



Orthographic variation < Macau>/<Macao > and Macau identity

Trin Qingyi Chen & Brian Hok-Shing Chan

To cite this article: Trin Qingyi Chen & Brian Hok-Shing Chan (2024) Orthographic variation < Macau>/<Macao > and Macau identity, Cogent Arts & Humanities, 11:1, 2360182, DOI: [10.1080/23311983.2024.2360182](https://doi.org/10.1080/23311983.2024.2360182)

To link to this article: <https://doi.org/10.1080/23311983.2024.2360182>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 31 May 2024.



Submit your article to this journal [↗](#)





View related articles [↗](#)



View Crossmark data [↗](#)

Orthographic variation <Macau>/<Macao> and Macau identity

Trin Qingyi Chen  and Brian Hok-Shing Chan 

Department of English, University of Macau, Av. da Universidade, Taipa, Macau, China

ABSTRACT

This study examines the orthographic variation between the place name <Macau> and <Macao> in English texts in Macau and its interaction with social factors. Using Sebba's framework of Orthographic Regimes, a statistical analysis was conducted to explore the usage of the two orthographic variants across three levels of orthographic regimes, namely, highly-regulated (HR) government writings, less-regulated (LR) school writings and lesser-regulated (LRR) online texts. As a means to collect data of LR regimes, a questionnaire was used to collect participants' identities and elicit potential indexical relationships between the two variants and aspects of Macau identity, which we see as two place identities. Results show a strong connection between orthographic regimes and variant usage, with <Macao> preferred in HR government writing while Macau is more commonly used in LR regimes where <Macao> is not required. Furthermore, this study discusses a connection between language practices and Macau identity based on orders of indexicality, that is, the use of <Macau>/<Macao> and the duality of Macau identity (i.e. a nationalistic identity vs. an ethnocultural identity).

ARTICLE HISTORY

Received 19 March 2024
Revised 9 May 2024
Accepted 22 May 2024

KEYWORDS

Sociolinguistics;
orthographic variation;
Macau; spelling;
indexicality; place identity

REVIEWING EDITOR

Jeroen van de Weijer,
Shenzhen University,
Shenzhen, Guangdong,
China

SUBJECTS

Applied Linguistics;
Sociolinguistics

1. Introduction

Orthographic variation, which can be viewed as visual representations of phonetic variation (e.g. <wuz> for *was*) or purely graphemic modifications (e.g. <kool> in *cool*) (Androutsopoulos, 2000: 514), is typically understood within the binary framework of standard/non-standard spellings. The convention of orthography ties it inseparably to the concept of 'correct spellings', as standard spellings do exist. Correct spelling is generally perceived as a sign of compliance with the standards and principles of the prevailing culture (Augst, 1988).

Previous studies have highlighted a dual role for non-standard orthography in informal discourse, namely, distancing and closeness. Distancing refers to its use in discourses that oppose standard language and associated ideologies, as seen in punk fanzines (Androutsopoulos, 2000), Italian hip-hop writing (Romiti, 1998), or British-Jamaican creole (Sebba, 1998). Closeness, on the other hand, aims to foster intimacy between the writer and the reader (e.g. Bell, 1984; Davies, 1987). Thus, non-standard orthography can symbolize a departure from the dominant culture. In certain contexts (like media discourse), these unconventional spellings are used deliberately to appeal to a specific audience, fostering a sense of shared group identity or enhancing the writer-reader relationship.

In this perspective, orthographic variations potentially reflect regional and stylistic differences (e.g. Eisenstein, 2015; Ilbury, 2020) and serve as indicators of various identities (e.g. the use of 'written dialect' in fiction to index the low social status of a character), or signify a language user's alignment with a particular social identity.

Sociolinguistic research has explored how orthographic variation can index different identities in publishing media, subculture posters, SMS and online social media platforms. For example, Heffernan et al. (2010) found that Canadian print media's use of American (AmE) or British (BrE) spellings reflected national sentiments. Androutsopoulos (2000) showed how German punk fanzine writers use non-standard

CONTACT Trin Qingyi Chen  trin.chen@connect.um.edu.mo  Department of English, University of Macau, Av. da Universidade, Taipa, Macau, China

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

spellings for displaying music-related subcultural identity. Similarly, Tagg (2016) demonstrated how texters use linguistic resources, including non-standard spellings, to convey meanings and emotions in texted interactions.

Indeed, the notion of a 'standard' spelling, which is generally accepted as correct and widely used, may not be able to fully capture all linguistic practices. It is entirely possible for multiple spellings of a word to be recognized and utilized by people (e.g. AmE and BrE spellings in Canadian print media (Heffernan et al., 2010), <bluish> vs. <blueish> or <Macau> vs. <Macao> in the current research), adding a layer of complexity that challenges the simple typology of standard vs. non-standard spellings.

This study delves into the link between orthographic variation and identity across writing discourse of various registers. It also scrutinizes the distribution of commonly accepted (not necessarily standard) spelling variants in different types of texts, aiming to comprehend how indexical value might influence people's spelling choices. Drawing on the framework of *orthography regimes* (Sebba, 2007:46–47) and the understanding that language choice is steeped in sociocultural meanings, this article zeroes in on the orthographic variation <Macau>/<Macao>. This pair of spellings, with a complex and disputed history, is used interchangeably to denote the Macao Special Administrative Region (MSAR) (Moody, 2021:20). Given their broad acceptance and unclear status, neither variant is definitively considered standard. Based on a statistical analysis, this article suggests that these spelling variants act as semiotic tools for constructing two facets of Macau identity – which refers to two different views or perceptions of Macau rather than two different groups of people in Macau – in other words, two *place identities* (Peng et al., 2020) of Macau, in our opinion. By investigating this unique case where both spellings are prevalent in institutional and individual discourse, this article offers a viewpoint on how popular orthographic variation can index two *place identities* (Peng et al., 2020) of the same place. Positioned within this context, this article also highlights the significance of treating orthographic variation (purely graphemically modified) as a sociolinguistic topic in its own right and contributes to the sociocultural study of orthography by providing some new data. Moreover, it challenges the binary typology of standard/non-standard, and demonstrates the relevance and utility of quantitative methods in exploring sociolinguistic issues such as identity and language policy, which are often approached with qualitative or ethnographic methods.

This article is organized in five sections. In the next section, we delve into the naming and meaning of <Macau>/<Macao>, as well as the language situation and identity formation in Macau. The methodology section outlines the data sources and methods used in this research, including a self-built corpus of Macau written English texts and its three sub-corpora corresponding to three levels of orthographic regimes (Sebba, 2007:47), which encompass both institutional and individual texts (see Section 3). In the data analysis section, we present the results of our corpus analysis, focusing on the distribution and correlation of the two spelling variants across different orthography regimes and social factors. In the discussion section, we interpret our findings based on orders of indexicality (Eckert, 2008; Johnstone, Andrus & Danielson, 2006; Silverstein, 2003), orthography regimes (Sebba, 2007:46–47) and aspects of Macau identity (Kaeding, 2010; Lam, 2010), discussing how <Macau> and <Macao> may index different conceptions and perceptions of Macau identity in which Macau is seen as a nationalistic Special Administrative Region or a former colony with unique, blended cultural heritage. In this article, we use <Macau> to refer to the region and maintain original spellings from cited resources.

2. <Macao/Macau>: the naming and the meanings

The terms <Macau> and <Macao> are widely accepted orthographic variants used interchangeably to refer to the MSAR (Moody, 2021:20). English is employed in MSAR as a language for inter-group and inter-cultural communication (Young, 2006:480), an inter-ethnic lingua franca (Mann & Wong, 1999:20), and a *de facto* official language in government and education (Moody, 2008:4–8), despite not being one of the *de jure* official languages, which are Chinese and Portuguese.

A common belief among many people in Macau is that <Macau> and <Macao> are Portuguese and English, respectively (Moody, 2008:13). However, this assumption lacks linguistic and historical evidence. Wu and Jin's (2013) study on the evolution of the two spellings concluded that both <Macao> and <Macau> are likely to have originated from the Portuguese language, which was based on an examination of early archival materials in Portuguese and other languages. Specifically, both spellings appeared in

the archives prior to the Portuguese Orthography Reform in the 19th century, which changed the spelling to <Macau> (Bray & Koo, 2005:10). However, the original spelling <Macao> has been preserved in English (Bray & Koo, 2005:10). Following Macau's return of sovereignty to China in 1999, Chinese authorities declared <Macao> as the official spelling (Bruning, 2001, as cited in Bray & Koo, 2005:10). Furthermore, according to Bray and Koo (2005:10), the MSAR government issued an internal circular instructing the use of <Macao> in English government publications/websites, implementing a policy of standardization which requires people to use <Macao> in official documents. This raises a question: why is <Macau> still used in English if people have already associated <Macau> with Portuguese, and <Macao> with English?

The prevalent use of <Macau> in English implies that this variant may be driven by certain sociocultural factors rather than being in the inventory of Portuguese vocabulary itself. This is similar to the instance of <x> or <z> replacing <s> as in <boyz>, where spelling variants are primarily motivated by their symbolic values as cues of different identity/subculture positioning (Androutsopoulos, 2000:527). In other words, these grapheme alternations are visual reminders to engage the reader with a subculture with sociocultural factors (e.g. subculture of German punk fans).

Over the past two decades, scholars have explored potential factors for the choice of <Macao>. Some factors that have been suggested including register, more specifically, governmental texts (Bray & Koo, 2005:10; Moody, 2008:13), institutional language preferences (Moody, 2021:21), language standardization attempts by the People's Republic of China, or PRC (Moody, 2021:21), a governance change from Portugal to PRC (Moody, 2021:21), and indications of non-local user origin (Bray & Koo, 2005:10). It is noteworthy that since Macau's return to China, <Macau> appears to have been quickly replaced with <Macao> in governmental discourse over the past decades. This prompted Bray and Koo (2005) to update all instances of spelling to <Macao> in their book about the territory. Overall, the use of <Macao> might be related to governmental, institutional texts and an intent to show alienation from the local identity, whereas <Macau> could index non-institutional discourse and engagement with the local community.

Previous research on orthographic variants primarily relies on the standard/non-standard typology. However, for <Macau>/<Macao>, categorizing either spelling as 'standard' is challenging, given their widespread use across all written discourses in Macau. Consequently, it becomes necessary to explore a different framework.

It is not a happenstance for orthographic variants to aggregate in certain types of written discourse like social media, text messages, fan culture or graffiti. There are also published texts like newspapers or official documents where one can rarely find variation in spelling as they are policed and proofread. Sebba (2007:43) stated that orthography is under various degrees of regulation, ranging from highly regulated orthography, where words have standard spellings imposed as in public, official documents, to unregulated orthography like graffiti. It is also possible that some variants are widely accepted within certain communities, but not necessarily standardized.

The indexical function of spelling choices is based on their contrast to a neutral orthographic context (Lakoff, 1982). Under the typology of standard/non-standard spelling, this unmarked orthographic context is typically the standard spelling, as it is often taken as the expected variant. After all, any signs of deviation from the 'correct' spelling could be understood as non-standard, irrespective of all linguistic facts. However, when spelling variants are both accepted and widely used, it becomes challenging to categorize them under the typology. This situation prompts the question: what serves as the contrast for such variants?

Such variants are considered in the study of Le Page and Tabouret-Keller (1985:115–116). Focussing, in their account, implies a degree of regularity in the linguistic code; the greater the regularity, the less variability it has. A focussed system is one that exhibits relatively little variability, but it does not automatically qualify as a standardized system (Sebba, 2007:46) (e.g. <Macau> is now not the officially recognized variant while being accepted as correct and widely used by people in Macau).

Combining the notions of focusing and degree of regulation, Sebba (2007:44–47) put forward the framework of orthographic regimes to explain the relationship between orthographic variations and text types. It is the writing type that allows variation on spellings, either focussed or non-focussed. It is possible for a group of language users to adopt certain non-standard orthographic variants (Sebba, 2007:46), hence focussed and non-standard.

The case of <Macau>/<Macao> seems amenable to this framework, as the use of the two spellings are distributed in all written discourses throughout Macau. By looking at different levels of orthography regimes, we are able to evaluate how the two spellings are being used under different levels of regulation and its degree of regularity (being focussed). Table 1 is adapted from Sebba (2007:47), including three levels of orthography regimes where the two spellings are studied.

3. Macau Written English Corpus

This section describes the methods used in this research, namely, statistical analysis of a self-built corpus and a questionnaire. The aim of this research is to examine the distribution of orthographic variation <Macao>/<Macau> at different levels of orthographic regimes and to elicit its potential correlation with broad social categories.

To provide text data of the orthographic variation of <Macao>/<Macau>, the most direct way of displaying the distribution of the two variants is *via* Macau Written English Corpus (MVEC), a self-built corpus for this current study. In order to select local Macau English texts to represent different levels of orthography regimes, data need to be collected from various sources.

Texts selected for MVEC mainly consist of three sub-corpora that represent three levels of orthography regimes, namely, a sub-corpus of highly-regulated (HR) government news from the MSAR Government Information Bureau (GIB) website (HR-GIB), a small portion of less-regulated school writings from a questionnaire (LR-school writings), and a small portion of lesser-regulated online tweets from the social media platform Twitter (LRR-online texts). By analyzing three levels of regimes and supporting data with the statistical instruments of SPSS (IBM Corp, 2021, version 28.0.1) (Armonk, NY) and R (; R Core Team, 2023), the occurrence of the two variants in Macau in different orthography regimes can be generated (see Table 2).

The first type of texts to include in the MVEC is from the HR regime. Although English is not one of the official languages in Macau, GIB has been providing news in English since December 2000. The GIB website (<https://gcs.gov.mo>) offers a public archive of those news ranging from 2000 to 2016. The texts are directly drawn from the MSAR government and thus they are considered formal and official. The nature of the texts from this website archive matches what Sebba (2007:47) described as HR regime; consequently, MVEC includes raw English texts from the GIB website. However, the encryption of GIB website refuses the use of command line to fetch text data, so all the texts need to be downloaded manually. It is impossible to manually download all the texts from the GIB website due to the enormous amount of news it provides. MVEC, therefore, includes the news articles for a single year in every fifth year from 2000 to 2016 (i.e. articles in 2001, 2006, 2011 and 2016). This MSAR news sub-corpus includes 147,494 tokens in total.

The second type of texts comes from Twitter. We utilized Twitter developer API and TwitteR package in R to collect data (Gentry, 2015; R Core Team, 2023). The data were collected *via* a filtered search including tweets containing tokens of <MACAU> or <MACAO>. To ensure that tweets are from Macau, the tweet location filter has been set to Macau only. The language of tweets has been set to English. The time of the tweets has been set from 2020 to 2023. Data of profile location are also collected and

Table 1. Orthographic regimes for different types of texts (Adapted from Sebba, 2007:47).

| Regime | Writing types | Institutional order | Readership | Orthography |
|-----------------------|---------------------------------------|---------------------------------------|----------------|--|
| Most highly regulated | Texts for publication/ circulation | government, publishing, journalism | General public | Focussed, on standard |
| Less regulated | 'school' writing | School | Teachers | Usu. focussed, not always on standard |
| Lesser regulated | Social media | Not institutional | Self/in-group | Least focussed zone |

Table 2. Number of tokens and hits of <MACAO>/<MACAU> in each subcorpus.

| Sub-corpora | Number of Tokens | Number of <MACAO>/<MACAU> (%) |
|--------------------------|------------------|-------------------------------|
| HR (government releases) | 147494 | 1637/993 (62/38) |
| LR (student writings) | 1458 | 16/54 (23/77) |
| LRR (online tweets) | 3271 | 45/106 (30/70) |

treated as a self-reported *place identity* (Peng et al., 2020, as it can be seen as users' voice to show how they perceive Macau as, or the degree of affinity/attachment to Macau) to compare with data from school writings (see more detail below). With a timespan of three years, a total of 151 tweets (3271 tokens) by randomly sampled individual users based in Macau are included in the MWEC to represent online texts of a LRR regime. A smaller number of tokens are expected here because the mining process only includes layers of filters, including location, language and tweets with tokens 'MACAU/MACAO'. Variants in each tweet will be counted only once.

The third type of texts was collected via a bilingual in-class questionnaire to collect undergraduate students' information about gender, place of origin, identity to a place (how students see themselves as in relation to Macau, which in return, show us in what ways these people value Macau), and daily language, their use of the two spellings in different text types (after choosing the variants), their awareness of differences between the two spellings in terms of meaning and text types (institutional vs. individual). This 11-item questionnaire includes ten alternative-answer questions and one open-response question which asks the participants to describe what they appreciate or love about Macau and Macau's cultural/historical heritage in one to two sentences. These extra texts produced in an academic setting enrich the existing MWEC. To provide a clear emphasis, the questionnaire targeted young generations in Macau, investigating their use of the two spellings. Through this questionnaire, this study shows how students in Macau would use the two spellings; furthermore, by constantly comparing students' language use and their social information, the study can draw connection between the use of spellings with rather broad social categories. Question items and options containing <Macau> might prime students to spell the <Macau> variant. In order to lower the priming effect, spelling used in question items were presented as <Macau>/<Macao> or <Macao>/<Macau> in the open-response question. Eighty-five students aged from 18 to 22, in the University of Macau (UM),¹ including 43 females and 42 males, were invited to fill out the questionnaire, and they were required to use the word <Macao>/<Macau> in their writings. Seventy valid responses ($N=70$) were eventually elicited. Considering that the population of Macau is 686,600 (DSEC, 2022), the sample size level is above per ten thousand (1.02 per thousand). However, high-literacy participants may not be able to accurately portrait Macau's population. As the setting of the questionnaire is in-class and semi-formal, the texts collected can be considered as school writings.

The students were undergraduates taking University English I (UE I) and University English II (UE II) at UM when the questionnaire was conducted. English is the sole medium of instruction for these courses. Students' English proficiency is at CEFR-A2 level or above. Due to practical constraints, convenience sampling was applied here to develop an initial understanding of the use of spellings by younger Macau students. The majority of students from UM are Macau or mainland residents. As there are more than one question items that target at the same variable, participants whose answers contradict themselves were considered invalid and their responses were disregarded.

The questionnaire was designed and distributed via an online data collection and analysis platform Qualtrics. To ensure that participants are close in age and levels of education, all the questionnaires were distributed in-class *via* QR code to undergraduate students who were taking the courses of UE I and UE II provided by the English Language Centre at UM in the academic year of 2022/2023. Students could use their electronic devices to answer the questionnaire. The questionnaire was distributed to six classes, more specifically, three classes of UE I and three classes of UE II.

Data cleaning process was imposed to avoid confounding data, given that there are many addresses in Portuguese, proper nouns that contain the spelling of <Macau>, and individual Twitter accounts that send Tweets on behalf of an institution (e.g. tweets from local journalists). By using the feature of 'N-Gram' and 'Collocate' on AntConc (2023, Version 4.0), this study located common expressions and most frequent word combinations (e.g. for the 2016 texts from GIB, there are 62 tokens of <MACAU> and 57 of them are parts of names of institutions or Portuguese addresses). After scanning, all tokens of Portuguese addresses and proper nouns were located and disregarded for the statistical analysis in the later part. Tweets were manually reviewed, and all institutional tweets were removed.

The MWEC is designed to present a comprehensive view of the usage of the two spelling variants under different levels of orthography regimes. However, it is unbalanced due to practical constraints. The current research cannot access corpus data of equivalent amount to represent each level of orthography regimes. Consequently, the first type of texts takes up a major portion of the MWEC. Also, MWEC cannot

ensure all texts in sub-corpora are from the same period of time. For tweets and school writings, they are only complementary to the HR news discourse to show the current use of the two orthographic variants by individuals in Macau.

Using the statistical tool SPSS (IBM Corp, 2021, version 28.0.1) (Armonk, NY), this study runs Chi-square tests of independence (χ^2) to gauge correlation between two categorical variables.

4. Data analysis

This section presents the results of a statistical analysis of data from both the MWECC corpus and the questionnaire. It is important to note that several categories were used to classify the independent variables (external linguistic variables) in order to establish a broad connection between orthographic practices in Macau at the initial stage of discovery. The MWECC corpus was constructed according to different orthography regimes, and other external linguistic variables were also collected during its construction. The HR-GIB sub-corpus provides data spanning over 15 years while the online text sub-corpus includes information on the time of posts and users' region (considered a self-reported identity). The questionnaire data covers gender, languages used for daily communication, language proficiency, and self-reported place identity. For χ^2 , significance levels have been set at $p < .05$.

By combining data from the three sub-corpora, we arrived at the overall distribution of the two spelling variants. It is noteworthy that while the three sub-corpora vary in size, both spelling variants are used in Macau, see Table 3. Overall, the figures do not show a significant difference in the usage of the two spelling variants. This is expected as most instances of <Macao> come from the largest sub-corpus (i.e. HR-GIB).

When the data are organized according to the three different orthography regimes represented by each sub-corpus – HR Regimes of government writings, LR regime of school writings, and LRR regime of online text – significant differences in the usage of the two orthographic variants can be observed. Table 3 presents the distribution of the two spelling variants according to these orthography regimes. Data representing the HR is only from the year of 2016 to reflect the latest distribution.

In Table 3, the distribution of <Macau> and <Macao> differs significantly between HR and LR/LRR. A chi-square test of independence was performed to evaluate the relationship between the orthographic variables <Macau>/<Macao> and orthography regimes. The relationship between the frequencies of the two spelling variants at all three levels of orthography regimes was found to be significant at a level of $p < .05$ with a p value of $<.00001$, indicating that there is a significant relationship between orthography regimes and spelling variants, i.e. <Macao> is more likely to appear in HR, whereas <Macau> is more likely to be used in both LR and LRR. In fact, <Macau> is almost never used in HR, while in LR and LRR where the use of <Macao> is not 'required' (see Section 2), most instances are spelled as <Macau>. This significant difference across orthography regimes suggests a strong connection between linguistic variables (<Macau> and <Macao>) and orthography regimes, warranting further investigation into each regime.

4.1. Highly-regulated regime

The distribution of the two orthographic variants in a HR regime (see Table 4) can be observed through the sub-corpus of governmental writings. In the GIB archive, the spelling <Macao> appears 1637 times while <Macau> appears 1251 times, a total of 2888 tokens (1.95% of the overall corpus) containing either spelling. However, due to the nature of the GIB archive, this data does not provide any additional information of social categorizations.

Table 3. Frequency of <Macau>/<Macao> according to orthography regimes.

| Orthography regimes | <Macao> | | <Macau> | |
|---------------------|-----------|----|-----------|----|
| | Frequency | % | Frequency | % |
| HR | 518 | 99 | 5 | 1 |
| LR | 16 | 23 | 54 | 77 |
| LRR | 45 | 30 | 106 | 70 |

Table 4. Longitudinal distribution of <Macau>/<Macao> in HR-GIB.

| HR-GIB | <Macau> | | <Macao> | |
|--------|-----------|----|-----------|----|
| | Frequency | % | Frequency | % |
| 2001 | 259 | 24 | 819 | 76 |
| 2006 | 348 | 73 | 131 | 27 |
| 2011 | 512 | 93 | 38 | 7 |
| 2016 | 518 | 99 | 5 | 1 |

As in Table 4, the use of <Macao> in GIB has increased over time. In a HR orthography regime, <Macao> is used in 62% (1637) of instances while <Macau> accounts for only 38% (993) of instances. Notably, there is a sharp decline in the use of <Macau> and a corresponding increase in <Macao> after 2001. In 2016, <Macao> accounted for 99% of instances while <Macau> made up only 1%. This suggests that <Macao> is becoming the only recognized English spelling in government writings. The data from 2011 and 2016 provide clear evidence that the use of <Macao> reflects a focussed and standardized orthography in HR.

4.2. Less-regulated regimes

To analyze the distribution of the two spelling variants in a LR regime, we delve into the two LR regimes. Both spelling variants are used in LR-school writings, indicating that orthography in this regime is less on standard and less focussed. Nonetheless, <Macau> (47, 67%) is used more frequently than <Macao> (33, 33%). Sorting the data by gender does not yield any significant results. However, when the data are reorganized according to participants' region and self-reported identity of a place, interesting patterns emerge, as shown in Table 5. Identities of participants from region of Macau (MO) and Mainland China (CN) are presented. Data from participants identifying themselves as Hong Kong people is too sparse and therefore excluded from the table.

Of the 41 participants from Macau, all identified themselves as Macau people and 9 also identified as Chinese (M+C). Among those identifying themselves as Macau people, 31 (76%) used <Macau> while only 10 (24%) used <Macao>. It is noteworthy that the majority of participants identifying themselves with the group of Macau people predominantly used <Macau> while only a minority used <Macao>. Of the nine participants who identified as both Macau people and Chinese, four (44%) used <Macao>, accounting for 40% of its total frequency. Additionally, there were twenty-four participants from Mainland China (CN) who all identified as Chinese. In this group, <Macao> was used slightly more frequently than <Macau> (thirteen vs. eleven). These results suggest that the increased use of <Macao> may be related to Chinese identity.

The distribution of orthographic variants in a LRR regime can be observed through the sub-corpus of online text from Twitter. A total of 151 tweets from 2020 to 2023 (3271 tokens) were included in the analysis.

The texts of tweets were included in the analysis to provide complementary data for the GIB sub-corpus. Despite the smaller size of this dataset compared to the GIB sub-corpus, a significant shift in spelling preference was observed within the LRR regime of Twitter. In this context, the <Macau> spelling accounted for 106 hits (70%), while the <Macao> spelling had only 45 entries (30%). It is notable that during this period both spellings were used within this LRR regime where orthography is less focussed (described as 'less focussed zone' – Sebba, 2007). Furthermore, a notable difference in the frequency of <Macau> was observed between online and governmental writings. In online texts, where orthography is less focussed, <Macau> accounted for 70% of usage compared to just 1% in HR government writing. This marked contrast suggests that changes in orthographic regimes significantly impacts on the use of these two spellings. Statistical evidence supporting this observation is found in Table 3.

Online texts from Twitter were also coded according to the user's region. The user's region refers to the location displayed on their Twitter personal profile, which can be selected by users and may be considered a self-reported identity. Four main regions were identified in the data, namely, Macau (MO), Mainland China (CN), Hong Kong (HK) and Others.

Table 5. Frequency of <Macau>/<Macao> in LR according to region and identity.

| Identity | MO* (N=41) | | CN (N=24) | |
|--------------|-------------|-------------|-------------|-------------|
| | <Macau> (%) | <Macao> (%) | <Macau> (%) | <Macao> (%) |
| Macau people | 26 (81) | 6 (19) | – | – |
| M+C | 5 (54) | 4 (44) | – | – |
| Chinese | – | – | 11 (46) | 13 (54) |

*In MO region, the relationship between identity and linguistic variables was not found to be statistically significant due to the small sample size ($p=112786$). However, if the sample size were to be increased, it is possible that these results could become significant at a level of $p < .05$ or less.

Table 6. Frequency of <Macau>/<Macao> in Twitter according to region.

| Regions | <Macao> | | <Macau> | |
|---------|-----------|----|-----------|----|
| | Frequency | % | Frequency | % |
| MO | 22 | 19 | 96 | 81 |
| CN | 18 | 82 | 4 | 18 |
| HK | 2 | 33 | 4 | 67 |
| Others | 3 | 60 | 2 | 40 |

Table 7. Frequency of <Macau>/<Macao> in Twitter according to self-reported region and orthography regime.

| Orthography regimes | MO* | | CN* | |
|---------------------|-------------|-------------|-------------|-------------|
| | <Macau> (%) | <Macao> (%) | <Macau> (%) | <Macao> (%) |
| LR | 31 (76) | 10 (24) | 11 (46) | 13 (54) |
| LRR | 96 (81) | 22 (19) | 4 (18) | 18 (82) |

* $p < .05$.

When the variable of users' self-reported identification to place (i.e. region) was considered, several significant patterns emerged. Table 6 presents the frequency of the two spellings according to users' region. This data provide further insight into the relationship between orthographic variation and social factors.

A chi-square test of independence was performed to evaluate the relationship between users' self-reported region and the linguistic variable <Macau>/<Macao>. The results were significant at $p < .05$ with a p value < 0.00001 . This indicates that users from the Macau (MO) region are more likely to use <Macau> in their tweets, while users from Mainland China (CN) are more likely to use <Macao>. When examining the distribution of the two orthographic variants by user's region, as shown in Table 6, we find that most tokens in either spelling (94 and 22 hits, respectively) come from users whose location is reported as Macau. Among these users, <Macau> is overwhelmingly preferred (81%). Interestingly, among users whose location is reported as Mainland China, <Macao> is predominantly used (82%).

Given the significant difference between MO and CN regions in their usage of these two spellings, further analysis was conducted to examine how these variants are distributed according to region within LR orthography regimes. Table 7 presents this data according to levels of orthography regime and region for MO and CN.

As Macau (MO) and Mainland China (CN) are the primary foci of the analysis and there are sufficient data for a chi-square test, only figures from these two groups were considered in the statistical test. A chi-square test of independence was therefore performed to evaluate the relationship between the linguistic variables <Macau> and <Macao> and self-reported regions. The results are significant at $p < .05$ for both variables (with p values of .015397 and < 0.00001 , respectively). This indicates that within LR and LRR orthography regimes, <Macau> is more likely to be used by individuals who self-reported as being from Macau. Conversely, individuals who self-reported as being from Mainland China are more likely to use <Macao>. Table 7 shows that overall usage patterns within these two LR levels of orthography regime are similar: both variants are used, with <Macau> being the focussed variant among people from Macau. However, among people from Mainland China, usage of <Macao> increases, surpassing 50% within this group.

In conclusion, the analysis of the distribution of <Macau> to <Macao> reveals that both spellings are used in Macau. When data are sorted according to orthography regime, a marked difference in the distribution of these variants is observed. At the HR level, the spelling preference of the MSAR government

has shifted from <Macau> to <Macao> since the handover in 1999. In recent years (2011 and 2016), orthography has become extremely focussed with almost all instances being spelled as <Macao> to align with the PRC's official English spelling of MSAR.

As we move down to LR and LRR, a significant increase in usage of <Macau> is observed. This reverses the trend witnessed at the HR level. Within these LR regimes, most people in Macau use <Macau>.

This dual pattern supports Sebba's (2007) framework of Orthography Regimes: at HR levels orthography is more focussed (i.e. only <Macao> is used), while at lower levels it becomes relatively less focussed (i.e. both <Macau> and <Macao> are used).

In LR and LRR where <Macao> is not required, analysis according to social categorizations (i.e. self-reported place identity and region) suggests correlations between the use of the two variants and language users' region and place identity. Specifically, individuals who identify themselves as Macau people or who self-report their Twitter profile location as Macau are more likely to use <Macau>. Conversely, individuals who identify themselves as Chinese or from Mainland China are more likely to use <Macao>. These results support discussions about the indications of <Macao> summarized in Bray and Koo (2005) and Moody (2008, 2021). Specifically, data from the HR-GIB sub-corpus indicates that usage of <Macao> is associated with official government writing. Data from LR and LRR sub-corpora further suggests that the usage of <Macao> may be somehow correlated with a non-local identity.

5. Discussion

The results of statistical analysis show two distinct patterns of orthographic variation. Specifically, <Macao> is becoming the only variant used in HR while both <Macao> and <Macau> are used in LR and LRR with a preference for the latter. Furthermore, the data from LR and LRR indicates an indexical connection between the frequency of <Macau> usage and local identity as Macau people (i.e. people whose self-reported identity includes 'Macau people'). Conversely, the frequency of <Macao> usage is correlated with national identity as Chinese (i.e. people whose self-reported identity includes 'Chinese'). This connection seems to tally with two aspects of Macau identity discussed by scholars (e.g. Kaeding, 2010; Lam, 2010).

5.1. Duality in Macau identity and orthography

Macau's identity is a complex issue due to its diverse population and strong ties to China, leading to an ambiguous local identity (Chan, 2015; Yan & Moody, 2010). Macau's identity is a hybrid construct of national (i.e. a belonging to China), local (i.e. an ethno-cultural local identity defined by pride in Macau's distinct cultural heritage), and international components (see Kaeding, 2010:134; Lam, 2010:666–668). Prior to the 1999 handover, Macau was seen as a nationalistic society due to its elder population's connections with their Chinese roots. Post-handover, both MSAR and the Chinese government have encouraged the development of a local Macau identity by promoting pride in its colonial and cultural heritage (Lam, 2010).

Studies show a strong sense of local identity among younger generations in Macau. A study by Leong (2009) found that more than half of the participants identified primarily as Macao people. A survey by Kaeding (2010) suggested an emerging local identity among younger generations, more ethno-cultural in nature than civic. However, Wang's (2017) study contradicted Kaeding's prediction that most Macau people would continue to identify with China, finding that most local college students did not identify with China. Despite some discrepancies, it is clear that a local identity has emerged among younger generations in Macau, closely tied to ethno-cultural factors, particularly Macau's colonial Portuguese heritage.

Language significantly influences Macau's identity, with Cantonese and Portuguese as official languages shaping local identity (Kaeding, 2010:134; Lam, 2010). Studies show that local language traits in a foreign language can serve as an identity index in a local community (Gao et al., 2000; as cited in Yan & Moody, 2010; Shen, 2008). Despite the limited use of Portuguese, its cultural and official status in Macau allows it to be used as a means of displaying local identity (Mann & Wong, 1999) to show attachment to the Portuguese culture. <Macau> also distinguishes it from the officially promoted spelling of <Macao>, providing residents a direct access to Macau's 'Portuguese-ness'.

However, given that the population of Macau has always been mobile and diverse, there has not been a strong sense of shared identity among the Macau residents (in contrast with, for instance, Hongkongers in neighboring Hong Kong). Even for the specific term ‘Macanese’, for many, it only stands for a very small group of people with mixed Chinese-Portuguese descents in Macau (Chan, 2015). Yet, Macau has a more easily noticeable uniqueness in relation to its colonial history, the East-meets-West culture, its special political status, languages, food, etc. Since there does not seem to be a strong sense of in-groupness and shared commonalities among the residents in Macau, what exactly is Macau identity as discussed in Kaeding (2010) and Lam (2010)?

This makes us wonder if this ‘Macau identity’ (Kaeding, 2010; Lam, 2010) really represents the majority group or groups of people in Macau. Instead, this ‘Macau identity’ is more likely ‘a social construction based on physical reality’ which distinguishes a place from other places (Peng et al. 2020:14). In other words, what Kaeding (2010) and Lam (2010) discussed might be better understood as *place identities of a place* (Peng et al. 2020), which can be defined as different characteristics or realities of a place (e.g. Macau) which people, locals or expatriates, current residents or those having migrated elsewhere, feel more attached to, or find more meaningful. In sum, there appear to be two place identities of Macau, one highlighting its unique cultural heritage and special political status as a former Portuguese colony (Kaeding’s ‘Macau identity’), and another emphasizing PRC-governance and Macau being an economy-driven and post-colonial international city (‘Lam’s identity’). These two place identities do not have to neatly map onto two different groups of people in Macau, but it is not surprising that a certain group of people may prefer to see Macau in one way or the other (e.g. a local resident may be more attached to the part of Macau as a former colony with unique cultural heritage, whereas a migrant from mainland China may tend to see Macau as a vibrant Chinese city).

Statistical results suggest a connection between these variants and duality in Macau identity within LR orthography regimes. Here, under the framework of orthography regimes, more specifically, the two orthographic variants’ indexes in relation to two aspects of Macau identity, namely <Macao>only (MOO) and *primarily*<Macau> (PMU). The MOO pattern includes <Macao> as the dominant variant situated in HR, given that the nature of this regime requires the least amount of orthographic variation and a high degree of standardization. The PMU pattern, on the other hand, includes both orthographic variants (i.e. <Macau> and <Macao>) and occurs in LR regimes such as school writings and online tweets where use of <Macao> is not required. Within this pattern the <Macau> variant is overtly preferred while a small portion of users continue to use the <Macao> variant (see Table 8).

Within the framework, the two patterns are largely distinct from one another. Orthography in MOO remains focussed and official with the exclusive use of the <Macao> spelling. Given HR’s nature and the spelling’s strong connection with PRC and Macau’s handover in 1999, it is likely that, in HR, the focussed <Macao> indexes a national identity for Macau or the MSAR government.

Texts within this HR regime are primarily drawn from government authorities (i.e. MSAR government). Consequently, HR automatically represents a voice from the MSAR government where nationalism has been reiterated (Lam, 2010). The MSAR government has exerted a subtle sense of nationalism on Macau identity. In multiple official discourses, it addresses nationalistic sentiment and reinforces Macau residents’ sense of belonging to China (Lam, 2010), thus strengthening the MSAR’s nationalistic standpoint. Within this regime where nationalism prevails, the required, official, and focussed <Macao> in HR potentially indexes an image of nationalistic, which shows a strong bond to motherland.

Furthermore, although <Macao> originated from Portuguese orthography following Portugal’s Orthography Reform, it was gradually phased out and retained only within some English texts. However, it has been re-appropriated by China in a way that it hints at a strong connection with China in a nationalistic way. <Macao>, derived from Portuguese and later used in English, had no significance other than

Table 8. Two patterns of distribution in Orthography Regimes.

| Patterns | Orthography regimes | Text types | Orthography | Variables |
|----------|---------------------|---------------------|-------------------------------------|---------------|
| MOO | HR | Government releases | focussed, standard | <Macao> |
| PMU | LR | Student writings | less focussed, not always standard | <Macau/Macao> |
| | LRR | Online tweets | least focussed, not always standard | <Macau/Macao> |

referring to Macau as a name that is more likely to be used by non-Macau people (see Bray & Koo, 2005), until it was officially recognized by the Chinese government. This act can be considered linguistic reappropriation (Brontsema, 2004) or recontextualization (Bauman & Briggs, 1990) where <Macao>, a foreign form previously signifying Macau's colonial history, is reused and imbued with new meaning. Specifically, this new meaning represents a change of status from a former colony to post-colony, and from Portuguese to Chinese sovereignty (Bray & Koo, 2005), differentiating it from the Portuguese spelling <Macau>. Accordingly, Bray and Koo (2005) argue that adopting <Macao> indicates a strong sense of belonging to China, as it is now 'Chinese<Macao>' and not 'Portuguese<Macau>'.

In addition, <Macao> is considered strongly connected with PRC and the 1999 handover. In 1999, Beijing announced that 'Macao' would be MSAR's official English spelling after its handover (Bruning, 2001, as cited in Bray & Koo, 2005). Consequently, MSAR has gradually changed their spelling from <Macau> to <Macao> to align with China's policy since then, prompting Bray and Koo (2005:10) to alter all instances of <Macau> to <Macao> in their book. <Macao> being the preferred spelling in English in MSAR is evident in various governmental websites (e.g. the MSAR official website in English and GIB).

Given its new meaning and exclusive use within HR official discourse from the MSAR government, the chosen variant under such regimes is regulated to be focussed and therefore highly likely to index the national identity represented by the texts in HR. Specifically, it is the only orthographic variant appearing in such regimes where connections with 'motherland' (i.e. China) are emphasized (Lam, 2010). HR correlates highly with this national component of Macau's identity represented by MSAR's government. As more texts within HR adopt this spelling through focusing (as shown in Table 3), its indexicality to the nationalistic Macau identity strengthens. HR's nationalistic nature and traits in the <Macao> variant allow <Macao> to index the nationalistic Macau identity in HR.

Another pattern of PMU occurs in LR regimes such as LR (school writings) and LRR (online tweets). The nature of these regimes differs significantly from that of HR, resulting in the use of both orthographic variants with a preference for <Macau>. Additionally, orthography in these regimes does not always adhere to the government's preference (i.e. <Macao>). Texts from LRs are of individual rather than institutional, allowing for an examination of people's use of the two orthographic variants and their indexical relationship with the ethnocultural Macau identity under LR regimes.

It is noteworthy that even within LR regimes where orthography is less focussed, there is still a preferred variant: <Macau> (as seen in LR and LRR at 77% and 70%, respectively; see Table 3). The high adoption rate of <Macau> within these two regimes suggests that it has become a highly focussed variant comparable to <Macao> in HR. This indicates the existence of a spelling norm in LR regimes, providing greater potential for this variant to index group traits than its weaker counterpart <Macao> (see Table 5).

Of the 41 participants from Macau surveyed in this study, only 9 identified themselves as both Macau people and Chinese. The majority of participants identified themselves solely as Macau people. The results suggest a strong emphasis on self-identification to the local-level Macau, among younger generations (in this case, university students), which aligns with findings in Leong (2009) and Wang (2017) in which most participants considered themselves as '*primarily Macao people*' or '*Macau people only*', both descriptions indicating a strong sense of local identity. Notably, data also shows that most individuals in this group (i.e. Macau people) prefer to use <Macau> rather than <Macao> (76%). Therefore, it is likely that the <Macau> variant indexes a local part of Macau identity.

Given its direct connection to Portuguese language and culture, it is highly likely that the spelling <Macau> serves as a means for Macau residents to display their cultural pride. Kaeding (2010) noted that an emerging local identity in Macau is characterized by pride in the region's cultural heritage and rich historical background. Lam (2010:664) also argued that Macau's image as a fusion of modern tourism and historical cultural heritage has become integrated into the local component of Macau identity. Macau's colonial history has not been suppressed by the MSAR government; rather, it has been actively promoted as a source of pride and an element of 'local identity' (Cardinal, 2008; Kaeding, 2010; Lam, 2010). One manifestation of this pride is the continued use of Portuguese as an official language in Macau following the 1999 handover. Street and road signs are written in both Chinese and Portuguese, despite the fact that only a small portion of the population use Portuguese as their daily language (see Figure 1).

This serves as a visual reminder of Macau's cultural heritage and reinforces its role as a constructive component of local identity (Lam, 2010). Among Portuguese vocabulary items, <Macau> can be



Figure 1. Road sign in Coloane, Macau (Photographed by the author).

considered one of the most accessible words for those who do not speak the language on a daily basis since it appears frequently throughout Macau. The use of this spelling thus potentially indexes that conveys pride in Macau's Portuguese cultural heritage and language users' respect to Macau's Portuguese history. Given its direct connection with Portuguese culture and focussed use within two LRs by individuals in Macau, this favored variant under such regimes has greater potential than <Macao> to index this ethno-cultural aspect of Macau identity.

For instance, consider the following examples in [Figure 2](#).

1. How about some pastel de nata? #EggTart #**Macau** [A photo of a box of pastel de nata] (LRR-2023/02/Macau).
2. Coloane, **Macau** SAR China. Welcome to Macau! [MSAR flag emoji] [four sightseeing photos of Coloane] #**Macau** #**macau** #China #Tourism #travel (LRR-2022/Chinese).

These examples are from LRR, a regime where the use of <Macau> is expected, as it is the focussed variant. In the first example, a tweet posted by a Macau resident, the user shows a picture of them waiting in queue for 'pastel de nata', the Portuguese name of a traditional Portuguese pastry. <Macau> in this context could be understood as a sense of affinity towards Portuguese culture, as it is in the Portuguese lexis, indexing a language user's ethnocultural pride or showing what she/he cares or values – the culture. It is worth noting that the hashtag <Macau> in Example 1, although potentially not the user's choice, conjures a sense of 'Portuguese-ness'. This aligns with the Portuguese phrase 'pastel de nata', which the user could easily substitute with 'eggtart'.

The second example is a post by a mainland student living in Macau. <Macau> in this context, as well as the inclusion of the MSAR flag emoji, suggests a recognition, affection and respect for Macau's unique status and cultural heritage. This indexes to the user's positive attitude towards Macau's ethnocultural identity and their anticipation of experiencing it firsthand. <Macau> in this context potentially indexes a sense of respect for Macau's Portuguese history and a recognition of its unique history and culture. This is similar to the first example, where the use of <Macau> indexes an affinity for Portuguese culture.

Example 1**Example 2**

Coloane, Macau SAR China.
 Welcome to Macau! 🇲🇴 #Macau #macau
 #China #Tourism #travel

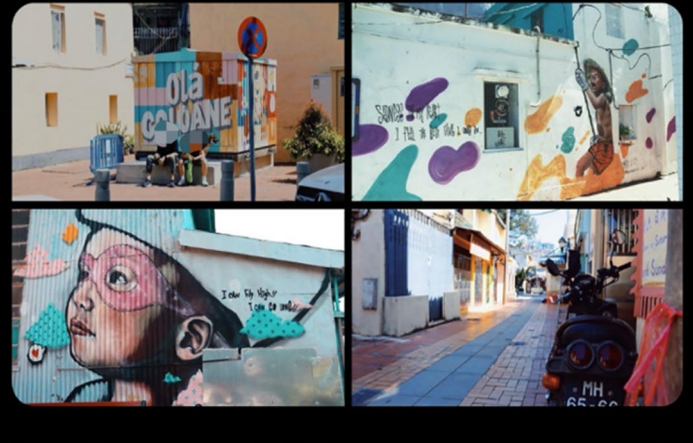


Figure 2. Two examples of <Macau> in LRR.

In both cases, <Macau> serves to reinforce and express pride and appreciation in Macau's unique cultural heritage. This could also suggest that the use of <Macau> could be something more nuanced than simply identifying as Chinese or Macau people (Section 5). For instance, this indexical value may be pride in or respect for Macau's unique culture and history. It is also plausible that the tweets about Macau, self-reported by the user as being in Macau, could actually be authored by tourists visiting the area. In such instances, the user may not be a resident of Macau, and thus is not self-identifying as such. However, they are still communicating a sense of Macau's uniqueness. In our conception of Macau/Macao as a place identity (Peng et al., 2020), this language user needs not identify herself/himself as a Macanese or Macau person in showing attachment to a place and perceiving it in a certain way (i.e. Macau/Macao).

Returning to Table 5, we also observe a significant increase in the use of <Macao> among individuals who identify as Chinese or M+C (Macau people and Chinese), which can be seen as people's self-identification to a place. This trend suggests a correlation with national identity (i.e. Chinese). Contrary to the use of <Macau> in LR, the focussed variant, the use of <Macao> can be seen as an act of conformity to the assumption and PRC standardization (i.e. <Macao> being an English term that aligns with the standard promoted by MSAR and PRC in HR). Again, the indexical value of spellings is based on a contrast to an expected variant (Lakoff, 1982). In LR, the focussed variant is <Macau>, suggesting that this spelling is likely to be expected in such contexts. Therefore, any deviation from this expected spelling could be perceived as a departure from the mainstream spelling, or even a distancing from the mainstream identity. In this connection, <Macao> is not the usual spelling in LR. Specifically, it is 'different' from <Macau> which is more commonly used. The difference in spelling automatically serves as a cue of distancing from the mainstream identity that LR uphold, namely, an affinity to Macau's Portuguese history and culture. By using <Macao>, language users in LR can potentially adopt a stance where <Macao>, as 'an English word', is emphasized. This stance subtly distances them from the mainstream historical and cultural sentiment in LR and aligns them with the standards set by the MSAR and PRC.

5.2. Indexicality of <Macau>/<Macao>

To elucidate the two spellings' indexical values in the patterns of MOO and PMU, this part applies Silverstein's (2003) concept of orders of indexicality. First-order indexicality (or *n*th-order indexicality) refers any form of language that may be associated with a sociodemographic identity (Eckert, 2008; van

Table 9. Levels of indexicality and indexical meanings of <Macao>/<Macau>.

| Orthographic variants | Indexical meanings | L. of I |
|---|--|-----------|
| <Macao> for governmental documents in English | registers / the mythical association (i.e. <Macao> is English and <Macau> is Portuguese) | 1st-level |
| <Macao> in English writings in LRs | aligning oneself with PRC standardization or nationalistic aspect of Macau identity | 2nd-level |
| <Macau> in English writings in LRs | showing affinity to the local identity or respect for the region's Portuguese history | 2nd-level |

Compennolle, 2011), whereas second-order indexicality (or $n+1$ th-order indexicality) involves stylistic use of forms that are associated with a particular group or context and can be used to index identity or stance (Eckert, 2008). Table 9 serves as an attempt to describe different levels of meanings in different orthography regimes (see Table 9).

In the case of <Macau>/<Macao> orthography, at the level of first-order indexicality, people are somewhat aware of <Macao> as the English word, and <Macau> as the Portuguese word, and people would tend to use <Macao> in governmental English writings to align with the official style. 2nd-level indexicality is the level where people use linguistic forms to perform certain identities or make a stance. In other words, a person might choose to use focussed language forms to show solidarity with a particular social group, or to resist mainstream norms (could be the officially promoted <Macao>). That is, to use <Macao> in LRs could be seen as aligning oneself with PRC standardization or national aspect of Macau identity, while using <Macau> in LRs might be seen as showing affinity to the local aspect of Macau identity or respect for the region's Portuguese history.

This section examines two pairs of indexical relationships suggested by the statistical results: <Macao> and the nationalistic identity (Lam, 2010) in Macau, and <Macau> and ethno-cultural identity (Kaeding, 2010; Lam, 2010) in Macau. At present, these two pairs of indexical relationships remain at the level of first-order indexicality due to two factors. First, the indexicality is 'presupposing' or context-dependent (Johnstone, Andrus & Danielson, 2006:82) as the indexical relationships are interpreted under certain circumstances (e.g. <Macao> is interpreted to be indexical to the national component of Macau identity under a HR regime). Secondly, this orthographic variation is at risk to be considered stylistic since most participants tend to be invariant users who do not alter their spellings in different writings, even though most of them are aware of the 'mythical' association. Still, the results of spelling choices according to orthography regimes shows that only seven participants self-reported that they would change their spellings to <Macao> in formal registers (e.g. government documents), whereas the rest of the <Macau> users (47) did not.

6. Conclusion

Based on the framework of Orthographic Regimes (Sebba, 2007), this article presents a quantitative analysis of the distribution of the two orthographic variants at three levels of orthographic regimes, namely, a HR regime of government writing, a LR regime of school writings, and a LRR regime of online texts. The study utilized a questionnaire to gather LR data and to understand participants' self-reported place identities and the potential indexical relationship between the two variants and different components of Macau identity. The findings indicate that the spelling variants <Macau> and <Macao> are not merely orthographic choices, but they are also potentially powerful linguistic tools that reflect and construct different aspects of Macau identity.

The study concludes that the distribution of <Macau> and <Macao> varies across three levels of orthography regimes, displaying two patterns of distribution, namely, Macao only (MOO) and primarily Macau (PMU). MOO refers to the pattern where <Macao> is the only variant used due to its expected use in HR. In contrast, for PMU, <Macau> is the preferred and focussed variant in LRs, where the use of <Macao> is not expected. The results demonstrate that the distribution of these two orthographic variants aligns with Sebba's (2007) orthography regimes, i.e. at HR, <Macao> is highly focussed and almost codified, whereas at LRs both <Macao> and <Macau> are used.

The study also reveals that <Macao>, predominantly used in official government writings (i.e. HR), could index a nationalistic aspect of Macau identity, emphasizing alignment with PRC and MSAR's

language standardization process. Conversely, <Macau>, preferred in school writings and online texts (i.e. LRs), could index a local aspect of Macau identity that embraces and respects Macau's unique colonial and cultural heritage.

Furthermore, under orders of indexicality (Silverstein, 2003) and Macau identity (Kaeding, 2010; Lam, 2010) in the lens of *place identity* (Peng et al., 2020), this study explores the indexicality involving <Macau>/<Macao>. Under first-order indexicality, the use of <Macao> may signify registers (formal and governmental), and it immediately suggests that the author is following the <Macao>-is-English assumption. For second-order indexicality, conversely, given its direct connection to Portuguese culture and its prevalent use among individuals in Macau within LR and LRR contexts, <Macau> could potentially index an ethno-cultural identity (Kaeding, 2010), showing appreciation and respect to Macau's Portuguese history and culture. The presence of <Macao> in LRs, however, serves as an obvious contrast. The nationalistic connotations associated with <Macao> allow it to index a national component within Macau's identity or to show one's alignment with PRC and MSAR's standardization as the spelling is used in governmental writings.

As part of the LR data, the questionnaire shows that most participants were aware of the mythical assumption (<Macao> is English; <Macau> is Portuguese). However, most users tend to be invariant as they did not self-report that they have the tendency of changing spellings in different written discourses, as only seven participants (out of fifty-four) self-reported that they would alter spellings in formal documents (HR). This result suggests that most of the time, the variants remain in the description of *n*th-order indexicality (Eckert, 2008; Johnstone, Andrus & Danielson, 2006; Silverstein, 2003).

In light of the findings, future research could explore how orthographic choices are influenced by various social factors and how they contribute to the construction of place identities. Such research would not only enrich our understanding of the sociocultural dynamics of orthography but also inform language policy and planning in multilingual societies like Macau.

Despite the limited attention given to orthographic variation in Macau studies, this research aims to provide an analysis of orthographic practices within the region. The study employs self-designed corpus data to examine three levels of orthography regimes. However, practical constraints may have affected the consistency of data sourcing. Specifically, data from the MWEC are drawn from multiple sources, leading to an unbalanced corpus. Furthermore, the limited sample size and the fact that all participants are university students may not accurately represent the demographics of Macau. Consequently, the findings may not fully encapsulate the language practices associated with these two orthographic variants in Macau.

At any rate, these findings highlight the complex interplay between orthography, society and identity. They also underscore the need for further research on orthographic variation in other multilingual and multicultural contexts. Additionally, this study showcases the relevance and usefulness of quantitative methods in inquiry of sociolinguistic issues such as identity and language policy which have more frequently been approached with qualitative or ethnographic methods.

Note

1. The University of Macau (UM) retains the so-called Portuguese spelling <Macau> rather than the English variant <Macao>, possibly due to its original Portuguese name. While <Macao> is used on the official website (see <https://www.um.edu.mo/about-um/rectors-message>) and documents (e.g. the Academic Calendar, see <https://reg.um.edu.mo/download/um-academic-calendar/2tu.pdf>), <Macau> is preserved in proper names like 'University of <Macau>' or 'Centre for <Macau> Studies', especially in logos. The reason behind this spelling choice is unconfirmed as no public documentation exists to explain UM's decision.

Author contributions

Chen, Trin Qingyi is the author of this submission. His contribution to this article includes conceptualization, data curation, formal analysis, investigation, methodology and original draft writing and revision. Chan, Brian Hok-Shing is the co-author of this submission and Trin, Qingyi Chen's supervisor. He helped review and edit the original draft into this submission.

Disclosure statement

No potential conflict of interest was reported by the authors.

About the authors

Trin Qingyi Chen was a graduate student from the University of Macau. He obtained his Master's degree in English Studies with a focus on sociolinguistics in 2023. His research interests include language variation, language and sexuality, language and identity.

Brian Hok-Shing Chan is an Associate Professor in the Department of English at the University of Macau. He obtained PhD in Linguistics from University College London in 1999. His current research interests include code-switching/translanguaging, bi/multilingualism, sociolinguistics and linguistic landscape.

ORCID

Trin Qingyi Chen  <http://orcid.org/0009-0000-5426-8369>

Brian Hok-Shing Chan  <http://orcid.org/0000-0001-9621-1950>

References

- Androutsopoulos, J. (2000). Non-standard spellings in media texts: The case of German fanzines. *Journal of Sociolinguistics*, 4(4), 1–17. <https://doi.org/10.1111/1467-9481.00128>
- Augst, G. (1988). Theorie und praxis der textlinguistik. In G. Augst (Ed.), *Textgrammatik und textlinguistik* (pp. 1–30). Niemeyer.
- Bauman, R. & Briggs, C. L. (1990). Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology*, 19(1), 59–88. <https://doi.org/10.1146/annurev.anthro.19.1.59>
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204. <https://doi.org/10.1017/S004740450001037X>
- Bray, M., & Koo, R. (2005). *Education and society in Hong Kong and Macao: Comparative perspectives on continuity and change* (2nd ed.). Springer.
- Brontsema, R. (2004). A queer revolution: Reconceptualizing the debate over linguistic reclamation. *Colorado Research in Linguistics*, 17, 1–17. <https://doi.org/10.25810/dky3-zq57>
- Cardinal, P. (2008). Macau: The internationalization of an historical autonomy. *Boletín Mexicano de Derecho Comparado*, 41(122), 637–689. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0041-86332008000200003
- Chan, B. H. S. (2015). A local voice of Macau: Traditional characters, code-switching and written cantonese in an internet forum. *Global Chinese*, 1(2), 281–310. <https://doi.org/10.1515/glochi-2015-1013>
- Davies, E. E. (1987). A contrastive approach to the analysis of politeness formulas. *Applied Linguistics*, 8(1), 75–88. <https://doi.org/10.1093/applin/8.1.75>
- DSEC. (2022). *Detailed results of 2021 population census*. https://www.dsec.gov.mo/getAttachment/6cb29f2f-524a-488f-aed3-4d7207bb109e/E_CEN_PUB_2021_Y.aspx (13 December, 2022).
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), 453–476. <https://doi.org/10.1111/j.1467-9841.2008.00374.x>
- Eisenstein, J. (2015). Systematic patterning in phonologically-motivated orthographic variation. *Journal of Sociolinguistics*, 19(2), 161–188. <https://doi.org/10.1111/josl.12119>
- Heffernan, K., Borden, A. J., Erath, A. C., & Yang, J.-L. (2010). Preserving Canada's ' honour. *Written Language & Literacy*, 13(1), 1–23. <https://doi.org/10.1075/wll.13.1.01hef>
- IBM Corp. (2021). *IBM SPSS statistics for macintosh [computer program]. Version 28.0.1*. IBM Corp.
- Ilbury, C. (2020). "Sassy queens": Stylistic orthographic variation in Twitter and the enregisterment of AAVE. *Journal of Sociolinguistics*, 24(2), 245–264. <https://doi.org/10.1111/josl.12366>
- Johnstone, B., Andrus, J., & Danielson, A. (2006). Mobility, Indexicality, and the Enregisterment of "Pittsburghese". *Journal of English Linguistics*, 34(2), 77–104. <https://doi.org/10.1177/0075424206290692>
- Kaeding, M. P. (2010). The evolution of Macao's identity: Toward ethno-cultural and civic-based development. *Journal of Comparative Asian Development*, 9(1), 133–168. <https://doi.org/10.1080/15339114.2010.482804>
- Lakoff, R. (1982). Persuasive discourse and ordinary conversation, with examples from advertising. In D. Tannen (Ed.), *Analyzing discourse text and talk* (pp. 25–42). Georgetown University Press.
- Lam, W. M. (2010). Promoting hybridity: The politics of the New Macau identity. *The China Quarterly*, 203, 656–674. <https://doi.org/10.1017/S0305741010000640>
- Le Page, R. B., & Tabouret-Keller, Andrée. (1985). *Acts of Identity: Creole-Based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.

- Leong, S. M. (2009). The impact of intergroup attitude on language: A survey study of Macau high school student 2009s. *Journal of Macau Studies*, 51(2009), 146–154.
- Mann, C., & Wong, G. (1999). Issues in language planning and language education: A survey from Macao on its return to Chinese sovereignty. *Language Problems and Language Planning*, 23(1), 17–36. <https://doi.org/10.1075/lplp.23.1.02iss>
- Moody, A. J. (2021). *Macau's languages in society and education: Planning in a multilingual ecology* (Vol. 39). Springer International Publishing AG.
- Moody, A. J. (2008). Macau English: Status, functions and forms. *English Today*, 24(3), 3–15. <https://doi.org/10.1017/S0266078408000242>
- Peng, J., Strijker, D., & Wu, Q. (2020). Place Identity: How Far Have We Come in Exploring Its Meanings? *Frontier in Psychology*, 11, 1–19. <https://doi.org/10.3389/fpsyg.2020.00294>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. R Foundation for Statistical Computing. www.r-project.org (June 2023)
- Romiti, M. (1998). Il testo come oggetto di analisi. In M. Romiti (Ed.), *Il testo* (pp. 13–36). Carocci.
- Sebba, M. (1998). A congruent approach to the syntax of codeswitching. *International Journal of Bilingualism*, 2(1), 1–19. <https://doi.org/10.1177/136700699800200101>
- Sebba, M. (2007). *Spelling and society: The culture and politics of orthography around the world*. Cambridge University Press.
- Shen, C. L. (2008). *In search of a medium of instruction: Macao secondary students' attitudes towards three English varieties*. University of Macau MA Thesis.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3–4), 193–229. [https://doi.org/10.1016/S0271-5309\(03\)00013-2](https://doi.org/10.1016/S0271-5309(03)00013-2)
- Tagg, C. (2016). Heteroglossia in text-messaging: Performing identity and negotiating relationships in a digital space. *Journal of Sociolinguistics*, 20(1), 59–85. <https://doi.org/10.1111/josl.12170>
- van Compernelle, R. A. (2011). Developing second language sociopragmatic knowledge through concept-based instruction: A microgenetic case study. *Journal of Pragmatics*, 43(13), 3267–3283. <https://doi.org/10.1016/j.pragma.2011.06.009>
- Wang, H. (2017). Media exposure and national identity formation among college youth in postcolonial Macau. *Issues & Studies*, 53(04), 1750009. <https://doi.org/10.1142/S1013251117500096>
- Wu, Z., & Jin, G. (2013). The evolution of spellings of 'Macau': An examination of early Portuguese and Western archival materials. In K. K. W. Wong, C.X. George. (Ed.), *Macao - cultural interaction and literary representations* (pp. 3–10). Routledge.
- Yan, X. & Moody, A. (2010). Language and society in Macao. *Chinese Language and Discourse. An International and Interdisciplinary Journal*, 1(2), 293–324. <https://doi.org/10.1075/cld.1.2.07xi>
- Young, C. M. Y. (2006). Macao students' attitudes toward English: A post-1999 survey. *World Englishes*, 25(3–4), 479–490. <https://doi.org/10.1111/j.1467-971X.2006.00468.x>