

Mandarin Mingled by Cantonese: A Phenomenon of Language Variation*

LI Jin-feng

School of Liberal Arts, Zhaoqing University, Zhaoqing 526061, China

Mandarin mingled by Cantonese is a phenomenon of Mandarin variation. This paper adopts heterogeneous words to explore its specific performances within the context of Mandarin by investigating 21 cities in Guangdong Province of China, and applies the average ratio of heterogeneous words to measure the variation degree of Mandarin influenced by Cantonese. The statistical results suggest that language variation is mainly reflected in pronunciation, vocabulary and grammar, especially in pronunciation. Variation can be found in all kinds of parts of speech, mostly nouns, verbs and adjectives. Furthermore, there is a certain internal correlation between variation and residents' individual characteristics: the elder, less educated and rural resident, the more serious the variation is; and male is more prone to variation than female. Additionally, the degree of variation varies greatly from city to city. The Mandarin level in Shenzhen and Dongguan is much higher than that in Yunfu and Jieyang, which is closely related to the level of regional economic development.

Keywords: language variation, Mandarin mingled by Cantonese, heterogeneous words, average ratio of heterogeneous words

Introduction

The communication between different countries or regions is accompanied by the contact of different languages. The long-term contact between languages may leave traces in each other and breed some new features, such as the phenomenon of diglossia (Hartmann & Stork, 1972, pp. 277-302). With the popularization of Mandarin and the development of society and education, the fierce competition between Mandarin and dialects has resulted in two kinds of language phenomena: one is Mandarin with regional dialect features, namely regional Mandarin, such as Mandarin with Minnan accent, Mandarin with Shanghai style, Mandarin with Hong Kong style, Mandarin with Cantonese style (Li, 2020a) or Mandarin mingled by Cantonese (Li, 2020b); the other is the dialect with Mandarin features (Li, 2019b).

This paper is concerned with Cantonese to explore the process of interaction and mutual penetration between Cantonese and Mandarin, that is, the variation of Mandarin influenced by Cantonese. Mandarin is the

* **Acknowledgements:** this paper is funded by Project: 2019 Youth Project of Philosophy and Social Sciences Planning in Guangdong Province "Research on the Inter-generational Inheritance of Cantonese: A Survey Based on the Pearl River Delta Region (GD19YYY07)"; 2019 Young Innovative Talents Project of Guangdong Ordinary Colleges and Universities "The Inter-generational Inheritance of Dialect of Foreign Population: Based on the Survey of Guangzhou and Zhaoqing (2019WQNCX124)".

LI Jin-feng, Ph.D., lecturer, School of Liberal Arts, Zhaoqing University.

official language of China, while Cantonese is widely used in Guangdong Province, some parts of Guangxi Province, Hong Kong and Macao, with a population of nearly 100 million. Additionally, many people are proud of Cantonese due to its long history and unique culture. Therefore, people in Cantonese speaking areas will inevitably be influenced by Cantonese when using Mandarin. That is to say, Mandarin will inevitably face changes or even variations in this process. This variation is Mandarin mingled by Cantonese (Li, 2019b), a concept used in this paper.

Literature Review

Language Variation

The long-term and frequent contact between different languages leads to the mutual influence between languages (Thomason, 2001), which may further lead to the mutual transfer of language elements (Lado, 1957; Odlin, 1989) and language variation (Thomson, 2001, 2008, 2009).

Language variation research is the core content of sociolinguistics. Social dialect variation and register study constitute two major parts of sociolinguistic variation research. The study of social dialect variation, such as the sociolinguistic stratification by William Labov and Peter Trudgill, lays a basic paradigm for the study of social dialect variation, focusing on the changes of language features in different social groups and in different situations. Compared with the study of social dialect variation, register variation focuses on the analysis of the influence of situational factors on language features, with the help of corpus and modern statistical techniques. These studies involve a variety of registers, promote the development of the study of variation in terms of methodology and achieve a comprehensive analysis of language features and a multidimensional description of language variation, which complements and expands the achievements of sociolinguistic variation studies.

Language variation discussed in this paper is limited to the variation in a narrow sense, and the synchronic differences in the use of Mandarin among different groups of people and in the context of different social attributes, as well as the variation in pronunciation, vocabulary, and grammar due to the influence of dialects.

Performance of Language Variation

There are different standards for the classification of language variation. The common classification includes phonetic variation, lexical variation, grammatical variation etc.. Language variation can also be divided into diachronic and synchronic language variation etc.. This paper focuses on the analysis of synchronic language variation, which is manifested in pronunciation, vocabulary and grammar aspects.

Much attention had been paid to pronunciation variation, such as Vineyard Island, Massachusetts (Labov, 1963), department store salesmen in New York City (Labov, 1966), Norwich dialect (Trudgill, 1974), the pronunciation of after vowel [r] in Detroit Black oral English (Trudgill, 1983), and the pronunciation of [ŋ], [t] and [h] in Norwich (Trudgill, 1983). Claire (2015) noted that the Gaelic spoken by elder Gaelic speakers is a word stress language similar to Swedish, and young people do not use traditional accents. In terms of Chinese, Liang (2018) argued that the Yinping tone of Dongguan Cantonese was adjusted and increased to high-flat tone, which may be due to the transfer from Mandarin with higher reputation or standard Cantonese. This reflects the variation of dialects under the influence of Mandarin. Conversely, Mandarin mingled by Cantonese is influenced by Cantonese when speaking Mandarin. Seong et al. (2019) conducted an interview and questionnaire survey in Kota Kinabalu, the capital of Sabah Prefecture and mentioned that the Hakka spoken by the ethnic group had

changes in pronunciation and vocabulary, showing the change from different ancestral Hakka varieties to regional Hakka varieties. In the Cantonese speaking areas, Mandarin mingled by Cantonese is expressed by the vocabularies in the Cantonese system when speaking Mandarin, such as “鸡公 (cock)” instead of “公鸡”, “得闲 (free)” instead of “有空” and “一个钟 (an hour)” instead of “一个小时”, etc..

From the perspective of language itself, grammatical variation is relatively slow. In contrast, there are relatively few studies in this area. This paper is based on the comparison of grammatical expression habits between Mandarin and Cantonese. For example, many Cantonese often say “你走先 (you go first)” and “给支笔我 (give me a pen)” when speaking Mandarin, which involves the heterogeneous grammar of Mandarin.

Influencing Factors of Language Variation

Language variation is the result of both intra-lingual and extra-lingual factors (Anttila, 2002). Intra-lingual factors mainly include language contact, while extra-lingual factors involve the language attitude and personal attributes of the speaker, as well as economic, social and policy factors.

In the area of Chinese dialects, the long-term contact between dialects and Mandarin leads to the variation of Mandarin under the influence of dialects, resulting in the phenomenon of “Mandarin Mingled by dialect”, which is the variation of standard Mandarin (Li, 2020b).

Language variation is the social variation of language. On one hand, language variation is restricted by the social class or status of communicators (Labov, 1966; Trudgill, 1974), which conforms to the “social linguistic patterns”, that is, people with high social status pronounce more accurately and use more standard variants than those with low social status. On the other hand, speakers’ gender (Biber & Burges, 2000; Rey, 2001), age (Cheshire et al., 2011), occupation (Trudgill, 1974), education level (Trudgill, 1974) are the key social factors leading to language variation. Moreover, the role of external factors in language variation can not be ignored, including economic factors (Dauenhauer & Dauenhauer, 1998), social factors (Kimberly, 2014), and policy factors (Patrick, 2017).

Result of Language Variation

As far as Chinese dialects and Mandarin are concerned, according to the implication relationship between language contact intensity and borrowing level (Thomson, 2001, pp. 70-71), they belong to deep contact. The influence of both sides will go deep into their respective phonetic, lexical and grammatical systems. Therefore, the result of variation may be the formation of the phenomenon of Mandarin and dialect diglossia, that is, the coexistence of Mandarin mingled by dialect and dialect mingled by Mandarin (Li, 2019a). In the Cantonese speaking areas of Guangdong Province, Cantonese is the main competitor of Mandarin, which greatly hinders the process of popularizing Mandarin (Liang, 2015). Therefore, the phenomenon of “Mandarin mingled by Cantonese” will exist for a long time.

Aims of This Paper

Previous researches mainly focus on the distribution characteristics and influencing factors of language variation. However, there is no quantitative researches to measure the degree of language variation in different groups, in different dimensions (gender, age, etc.), and in different regions. Therefore, Aims of this study are twofold. First, we introduce the average ratio of heterogeneous words in Mandarin mingled by Cantonese to measure the variation degree of Mandarin influenced by Cantonese and quantitatively analyzed the specific

performances of Mandarin variation in pronunciation, vocabulary and grammar. Second, the aim was to reveal the distribution of Mandarin mingled by Cantonese in different dimensions from the perspective of the part of speech, speaker's individual characteristics and region he belongs to, so as to expand the relevant research on language variation.

Research Design

This paper explores the variation of Mandarin in Cantonese speaking areas and obtains the first-hand information from the speakers. The field investigation was conducted from July to August 2019 through interviews and WeChat.

The survey was limited to 21 prefecture-level cities in Guangdong Province. To avoid the extreme distribution of samples, stratified sampling method was adopted, that is, according to a certain standard, a fixed sample value or range was artificially set for a certain city. The stratification criterion was the number of registered population of each city, namely, the proportion of total population with residence registration at year-end by city (2018) in the total registered population of Guangdong Province. The demographic data were obtained from the Statistical Yearbook of Guangdong Province in 2019. A total of 2000 people participated in the survey, they were registered in Guangdong Province, aged between 7 and 60, and can speak Cantonese and Mandarin.

A total of 200 investigators were involved in this survey, the condition was consistent with that of the respondent. Additionally, investigators were all undergraduates from Guangzhou and Zhaoqing City, Guangdong Province and required to have the Certificate of Mandarin Level 2 or above. In the process of the survey, the investigators used standard Mandarin to communicate with a single interviewee on a certain topic within 5 minutes, and after obtaining the consent of the interviewee, the interview content was recorded.

According to the above proportion of population, the distribution of the investigated samples and investigators is shown in Table 1.

Table 1

Distribution of Survey Samples and Investigators

City	Proportion of population (%)	Number of respondents	Number of investigators	Effective sample
Guangzhou	9.76	195	20	195
Shenzhen	5.24	105	10	101
Zhuhai	1.34	28	3	26
Shantou	5.99	120	12	112
Foshan	4.60	92	9	90
Shaoguan	3.54	71	7	55
Heyuan	3.92	78	8	59
Meizhou	5.77	115	12	110
Huizhou	4.01	80	8	80
Shanwei	3.83	77	8	63
Dongguan	2.44	49	5	49
Zhongshan	1.86	37	4	37
Jiangmen	4.20	84	8	80
Yangjiang	3.16	63	6	57

Zhanjiang	8.92	178	18	153
Maoming	8.53	171	17	149
Zhaoqing	4.74	95	9	95
Qingyuan	4.66	93	9	82
Chaozhou	2.90	58	6	57
Jieyang	7.42	148	15	131
Yunfu	3.17	63	6	62
Total	100	2000	200	1843

Descriptive Statistical Analysis

Basic Statistical Results

According to statistics, 244,817 words and 25,131 heterogeneous words in Mandarin mingled by Cantonese were obtained, that is, the proportion of heterogeneous words is about 10.27%, which is called “average ratio of heterogeneous words” in this paper. This illustrates that there are 10 variations on average per 100 Chinese words. According to the sample size, there are about 13~14 heterogeneous words per person in this survey. Furthermore, for a single respondent, the maximum proportion of heterogeneous words is 48.23%, and the minimum is 0.00%, indicating that the Mandarin level of respondents varies greatly.

Reflection of Heterogeneous Words

In terms of the specific reflection of heterogeneous words in pronunciation, vocabulary and grammar, the proportion is 82.53%, 14.36% and 3.11% respectively. As shown in Figure 1.

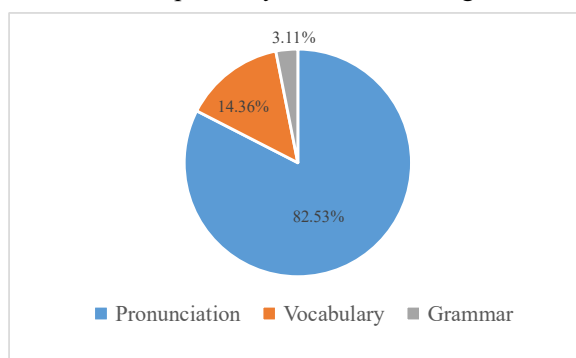


Figure 1. Distribution of Type of Heterogeneous Words.

As shown in Figure 1, the proportion of pronunciation variation in Mandarin mingled by Cantonese far exceeds that in vocabulary and grammar, which is consistent with the conclusion of mainstream research.

Part of Speech

Nouns in the heterogeneous words account for 43.33%, verbs accounted for 21.98%, adjectives accounted for 8.13%. As demonstrated in Figure 2.

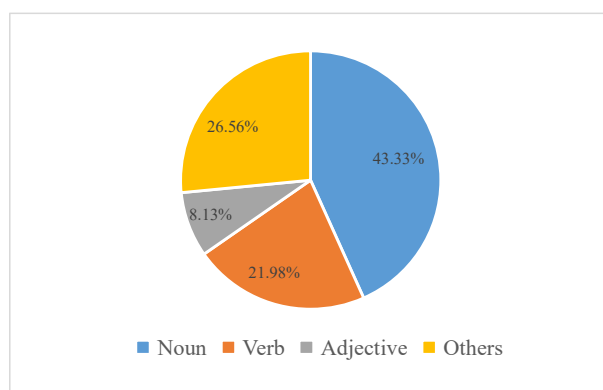


Figure 2. Distribution of Parts of Speech of Heterogeneous Words.

Figure 2 demonstrates that the proportion of noun variation in Mandarin mingled by Cantonese is significantly higher than that in other parts of speech. Few studies have made statistics on the part of speech of language variation. The above statistical results can provide a general understanding of the distribution of part of speech of language variation.

Individual Characteristics

Next, the variation of Mandarin was discussed from the perspective of speaker's individual characteristics, i.e. age, gender, education level, registered permanent residence. The full sample was divided according to this perspective, and the corresponding average ratio of heterogeneous words was calculated, as shown in Table 2.

Table 2

Average Ratio of Heterogeneous Words: Speaker's Individual Characteristics

Individual characteristics		Average ratio of heterogeneous words (%)
Age	≤30	7.39
	30~50	11.68
	≥50	21.20
Gender	Male	13.14
	Female	8.96
Education level	Secondary school and below	15.85
	Junior college and above	8.03
Registered permanent residence ¹	Rural area	16.22
	Urban area	7.67

Table 2 illustrates that, under the different dimensions of individual characteristics, the average ratio of heterogeneous words in Mandarin mingled by Cantonese present certain regularity. Firstly, the elder the speaker, the more susceptible their mandarin is to Cantonese, and the deeper the degree of language variation. Secondly, the average ratio of heterogeneous words of the female is lower than that of the male, which proves that the Mandarin level of female is higher than male on the whole, which may be related to the ability of the female to learn the language (Mandarin) (Hyde & Linn, 1988). Thirdly, it confirms that the better educated, the stronger the ability to use Mandarin. Language variation is more likely to occur in people with lower education level. Last but

¹ Shenzhen is the only city with a 100% urbanization rate, so all respondents are registered as urban residents.

not least, the average ratio of heterogeneous words in rural areas is much higher than that in urban areas, and the former is twice as much as the latter, due to the low education level of rural residents and the lack of opportunities to communicate with the outside world.

Geographical Distribution

In this paper, the full sample is divided into 21 sub samples according to the city dimension. Then the average ratios of heterogeneous words in Mandarin mingled by Cantonese of 21 cities are calculated, as shown in Figure 3.

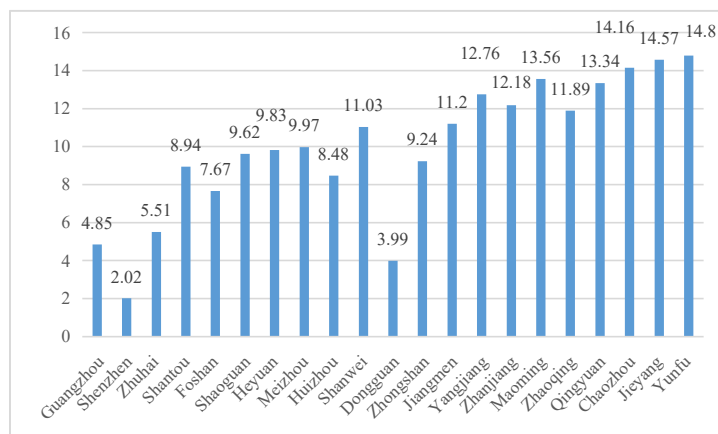


Figure 3. Average Ratio of Heterogeneous Words: City.

Figure 3 demonstrates that Shenzhen has the lowest ratio of heterogeneous words in Mandarin mingled by Cantonese, namely 2.02%, followed by Dongguan, about 4 heterogeneous words per 100 words. It suggests that these two cities have the least degree of language variation. Conversely, the city with the highest ratio of heterogeneous words (14.80%) is Yunfu, followed by Jieyang. That means that the degree of variation in these two cities is relatively deep. Apparently, among the 21 cities in Guangdong Province, the degree of language variation varies greatly. Yunfu, the city with the deepest degree, is nearly seven times than Shenzhen. This shows that the Mandarin level of the two cities is quite different. Intuitively, the macro level of economic development is an important factor. Shenzhen is China's special economic zone and the first-tier city; Dongguan is the second-tier city, once the "world manufacturing center", its developed economy has attracted a large number of migrant workers. The integration of population has further promoted the popularization of Mandarin. However, Yunfu, Jieyang and some cities are located in remote areas and many residents have little communication with the outside world, so the impact of Mandarin on Cantonese is relatively weak. Conversely, the traces of "Cantonese flavor" left in Mandarin are relatively more.

Conclusion and Discussion

Conclusion

Based on the investigations on the use of Mandarin by residents in 21 cities of Guangdong Province, China, this paper finds that the phenomenon of language variation is common in Cantonese speaking areas, namely Mandarin mingled by Cantonese. This phenomenon is mainly reflected in pronunciation, vocabulary and

grammar, especially in pronunciation, accounting for 82.53%. It is reflected in all kinds of parts of speech, but mainly reflected in nouns, verbs and adjectives, accounting for 73.44% in total.

Meanwhile, this phenomenon shows great differences in different groups and regions. Specifically, in terms of individual characteristics of the respondents, the elder and the less educated, the deeper the variation degree of Mandarin; the variation degree of male is slightly deeper than that of female; the variation ratio of Mandarin is much higher in rural areas than in urban areas. What's more, the degree of Mandarin mingled by Cantonese in Shenzhen and Dongguan is relatively weaker, while that in Yunfu and Jieyang is relatively deeper.

Discussion

The above statistical results have proven that there is a close correlation between the level of regional economic development and the degree of Mandarin mingled by Cantonese. Studies indicate that economic development is an important factor in language development (Dauenhauer & Dauenhauer, 1998). Generally speaking, the more developed the regional economy, the more effective the language promotion is. In order to reveal the relationship between regional economic scale and residents' Mandarin level, this paper collates the GDP ranking (amount from high to low) of each city in 2018, and compares it with the rank of average ratio of heterogeneous words in Mandarin mingled by Cantonese in each city in Figure 3 (from weak to deep), as shown in Figure 4.

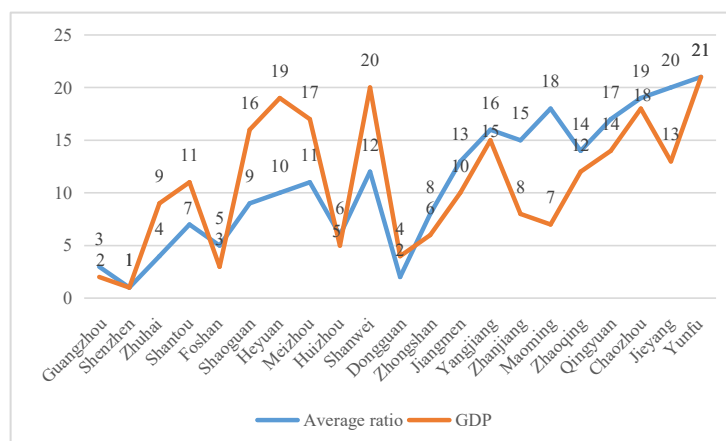


Figure 4. Comparison between Average Ratio of Heterogeneous Words in Different Cities and GDP Ranking

Note: Data of GDP is from the Guangdong Statistical Yearbook (2019).

Figure 4 indicates that the degree of Mandarin mingled by Cantonese in 21 cities is relatively convergent with the corresponding GDP, that is, the larger the economic scale of the city, the better the Mandarin level of residents, and the weaker the variation degree of Mandarin.

The divergence of the two curves indicates that the level of regional economic development is not the only determinant, other external factors such as government policy, geographical location also play important roles. From the perspective of micro level, internal factors such as age, gender, education level, registered permanent residence are important factors of language variation. Therefore, the variation of Mandarin is a qualitative change in the process of long-term language contact and integration, which is caused by a series of internal and external factors.

References

- Anttila, A. (2002). Variation and phonological theory. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 206-243). Oxford: Blackwell.
- Biber, D., & Burges, J. (2000). Historical change in the language use of women and men: Gender differences in dramatic dialogue. *Journal of English Linguistics*, 2(1), 21-37. <https://doi.org/10.1177/00754240022004857>.
- Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of multicultural London English. *Journal of Sociolinguist*, 15(2), 151-196. <https://doi.org/10.1111/j.1467-9841.2011.00478.x>.
- Claire, N. (2015). Intonational variation and change in Scottish Gaelic. *Lingua*, 160, 1-19. <https://doi.org/10.1016/j.langsci.2016.02.005>.
- Dauenhauer, M., & Dauenhauer, R. (1998). Technical, emotional and ideological issues in reversing language shift: Examples from southeast Alaska. In L. A. Grenoble & L. J. Whaley (Eds.), *Endangered languages: Language loss and community response* (pp. 57-98). Cambridge: Cambridge University Press.
- Hartmann, R. K., & Stork, F. C. (1972). *Dictionary of language and linguistics*. London: Applied Science Publishers.
- Hyde, S. J., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. Paper presented at the *Annual Meeting of the American Educational Research Association* (New Orleans, LA) (pp. 5-9).
- Kimberly, A. (2014). Language variation and ethnic identity: A social psychological perspective. *Language & Communication*, 35, 88-96. <https://doi.org/10.1016/j.langcom.2013.12.001>.
- Labov, W. (1963). *The social history of a sound change on the Island of Martha's Vineyard*. Massachusetts.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington. D. C. Center for Applied Linguistics.
- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press.
- Li, J. (2019a). Dialect attitude, dialect environment and dialect degradation: Evidence from Hukou dialect in China. In *3rd International Conference on Culture, Education and Economic Development of Modern Society* (ICCESE) (pp. 11-14). Atlantis Press.
- Li, J. (2019b). Investigation analysis of "dialect mingled by Mandarin" phenomenon: Evidence from Hukou dialect in China. In *4th International Conference on Contemporary Education, Social Sciences and Humanities* (ICCESSH) (pp. 1064-1067). Atlantis Press.
- Li, J. (2020a). A study on the differences of Mandarin mingled by Cantonese in the Pearl River Delta of China. *International Journal of Social Science, Innovation and Educational Technologies*, 1(3), 316-323.
- Li, J. (2020b). Distribution of Mandarin mingled by Cantonese: Survey based on Pearl River Delta in China. *Journal of Literature and Art Studies*, 10(7), 100-108. <https://doi.org/10.17265/2159-5836>.
- Liang, S. (2015). *Language attitudes and identities in multilingual China: A linguistic ethnography*. Cham, Heidelberg, New York, Dordrecht, London: Springer.
- Liang, Y. (2018). Merger and transfer: Tone variation and change of Dongguan Cantonese. *Lingua*, 208, 19-30. <https://doi.org/10.1016/j.lingua.2018.03.003>.
- Oldin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Shanghai: Shanghai Foreign Language Education Press.
- Patrick, N. (2017). *A study of attitudes of dialect speakers towards the speak Mandarin campaign in Singapore*. Springer Singapore.
- Rey, J. M. (2001). Changing gender roles in popular culture: Dialogue in Star Trek Episodes from 1966 to 1993. In S. Conrad, & D. Bibier (Eds.), *Variation in English: Multi-dimensional studies*. Essex: Pearson Education Limited.
- Seong, L. D., Chong-chieh, W., & Kim, L. G. (2019). In quest of a new identity? Language variation in Sabah. *Lingua*, 227, 102703. <https://doi.org/10.1016/j.lingua.2019.06.004>.
- Thomason, S. G. (2001). *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason, S. G. (2008). Social and linguistic factors as predictors of contact-induced change. *Journal of Language Contact*, 2(1), 42-56. <https://doi.org/10.1163/000000008792525381>.
- Thomason, S. G. (2009). *Contact explanations in linguistics*. Malden: Blackwell.
- Trudgill, P. (1974). *The social differentiation of English in Norwich*. London: Cambridge University Press.
- Trudgill, P. (1983). *Sociolinguistics: An introduction to language and society*. Penguin Books.