, Qianzhou Du, Yufeng Ma, and Weiguo Fan. 2017. "A Comparative Analysis of Major Online Review Platforms: Implications for Social Media Analytics in Hospitality and Tourism." *Tourism Management* 58: 51–65.

Yagi, Noriko, and Jill Kleinberg. 2011. "Boundary Work: An Interpretive Ethnographic Perspective on Negotiating and Leveraging Cross-cultural Identity." *Journal of International Business Studies* 42(5): 629–653.

Zhang, Ziyuan. 2021a. "Content Analysis of Language-Sensitive Recruitment Influenced by Corporate Language Policy Using Topic Modeling." *HERMES – Journal of Language and Communication in Business*, no. 61: 77–91. <https://tidsskrift.dk/her/article/view/127928>.

Zhang, Ziyuan. 2021b. "It is All about Topic: Discovering Topics and Trends in Employee Perceptions of Corporate Language Policy." *Journal of Multilingual and Multicultural Development*: 1–19.