universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Chinese as a Pluricentric Language

verfasst von / submitted by

Sandra Maria Kaltenegger, BA BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 899
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Angewandte Linguistik
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Eva Vetter

Danke an meine Mama, die mich all die Jahre ohne Vorurteile frei herumstudieren ließ, und an meinen Papa, der meine akademischen (Um-)Wege zwar nicht mehr mitverfolgen konnte, aber mir schon früh den Weg dafür geebnet hat.

Kiitos suomalaiselle isäntäperheelleni, joka otti minut avosylin vastaan. Lisäksi kiitos suomalaisille ystävilleni, jotka näyttivät minulle kuinka ihanaa elämä on Itävallan ulkopuolella.

Danke an meinen Lebenspartner Gregor, der all diese akademischen (Um-)Wege hautnah miterlebt hat, mir stets diskussionsfreudig zur Seite gestanden ist und sich derart mit meinen Auslandsaufenthalten arrangiert hat, dass er mich nach Taiwan begleitet hat. Ein Mensch, dem man die Welt einfach zu Füßen legen muss.

感謝我的老師，語言伙伴，朋友們向我帶來許多接觸中文世界的機會。

Thank you to my advisor and employer Eva Vetter, whom I've learned from so much over the past two years and to my wonderful colleagues at the Center for Teacher Education – the pleasure of working with all of you cannot be put into words!

Table of Content

Introduction 1

1. Language Policy 3

1.1. Language Planning – Solving Language Problems 3

1.2. Language Policy – Influencing Language Behavior 6

1.3. De Facto Language Policy – Uncovering Hidden Mechanisms 9

1.4. Terminology Shift in Language Policy – Specifying Denotations 12

2. Standard – Language – Variation 14

2.1. Language 14

2.1.1. Kloss' Understanding of Language – Abstand and Ausbau 15

2.1.2. Fishman's Adaptation of Kloss – Ausbau and Einbau 17

2.1.3. Croft's Alternative – Language Definition Based on Language Attitudes 19

2.1.4. Complementing Paradigms – The Two Approaches Combined 23

2.2. Standard Language 23

2.2.1. Defining Standardization 24

2.2.2. Standard Language Ideology 27

2.2.3. Standard Language (Ideology) and the Nation-State 29

2.2.4. Study on Attitudes on Standard Language 31

2.3. Standard Language Variation 34

2.3.1. Defining Pluricentricity 34

2.3.2. Categorizations of Pluricentric Languages 37

2.3.3. The Asymmetrical Relation between Pluricentric Standard Varieties 40

2.4. Looking Back and Looking Forward 43

3. The Chinese Language 47

3.1. Contextualization and Conceptualization 47

3.2. Chinese Prior to 1912 53

3.3. Standardizing Chinese 57

3.3.1. Modern Spoken Chinese 59

3.3.2. Modern Written Chinese 62

3.3.3. The Modern Chinese Writing System 65

3.3.4. Phonetization of Modern Chinese 66

3.3.5. Standardization of Other Fāngyán 67

3.3.6. Developments of Mandarin Outside Mainland China 69

3.4. Chinese Pluricentricity 73

3.4.1. Indications of and Elaborations on Chinese Pluricentricity in the Literature 73

3.4.2. Step I: What Is Chinese? 80

3.4.3. Step II: Which Fāngyán Have Explicit and/or Implicit Norms? 82

3.4.4. Step III: Which Fāngyán Are Pluricentric? 85

3.5. The Chinese Language in Retrospection 89

4. The Applicability of Pluricentricity to Chinese – Conclusion and Outlook 91

Bibliography 95

Appendix 103

English Abstract 103

German Abstract 104

List of Figures

Figure 1: Motivational factors for corpus planning 8
Figure 2: Geographical distribution of the seven major Fāngyán..... 50
Figure 3: Map of the Yellow River area..... 54
Figure 4: The four basic tones in Standard Mandarin..... 61
Figure 5: The pluricentric varieties of Chinese..... 89
Figure 6: The pluricentricity of scripts and phonetization systems in Mandarin 90

List of Tables

Table 1: Haugen's four aspects of language development..... 26
Table 2: Characteristics of the standard language..... 32
Table 3: The inclusive and exclusive standard language..... 33
Table 4: The Sino-Tibetan language family 48
Table 5: Examples for lexical variation in Mandarin 70

Introduction

The Chinese language – the first language of far more than a billion people – exhibits great linguistic variety that is a matter of continuous discussion. Chinese is commonly divided into seven major varieties that Chinese speakers themselves refer to as Fāngyán (方言)[1]. These Fāngyán are perceived of either dialects or independent languages in Western linguistics. However, Chinese does not only show great linguistic variety in local non-standard varieties, the standard varieties differ as well, depending on the region they are used in, i.e. Chinese presents characteristics typical for a pluricentric language.

The motivation for choosing Chinese pluricentricity as the topic of this Master's thesis is twofold – there is personal experience, on the one hand, and academic curiosity, on the other. Growing up with a non-dominant variety of German – namely Austrian German – as my first language, pluricentricity has been a topic in my life from the very beginning. Being exposed to English and Spanish in school, this further manifested my impression that pluricentric languages are not the exception but rather the norm. Yet, when I started attending Mandarin language courses, Chinese was depicted as a very homogenous language, making me believe that if I could acquire the standard variety taught in these classes, I could communicate with the entirety of the aforementioned one billion people. Nothing could be further from the truth. Extensive travels within Mainland China to areas where Fāngyán such as Cantonese and Wu are commonly spoken painfully illustrated to me that Chinese is a much more heterogenous language than I have learned to believe. Additionally, whilst my stay in Taiwan[2], I have come to realize that not even the standard variety used and taught in Taiwan is the same as I have acquired over the years in Vienna and Beijing. The knowledge gained about the linguistic situation of Chinese is frustrating, yet fascinating for me as a Mandarin language learner and highly insightful for me as a researcher. To attain a better understanding

[1] This Master's thesis uses the traditional script for writing Mandarin characters and Hànyǔ Pīnyīn for annotating the pronunciation of these characters. This combination is chosen intentionally, in order not to advocate for any particular Mandarin standard variety (the traditional script is typical for, e.g. Taiwan, whereas Hànyǔ Pīnyīn is used, for instance, in Mainland China).

[2] This Master's thesis does not attribute great relevance to the official recognition of certain regions. If this Master's thesis adhered to such a perspective, the adaption of pluricentricity would not align with the linguistic and social reality of these regions. Hence, more importance is attributed to whether such regions are self-governing entities. If they are, they can implement their own language policies and codify their varieties independently from other regions. For this reason, this Master's thesis differentiates between Taiwan and China, without positioning itself in the ongoing debate regarding Taiwan's independence.

of the Chinese language situation, I turned to the academic world, in which the theory on pluricentricity is well-established. However, Chinese is hardly ever incorporated in it. The few papers dedicated to Chinese pluricentricity reveal two major flaws. They either do not apply pluricentricity to Chinese as a whole (but only to Mandarin), or new categories and conceptualizations are developed that can describe only the phenomenon of Chinese pluricentricity, making it difficult to compare Chinese with other pluricentric languages. The unsatisfying treatment of Chinese pluricentricity was emphasized by Muhr (2016: 32) in his list of research desiderata on pluricentric languages.

Therefore, this Master's thesis poses the following research question:

To what extent is the notion of pluricentricity applicable to the Chinese language?

In order to answer this research question, various aspects need clarification, some of theoretical and some of practical nature. Theoretical aspects require discussion, such as the definition of languages and standard varieties. More specific questions regarding Chinese are, for instance, what the term *Chinese* incorporates and how to deal with the many Fāngyán in a pluricentric framework for Chinese. Chapter 1 provides the tools for a systematic approach to Chinese pluricentricity. It is dedicated to language policy, describing the shift in research focus from explicit language planning to implicit language beliefs and behavior. Chapter 2 constructs the theoretical scaffold of this Master's thesis, covering the ways of defining language and analyzing the notion of standard language before going into depth on the concept of pluricentricity. The centerpiece of this Master's thesis is chapter 3, in which the theoretical elaborations in chapter 1 and 2 are applied to the Chinese language. At first, a contextualization and conceptionalization of the Chinese language is provided, including an overview of the history of the Chinese language. Thereafter a detailed description of the standardization processes of Mandarin Chinese is presented, since it has had a great impact on the Chinese language as a whole. Additionally, this chapter is also concerned with the standardization of other Fāngyán and the developments outside Mainland China. Lastly, this Master's thesis discusses the indications of and elaborations on Chinese pluricentricity in the literature before applying its own framework to it. Chapter 4 concludes the outcomes of this Master's thesis and offers a list of possible future research foci regarding Chinese pluricentricity.

1. Language Policy

This chapter provides the theoretical scaffold and conceptual toolkit for this Master's thesis. The historical developments of the research field are covered chronologically. Chapter 1.1 is devoted to the origins of language policy and, hence, describes the context in which this new research field could thrive in the 1960s. Chapter 1.2 focuses on the opening of the field in the late 80s and early 90s, which allowed for more diverse research questions and took into account various influential factors that had been overlooked previously. Chapter 1.3 covers the hidden structures of language policy, namely language ideologies, which are at the center of attention nowadays. Lastly, chapter 1.4 provides a short elaboration on the choice of terminology for this Master's thesis. The last chapter is necessary since a conglomerate of terms has accumulated over the decades for the reference to the research field. Throughout this Master's thesis the term *language policy* is used for referring to the entire field, *language planning* is only used to address the early days of the field covered in chapter 1.1.

1.1. Language Planning – Solving Language Problems

The activity of influencing language and its usage – whether deliberate or not – is as old as humanity itself. Yet, it gained importance when nations started to develop, and national identities were constructed alongside them. This is the cause for the world experiencing a first wave of language planning – the way academia defined it at first – in the 1920s when dozens of languages were modernized under Stalinist rule. The scientific interest in the language planning activity, however, only emerged in the 1960s when the colonial system started to crumble, and new nations emerged, mainly on the African and Asian continent. (e.g. Nekvapil 2006: 92–93)

The mindset behind early work on language planning was rather optimistic. It was perceived that new nations are confronted with language problems for which researchers could offer a solution through planning, which is predominantly administered by governmental institutions. *Language Problems of Developing Nations* by Fishman, Ferguson and Das Gupta (1968) is a pivotal work from this period. Generally speaking, this first period of language policy research was characterized by the assumption that language planning activities are ideologically neutral and enhance nations by driving modernization. Yet, the one nation–one language ideology prevailed strongly in the academic discourse. (Ricento 2000: 197–200)

The term *language planning* per se was introduced to scientific literature in Haugen's (1959) study on language standardization in Norway, though the term was first used in a seminar at Columbia University by Uriel Weinreich two years prior (Cooper 1989: 29). According to Haugen, language planning is:

> the activity of preparing a normative orthography, grammar, and dictionary for the guidance of writers and speakers in a non-homogenous speech community. In this practical application of linguistic knowledge we are proceeding beyond descriptive linguistics into an area where judgment must be exercised in the form of choices among available linguistic forms. (Haugen 1959: 8)

This definition emphasizes two aspects that are crucial for the early days of the research field – namely language planning as a prescriptive activity and the strong focus on corpus planning.

Corpus planning belongs to the dichotomic categorization of language planning activities, initially introduced by Kloss (1969). His elaborations on corpus planning are very specific and are best summarized by Cooper as "the creation of new forms, the modification of old ones, or the selection from alternative forms in a spoken or written code" (1989: 31). *Status planning* – the 'counterpart' of corpus planning – is defined by Kloss only vaguely: "They [=those concerned with this type of language planning] are primarily interested in the status of the language whether it is satisfactory as it is or whether it should be lowered or raised." (1969: 81) In more specific terms, status planning concerns "the allocation of languages or language varieties to given functions, e.g. medium of instruction, official language, vehicle of mass communication" (Cooper 1989: 32). Even though this bipartite division implies a clean line between its two parts, especially since Kloss (1969: 81) points out that different groups of people are involved in the planning process of each. Corpus and status planning are not truly counterparts. In fact, distinguishing between the two is much easier in theory than in practice (e.g. Fishman 1983: 111–112).

Following his analysis of twelve different definitions of language planning, Cooper (1989) adds a third category to this dichotomic understanding of language planning, namely acquisition planning. The aim of acquisition planning is to increase the number of users of a certain language. Cooper's reasons for the introduction of a third category are twofold. On the one hand, "considerable planning is directed toward *language spread*" (1989: 33; emphasis in original), whereas on the other, "the changes in function and form sought by status and corpus planning affect, and are affected by, the number of a language's users"

(1989: 33). This addition to the language planning categorization – creating a trichotomic classification – is widely accepted in the literature which has followed. Spolsky and Shohamy (2000: 10) later introduced a subcategory to acquisition planning that described the efforts of diffusing a language beyond its original speech area, hence calling it *diffusion planning*. There are many institutions engaged in diffusion planning – the Confucius Institutes for Mandarin Chinese, to name only one.

But why conduct language planning – may it be corpus, status and/or acquisition planning – in the first place? In the early days of the research field, language planning was associated with linguistic goals only, not considering the various nonlinguistic factors that may spark incentives for language policy (see chapter 1.2 and 1.3). Several researchers deal with the goals of language planning, such as Cooper (1989) and Haugen (1983). To focus on only one, Nahir (1984) identifies eleven linguistic goals that are listed below including a short description of each (as in García 2015: 356; emphasis in original).

1. *Language purification*, that is, maintaining linguistic consistency and standards of a language, usually through the development of prescriptive grammars and dictionaries.
2. *Language revival, language revitalization and language reversal*, encompassing efforts at restoring the language.
3. *Language reform*, that is, changing the orthography, spelling, lexicon, or grammar of a language in order to facilitate language use.
4. *Language standardization* for effective communication, accomplished usually through pedagogical grammars and dictionaries.
5. *Language spread*, an attempt to increase the number of speakers of a language, usually by having speakers shift to another language.
6. *Lexical modernization*, that is, expanding the capacity of a language to deal with new concepts and technology;
7. *Terminological unification*, also known as term planning, and having to do with development of equivalent terminology across geographic areas, especially terms having to do with medicine, science, industries, aviation and maritime navigation, and technology.
8. *Stylistic simplification*, attempts to make text more readable and less complex in lexicon and syntax.
9. *Interlingual communication* to facilitate communication between members of different speech communities.

10. *Language maintenance*, having to do with the preservation of a language.

11. *Auxiliary code standardization*. Here Nahir (1984) refers to "signs for the deaf, place names, and rules of transliteration and transcription" (p. 318).

The attempt to sort these linguistic goals of language policy into the previously discussed types (corpus, status, and acquisition planning) quickly proves Fishman's point: it is much harder to categorize them in practice than in theory. The majority of Nahir's goals are addressing matters of corpus planning (e.g. language purification, lexical modernization, and terminological unification) but one soon stumbles upon goals that are ambivalent in nature. Language standardization, for instance, has a strong focus on corpus planning. Yet – as will be discussed in chapter 2.2 – it plays a crucial role in the status of a language as well. Another ambivalent example would be language revival since it encompasses all three types of language policy – corpus planning is necessary to enhance the language's status, which then needs to be spread in order to reach the goal of language revival. As García (2015: 355) points out, "[g]oals refer to what the action of language planning directs us toward" – a question that was dealt with already in the early days of the research field – whereas motivation, the flipside of it, "is why we're impelled toward a goal" – a question that emerged and was covered only much later (see chapter 1.2).

1.2. Language Policy – Influencing Language Behavior

The critical voices in language planning research grew louder and louder, disapproving of its isolating approach towards language. Tollefson's critique highlights the core issue: "The study of decontextualized language ignores the inherent dynamic relationship between language diversity and human social organization." (1991: 17) Thus, the field excluded many influential variables from the equation. Cooper's pivotal work from 1989 started shifting the common paradigm. By asking "Who plans what for whom and how?" (1989: 31; emphasis in original) he opened up the field to yet unforeseen research topics.

Firstly, Cooper expands the agents of language policy because "to restrict language planning to the work of authoritative institutions is to be too restrictive" (1989: 31). Grass-root level movements, the works of individuals, schools, religious institutions, families – all these agents need to be included in the research on language policy, otherwise it would "impoverish the field" (1989: 38). Secondly, Cooper adapts the categorization of language planning activities by adding a third category, acquisition planning, to the bipartite system (as

mentioned in chapter 1.1). Thirdly – parallel to the agents of language planning – Cooper also broadens the understanding of the target group of language planning activities. Previously, only aggregates at the level of entire states played a role in the research field, whereas he advocates the focus on smaller entities, such as the above-mentioned schools, religious institutions etc. Last but not least, Cooper is concerned with the how of the research field, endorsing a descriptive approach to language planning research that studies "what actually happens" (1989: 42). In his conclusion, Cooper summarizes his elaborations with another definition of language policy, incorporating and expanding the twelve definitions he has analyzed in his article: "Language planning refers to deliberate efforts to influence the behavior of others with respect to the acquisition, structure, or functional allocation of their language codes." (1989: 45)

The study on influencing language behavior became pivotal in the field of language policy and the complex social settings in which language planning actions take place were put into question.

> [I]t is primarily through language that the battles between homogenous ideologies, hegemony and power vs. diversity, voice, representation and inclusion continue to take place. Language policy […] is the major tool through which such battles and manipulations take place. It serves as a device to perpetuate and impose language behaviors in accordance with the national, political, social and economic agendas. (Shohamy 2006: 2–3)

But what are these agendas and what is the driving force behind language policy? Fishman (2000; see also Figure 1) identifies two overriding motivational factors – an urge for independence versus a longing for interdependence, consisting of four factors each. The four motivational reasons driving for independence are *ausbau*, *uniqueness*, *purification*, and *classicization*. Their four counterparts advocating interdependence are *einbau*, *internationalization*, *regionalization*, and *vernacularization*. In his elaborations, Fishman (2000: 48) emphasizes the connection between corpus and status planning, foreshadowing the research focus on covert language policies (see chapter 1.3): "[I]t [=the opposition between independence and interdependence] clearly implies that corpus planning, in itself, is an expression of a status planning agenda, albeit in more muted, disguised, or indirect terms than those openly avowed in governmental or other authoritative declarations."

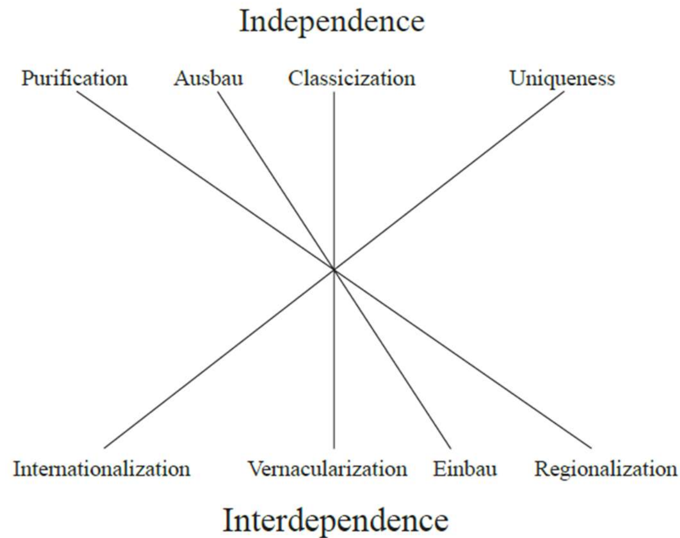

Figure 1: Motivational factors for corpus planning (Fishman 2000: 49)

The first process driving for independence, *ausbau*, is based on the work of Kloss (1967) who differentiates between *Abstandsprache* (an independent language due to its linguistic distance to other languages) and *Ausbausprache* (an independent language for reasons of language planning) – this distinction is pivotal for the development of theories on pluricentricity, hence, more on that in chapter 2.1.1. Having *ausbau* as a motivation for language policy manifests in the attempt to distance a language or a variety from another. *Uniqueness*, a process that often coincides with *ausbau*, is the endeavor to add characteristics to a language that clearly distinguishes it from neighboring/rival languages. Whereas the latter is driven by the wish to stand out, *purification* is based on the fear of contamination, usually from a specific rival language. *Classicization* processes strive to keep or make a language more coherent to its source language.

In contrast to *ausbau*, *einbau* processes attempt to lessen the differences between languages, maybe, eventually, causing them to fuse into one. *Internationalization* tries to adapt to the international community by opting for a common ground. *Regionalization* relates to the concept of *Sprachbund*[3], allowing for languages in the same region, no matter their linguistic distance, to borrow from each other. Lastly, v*ernacularization* – in stark opposition to *classicization* – is oriented towards peoples' actual language usage.[4]

---

[3] "A linguistic area (or *Sprachbund*) is generally taken to be a geographically delimited area including languages from two or more language families, sharing significant traits (which are not found in languages from these families spoken outside the area)." (Aikhenvald & Dixon: 11; emphasis in original)

[4] Examples can be found in e.g. Fishman (2000) or García (2015).

The above mentioned motivational factors are diverse and useful for the description of language policy activities. However, their commonality with Nahir's eleven goals of language planning (see chapter 1.1) is that they still revolve around language being the main target of language policy. Yet, Cooper (1989: 35) already pointed out that language policy "is typically, perhaps always, directed ultimately towards nonlinguistic ends".

By putting the speakers of a language in the center of attention instead of the language itself, García (2015: 357–358) identifies three additional motivations for language policy which may take on various forms when implemented. These motivations either strive for *exclusion*, *inclusion* or *cultural autonomy*. *Exclusion* may be reached in three ways; one way is by *shaming* speakers of a language for their language use – often occurring in educational settings. *Othering* refers to the systematic discrimination of language minorities, e.g., by limiting their access to certain opportunities based on their language competencies. Another way to exclude people is through *silencing*, essentially taking away their voice. This may be realized through the prohibition of using a language in certain contexts. Although most language policy activity is driven by the desire to exclude people groups, García also recognizes three means for inclusion – namely through *bilingualism/multilingualism*, *language teaching* (language spread and translanguaging), and *diversity and pluralism*. With *language policy for cultural autonomy,* she refers to activities with which people do not strive for independence per se but merely want to remove dominant outside forces.[5]

Defining language policy as the research of language behavior and opening up to the highly complex social, economic, political etc. settings language policy takes place in, lead to an empowerment of the entire field. As García (2015: 359) puts it: "The field has evolved from one where descriptions abounded and the state was in charge to one that is developing a theory of social action motivated by the struggle created by power and inequality." This becomes even more apparent in the following chapter, which presents the shift towards the underlying processes of language policy driven by language ideologies.

1.3. De Facto Language Policy – Uncovering Hidden Mechanisms

Differentiating between corpus, status and acquisition planning is still common today because this trichotomy allows for an initial patterning of the various language policy activities. Yet, it

[5] Again, hands-on examples in García (2015).

has been criticized from early on that, even though the different aspects of language policy are easily distinguishable in theory, in practice, they take on similar appearances and often overlap with each other (see chapter 1.1). Additionally, language planning is – in contrast to the common belief in the early days (see chapter 1.1) – not ideologically neutral. As a consequence to this critique and to the paradigm shift towards defining language policy as the study of language behavior (see chapter 1.2), the research field has turned its focus on covert language policy, which this chapter is dedicated to.

Explicit policy statements regarding language are the appearance of language policy especially (but not only) early language planning researchers were/are most concerned with. Such a policy statement typically consists of the following parts:

> A specified group (e.g., all native speakers of any named language, L; anyone who finished secondary education; or any applicant for a position in the diplomatic service) should use/acquire/have the ability to read/speak/write/understand a specific variety (or specific varieties, or even, specific features of a variety) of L for at least one defined role or function (e.g., as citizens, for employment, or for community use). (Spolsky/Shohamy 2000: 9)

The implementation of such policy statements may have far-reaching consequences for the affected population. Yet, language policies that are derived from language practice are just as powerful and consequential. Hence, the differentiation between *overt* and *covert* language policies (Schiffman 1996: 2) or *explicit* and *implicit* language policies (Shohamy 2006: 50–51) is introduced to the field. With *overt*/*explicit* Schiffman and Shohamy refer to “those language policies that are explicit, formalized, de jure, codified and manifest” (Shohamy 2006: 50), such as the declaration of the official/national language/s of a nation-state. These policies may have unexpected effects, though – Baldauf (1994) calls this *unplanned language planning*. *Covert*/*implicit* language policies, on the other side, concern “language policies that are implicit, informal, unstated, de facto, grass-roots and latent” (Shohamy 2006: 50), e.g. the officially undefined status of the English language in the USA, even though its dominance is substantial in most domains.

Schiffman (1996: 13) criticizes that researchers do not recognize the importance of covert language policies, focusing instead on overt language policies only. However, the perception within the field slowly started to shift towards the underlying structures of language policy. This becomes evident in Spolsky’s proposed framework of language policy,

consisting of practice, beliefs, and management. *Language practice* is “the habitual pattern of selecting among the varieties that make up its [=a speech community’s] linguistic repertoire” (Spolsky 2004: 5). With *language beliefs* Spolsky (2004: 14) refers to “a general set of beliefs about appropriate language practices, sometimes forming a consensual ideology, assigning values and prestige to various aspects of the language varieties used in it [=a speech community].” Lastly, *language management* is “the formulation and proclamation of an explicit plan or policy” (Spolsky 2004: 11). These three components are in a dynamic relationship in which they constantly influence one another. For instance, certain language practices create new beliefs about correct language use which then, as a consequence, manifest in a policy statement.

Shohamy (2006: 54) positions herself within Spolsky’s framework, devoting her book *Language Policy: Hidden agendas and new approaches* to the underlying links between language practice and language beliefs. She refers to those links as *mechanisms*, i.e. “rules and regulations, language educational policies, language tests, language in the public space as well as ideologies, myths, propaganda and coercion” (2006: 56). These mechanisms are overt as well as covert tools to influence language practices. However, the motivational force driving these mechanisms is always language ideology (Shohamy 2006: 57).

Weber and Horner (2012: 16–20) identify five different kinds of language ideology, which are closely interrelated. First, the *hierarchy of languages* refers to the assumption that linguistic practices can be divided into, e.g. ‘languages’ and ‘dialects’, which have a defined place in a hierarchical order, making a linguistic practice more prestigious in comparison to another. The most commonly used argument for such differentiation is that of mutual intelligibility – if two varieties are mutually intelligible, they are dialects of the same language, if they are not, they are separate languages. Chinese is frequently instanced as an example, for which this argument does not apply (more on that in chapter 3). Second, the *standard language ideology* assumes that languages are “internally homogeneous, bounded entities” (2012: 17), but in fact, “[l]anguage is open, dynamic, energetic, constantly evolving and personal. It has no fixed boundaries but is rather made of hybrids and endless varieties” (2006: 5), as Shohamy shows in the first chapter of her publication. The standard of a language prescribes what is correct and what is not. The standardization process is defined by the socio-political circumstances in which it takes place and has nothing to do with an “inherent

superiority" (2012:17) of one variety in contrast to another. Third, the *one nation–one language ideology*, already shortly touched upon in chapter 1.1, represents the belief that a certain language corresponds to a certain territory. Through the European nation-building processes in the 18th and 19th century language additionally became a pivotal part of national identity. Fourth, the *mother tongue ideology* assumes that humans have only one 'mother tongue', through which the norm is constructed that the majority of people are monolingual, which is, in fact, at odds with the truth. Fifth, the *ideology of purism* is inherent in one aspect each of the linguistic goals by Nahir (1984) presented in chapter 1.1 and the motivational factors for language policy by Fishman (2000) in chapter 1.2. It denies the fact that language is constantly changing and, similarly to the standard language ideology, prescribes what is 'good language'.

The elaborations in this chapter show how closely intertwined language policy and ideology are, contrary to the belief at the beginning of the research field in which it was assumed that language planning is an ideology-free zone. The many different language ideologies will accompany the elaborations on Chinese as a pluricentric language throughout this Master's thesis and, hence, will be touched upon again in chapter 2 and 3. But beforehand, some issues in terminology regarding language policy need clarification (see chapter 1.4).

1.4. Terminology Shift in Language Policy – Specifying Denotations

As the first chapter of this Master's thesis demonstrates, the field of language policy can look back on a rich research history. In this subchapter, the historical developments of the research field are interlocked with the changes in terminology, specifying the terminology used for this Master's thesis.

In the 1960s, language policy started to be on the radar of researchers. It was believed that especially the new nations that were emerging in Africa and Asia are confronted with language problems, which they could solve through planning. Language planning was perceived as an ideologically neutral activity that leads to modernization. Hence, language planning goals were mere linguistic goals. The dichotomy of corpus and status planning was crucial for the description of language planning activities. Jernudd and Neustúpny (1987: 71) suggest using the term *language planning* only for this early period of the field, so to differentiate it from the period after the paradigm shift. For the purpose of a more precise description of the field, this Master's thesis follows suit.

Over the decades, critical voices grew louder, disapproving of the study of isolated language. Cooper's pivotal work of 1989 led to a paradigm shift, moving away from focusing on language problems to a study of language behavior. This allowed for the inclusion of the complex social settings in which language is used and smaller entities than the nation-state could be put under scrutiny. Alongside this shift, the name of the research field changed to *language policy* (Ricento 2006; Shohamy 2006; Spolsky 2004).

Nowadays, the focus of language policy lies on the hidden structures of language policy. Hence, the differentiation between overt and covert, explicit and implicit language policy has become crucial and manifests itself in Spolsky's framework that consists of language practice, language beliefs and language management. Language ideologies have a strong influence on language behavior and are, therefore, of great interest to researchers today. Alternative terminology has been suggested (e.g. *language policy and planning* (Hornberger 2006) and *language policy and language planning* (Wright 2004)) to cover both the explicit and implicit aspects of the research field – this Master's thesis, however, uses the term *language policy* as an umbrella term for both, perceiving of language planning as a subcategory of it.

By presenting the historical development and research interests of language policy, chapter 1 should provide the theoretical background and the conceptual toolkit for dealing with Chinese as a pluricentric language. The terminology and categorizations presented here provide the foundation for dealing with standard language variation (chapter 2) and offer the tools for describing the Chinese language in chapter 3.

2. Standard – Language – Variation

> [W]e can recognize the ‘‘languageness’’ of X and Y, but we cannot pinpoint where one ends and the other begins. They are recognizable if seen from a convenient distance, but trouble begins when we zoom in and (the devil is in the detail!) find all kinds of disturbing elements: diglossias involving so-called dialects (which often dialects are not), bi- and multilingualism and their by-products of intermediate and mixed varieties, code switching, and dialect continua.
> (Tosco 2008: 1)

This chapter provides the theoretical core for this Master’s thesis which will be applied to Chinese in chapter 3. Tosco boils the matter down to its essence in the quote above; when dealing with language, standard language and standard language variation from afar, it may appear that there were easy answers to the questions at hand. However, once one gets closer, previously clear lines get blurred and unreflected presuppositions need to be put under scrutiny. Chapter 2.1 deals with the definition of language proceeding from Kloss’ understanding of language since the concept of pluricentricity can also be found in his paradigm. Kloss’ bipartite system for language definition will be discussed, Fishman’s adaptation included and an alternative approach by Croft will be offered. Chapter 2.2 evolves around the concept of standard language. The focus of this chapter lies on standard language ideology and its connection to the nation-state. In chapter 2.3, the theoretical background on pluricentricity is provided. The concept of standard language variation will be defined, various categorization approaches discussed, and a special focus will be put on the asymmetrical relation between pluricentric standard varieties. Chapter 2.4 offers a short summary and puts the provided theory into context for chapter 3.

2.1. Language

Before going into detail on pluricentricity (see chapter 2.3), one needs to elaborate on the definition of language. Questions such as ‘What is language?’ and ‘Where to draw the line between language and dialect?’ will never be fully answered, yet – or even more so: therefore –, a discussion is needed to find a suitable definition for the scope of this Master’s thesis that is not only compatible with the state of the art of language policy (see chapter 1) but also complies with the concept of pluricentricity (see chapter 2.3). Finding a suitable definition of language is the purpose of this subchapter.

Since this Master's thesis is devoted to pluricentricity, Kloss' dichotomic understanding of language (Abstand and Ausbau) is discussed in detail in chapter 2.1.1., followed by Fishman's critique and adaptation (Ausbau and Einbau) in chapter 2.1.2. In chapter 2.1.3, an alternative approach to language definition is presented that puts the human factor as the center of attention by relying on attitudes of speakers for language definition. Croft's conceptualization of language is in accordance with the theories presented in chapter 1 since human interference is paramount in language policy as well. Last but not least, chapter 2.1.4 points out why the two approaches – Fishman's Ausbau and Einbau as well as Croft's population theory – complement each other and allows for a more comprehensive description of Chinese in chapter 3.

2.1.1. Kloss' Understanding of Language – Abstand and Ausbau

Working with the concept of pluricentricity, there is no way around Heinz Kloss, who is one of the first researchers mentioning the term *pluricentric* and has influenced the field of linguistics for decades with his dichotomic understanding of language – *Abstand* and *Ausbau* (Kloss 1967, 1978). His basic interpretation of *language* is that it stands in a contrasting juxtaposition to *dialect*, making clear that a variety can only be one of the two, never both at the same time (1978: 23). This contrasting juxtaposition is also always necessary when dealing with the concepts of Abstand and Ausbau, since classification of a variety into either of the two categories is not possible without a comparison language.

Even though Kloss introduces the category of Abstand languages, he does not dwell on its definition for long since – in his opinion – it is a "predominantly linguistic concept" that linguists need to deal with (Kloss 1967: 30). The definition that he gives in his later work (1978) does reveal an important aspect of his conceptualization, though:

> Manche Idiome werden als „Sprachen" bezeichnet, weil sie von jeder anderen Sprache in ihrer Substanz, ihrem „Sprachkörper", so verschieden sind, daß sie auch dann als Sprache bezeichnet werden würden, wenn es in ihnen keine einzige gedruckte Zeile gäbe. (Kloss 1978: 24)

The important aspect is that having a written tradition and, as a consequence, having undergone a standardization process is irrelevant for the definition of an Abstand language. In other words, any language form – may it be a standard language or a dialect – may be defined as an Abstand language. Kloss' examples of typical Abstand languages always make use of comparisons across language families, such as the juxtaposition of Lower Saxon

(Germanic), Occitan (Romance) and Sorbian (Slavic) (1978: 24–25). Another noticeable facet in his definition is the attempt to hide the agent/s who decide on what a language is. Through the passive phrasing of "daß sie auch dann als Sprache bezeichnet werden würden", the question 'Who labels the variety under scrutiny?' is left unanswered. This way Kloss implicitly concedes the importance of language attitudes, whilst only two pages later explicitly rejecting it:

> [D]ie Einstufung durch die Sprecher der Sprache […] scheint mir entbehrlich, ja bedenklich zu sein. Wie schon dargelegt, hörten Okzitanisch und Niedersächsich, als ihre Sprecher begannen, sie als Dialekte zu empfinden und zu bezeichnen, nicht auf, Sprachen zu sein. (Kloss 1978: 27)

Again, the question remains who decides on the perseverance of Occitan and Lower Saxon as languages in the given example.

Kloss puts more focus on his second category, namely Ausbau languages, which is, hence, covered in more detail. Ausbau languages are not independent languages unless they have been altered through human intervention – this stands in stark contrast to Abstand languages that gain the label *language* merely through their being different enough from another speech form defined as such. This is why Abstand is defined as a linguistic and Ausbau as a sociological concept. The above mentioned human intervention interferes with the natural process of language change, mainly through explicit language policy – an example would be Czech and Slovak, two languages that bear this label only for Ausbau reasons (Kloss 1978: 25).

The two categories also differ in their elemental character. Abstand is a bipartite division – either a language is an Abstand language or not – whereas Ausbau languages undergo a process, hence the concept needs to be dealt with as a continuum. Kloss (1978: 32) suggests that the minimum degree of Ausbau – in other words, the point at which one can speak of an independent Ausbau language – is reached when two of the following three requirements are fulfilled:

1. the variety in question is used in at least one magazine for non-belletristic content
2. the variety in question is a subject in school, acts as a medium of instruction and is used in the teaching material in primary education

3. the variety in question is frequently used in situations such as preaching, broadcasting and presentations in educational settings

(Kloss 1978: 32; shortened and translated by author)

The aim of languages that have fully undergone the Ausbau process is to be "standardisierte Werkzeuge literarischer Betätigung" (Kloss 1978: 25). This implies that there is a finite stage of the Ausbau process. In order to reach this stage, the population needs to become literate as a fundamental prerequisite. Subsequently, a variety has to undergo the following steps:

1. Standardization of spelling
2. Standardization of speech form (morphology, lexicon):
   either monodialectal or pluridialectal basis
3. Ausbau:
   new rhetorical devices and new fields of application

(Kloss 1978: 37; shortened and translated by author)

Furthermore, Kloss (1978: 31) separates "Nur-Abstandsprachen" (Abstand-only languages) from "Nur-Ausbausprachen" (Ausbau-only languages) and languages that are defined as independent languages on the grounds of both their Abstand and Ausbau. This distinction lacks awareness regarding the necessity of a comparison language for classifying. By consciously choosing certain language combinations, one may influence the classification. For instance, Kloss assumes that Slovak is an Ausbau-only language in comparison to Czech (1978: 26). This may be correct, yet when substituting Czech with e.g. Chinese, Slovak needs to be classified as an Abstand language. So, Kloss' dichotomic understanding of language is actually a matter of comparison.

2.1.2. Fishman's Adaptation of Kloss – Ausbau and Einbau

Fishman (2008) revisits Kloss' theoretical contribution in his article *Rethinking the Ausbau–Abstand dichotomy into a continuous and multivariate system* and comes to the acknowledgement that

> [p]erhaps our own relatively poor German and the masterful reputation of Kloss, both as a typologist (with a particular fondness for dichotomies) and as a Germanist, have long misled us into not recognizing the logical inconsistency into which he misled us by stressing both *Ausbau* and *Abstand* as if they were two equally prominent and legitimately opposite ends of one and the same dimension. (Fishman 2008: 18; emphasis in original)

In Kloss' defense, he did mention the fundamental difference of the two categories, implicating that the two are not in the same sphere:

> Das Problem des sprachkörperlichen Abstandes zwischen zwei Sprachvarianten ist primär ein rein linguistisches und gehört als solches in diese rein sprachsoziologisch ausgerichtete Schrift nicht ohne weiteres hinein. (Kloss 1978: 63)

Yet, the way the two concepts were generally presented by him and how his works were interpreted, did just that – putting Abstand and Ausbau on 'one and the same dimension'.

In a theory on language policy – more precisely on corpus planning – there is no space for the Abstand concept because it does not incorporate any human interference. Instead, Fishman suggests that Ausbau is "the concern for fostering dissimilarity-focused interventions" and its counterpart Einbau "the concern for fostering *similarity*-focused emphases" (2008: 18; emphasis in original). This allows for a continuum that consists of various "'degrees' of *ausbauness* and degrees of *einbauness*" (Fishman 2008: 19; emphasis in original). Fishman made use of this dichotomy in a previous work on motivational factors for language policy (independence/interdependence), already elaborated on in chapter 1.2.

Fishman (2008: 22) comes to the conclusion that the occurrence of einbauization is "a much rarer phenomenon" in Europe than ausbauization – suspecting that the literacy factor may have led to stronger Ausbau desiderata on the continent and speculating that in regions with a rather short history of nation building, such as Africa and Latin America, one may find more examples of einbauization. Tosco criticizes this assumption with a rather negative standpoint on humanity: "I suspect that the quest for distinctiveness and separateness are much more entrenched in our behavior (linguistic and not) than the aspirations to likeness, homogeneity, and uniformity." (2008: 2) Yet, Fishman is able to name some European examples for einbauization:

> The Samnorsk efforts in Norway, the current Romanian treatment of Moldavian, the current Serbian treatment of Bosnian (or the former Serbian treatment of Croatian before the collapse of the larger Yugoslavia), the former Soviet treatment of both Ukrainian and Belarussian vis-à-vis Russian (continuing now in independent Belarus) are all examples of such *Einbau* efforts." (Fishman 2000: 45; emphasis in original)

Chinese may also be a case of einbauization (García 2015: 357) – more on that in chapter 3. Furthermore, a straightforward reason for finding more examples of ausbauization than

einbauization may simply be that making out something that wants to stand out is easier than noticing something that wants to fit in.

2.1.3. Croft's Alternative – Language Definition Based on Language Attitudes

This Master's thesis has thus far only presented internal definitions of language, for which language-internal characteristics and the factor of mutual intelligibility are paramount. This holds true for both Kloss and Fishman, even though human interference does have a place in their paradigms, too. External definitions, in contrast, put the human factor in the center of attention by relying on attitudes of speakers for language definition. In this subchapter an alternative to the previously discussed internal definitions will be offered (Croft 2000) that forgoes the static perception of language, the focus on language-internal characteristics and mutual comprehensibility.

William Croft (2000) takes a theory of evolutional biology − namely the population theory of species (Hull 1988) − and applies it to linguistics for an evolutionary model of language change. He criticizes the contrasting essentialist view on species due to its focus on structural characteristics. In order to define a species, the essentialist approach requires the description of its prominent structural features – simply put, if an individual organism has the prominent structural features of a defined species, it belongs to this species; if it does not have them, it does not belong to the species. This approach meets several issues, though:

> The essentialist view ran into problems due to various sorts of structural variation among species, including high degrees of structural variation among individuals in a population and also among different life-stages in an individual in a population (for example, a caterpillar and the butterfly it turns into, or a species that changes sex over its lifetime […]). (Croft 2000: 13)

In the population theory of species, on the other hand, the only relevant property is that of reproductivity – if individual organisms interbreed, they belong to the same species; if they do not, they belong to different species. This makes the internal characteristics of individual organisms irrelevant, but focus is put on its interaction with other individuals instead. The advantages of this approach are tremendous:

> Individuals can vary in enormous ways in physical structure (and behavior), but as long as they form a population in the evolutionary sense, they are members of the same species. Conversely, individuals may be structurally extremely similar, but if they come from two distinct reproductively isolated populations, they are members of different species. (Croft 2000: 13)

The first-mentioned species – a species that contains individual organisms with great variety, yet they interbreed – may also be called a *polytypic species*. From an essentialist standpoint, such a species would be identified as several independent species due to their structural differences. The species mentioned later in the quote – individual organisms are very alike, yet they do not interbreed – can be called *sibling species*. An essentialist biologist would classify the individual organisms as part of one and the same species, even though they are highly intersterile. (Croft 2000: 14–15)

Another important concept in the population theory is that of *variety*. Varieties are subparts of species that may stop interbreeding because of structural and behavioral differences (developed over time due to e.g. geographical isolation). This is when a species splits up into two *daughter species*, provided that the original species does not exist anymore and, hence, there is no need for discussion on which of the two daughter species is the continuation of the original species. From this standpoint one may conclude that "[v]arieties are merely incipient species. Not all varieties become species, but all species at one time were varieties" (Hull 1988: 96).

In order to transfer the population theory to the field of linguistics, Croft (2000: 17–18) was in need of a language definition that does not work with language-internal characteristics as the paramount factor for defining a language – this approach is identical to the essentialist theory of species in biology. Instead, he uses a social definition of language postulated by Chambers and Trudgill (1980). They object to the widespread assumption that "a language is a collection of mutually intelligible dialects" (1980: 3) because mutual intelligibility is not a suitable benchmark for language definition. It is a matter of degree (and not a bipartite classification), mutual intelligibility may not exist on both sides equally and most crucially:

> Mutual intelligibility will also depend, it appears, on other factors such as listeners' degree of exposure to the other language, their degree of education and, interestingly enough, their willingness to understand. (Chambers & Trudgill 1980: 4)

As an alternative, Chambers and Trudgill (1980: 9–12) introduce the two concepts *autonomy* and *heteronomy*, both revolving around speakers' language attitudes. The speakers of an autonomous variety perceive their variety as an independent language, disregarding its structural similarity to other varieties, whereas speakers of a heteronomous variety are dependent on the former, perceiving their variety as part of it, no matter their structural

differences. The two authors emphasize that "[s]ince heteronomy and autonomy are the result of political and cultural rather than purely linguistic factors, they are subject to change" (Chambers & Trudgill 1980: 9).

With this language definition in mind, Croft (2000: 17–18) suggests that the linguistic equivalent to the biological factor of interbreeding is *communicative interaction*. In other words, linguistic interbreeding takes place when "every speaker perceives every other speaker as someone he or she should be able to communicate with by using what they perceive as the same language" (Croft 2000: 18).

On the basis of these assumptions, Croft's application (2000: 16–17) of the biological theory onto linguistics is easily comprehensible. *Sibling languages*, just as sibling species, are two varieties that are structurally so similar that they would be defined as varieties of the same language if an internal language definition was applied. Their speakers, however, perceive their varieties as distinct languages. This perception might not be shared amongst all speakers of the variety and/or amongst the speakers of the various sibling languages, which can cause conflicts – e.g. Macedonian and Bulgarian, Hindi and Urdu, or Malay and Indonesian. A *polytypic language*, just as a polytypic species, would, on the other hand, be identified as several distinct languages were it for structural characteristics, but its speakers perceive of it as a single language. The two examples for polytypic languages given by Croft are both essential for this Master's thesis: firstly, Chinese and its so-called dialects, and secondly, diglossia.

The criticism evoked by Croft's approach to language definition shall not be withheld at this point since it would lead to a biased presentation. Tosco (2008) introduces Croft in his article *Introduction: Ausbau is everywhere!* only to reject his ideas at the end and stick to the internal definition of language, heavily depending on mutual intelligibility, instead. His main point of criticism is that, in his opinion, Croft's approach bases its language definition merely on language attitudes – something that may always be inconsistent amongst communities and bound to change frequently (Tosco 2008: 2–5). This does not necessarily hold true – language attitudes are in the center of Croft's approach, but the single most important entity is that of communicative interaction. This does not imply that all speakers of a language need to interact with each other in order to prove that they speak the same language (similar to that not all individual organisms need to interbreed in order to verify that they all belong to the same

species) (Croft 2000: 17–18). The mere possibility of communicative interaction is key – which is highly dependent on the speakers' perception of their variety. If both participants in a dialog are convinced that they speak different languages, i.e. that there is no chance for communicative interaction, this does hinder mutual intelligibility, no matter the structural similarities between the two varieties in question. So, in fact, Croft's theory is not based on the speakers' perception but rather on the possibility of communicative interaction – however, the latter is expressed through the former.

The second piece of Tosco's criticism refers to the concept of polytypic languages. He argues that Croft oversimplifies the identity construction of its speakers (Tosco 2008: 4). Whereas individuals may identify themselves as speakers of a certain (polytypic) language, they may still be well aware of the fact that mutual intelligibility, i.e. communicative interaction, is not possible amongst all its speakers. This discrepancy is indeed an issue, however, – and this is just an assumption – the common belief of belonging to the same speaker community may override the impossibility of communicative interaction.

Tosco's example, cited below, further evokes questions in regard to minority languages:

> Let us imagine a particularly "aggressive" and demographically powerful community of X-speakers which, any linguistic (structural) difference notwithstanding, considers the neighboring, demographically weaker variety Y as a "dialect of the same language" (X, obviously). Would we still have here polytypic languages? (Tosco 2008: 4)

In this case, the willingness/belief of belonging to the same speaker community is forcefully inflicted on the speakers of the weaker variety Y by the speakers of the powerful variety X. The result might be that at some point speakers of the weaker variety Y themselves believe that they speak a dialect of variety X[6], creating a polytypic language in Croft's sense since variety X shows great diversity in its structure. This forceful infliction of beliefs seems unfair to the members of the minority group, but this is what language policy is all about. Furthermore, what reason is there to stick to the definition of a variety as a language when not even its speakers believe any longer that it is, in fact, a distinct language? In Croft's paradigm there is no reason for doing so, hence, the variety is not an independent language anymore. Kloss objects to this, insisting that e.g. Occitan and Lower Saxon remain distinct

[6] An explanation for this result offers e.g. the semiotic processes in ideologies of linguistic differentiation – iconicity, recursiveness and erasure – by Gal and Irvine (1995).

languages, even though their speakers believe they speak a dialect (see quote in chapter 2.1.1, Kloss 1978: 27). As criticized before, he does not give any reasons for this classification, though.

2.1.4. Complementing Paradigms – The Two Approaches Combined

Even though the paradigms presented in this subchapter may be perceived as opposing each other, they do actually complement each other. Kloss' concept of Ausbau and Fishman's adapted notion of Einbau have nothing to do with language definition per se but rather describe processes of language policy that, however, do have great influence on language definition because Ausbau and Einbau attempts change the perception of the varieties in question.

The advantage of Croft's language definition for this Master's thesis is obvious. His approach incorporates the human factor, whereas language-internal definitions exclude human intervention from their paradigms, presenting humans as if they were meddling with the otherwise easy task of defining languages – but without humans there would be no language. When reading about Chinese, one often finds statements such as 'Chinese consists of several varieties that could be defined as distinct languages but…' and a list of language political reasons follows (e.g. Crystal 2010: 322). Incorporating these reasons in the approach of language definition may allow for a more comprehensive description of languages.

One way of offering a more comprehensive description of languages is that "[t]he social definition makes predictions of likely historical developments whereas the structural definition does not" (Croft 2000: 18). It is likely that sibling languages develop more structural differences over time since there is usually a lack of communication between speakers of the sibling languages. Polytypic languages, on the other hand, may split up either due to their great inherent diversity, or become more homogenous. The analysis of Ausbau/Einbau attempts, in addition, can indicate in what way the perception of the affected variety or varieties may change in the future.

2.2. Standard Language

In order to deal with pluricentricity – a concept devoted to standard language variation –, not only does language need to be defined (see chapter 2.1), but also standard language. Linguistics has considered itself a descriptive discipline for many decades – if not centuries –, being oriented towards natural sciences. Hence, the focus was put on the internal description

of languages, i.e. their form and system, whereas linguistic prescription was deemed irrelevant (Milroy & Milroy 1985: 5–7). However, "[w]hen we view language as fundamentally a social phenomenon, we cannot then ignore prescription and its consequences" (Milroy & Milroy 1985: 11). Therefore, chapter 2.2.1 firstly looks at how standard and standardization can be defined. Chapter 2.2.2 then puts the previously mentioned standard language ideology (see chapter 1.3) under scrutiny. One very crucial factor in the standardization process will be emphasized in chapter 2.2.3 – the nation-state. Lastly, a study is presented in chapter 2.2.4 that shows what lay people from various countries perceive as standard and what characteristics they attribute to their standard languages (Smakman 2012).

2.2.1. Defining Standardization

As with how to define language, there are many ways how to define standard language. One definition has already been presented in this Master's thesis – namely that of Kloss in chapter 2.1.1 including the minimum degree of Ausbau, the aim of Ausbau languages and the steps towards a standard language. The concept of Ausbau and standardization are closely intertwined since without a standard language there would be no Ausbau language. Many different definitions for standard language have been proposed, however, many of them are interwoven with certain ideological beliefs (see chapter 2.2.2) that are themselves connected to the phenomenon at question. To name only one, Crystal (2008: 450) defines standard languages by their prestige.

In chapter 2.1 it was necessary to move away from language-internal approaches of language definition, since they contain the underlying belief that a variety must be defined as a distinct language if its characteristics are different enough in comparison to another variety, i.e. if mutual intelligibility is not possible, no matter how people who use that variety perceive of it. This does by far not imply that language attitudes are an ideology-free zone – quite the contrary, as discussed in chapter 1.3. It is simply important to remember that "any enterprise which claims to be non-ideological and value-neutral, but which in fact remains covertly ideological and value-laden, is the more dangerous for this deceptive subtlety" (Joseph & Taylor 1990: 2).

In the case of standard language, this hidden ideological thinking is not inherent in language-internal methods but comes to the surface in language-external approaches. A definition of standard language that escapes the grasps of the standard language ideology by

focusing on language-internal factors is that of Milroy and Milroy in their book *Authority in language: investigating language prescription and standardisation* (1985). They state that "the process of language standardisation involves the suppression of optional variability in language" (1985: 8). In other words, in a language there are naturally oftentimes several options for referring to the same meaning, e.g. due to regional variability. Through the standardization process only one of these options is deemed correct – creating the standard, on the one hand, but also the non-standard[7], on the other. The reason for this is very simple: "[D]ialects cannot be labeled 'non-standard' unless a standard variety is first recognized as definitive and central." (Milroy 2001: 534) Language-internally speaking, the variability of the non-standard and the lack thereof in standard varieties is a paramount difference between the two (Milroy & Milroy 1985: 8).

The aim of creating a standard and of suppressing the optional variability in language is functional efficiency: "Ultimately, the desideratum is that everyone should use and understand the language in the same way with the minimum of misunderstanding and the maximum of efficiency." (Milroy & Milroy 1985: 23) Yet, Standard English – and this may hold true for standard varieties of any language – is not as well understood as commonly assumed (e.g. Milroy 1984). Kloss, as cited in chapter 2.1.1, points out that the purpose of a standard language is to offer tools to create great literature, whereas Milroy (2001: 535) denies that and refers to the "economic, commercial and political [goals]" that are being served by the standardization process. Chapter 1 has offered some insight especially into the political aspirations of creating a standard language and chapter 3 will show that the goals regarding economy, commerce and politics were paramount for the development of the Chinese standard varieties.

As with Ausbau languages, standard languages also go through various stages until they reach their status as a standard language. For that matter two conceptualizations shall be presented here that show that constructing such stages depends strongly on the standpoint and the focus of their developers. One of them (Haugen 1966) aims at describing the stages of the standardization process itself, whereas the other (Milroy & Milroy 1985) perceives its categorization as the stages of implementation of the standard language, hence

[7] It needs to be pointed out that *non-standard*, similar to *dialect*, is a word that carries negative connotation (Weber & Horner 2012: 6). For lack of a neutral reference this Master's thesis uses both terms in contrast to its 'counterpart' *standard*, whilst being aware of the ideological aspects it conveys (see also chapter 2.2.2).

they describe slightly different aspects. What pertains for both is that their stages do not necessarily succeed one another but oftentimes overlap, temporally speaking.

The first of the two is by Haugen (1966) who proposes a simple, yet comprehensive conceptualization. He assumes that standard languages go through four stages. The *selection of norm* is the first and a highly relevant step – it not only determines what variety (or varieties) is (are) chosen as a standard (similar to the second stage of the development of Ausbau languages by Kloss (see chapter 2.1.1) – the standardization of speech form – that can either have a monodialectal or pluridialectal basis) but it also decides on which group/s of people will gain an advantage through the standardization process and who will be confronted with disadvantages. Haugen states that "[i]t may often be a question of solidarity versus alienation: a group that feels intense solidarity is willing to overcome great linguistic differences, while one that does not is alienated by relatively small differences" (1966: 932). Stage 2 and 3 are that of *codification* and *elaboration*, which he describes concisely as follows: "[C]odification may be defined as *minimal variation in form*, elaboration as *maximal variation in function*" (Haugen 1966: 931; emphasis in original). Here it is worth noting that stage 2 is identical with the important aspect of standard language emphasized by Milroy and Milroy (1985) above. Last but not least is the stage of *acceptance*, without which there would be no standard language. From Table 1 it appears that two stages each are concerned with either form (*selection* and *codification*) or function (*acceptance* and *elaboration*). Additionally, *selection* and *acceptance* are dependent on language-external factors whereas *codification* and *elaboration* are attributed to language-internal aspects.

	Form	**Function**
Society	Selection	Acceptance
Language	Codification	Elaboration

Table 1: Haugen's four aspects of language development (1966: 933)

The second conceptualization that shall be addressed is by Milroy and Milroy (1985: 27–28). They assume that the implementation of standard language undergoes three stages – *selection*, *diffusion* and *maintenance* –, all of which are kick-started by "a need for uniformity that is felt by influential portions of society" (1985: 27). The Milroys' first step *selection* incorporates Haugen's stages of both *selection* and *acceptance*, so the two stages that depend on societal factors. In this stage a variety is firstly selected as the standard – at the beginning

there might be competing standards, since different groups of people may choose different varieties as their standard, however, eventually one standard variety prevails – and secondly, the standard is accepted by influential individuals. The second stage, *diffusion*, in which the previously chosen standard is spread on a geographical as well as social scale (in schools, through media etc.) is not regarded at all in Haugen's conceptualization. However, as pointed out in chapter 1.1, language spread is an important part of language policy actions. *Maintenance* is the last of the three implementation stages by Milroy and Milroy (1985). Their examples of how a standard is maintained (through the elaboration of function, prestige, the idea of 'correctness' etc.) show that the standard language ideology plays the single most critical role in maintaining the standard's status quo. The *codification* (as well as the *elaboration*) process mentioned in Haugen's conceptualization are included in this last stage by Milroy and Milroy, since it only further manifests language beliefs by offering e.g. dictionaries that can 'prove' a variant's correctness.

As became evident in this subchapter, defining standard language without reference to its inherent ideology is a sheer impossibility. It is all there is – a standard language without anybody believing in it simply does not exist. Due to its importance not only in regard to standard language itself but also to pluricentricity, the following subchapter is dedicated to standard language ideology.

2.2.2. Standard Language Ideology

As one could notice in chapter 2.2.1, standard language is so tightly interwoven with its ideology that a description of the matter at hand is impossible without reference to ideological aspects. Milroy and Milroy (1985) point out that,

> [t]he ideology of standardisation, whatever merits there may be in it, tends to blind us to the some-what ill-defined nature of a standard language, and may have some undesirable consequences in that it leads to over-simplified views of the nature of language, evidently held even by highly educated speakers. (1985: 26)

For that reason, this subchapter discusses the phenomenon *standard language ideology* in more detail.

Deumert (2003) concisely defines standard language ideology as "a metalinguistically articulated and hegemonistic belief in linguistic uniformity, correctness and authority" (2003: 37). First recordings of such a discourse are from the first half of the 16th century, however, at

that time this belief was only shared by some individuals. Only by the late 17th and early 18th century was the standard language ideology fully established – in Europe that is to say. Phenomena that developed alongside with it are what Deumert calls *rituals*: codification rituals, pedagogical rituals and communication rituals. These rituals would fall under standard language maintenance in Milroy and Milroy's conceptualization mentioned in 2.2.1, since the standard is being further manifested by creating dictionaries, grammars etc., spreading the knowledge of the standard to more people and lastly diffusing the standard across social groups. The three rituals play a crucial role in normalizing the development of the standard language ideology (Deumert 2003: 37–48). This further shows, that the stages of the standardization process presented in 2.2.1 are not sequential but occur simultaneously.

Milroy and Milroy (1985: 22–23) speak of standardization "as an *ideology*, and a standard language as an idea in the mind rather than a reality – a set of abstract norms to which actual usage may conform to a greater or lesser extent" (emphasis in original). This continuum of conformity refers to the written and spoken form of the standard. The authors (1985: 22) assume that full standardization can only be achieved in written form, namely in the spelling system, as neglecting the norm leads to penalties and disadvantages for the users (e.g. bad school grades). With this perspective in mind, the consequence is that "the only fully standardised language is a dead language" (1985: 22). On the opposite side there is the spoken form, in which many deviations of the norm are tolerated, hence, it is legit to speak a standard language with various accents or even speak different standard varieties altogether. The different variants can be traced back to "at least three dimensions: geographical, social and situational" (Milroy & Milroy 1985: 54). In contrast to the assumption that standard language suppresses optional variability, as discussed in chapter 2.2.1, this opens up the possibility of having multiple standard language varieties in a language – the starting point for pluricentricity (see chapter 2.3).

Even though variation in the standard language is tolerated to some extent, the ideology contains a firm belief in correctness. This belief manifests itself in discussions about which of the several variants at choice is correct – it is common sense that some variants are correct, and others are not. The foundation for discussion is provided by the standard language ideology. Arguing with 'common sense' makes any justification unnecessary because 'common sense', in itself, contains the assumption that everybody possesses the same

knowledge regarding the topic in question. If one were to bring up an argument that contradicts that common sense, one would simply be perceived as not being part of the common culture and, hence, ignored. (Milroy 2001: 535–536)

Another aspect that is very distinctive for the standard language ideology is that of prestige. As mentioned in 2.2.1, some researchers define standard language through its prestige, oftentimes equating the standard language with the most prestigious variety. However, prestige is not an inherent property of language, in fact, language itself cannot carry any degree of prestige – it only receives prestige through the social positioning of its speakers (Milroy 2001: 532–533). So, "although involved in the standard *ideology*, [prestige] has nothing whatsoever to do with the process of *standardization*" (Milroy 2001: 533; emphasis in original).

So far, the problematic endeavor of defining standard language has been discussed and the reason for that difficulty addressed – namely the ideology behind standard language. But what role does the nation-state play in the creation of a standard language and the construction of a standard language ideology? The next subchapter (chapter 2.2.3) is dedicated to this question.

2.2.3. Standard Language (Ideology) and the Nation-State

The discussion in previous chapters shows that the nation-state is one of the most pivotal factors for the distinction of different languages and the development of standard varieties – if not *the* most pivotal one altogether. Milroy (2001) contemplates that "it may not be a linguistic universal that language necessarily splits up into different languages as part of its intrinsic nature as language" (2001: 541) and that "we may have been forcing languages into greater states of orderliness and definitiveness than they actually possess qua languages" (2001: 540). The reason for that is the "politics of state-making" (Weber & Horner 2012: 29) that requires "a private-property language", as Anderson (2006: 68) puts it.

Anderson (2006: 5) suggests to not perceive of nationalism in the realm of political philosophies such as liberalism and fascism, but rather in connection with communal phenomena, like kinship and religion. He defines nation as an "imagined political community" (2006: 6) – an approach that fits very well into the overall positioning of this Master's thesis in which the definition of language is dependent on its speakers (see chapter 2.1.3) and standard language is identified as a mere ideology (see chapter 2.2.2). The nation is imagined,

Anderson (2006: 6) further elaborates, because its members do not know most of their fellow members, yet they believe in their connection to each other. There are two characteristics of these imagined communities: they are limited and sovereign. Its limitedness stems from the imagination that one own's nation is restricted in territory by the existence of other nations. Furthermore, the aim of nations is never to include the entire world population in their notion of community – limits are the logical consequence to that. The idea of sovereignty originates from the period of Enlightenment in which the original world order was in turmoil and new entities had to be formed. (Anderson 2006: 7)

As mentioned in chapter 2.2.2, the urge to standardize languages was a predominantly European one and literacy – an areal phenomenon, as Fishman (2008: 21–22) points out – has strong roots in Europe as well. Mühlhäusler (1996) argues that the entire field of linguistics adheres to Eurocentric assumptions, namely:

1. the belief in distinct word classes
2. the belief in the possibility of using the same descriptive labels for all languages
3. the belief in the separability of language and other non-linguistic phenomena
4. the belief in the existence of separate languages (Mühlhäusler 1996: 328)

Additionally, many European languages have gone through a *natural* process of dialect mixture between the 13th and 16th century due to non-linguistic factors (mainly economic and social reasons). Creating a standard language on varieties that have been in touch and merging for centuries is much easier than *artificially* mixing dialects, on which to base a standard language – the latter being the case in many non-European contexts. (Thomason & Kaufman 1988: 209–210)

In contrast, one needs to point out that the research on and description of diachronic varieties of a language have a crucial role to play in the construction of standard languages in that they provide a history to the standard language. In other words, the historicization of a standard language ensures its legitimization. It "requires that it [=the language] should possess a continuous unbroken history, a respectable and legitimate ancestry and a long pedigree" (Milroy 2001: 548–549). Milroy (2001) further elaborates that "*these codifications are themselves part of the process of the legitimization of the standard language in its function as the language of the nation state*" (2001: 548; emphasis in original). Chapter 3 will present in detail how Chinese is depicted for these purposes.

2.2.4. Study on Attitudes on Standard Language

Since the attitudes of lay persons are crucial not only in the approach to language definition chosen for this Master's thesis (see chapter 2.1) but also in the context of standard language (ideology) (see chapter 2.2), the study conducted by Smakman (2012) will be briefly presented in this chapter. His study objective was to find out how lay persons define 'standard', therefore the attitudes of 1,014 non-linguists from seven different countries (England, Flanders, Japan, the Netherlands, New Zealand, Poland and the USA) were collected with a questionnaire. This questionnaire contains open and multiple-choice questions that were adapted to each national context and it was answered via e-mail or through a paper survey. Most participants in the survey were university students. In roughly half the countries the participants were located in only one city, whereas in the other half they lived in various cities. The two research questions were "What are the intrinsic qualities of the standard language?" and "What are the characteristics of speakers of the standard language?" (Smakman 2012: 30).

The following characteristics of the standard from the viewpoint of non-linguists could be made out to a varying degree in each country: lingua franca, correct, media language, opposite of dialect, non-regional, formal, qualitative features (e.g. "It does not have those illogical double negations" (2012: 39)), rare/non-existent, accepted, external language, informal, different, capital and social class. The distribution of answers is shown in Table 2, visualizing how varied the answers are depending on the nation-state. The only quality of the standard language that was indicated by at least 20% of the participants of each country and was a popular standard language description in all the countries is 'lingua francaness'. There are several qualities that are commonly ascribed to standard language, yet, this survey shows, that hardly any are universal – one example would be that of the strong connection between the standard and a certain city/area. In contrast, the author also assumes that "[t]he common association of standard languages with non-regionality may only be true for old standard languages" (2012: 25).

Characteristic	EN	FL	JP	NL	NZ	PL	US
lingua franca	24.0	25.0	35.1	20.6	*33.7*	38.1	*22.1*
correct	32.0	11.0	14.9	38.2	–	*55.8*	20.0
media language	10.5	14.0	*47.0*	10.3	1.1	6.2	3.2
opposite of dialect	20.0	18.0	0.6	19.1	2.2	27.4	5.3
non-regional	*44.5*	*27.0*	5.4	*39.7*	5.6	8.0	17.9
formal	7.5	18.0	19.6	5.1	1.1	–	6.3
qualitative features	9.5	7.0	–	4.4	5.6	17.7	11.6
rare/non-existent	5.0	21.0	–	2.9	6.7	0.9	3.2
accepted	–	11.0	–	8.8	3.4	1.8	2.1
external language	–	19.0	–	–	14.6	–	2.1
informal	20.5	–	–	–	27.0	–	6.3
different	–	–	–	–	15.7	–	2.1
capital	24.0	–	18.5	–	–	–	–
social class	32.0	–	–	14.0	–	–	3.2
various	10.5	18.0	6.5	30.9	21.3	46.0	52.6

* The abbreviation for each country is as follows: EN = England; FL = Flanders, JP = Japan, NL = Netherlands, NZ = New Zealand, PL = Poland, US = United States. The number of respondents who answered this question: EN = 200; FL = 100; JP = 168; NL = 136; NZ = 89; PL = 113; US = 95.

Table 2: Characteristics of the standard language (Smakman 2012: 36)
The most popular answer in each country is written in Italics, multiple answers were allowed per participant.

The outcome of the study regarding its second question, the characteristics of speakers of the standard language, is similarly dependent on the respecting countries. Four aspects were put under scrutiny: residence of the speakers, profession of the speakers, famous speakers, and sex of the speakers. Participants of five out of seven countries stated that the speakers of the standard language are locals of a specific city and/or region. The profession that represents the standard language the most is newsreaders, which is the most frequently given response in three countries, but also teachers (in the USA) and the queen (in England) made it to the top of the list. Similar to this outcome, famous speakers mentioned as speakers of the standard language are oftentimes TV newsreaders and presenters, but again, not only. Respecting the sex of standard language speakers, about two thirds of participants in each country chose the answer that both men and women speak the standard. For more detailed information on the outcome regarding the second research question see Smakman (2012: 43–49).

Overall, the conductor of the study could make out two types of standard language. The inclusive standard language represents a sociological perspective in that it allows for

widespread communication, whereas the exclusive standard language serves a symbolic purpose since it is the language of only a minority.

Category of characteristics		EN	FL	JP	NL	NZ	PL	US
inclusive	lingua franca media language accepted informal	22.9	26.5	*55.6*	22.5	*47.2*	22.8	21.3
exclusive	correct opposite of dialect non-regional formal rare/non-existent social class	*58.8*	*50.3*	27.4	*61.4*	11.4	*45.6*	*35.3*
other		18.3	23.3	16.9	18.2	41.5	31.6	43.3

* The number of characteristics put forward per country: EN = 480; FL = 189; JP = 248; NL = 245; NZ = 123; PL = 228; US = 150.

Table 3: The inclusive and exclusive standard language (Smakman 2012: 50)
Again, the most frequent answer in each country is written in Italics.

Table 3 shows that these two types are not mutually exclusive, yet it indicates that older standard languages have the tendency to be rather exclusive, whereas relatively newly established standard languages follow the inclusive aspects. Smakman (2012) explains how these seemingly opposing categories of characteristics can be comprised under one roof:

> The inclusive standard language is sociologically standard in the sense that many people speak it, but it is far from homogenous and thus not standard from a structural linguistic point of view. And while the exclusive standard language is highly homogenous and thus linguistically standard, it is not standard in the sense of widely adopted. These two views can nevertheless be reconciled by regarding the exclusive standard language as the prototypical standard language, which is part of the wide range of varieties in the inclusive realm of varieties, i.e. varieties that are convenient tools of communication with anyone in the speech community or country. (2012: 52)

For the scope of this Master's thesis, this study displays that the tension between language-internal attributions and language-external factors does not only exist in theory – in the definition of language (chapter 2.1) and in the dealings with standard language (chapter 2.2) – but also in practice, surfacing in language attitudes. Furthermore, it tells a lesson that is important to keep in mind – attributions to standard language are highly dependent on the standard language in question and they are manifold and oftentimes contradicting, yet they exist side by side.

2.3. Standard Language Variation

As a third step, after dealing with the concepts language (chapter 2.1) and standard language (chapter 2.2), this subchapter is dedicated to standard language variation and, hence, pluricentricity. At first glance, standard language variation stands in stark contrast to the elaborations presented in chapter 2.2, in that "the process of language standardisation involves the suppression of optional variability in language" (Milroy & Milroy 1985: 8). It follows that "a standardized variety must be invariant" (Milroy 2001: 534). To resolve this contradiction, Milroy (2001: 534) points out that, even though invariance is a linguistic goal of the standardization process, it cannot ever be fully reached in practice. This fact in combination with the complex dynamics of language policy, described in chapter 1, and the paramount factor of nation-states (see chapter 2.2.3) leads to the phenomenon of pluricentric languages. Chapter 2.3.1 introduces the concept of pluricentricity by addressing its history, some key aspects (e.g. centers) and terminological differentiations. In chapter 2.3.2 various conceptualizations of pluricentric languages are discussed, such as types of pluricentric languages and levels of pluricentricity. Lastly, chapter 2.3.3 is dedicated to the asymmetrical relation between pluricentric standard varieties and its consequences.

2.3.1. Defining Pluricentricity

Pluricentricity describes the variation of standard languages that occurs when they are being used in more than one region. These regions exert influence on their standard varieties, which develop "grammatical, lexical, phonological, graphemic, prosodic, and pragmatic" (Clyne 2004: 297) differences to their fellow regions' varieties over time. Such centers of pluricentric varieties

> can refer to defined sub-populations of a language community without a delimitable region (e. g. nomads like the Roma and the Sinti in Europe) or with dispersed settlements (e. g. Jews before the foundation of the state of Israel), as long as the language is standardized. Another possible reference are regions, mostly within a nation […]. Mostly however, the term center refers to nations or states. (Ammon 2005a: 1536)

Hence, the definition of standard varieties is frequently attached to the existence of independent countries as a prerequisite (see e.g. Ammon et al. 2004: XXXI; Muhr 2016: 20; Glauninger 2013: 2). Since the independence of certain regions in the Chinese-speaking world is a controversial topic (see introduction to chapter 3), to say the least, such an oversimplification must be avoided in this Master's thesis.

For doing so, Ammon's terminological differentiation (2005a: 1537) is very useful. He employs the term *plurinational* for varieties that are used in different nations (e.g. Spain and Colombia in the case of Spanish), *pluriregional* for varieties spoken in different regions, usually within the same nation (such as Bokmål and Nynorsk in Norway), and *pluristatal* for varieties which occur in several states that are not defined as nations (e.g. North and South Korea). *Pluricentric* forms the umbrella term for these different kinds of standard language variation. Additionally, these differentiations are not mutually exclusive but may apply simultaneously to a standard language, depending on the focus of research and the standpoint of the researcher. German provides an excellent example, since it is plurinational (Austria, Germany, Switzerland), pluriregional (e.g. North and South Germany) and it used to be pluristatal (East and West Germany). Due to the continuous speech area of German, the *pluriareal* approach is developed especially in the German context (e.g. Pohl 1997) and focuses on innerstatal differences and similarities across borders in the standard language.

Ammon (2005b: 30) also offers the tools to talk more precisely about the lexical and semantic differences in standard varieties. He suggests *national variant* to denote a specific speech form of a nation (e.g. *Karfiol* (cauliflower) for Austrian German), *national variable* for the entirety of variants existent in the standard varieties to denote the same meaning (e.g. 'stapler' – *Klammermaschine* (Austrian German), *Tacker* (German German), *Bostitch* (Swiss German)). Lastly, *national variety* refers to the entire standard language system. If the meaning stays the same across standard language varieties, it is an onomasiological variable (such as the stapler example); if the meaning changes while the expression stays the same, depending in which standard variety it is used, the variable is a semasiological one (e.g. the German verb *wischen*; in Switzerland this cleaning activity is done with a broom, in Austria and Germany with a cloth). In connection with the critique expressed above regarding the problematic nation terminology for the application of pluricentricity to the Chinese context, Ammon's notations are not ideal, even though the concepts behind them are very useful. Hence, this Master's thesis uses the terms *regional variant*, *regional variable* and *regional variety*, instead. That way, the issue of nationhood/statehood can be avoided while at the same time still allowing for the incorporation of Ammon's differentiation in this Master's thesis. Yet, it needs to be pointed out that *regional variant* does not refer to local non-standard expressions, as the term might imply, but implicates only standard speech forms, since pluricentricity is located within the realms of standard language/s.

As mentioned in chapter 2.1.1, Heinz Kloss (1967: 66–67) was one of the first to describe the phenomenon of pluricentric languages by its now commonly known name. He puts pluricentric languages in contrast to Ausbau languages, in that variation does not necessarily lead to independent Ausbau languages but may also cause the emergence of several standard varieties. Hence, pluricentric languages take up an intermediate position between languages with only one standard variety and separate Ausbau languages (or, in reference to the elaborations of Fishman (2008) in chapter 2.1.2: the intermediate position between highly diverse languages and the existence of a single standard variety through successful Einbau attempts). Clyne (2004) emphasizes that position as follows: "Any national variety of a pluricentric language is potentially a separate language." (2004: 296) Kloss points out the correlation between independent states and pluricentric languages, as many have done so after him: „Hochsprachen sind besonders dort häufig plurizentrisch […] wo sie die Amts- und Verwaltungssprache mehrerer größerer unabhängiger Staaten ist" (1967: 67).

However, Heinz Kloss was not the first to dedicate his work to the description of the pluricentric phenomenon, many researchers have done so ever since the 1950s. Einar Haugen describes the Norwegian language with its two standard varieties as early as 1959, referring to Bokmål und Nynorsk as two different "stylistic norms" (1959: 15). Russian, Marxist-oriented linguists were especially active in the field of pluricentricity, both regarding its conceptualization (e.g. national language versus nationality language (Ising 1987) – a highly relevant issue in the former Soviet Union) and its language-specific application (Stepanov on Spanish in Latin America and Spain (e.g. 1957) or Riesel on German, with a special focus on the Austrian variety (e.g. 1953), to name only a few).

The above-mentioned potential of a pluricentric language becoming a separate language is dependent on an interplay of various factors – interestingly, geographic distance is not a decisive one, since interaction and especially distancing attitudes play a more pivotal role. Ammon (2005a) describes the outcome depending on the existence (+) or non-existence (–) of factors with examples as follows:

> –Interaction → linguistic division (Netherlands/South Africa); +Interaction –distancing attitude → no linguistic division (Britain/USA, Austria/Germany); +Interaction +distancing attitude -> linguistic division (Luxemburg/Germany). Ammon (2005a: 1540)

Additionally, Ammon (2005a: 1540) points out that attitudinal factors (especially competition and tension) are so influential that “they can even be mirrored in the attitudes of learners of the respective languages”.

2.3.2. Categorizations of Pluricentric Languages

Many schemes have been proposed for categorizing pluricentric languages which look at the characteristics of pluricentric languages from various angles. The categorizations presented here may best help to put Chinese into the framework of pluricentricity, yet, they are not flawless, hence a discussion is added where needed.

Ammon (2005a) talks about *divided languages*. He states that divided languages can refer to two opposing meanings, only one of which representing pluricentric languages. The possibility of a pluricentric language becoming an own language (as mentioned in chapter 2.3.1) is inherent in his depiction of divided languages. The first interpretation of the term is “two different languages deriving from a single language” (Ammon 2005a: 1537), with an example being Czech and Slovak – congruent with the concept of Ausbau language (see chapter 2.1.1). The other, more important interpretation for this context is “[a] single language whose speakers have been separated politically”, the examples given are “Chinese since 1949, Korean since 1954, or German 1949–1990” (Ammon 2005a: 1537). The lack of traffic and communication across borders leads to linguistic differences. Usage and codification drifts apart, in the Chinese case an entire script reform was conducted on one side of the division (more on that in chapter 3.3.3) (Ammon 2005a: 1541–1542).

Clyne (2004: 296) points out as well that political divisions alongside the Capitalist-Communist spectrum have led to the development of different standard varieties (e.g. Chinese, German) and that, whenever a region was reunified, those differences were downplayed (such is the case with East and West German as well as with North and South Vietnamese). However, political division makes up only one of many reasons for pluricentricity in Clyne’s framework. Other reasons are e.g. colonization (e.g. English, French), immigration (e.g. Spanish, Tamil), emigration (e.g. Kurdish, Yiddish), the redrawing of borders (e.g. Dutch, Hungarian) and religion and nationalism (e.g. Hindi and Urdu). Another scale on which one may differentiate between pluricentric languages is whether they are geographically contiguous (e.g. German, Swedish, Arabic) or not (English, French, Chinese).

Muhr (2016) provides yet another way of categorizing pluricentric languages. He proposes six criteria, of which at least some need to be met for a language to be pluricentric. Based on these, he then suggests ten types of pluricentricity, especially one of which will be crucial in chapter 3.4. In the following list of six criteria the first and second need to be met by a language as a minimum in order to be a pluricentric language:

Criterion 1: Occurrence
Criterion 2: Official Status or strong ethno-linguistic awareness
Criterion 3: Linguistic distance (Abstand)
Criterion 4: Acceptance of pluricentricity
Criterion 5: Relevance for identity
Criterion 6: Codification of norms (Muhr 2016: 20–21)

Criterion 1 refers to the occurrence of the standard language in at least two nations that are pluricentric centers. This criterion is difficult to apply in the Chinese case (see chapter 3), hence it is proposed for this Master's thesis to perceive of nations, as in criterion 1, as self-governing entities (see also chapter 3.4.1). The mentioning of Abstand in Criterion 3 is clearly a reference to Kloss' bipartite conceptualization of language, yet, as emphasized in chapter 2.1.1, Abstand has nothing to do with standard varieties and, hence, shouldn't play any role in the defining of pluricentric languages that revolve around standard language. The ten types of pluricentricity Muhr (partially) draws from these criteria are the following:

Type 1:	Nationless pluricentricity
Type 2:	Formal pluricentricity
Type 3:	PCLs [Pluricentric languages] with varieties lacking the appropriate formal status and waiting for recognition
Type 4:	Languages where the status of pluricentricity is denied by the dominant variety or by the language as a whole
Type 5:	Languages where the status of pluricentricity is acknowledged by the "dominant/mother"-variety
Type 6:	Languages where the pluricentricity is deliberately practiced by model speakers of the respective NV [national variety]
Type 7:	PCLs where the NVS (a) are taught in schools and (b) the linguistic differences are made aware of
Type 8:	PCLs that act as a "dachsprache" (roof language) for (a) many so-called "mother tongues" and (b) as a PCL towards the other standard varieties
Type 9:	Nativized pluricentricity
Type 10:	Migrant pluricentricity – PCLs in a migrant context (Muhr 2016: 22–24)

Some of these types are not self-explanatory, hence, further information is provided. Type 1 describes pluricentric varieties that "have no territory of their own and no official recognition"

(Muhr 2016: 22), whereas formal pluricentric languages (type 2) meet criterion 1 – they occur in at least two separately administered regions. Type 8 is the most pivotal one for this Master's thesis, since Muhr (2016) creates this type for only two languages – Hindi and Chinese.

> Type 8: PCLs that act as a "dachsprache" (roof language) for (a) many so-called "mother tongues" and (b) as a PCL towards the other standard varieties. […] A similar case is "Chinese" which can also be seen as a dachsprache (Tien, 2016, Clyne/Kipp, 1999) as it is pluricentric in respect to a large number of mutually unintelligible *fangyan* varieties (Cantonese, Hokkien, Mandarin, etc.) and by *Mandarin*, the standard variety, that is pluricentric in respect to the different Chinese-speaking countries. (Muhr 2016: 23–24; emphasis in original)

Since the concept of *Dachsprache* is proposed for the description of Chinese pluricentricity, a short digression is in order here.

Dachsprache is a term that is usually attributed to Kloss. He did work with the roof metaphor (e.g. *dachlose Außenmundarten* (Kloss 1978: 60)), yet he did not, in fact, coin the term Dachsprache. Muljačić (1989: 260) presumes that Hans Goebl (e.g. 1975) was the first to use the term. A Dachsprache must always be an Ausbau language (Muljačić 1989: 264), since it provides a standard for its many non-standard varieties – hence, a Dachsprache is per definition a standard language. Muljačić' (1989) elaborations on Dachsprache would still be adequate if one were to replace the term *Dachsprache* with *standard language*, so including the concept Dachsprache in pluricentricity does not add any value but only terminological confusion.

The assumption of pluricentricity having several levels is inherent in Muhr's description of Chinese pluricentricity and he elaborates on it elsewhere (Muhr 2016: 20 and Muhr 1997: 53). "External pluricentricity" or "first level-pluricentricity [sic!]" describes the variation of a standard language used across different nations, whereas "internal pluricentricity" or "second-level pluricentricity" refers to standard language variation within these nations (Muhr 2016: 20). This approach incorporates the pluriareal conceptualization mentioned in chapter 2.3.1, and stems from the description of German pluricentricity as well.

Muhr also encompasses diglossia in second-level pluricentricity, however, the foundations for this assumption are not apparent. A diglossic situation exists when a speech community uses more than one language variety for different purposes, certain registers can only be expressed in certain varieties without arising attention. There is usually a H(igh) variety

that covers official functions whereas the L(ow) variety is the linguistic tool for conversational situations. Since it is always the H variety that is standardized, its characteristics are congruent to that of standard languages in general (see chapter 2.2.1) – e.g. its functional diversity, its prestige and its literary heritage. (Ferguson 1959) Muhr is not the only one to connect pluricentricity to diglossia, yet, those two concepts describe different phenomena, even though they often occur simultaneously, like in Chinese or German. One focuses on the differences in standard varieties, their development and asymmetrical relationship, the other concentrates on the registers covered by at least two varieties that differ e.g. in their degree of standardization and social attributions.

Another approach that perceives of pluricentricity in levels is by Glauninger, who assumes that pluricentric languages can either be genetically inherent pluricentric ("genetisch inhärent plurizentrisch") or secondary pluricentric ("sekundär plurizentrisch"). The first refers to pluricentric languages whose varieties have never been encompassed by only one center, the latter to languages that have gone through a monocentric development until their standard variety was spread through colonialism (Glauninger 2013: 3). His elaborations refer to German as a pluricentric language.

2.3.3. The Asymmetrical Relation between Pluricentric Standard Varieties

The concept of *centers* is crucial for the theory on pluricentricity – as the name already hints at it – and it also plays a pivotal role in the oftentimes asymmetrical relation between pluricentric standard varieties. Stewart (1968) was the first to introduce the term *center* for the description of pluri*centric* languages, whereas its origination is often mistakenly attributed to Kloss (1967, 1978) (see e.g. Clyne 1992: 1). Both authors assume that a language may have only one set of universally accepted norms (monocentric) or multiple sets (pluricentric). Stewart also elaborates on the norm orientation of standard varieties: "[T]he form of standardization prevalent in any one country may be either *endonormative*, when it is bases [sic!] upon models of usage native to that country, or *exonormative*, when it is based upon foreign models of usage." (1968: 534; emphasis in original) Ammon (1989) refined the conceptualization of in- and outward oriented standard varieties with a five-step scale, allowing for a more detailed description of the relation between standard varieties. The two pivotal aspects are codices (e.g. dictionaries, grammars, pronunciation guides) and model speakers (e.g. teachers, ministers, broadcasters) (Ammon 1989: 90):

(1) Full endonormativity: the models and the codex are entirely from within C_a [Country A].

(2) Predominant endonormativity: the codex comes entirely from within but the models come in part from outside C_a.

(3) Semi-endonormativity: the codex as well as the models come in part from within and in part from outside C_a.

(4) Predominant exonormativity: the codex comes entirely from outside but the models come in part from within C_a.

(5) Full exonormativity: the models and the codex come entirely from outside C_a.

Ammon (1989: 90) adduces that American English used to be predominantly endonormative since British models were accepted alongside own models. The Australian variety of English is in a different position, both British and own codices and models are accepted and is, hence, an example for semi-endonormativity. Many African varieties are nowadays examples for predominant exonormativity and used to be fully exonormative in times of colonialism with no codices and models to call their own. Parallel to this scale of endo- and exonormativity, Ammon (1989) suggests corresponding terms for the reference to centers:

Full centre (or *fully-fledged centre*): fully endonormative.
Nearly full centre: predominantly endonormative.
Semi-centre: semi-endonormative.
Rudimentary centre: predominantly exonormative. (1989: 91; emphasis in original)

His conceptualization lacks a fifth center term for full exonormativity due to the simple reason that one cannot speak of a center at all if the models and codices are taken entirely from outside the region in question.

Ammon's (1989) five step scale regarding the orientation of pluricentric varieties emphasizes the asymmetrical relationship between varieties of pluricentric languages. Their status and prestige differ, depending on their social attributions made from within and without the speaking community. Clyne (2004: 297) differentiates between D(ominant) and O(ther) varieties and provides a list of ten aspects in which he describes the differences between them:

(1) The nation(s) using the D variety(ies) have difficulty in understanding that even a small number of differences between the national varieties can be significant in that they play an important part in marking national identity.

(2) The D nations tend to confuse national and regional variation because of overlapping linguistic indices without considering the function, status and symbolic character of these indices in the national variety.

(3) The D nations generally consider their natioinal [sic!] variety to be the standard and the varieties of the O nations as deviant, non-standard, exotic, cute and somewhat archaic.

(4) In the O nations, cultural élites tend to defer to the D norms since the more distinctive D varieties are dialectally and sociolectally marked.

(5) In both D and O nations, it is believed that norms are less rigid in D nations than in O nations.

(6) Convergence in communication between participants from D and O nations is generally towards the D variety.

(7) D nations have better resources to export their variety in language teaching programs.

(8) D nations have better resources to codify their language since the publishers of grammars and dictionaries tend to be located in those countries.

(9) There is a prevalent belief especially in D nations that variation exists only in the spoken norms.

(10) In some cases, members of D nations are not even familiar with or do not understand (all or some) O varieties. (Clyne 2004: 297; paragraphs added by author)

Clyne (2004: 297) points out, though, that not all characteristics apply to every pluricentric language. The *International Working Group on Non-Dominant Varieties of Pluricentric Languages* later renamed the O varieties to *non-dominant varieties* for having an unambiguous antinomy (Muhr 2016: 25).

As can be seen in Clyne's aspect number 3, speakers of dominant varieties oftentimes believe that their standard norms are the only ones that exist or that are valid, hence, the perspectives and opinions that speakers of dominant varieties have are often shared with those who speak actual monocentristic languages. Muhr (2012: 28) sums up their attitudes as follows: "(1) centralist; (2) elitist; (3) monolingual (= mono-varietal); (4) mono-normative and (5) derogatory towards non-core-norm speakers".

One last aspect needs to be mentioned in regard to the asymmetrical relation between pluricentric varieties – that is how such asymmetry can be maintained over a long period of time. Muhr (2016) regards the following five levels as crucial for the maintenance of dominance:

> Level (1) – Political and economic power […]
>
> Level (2) – Language spread via electronic media and international language teaching organisations […]
>
> Level (3) – The transnational level of centralisation via international language organisations that promote the dominant norm […]
>
> Level (4) – Codifying institutions that ensure the control and centralisation of norms […]
>
> Level (5) – Exonormative codification practices in NDVs […] (Muhr 2016: 29–30)

The asymmetrical relation between the standard varieties of a language describes the situation of the Chinese standard varieties very well since one of them has the powerful People's Republic of China as its center. Hence, the elaborations in this chapter will play a crucial role in the description of Chinese pluricentricity in chapter 3.

2.4. Looking Back and Looking Forward

At the very beginning there was Kloss (1967, 1978) with his dichotomic understanding of language – *Abstand* and *Ausbau*, the former defining a language by simply being different to other languages, the latter creating new languages through human intervention. Therefore, a variety could be called a language on the basis of the Abstand principle when its linguistic characteristics distinguished itself enough from another variety; however, an Ausbau language needs a defined standard language in order to be an independent language – through which it also creates the non-standard. Fishman (2008) points out much later that these two principles are in fact not two opposing endpoints of one and the same continuum but lay on different dimensions altogether. Hence, he suggests that the opposite of *Ausbau* (language policy activities that strive to create more differences) shall be *Einbau* (language policy activities that aim for creating more similarities), excluding the concept Abstand from the framework entirely. Including Einbau in the conceptualization of language definition is a first step, yet the framework is still closely intertwined with the description of language-

internal characteristics, which does not correspond with the state of the art of language policy research that focuses on de facto language policy and language attitudes. Croft (2000) offers an alternative that provides a way of defining language that does not rely on the static perception of language, the focus on language-internal characteristics and mutual comprehensibility. He takes a theory of evolutional biology – namely the population theory of species (Hull 1988) – and applies it to linguistics for an evolutionary model of language change. The social definition of language that provides the basis for his elaborations is postulated by Chambers and Trudgill (1980) whose two key concepts are *autonomy* and *heteronomy*. The speakers of an autonomous variety perceive their variety as an independent language, disregarding its structural similarity to other varieties, whereas speakers of a heteronomous variety are dependent on the former, perceiving their variety as part of it, no matter their structural differences. Basing his elaborations on these concepts, Croft assumes that communicative interaction is the linguistic equivalent to the biological factor of interbreeding. Linguistic 'interbreeding' occurs when "every speaker perceives every other speaker as someone he or she should be able to communicate with by using what they perceive as the same language" (Croft 2000: 18). *Sibling languages*, just as sibling species, are two varieties that are structurally so similar that they would be defined as varieties of the same language if an internal language definition was applied. Their speakers, however, perceive their varieties as distinct languages. A *polytypic language*, just as a polytypic species, would, on the other hand, be identified as several distinct languages were it for structural characteristics, but its speakers perceive of it as a single language. The advantage of both Kloss'/Fishman's conceptualization and Croft's framework is that they are not mutually exclusive. Both approaches need to be kept in mind since they influence each other but describe different things: Kloss'/Fishman's conceptualization emphasizes language-internal aspects whereas Croft's approach incorporates the human factor.

In the case of standard language, this hidden ideological thinking is not inherent in language-internal methods but comes to the surface in language-external approaches. A definition of standard language that escapes the grasps of the standard language ideology by focusing on language-internal factors is that of Milroy and Milroy (1985) who state that "the process of language standardisation involves the suppression of optional variability in language" (1985: 8). Through the standardization process only one of many options is deemed correct – creating the standard, on the one hand, but also the non-standard, on the other.

Reaching functional efficiency is one of the main goals of standardizing a language. Haugen (1966) describes the process of standardization in four stages, namely the selection of norm, codification, elaboration and acceptance, whereas Milroy and Milroy (1985) focus on the stages of implementation of a standard language, since they believe that a language can never be fully standardized, especially in its spoken form. They assume three stages – selection, diffusion and maintenance –, all of which are kick-started by "a need for uniformity that is felt by influential portions of society" (1985: 27).

These elaborations lead to the pivotal part of this Master's thesis: standard language variation. At first glance, variation in standard language seems like a contradiction, since the process of standardization goes hand in hand with the suppression of variation. Yet, language political factors create a loophole that allows for some variation. The different standard varieties of a language are attached to centers, which can refer to language communities, regions etc. but mostly it refers to the political entity of a state. Hence, a terminological differentiation is needed for the different kinds of pluricentric languages: Ammon (2005a) suggests *plurinational*, *pluriregional* and *pluristatal* and *national variant, national variable* and *national variety* (2005b) for the lexical and semantic dimension. Since the nation terminology is problematic in the Chinese case, as will become obvious in chapter 3, this Master's thesis uses *regional variant*, *regional variable* and *regional variety*, instead – without referring to local non-standard expressions but implicating only standard speech forms. The reasons for the development of pluricentric languages are manifold, such as political division, colonization, immigration and the redrawing of borders (Clyne 2004). Hence, there are many kinds of pluricentricity, e.g. nationless pluricentricity, formal pluricentricity, or pluricentricity including a roof language (Muhr 2016) – the latter being crucial for Chinese. One may perceive of pluricentricity in levels as well, such as external and internal pluricentricity (Muhr 2016) or genetically inherent and secondary pluricentricity (Glauninger 2013). The asymmetrical relation between pluricentric standard varieties is another aspect that will be important in chapter 3. The two pivotal aspects are codices (e.g. dictionaries, grammars, pronunciation guides) and model speakers (e.g. teachers, ministers, broadcasters), which create a spectrum of full endonormativity (models and codices are entirely from within a center) to full exonormativity (models and codices are entirely from outside the center) (Ammon 1989). The corresponding center terms are *full center*, *nearly full center*, *semi-center* and *rudimentary center*. Clyne (2004) also offers a list of ten aspects that describes the difference between

dominant and other (later renamed to non-dominant) varieties and Muhr (2016) provides a list of ways how this inequality between varieties can be maintained over a long period of time.

The following step in this Master's thesis is to take a closer look at the Chinese language. How is the Chinese language defined? Which variety is the standard variety based on? In how far do the standard varieties differ from each other? What is the standpoint of academia on the matter? This Master's thesis will follow up on these questions in chapter 3.

3. The Chinese Language

This chapter is the centerpiece of this Master's thesis, in which the theoretical elaborations presented throughout chapter 1 and 2 are applied to the Chinese context. Before doing so, a short discussion on the denomination of certain regions, in which Chinese is used, needs to be anticipated. As pointed out in chapter 2.2.3, standard languages, and hence, pluricentricity revolves around nation-states, which is a delicate topic in the Chinese case. This Master's thesis does not attribute great relevance to the official recognition of certain regions. If this Master's thesis adhered to such a perspective, the adaption of pluricentricity would not align with the linguistic and social reality of these regions. Hence, more importance is attributed to whether such regions are self-governing entities. If they are, they can implement their own language policies and codify their varieties independently from other regions. For this reason, this Master's thesis differentiates between Taiwan and China, without positioning itself in the ongoing debate regarding Taiwan's independence. Therefore, when this Master's thesis refers to China, it denotes Mainland China, excluding regions such as Taiwan and Hong Kong. Parallelly, China may also be referred to as *Mainland China* or *People's Republic of China (PRC)*.

Chapter 3.1 offers a contextualization and conceptualization of Chinese, which is crucial for the application of pluricentricity to Chinese. Chapter 3.2 gives a rough overview about the early developments of Chinese before the fall of the Qing dynasty in 1912. The next subchapter (chapter 3.3) is dedicated to the standardization of Chinese, while chapter 3.4 goes into depth regarding the pluricentricity of Chinese. Chapter 3.5 summarizes the elaborations and concludes the findings.

3.1. Contextualization and Conceptualization

Chinese is the first language of far more than a billion people. It is the only official language in China and Taiwan, and one of several in Hong Kong (besides English) and Singapore (alongside English, Malay and Tamil) – providing a standard language for roughly 1,4 billion Chinese, 23,7 million Taiwanese, 7,4 million Hongkongers and 5,8 million Singaporeans (World Population Review n.d.). The scope and classification of the language family Chinese belongs to is still in dispute (Crystal 2010: 320). Mair (1991: 8–9) provides a classification of the Sino-Tibetan language family originally taken from the Encyclopedia of China (1988) (see Table 4).

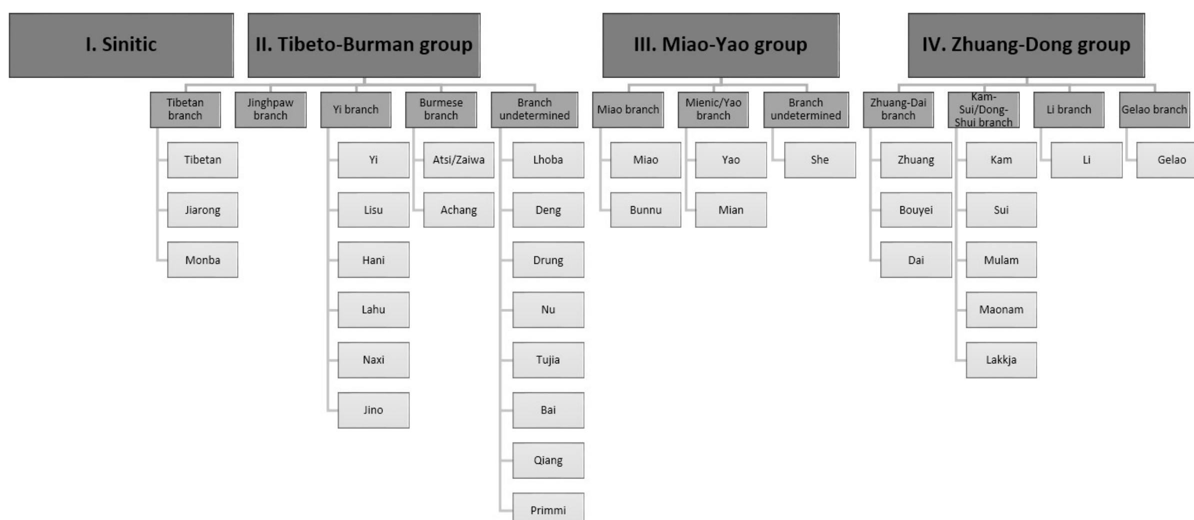

Table 4: The Sino-Tibetan language family (Mair 1991: 8–9; visually edited by author)

A matter of dispute is, e.g. whether the Miao-Yao group belongs to the Sino-Tibetan language family or forms a language family on its own (Crystal 2010: 321). However, regardless what stance one takes about these issues, what remains the same is the fact that the language group with by far the most speakers – Sinitic, i.e. Chinese – is not categorized any further. Hence, the term *Chinese* – commonly referring to a single language – is located on the level of language groups alongside, e.g. Tibeto-Burmese. To better grasp this unique situation a comparison is often drawn between Chinese and the classification of Western languages. In the Indo-European language family, for instance, there is the language group Germanic that splits up into two language branches (West and North) to which several languages are attributed each – English, amongst others, belongs to the Western branch of the Germanic language group in the Indo-European language family. In that context, Chinese is an equivalent to the Germanic language group and it lacks further categorization – at least officially and with the denomination *language*. With the great number of Chinese speakers and the difficult terrain especially in the South-East of China, which kept communication between villages to a minimum for millennia, there is no wonder that Chinese features tremendous language variation. DeFrancis (1984: 39) calls for further categorization of Chinese, stating that "[t]o speak of *the* Chinese language is to suggest a uniformity which is far from being the case" (emphasis in original).

Chinese is commonly divided into seven varieties that Chinese speakers themselves refer to as *Fāngyán* (方言) – Fāng (方) meaning *region* or *local* and Yán 言 *speech*. Due to the direct translation *local speech*, Fāngyán has frequently been equated with the English word

dialect (e.g. Mair 1991: 4 criticizes that), whereas at the same time being presented as different languages. This is often caused by unprecise and/or nonreflective usage of terminology:

> Because there has long been a single method for writing Chinese, and a common literary and cultural history, a tradition has grown up of referring to the eight main varieties of speech in China as *'dialects'*. But in fact they are as different from each other (mainly in pronunciation and vocabulary) *as French or Spanish is from Italian*, the *dialects* of the south-east being linguistically the furthest apart. The mutual unintelligibility of the *varieties* is the main ground for referring to them as separate languages. (Crystal 2010: 322; emphasis by author)

As highlighted in the quote above, Crystal first takes a distant stance on the usage of the term *dialect* for Fāngyán. He points out that they are actually as distant as separate languages, only to refer to them as dialects without quotation marks in the same sentence and as varieties in the next. This inconsequent usage of terms is very common in the description of Chinese but shall be avoided in this Master's thesis. Hence, when non-standard varieties, i.e. dialects, are addressed in this thesis, this is made explicit by referring to them as such. In Chinese, non-standard varieties are called *Tǔyǔ* (土語). The issue even with the Chinese term *Fāngyán* is that it has partially adopted the meaning of the Western misinterpretation (i.e. dialect), therefore sometimes denoting the Fāngyán varieties, sometimes non-standard varieties (Mair 1991: 14). In the framework of this Master's thesis, *Fāngyán* is only used to refer to the major varieties (depicted in Figure 2) commonly used by Chinese speakers to differentiate their language further.

The division of Fāngyán is a highly disputed matter, as Kurpaska (2010: 25–62) covers in depth. This Master's thesis follows the most commonly used division that accounts for seven major Fāngyán groups in Chinese (taken from Kurpaska (2010: 58), including a rough estimate of the Fāngyán distribution amongst Chinese speakers in percent (from Ramsey 1987: 87)):

Mandarin	官話, Guānhuà	71,5%
Wu	吳, Wú	8,5%
Yue/Cantonese	粵, Yuè	5,0%
Xiang	湘, Xiāng	4,8%
Min/Hokkien	閩, Mǐn	4,1%
Hakka	客家, Kèjiā	3,7%
Gan	贛, Gàn	2,4%

Figure 2: Geographical distribution of the seven major Fāngyán (Ramsey 1987: figure 6)

Additionally, it is disputable how to refer to Fāngyán in English and whether they fit into the Western conceptualization of languages, which is a language-internal approach (as discussed in chapter 2.1). Mair (1991) proposes the term *topolect* to be used as the English translation of Fāngyán, similar to DeFrancis' (1984: 57) not fully Greek-derived term *regionalect*. No matter which translation is used, the disadvantage is, as Mair (1991: 7) accurately points out, "that they do not fit into established Western schemes for the categorization of languages". In contrast, there is a term that has developed within these "Western schemes", namely *regiolect*, which refers to "standard-divergent varieties with broader regional distribution that can be located between dialects and standard varieties" (Lenz 2009: 302). As will be discussed in depth in chapter 3.3, Fāngyán cannot be placed anywhere on the continuum between non-standard and standard (except for the Fāngyán Mandarin, since the standard is based on it), they are something else entirely. Using terms such as the above mentioned *topolect*, *regionalect* or *regiolect* would yet again imply a uniformity within Chinese that does not reflect reality accurately. Hence, this Master's thesis uses the term *Fāngyán* in its untranslated form. Chinese does not fit into the Western approach to language definition because the way Chinese speakers define their language is not through language-internal factors as the Western world does but through language-external factors (see chapter 2.1.3). Applying Western terminology blindly and ignoring how Chinese speakers define their language only leads to more misinterpretations. Mair (1991: 10) suggests defining Chinese irrespective of how Chinese speakers do – not to cause turmoil in the Chinese world, but to provide a definition of Chinese to the Western world only, with its own terms. This suggestion would enhance the already existing contradiction of the two approaches to language definition and can, therefore, not be supported.

As established above, Fāngyán are more than regiolects. Mandarin is by far the most dominant Fāngyán, in terms of number of speakers as well as in terms of power. The standard variety of Mainland China is based on Mandarin (see chapter 3.3), hence, it can be assumed that there exists a regiolect in between the standard variety and local non-standard varieties. However, the other Fāngyán (also commonly referred to as *Southern Fāngyán*) are very distant from Mandarin (referred to as the only *Northern Fāngyán*). To illustrate these differences, a study conducted by Cheng and Zheng (1987) shall be presented here exemplarily, in which the researchers compare content and function words of Mandarin and Hokkien (as spoken on Taiwan). The corpus contained various genres, namely: prose, fiction and poetry. Additionally,

the research team looked at proverbs and sayings (1987: 109). Overall, 70% of the content and function words are etymons in Mandarin and Hokkien, i.e. the same, but a great 30% differs. When looking at function words only, this difference even grows to fifty-fifty. Very generally speaking, the differences between the various Fāngyán "amount, very roughly, to 20 percent in grammar, 40 percent in vocabulary, and 80 percent in pronunciation" (DeFrancis 1984: 63; based on Xu 1982). However, some Fāngyán are closer to each other than others. Ramsey (1987: 7), for instance, points out that one can find a non-standard continuum between Pekingese Mandarin and Shanghainese Wu, comparable to the continuum found in the border region of France and Italy.

Since the policy for standardizing Chinese in the first half of the 20th century (see chapter 3.3) was to *force the South to follow the North* (強南就北; qiǎng nán jiù běi), speakers of Southern Fāngyán have been put in a position of severe disadvantage. Southern Fāngyán speakers show a strong correlation "between an inability to speak good Modern Standard Chinese and the failure to write and comprehend written Chinese" (Chen 1999: 56; referring to Li 1995), contrary to the common myth that all Chinese varieties share the same writing system and, hence, do not break apart into several independent languages. In fact, written Chinese has always been attached to some form of Old Chinese or the Fāngyán Mandarin. The Chinese Fāngyán do not share the same written language, but the Southern Fāngyán never got the chance to develop their own, so they rely on an imposed written language that is very distant to their spoken varieties. DeFrancis (1984) describes the extra effort Southern Fāngyán speakers must endure to achieve literacy in comparison to Mandarin-speaking counterparts with a concise list of steps each group has to master. To understand this list, it is necessary to know that literacy in Chinese is not measured by the number of words one knows but by the amount of characters – roughly 2000 characters are said to be sufficient for e.g. reading newspapers. Primary school graduates have acquired around 2500 characters, whereas college graduates can read and write about 3500 characters. (Chen 1999: 136)

A. Tasks for Mandarin-speaking illiterates learning to read and write Standard Chinese:
   1. Learning the basic structure of characters
   2. Learning three thousand characters
   3. Learning the differences between spoken Mandarin and written Standard Chinese
   4. Practice in reading and writing

B. Tasks for Cantonese-speaking illiterates learning to read and write Standard Chinese:

1. Learning the basic structure of characters
2. Learning written Standard Chinese grammar different from Cantonese
3. Learning written Standard Chinese vocabulary different from Cantonese
4. Learn three thousand characters
5. Practice in reading and writing (DeFrancis 1984: 151)

As becomes apparent from the cited list, Mandarin speakers merely need to deal with minor differences in grammar and vocabulary between their spoken non-standard variety of Mandarin and the written standard variety whereas Southern Fāngyán speakers, in this case Cantonese speakers, need to deal with different grammatical structures and a great amount of previously unknown vocabulary to reach literacy in their standard variety. Yet, learning another Fāngyán (which it essentially is for Southern Fāngyán speakers when they acquire the standard variety) is not perceived of as learning another language but as "just 'picking up' pronunciations different from one's own 'dialect'" (Ramsey 1987: 16).

Now that Chinese and its Fāngyán have been conceptualized, it is time to take a closer look at where their origins lie and what the basis was for the standardization attempts in the first half of the 20th century (chapter 3.2).

3.2. Chinese Prior to 1912

This chapter mostly relies on Ping Chen's elaborations in *Modern Chinese: History and Sociolinguistics* (1999) that do not only treat the topic with high precision but also provide the structure of this subchapter (and partially of chapter 3.3) since Chen differentiates between spoken Chinese, written Chinese and the Chinese writing system. This differentiation is pivotal, for their development was separate from each other for a long time, as will become apparent in this subchapter.

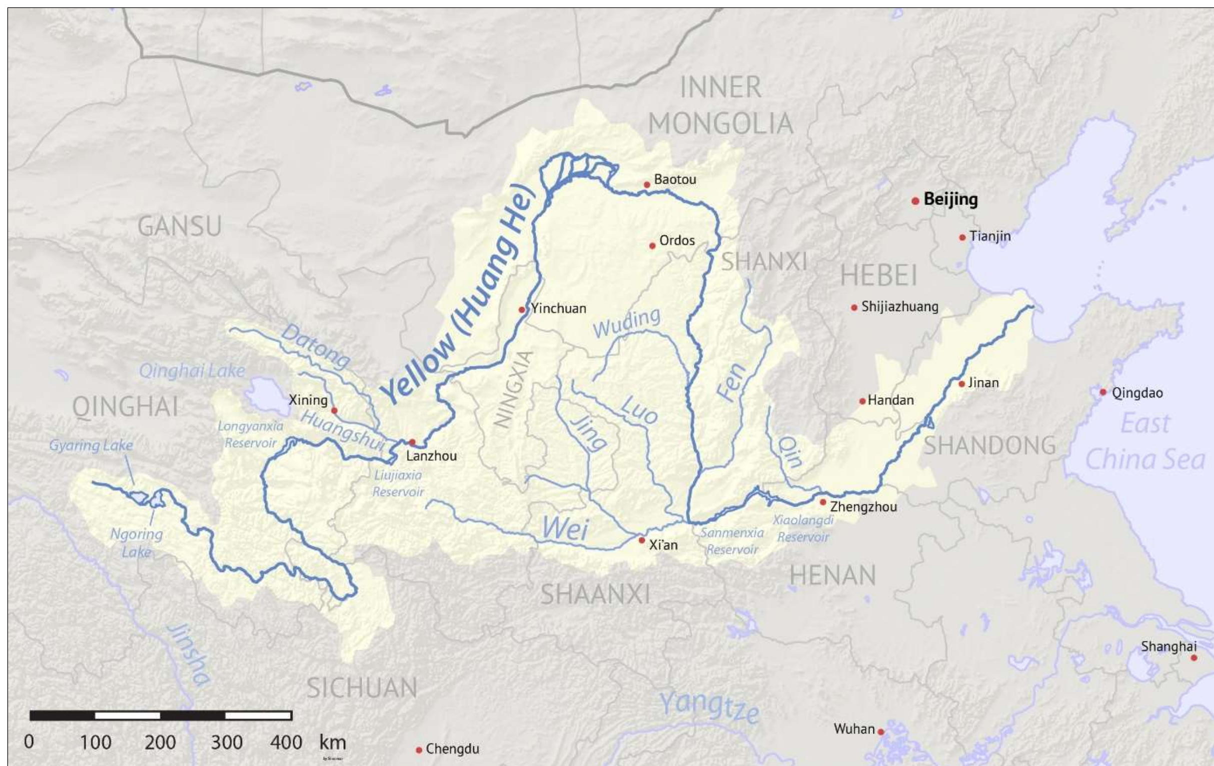

Figure 3: Map of the Yellow River area (Wikimedia Commons 2017)

The Chinese civilization has its origins in the Yellow River area (see Figure 3). By the time of the Zhou dynasty (roughly 1100–771 BC), the language spoken by the people varied widely, there arose a need for a lingua franca, on the one hand for administrative and diplomatic purposes (e.g. providing a means of communication between the central government and the local states), and on the other for enhancing cultural and economic exchanges. The spoken standard variety that has evolved to satisfy that need is called Yǎyán (雅言; *elegant speech*) in the *Analects of Confucius* and was based on the dialect of the city of Zhengzhou. Since Luoyang and Kaifeng (both cities are nearby Zhengzhou) served as capitals in many dynasties, Yǎyán became the spoken standard for the whole empire for roughly two millennia. The spoken standard was spread farther in the Wei-Jin period (220–420 AD), when the royal court moved to Nanjing (a city to the South of the Yangtze River, roughly 600 kilometers away). In the year 530 AD, the imperial examination system (科舉; Kējǔ) was introduced for bureaucratic appointments that lasted until 1905. Rhyme writing was a pivotal component of this examination, hence the examinees had to abide to a pronunciation standard that was compiled in the famous dictionary Qièyùn (切韻; 601 AD). In it, the pronunciation typical for Zhengzhou is dominant, with some Nanjing influences. The standard pronunciation was oriented towards the Nanjing dialect for many centuries. The Beijing dialect became the base

of standard pronunciation only in the 19th century. One factor was that the city had served as the capital of several dynasties by then and, therefore, gained in prestige. Another factor was that Nanjing's population decreased rapidly due to the Taiping rebellion in the previous century (1850–1864). In general, the spoken standard was ill-defined until the 1920s, and several dialects (those of Zhengzhou, Nanjing, and Beijing, consecutively) provided the basis for the implicit norms. The pronunciation standard which officials abided to was generally called 官話 (Guānhuà; *official speech* or *the speech of public servants*). (Chen 1999: 7–12)

The terms *Guānhuà* and *Mandarin* – the name nowadays used to refer to the Northern Fāngyán variety – are intriguingly connected. Mair (1991: 11) states that the English word *Mandarin* was derived from the Portuguese word *mandarim*, which – even though influenced by *mandar* (meaning *command*, *order*) – was taken from Malay (*mantri*), which in turn was borrowed from Hindi-Urdu and goes back to the Sanskrit word *mantrin* (meaning *counsellor*). This makes obvious that Mandarin has long been intertwined with standardization.

When the need arose for a spoken standard, so did the need for a written standard. Many different writing styles and systems evolved over the time but Qin Shihuang, the emperor who unified China for the first time in 221 BC, put an end to that. He legitimized only one writing system (小篆; Xiǎozhuàn; *Small Seal script*) and prohibited all the others by harsh means. By doing so he laid the foundation for the standardization of the written Chinese language. In the previously mentioned imperial examination system, the writing style of the pre-Qin classics were emulated. It must be remembered that by the time the examination system was introduced, these classics were already over 700 years old. Hence, the great gap between spoken and written language has existed in Chinese from the very beginning. This rather isolated writing style is referred to as Wényán (文言; *classical literary language*) and functioned as the written standard for two millennia. Even though Wényán was a time-consuming task to learn, the logographic writing system ensured "a degree of accessibility across time and space" (Chen 1999: 68) that would have not been possible with a phonographic script, in which case the written language would have followed the spoken language much more closely. However, Wényán was not the only writing style that existed – over the centuries Baíhuà (白話; *vernacular*) developed, a writing style that was much closer to the contemporary spoken language than Wényán. It was fully established by the end of the Tang dynasty (907 AD). Vernacular expressions found their way into writing especially through

Buddhist translations and folk stories. Hence, the Chinese had a diglossic situation in written language – one variety was used for highly prestigious functions, the other for less prestigious ones. Like the developments in the spoken standard, in its early days, Baíhuà was based on the Zhengzhou dialect, whereas it shifted to dialects spoken in the North by the 19th century. It provided the written version of the above mentioned spoken standard of Guānhuà. (Chen 1999: 67–70)

The Chinese script is one of the oldest writing systems in the world (but not the oldest, the Sumerian script, for instance, is much older, still (DeFrancis 1984: 40)), and the oldest one that is still in use today. The general characteristics of the Chinese script are that it is logographic and morpho-syllabic. Hence, Chinese characters are not made to indicate their precise pronunciation, but they contain information on the semantic level instead. By the time the Chinese came in contact with the people of Japan, the Korean peninsula and from today's Vietnam two millennia ago, they already had a well-developed writing system whereas the others were still unfamiliar with any kind of writing. So, they started borrowing the writing system (and vocabulary) which can still be encountered today – e.g. Kanji in the Japanese writing system, Hanja in Korean or the former script of Vietnamese, Chữ Nôm (DeFrancis 1984: 65). The Chinese script, contrary to common belief, is a vivid system that has changed and evolved over the centuries. It has constantly adapted to new requirements and often underwent periods of simplification. The Chinese characters can be classified in four types. Xiàngxíng (象形) are characters that are pictographs, i.e. they resemble the object they refer to – e.g. 日 (rì; *sun*) and 木 (mù; *tree*). Ideograms (指事; Zhǐshì) are characters that express some abstract idea – e.g. the numbers *one, two, three* (一 二 三; yī ér sān) and 本 (běn; *root*; similar to the character for *tree* mentioned above in which an extra line marks the lower part of the character). These two types were the first characters that existed and on the basis of which the third type evolved, namely Huìyì (會意). An example would be the simplified character for *to follow* that consists of two *persons* (人; rén), one following the other: 从. The last and by far most important character type are the Xíngshēng (形聲) characters that consist of two parts – a phonetic and a semantic component. 沐 (mù; *to bathe*) shall be used as an example here. The character is pronounced in the exact same way as the tree mentioned above, hence, *to bathe* could be written with the same character but that would lead to misinterpretations. Therefore, the character 木 is combined with the semantic component 氵

that means *water*. In combination, the character 沐 carries the semantic meaning *water* and the phonetic content that it is pronounced like 木. Due to sound changes in the Chinese language the phonetic implications are, however, by far not always as clear as in the given example. In fact, only 26% of Xíngshēng characters are pronounced exactly the way the phonetic component implies, yet over 90% of the characters in the Chinese script belong to this last type. The old writing style Wényán had a strong tendency to rely purely on monosyllabic characters, i.e. words that consist of only one character. Homophony grew to be a major problem in Chinese, hence, even though the written language still used only monosyllabic characters, di- or multisyllabic words gained in popularity in spoken Chinese. Nowadays, monosyllabic characters play a minor role, making up just 5% of all entries in the dictionary Xiàndài Hànyǔ Cídiǎn (現代漢語詞典). Even though the Chinese script has its advantages (for one, it can handle homophonous morphemes well), there are two major disadvantages. On the one hand, the Chinese script is very difficult to learn. School children require roughly 30% of their entire class hours in primary and secondary school to acquire literacy, whereas only three to four months are needed for becoming literate in a phonographic script. On the other hand, the Chinese script is also difficult to use once learned. Due to the lack of an alphabet, the Chinese script poses a great hurdle for indexing, compiling dictionaries etc. There are four major systems for sorting characters (the radical and stroke number system, the beginning stroke system, the four-corner system and the phonetization system), all of which take longer to use in comparison to systems for phonographic scripts. (Chen 1999: 131–147)

This chapter provided a very concise overview of the Chinese language and its historical developments with a special focus on the spoken and written standard, as well as on the script prior to 1912. The next chapter (chapter 3.3) is dedicated to the pivotal changes the Chinese language went through in the first half of the 20th century.

3.3. Standardizing Chinese

The first half of the 20th century posed turbulent times for China and the Chinese language was in the center of attention for several decades. The starting point for the major changes in China was the defeat in the Opium Wars in the mid-19th century by the British Empire, which painfully illustrated the Chinese that they were not the powerful empire they envisioned themselves to be. Already in the late Qing dynasty but especially after its fall in 1912, the

modernization of the country was made a priority and reforming the Chinese language was one of the most crucial endeavors, since Chinese was perceived as the major cause for the empire's deficits. (Chen 1999: 13–14) The Chinese elite was inspired by the rapid standardization process Japanese went through, which itself was emulating Western models of standardization (Klöter 2016: 58). The opinions ranged from very moderate adaptations all the way to radical approaches such as replacing Chinese entirely with Esperanto or English (Ramsey 1987: 50).

To draw on the elaborations in chapter 1, China was keen on corpus planning in the first half of the 20th century. Its aim was also to expand the status of the Chinese standard variety – moving away from a very elitist stance towards providing a communication tool to the masses –, which had to be accomplished through intensive acquisition planning, i.e. teaching the new standard variety in schools. The ultimate non-linguistic goal of the Chinese government was to modernize the country and, thus, catch up with the West, yet it manifested in various linguistic goals and overt policies. At least half of Nahir's (1984) eleven linguistic goals (see chapter 1.1) pertain in the Chinese case, namely: language reform, language standardization, language spread, stylistic simplification, interlingual communication, and auxiliary code standardization. The overriding motivational factor for China, in Fishman's (2000) terms, was vernacularization, i.e. moving away from Wényán to a written variety that resembles the spoken language more. Through extensive overt language policies, the Chinese state managed to change the language practices and beliefs of its population (Spolsky 2004).

This chapter will discuss in detail, which notions prevailed and how Chinese was standardized in the 20th century, putting a special focus on the tension between the various Southern Fāngyán and Mandarin. Different roads were taken for the various aspects of the Chinese language, hence, this chapter is further divided into six subchapters that are dedicated to Modern Spoken Chinese (chapter 3.3.1), Modern Written Chinese (chapter 3.3.2), the Modern Chinese Writing System (chapter 3.3.3), the phonetization of Modern Chinese (chapter 3.3.4), the standardization of other Fāngyán (chapter 3.3.5) and the developments of Mandarin outside Mainland China (chapter 3.3.6).

3.3.1. Modern Spoken Chinese

The starting point for a standard in Modern Spoken Chinese was a difficult one, Guānhuà was still commonly used by government employees at the beginning of the 20th century, yet it allowed for great variation since it did not have strict phonological rules. Voices were raised to base the new standard pronunciation on the common pronunciation of various cities (Nanjing, Wuhan, Shanghai, Beijing) or to create a more general form that incorporates Fāngyán pronunciation. As will be discussed in chapter 3.3.4, some advocated for the abolition of the Chinese script and replacing it with a phonetic script, which would have facilitated the diffusion of the new norm. The foundation for standardization was already laid in the late Qing dynasty, more precisely in 1911, when the *Act of approaches to the unification of the national language* was passed, which addressed several aspects of it. Research on Fāngyán was ordered on the basis of which the norms for Guóyǔ (國語; *national language*) were to be decided – namely a standard pronunciation and a standardized phonetization system. Additionally, instruction schools for teachers were to be established to instruct the teachers in the new standard for language diffusion. Only two years later, in which much had changed (the falling of the Qing dynasty), Zhùyīn Zìmǔ (注音字母) was chosen to provide the phonetic script that was mostly oriented towards the phonology of the Beijing dialect, yet it incorporated some characteristics of other Fāngyán, especially those of Wu. Additionally, a dictionary (國音字典; Guóyīn Zìdiǎn) was compiled, published in 1919, that contained the new national pronunciation, which constituted the pronunciation standard for several years. However, the discussions continued about what basis the new standard shall have. It took over ten years for the advocates of the Beijing pronunciation (京音; Jīngyīn) to prevail. In 1926, Zhùyīn Zìmǔ was renamed to Zhùyīn Fúhào (注音符號), a romanized script was introduced in addition to it (Guóyǔ Luómǎzì; 國語羅馬字; *national language romanization*) and most importantly – the pronunciation was detached from the literary pronunciation still common in Guānhuà and was oriented towards the vernacular pronunciation for the first time in spoken Chinese. Changing from the highly prestigious literary pronunciation to the less prestigious vernacular pronunciation posed a major shift in the development of Chinese, since literacy was easier to acquire (for those with Mandarin as their first Fāngyán) but put Southern Fāngyán speakers at a disadvantage (as already elaborated in chapter 3.1). The Ministry of Education was even faster than the national standardization process – ever since 1920 primary

school children learned Chinese not via the literary style but through the vernacular standard. (Chen 1999: 13–23)

By choosing the basis for the newly introduced norms, the first big step towards a standardized language was taken – in the framework of Haugen (1966) and Milroy and Milroy (1985), as presented in chapter 2.2.1, this step poses the first towards (the implementation of) standardization. The next step was codification. The Chinese Communist Party (CCP) took over in 1949, the Guómíndǎng – the party that ruled China until that point – withdrew to Taiwan and took their concept Guóyǔ with them. The CCP discredited the term Guóyǔ for its alleged emphasis on the Han population, disregarding the 55 officially recognized ethnic groups in China, hence, the term Pǔtōnghuà (普通話) was introduced in its stead, denoting *common language* (Chen 1999: 25). Two major conferences convened in Beijing in 1955, the National Conference on Script Reform (全國文字改革會議; Quánguó Wénzì Gǎigé Huìyì) and the Symposium on the Standardization of Modern Chinese (現代漢語規範化學術會議; Xiàndài Hànyǔ Guīfànhuà Xuéshù Huìyì), in which a definition of Pǔtōnghuà was proposed:

> *Pǔtōnghuà* is the standard form of Modern Chinese with the Beijing phonological system as its norm of pronunciation, and Northern dialects as its base dialect, and looking to exemplary modern works in *báihuà* 'vernacular literary language' for its grammatical norms. (Wang 1995, taken from Chen 1999: 24; emphasis in original)

With this definition, the struggle for representation of Fāngyán in the standard language was over – and lost. The Northern Fāngyán Mandarin provides the sole basis for the standard language with the specific pronunciation of China's capital Beijing. Baíhuà – the writing style with little prestige that is closer to spoken Chinese (in particular the Fāngyán Mandarin) in comparison to Wényán (see chapter 3.2) – functions as a template for the (from now on) correct usage of grammatical norms. Hence, three aspects were defined: lexicon, phonology, and grammar. Additionally, the scope of functions was expanded – prior to the fall of the Qing dynasty, knowledge of and competence in the standard language was reserved for the elites, whereas the newly defined standard language by the CCP provides a tool for the general public (Chen 1999: 3). With this definition, Haugen's (1966) stage 2 and 3 (*codification* and *elaboration*) in his four-stage conceptualization of the standardization process are fulfilled. In the course of the second half of the 20th century, Haugen's fourth stage, *acceptance*, was also reached, as a report shows: "*Putonghua* as the language of official business has […] become deeply rooted in the hearts of the people" (Wang & Yuan 2013: 32; emphasis in original).

Defining the standard language in such a way also entails the restriction of the spoken standard to having only four tones, even though many Fāngyán have more than that. Those four tones are commonly depicted in a scale with five different pitch levels (see Figure 4).

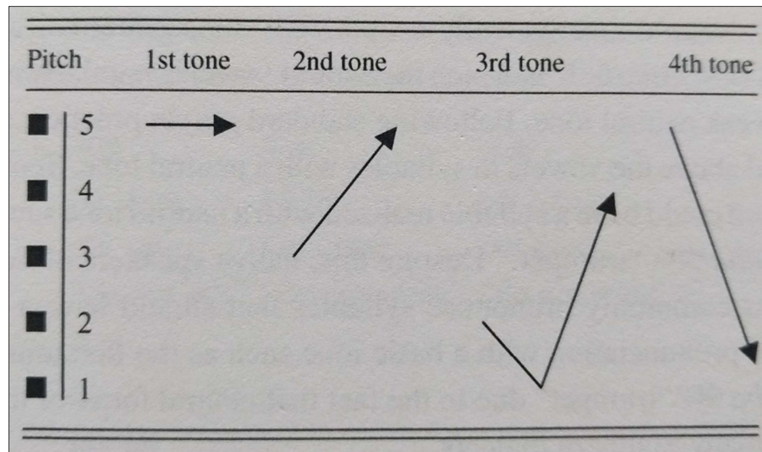

Figure 4: The four basic tones in Standard Mandarin (Sun 2006: 39)

Depending on the pitch, a string of phones (e.g. *ma*) conveys different meanings (mā 媽 *mother*, má 麻 *hemp*, mǎ 馬 *horse*, mà 罵 *scold*). Additionally, there is an unmarked neutral tone that is frequently used for particles and the second syllable of a compound word such as in lǎba 喇叭 (*trumpet*). In Southern Fāngyán there are no neutral tones in such a manner, hence speakers of Southern Fāngyán often add a pitch to the neutral tone (e.g. a first tone to lǎbā). Another phenomenon specific to the Mandarin speech area, more precisely that of Beijing, is rhotacization, in which many finals in spoken Mandarin are changed by suffixing a diminutive marker *-er* (hence also known as érhuà 兒化). Rhotacization is also uncommon in Southern Fāngyán, therefore often omitted by Southern Fāngyán speakers when using their standard variety. (Sun 2006: 38–40)

Once the CCP had defined Modern Chinese, acquisition planning was in the center of attention. Surveys were conducted on Southern Fāngyán on the basis of which pamphlets were compiled as an aid for learners to understand the differences and similarities between their Fāngyán and the new standard language. Pǔtōnghuà was introduced as the medium of instruction in all schools and the language of mass media. After the initial years of introduction in the late 1950s the promotion of Pǔtōnghuà ebbed and the CCP turned to other matters. Only after the Cultural Revolution (1966–1976) the objective of language standardization and its spread was resumed, but less rigorously than before. (Chen 1999: 25–27) However, the Ministry of Education and its State Language Commission regularly conduct inspection tours,

survey the knowledge of the standard variety in various cities and publish the outcome ever since 2005 in annual reports, which show that the spread of Pǔtōnghuà is far-reaching. (Klöter 2016: 62–63) However, the efficacy of Pǔtōnghuà promotion is dependent on various factors, such as the existence of a prestigious Fāngyán (mainly Cantonese and Wu) in an area, the linguistic heterogeneity of a region, the degree of mobility and education of a region's population, and, last but not least, the consumption of mass media (Chen 1999: 28–30). Even though the Southern Fāngyán have a negative impact on the spread of Pǔtōnghuà, the PRC has never attempted to completely ban the usage of Southern Fāngyán (it does, however, restrict the usage of Fāngyán in some cases, such as in Guangdong). Instead, the CCP envisages a situation in which Pǔtōnghuà plays the central role for all public activities, whereas Fāngyán are used amongst friends and family and for unofficial occasions. (Chen 1999: 57–59)

This chapter shows what far-reaching changes spoken Chinese has undergone, especially in the 1950s. It also paves the way for the description of the other aspects of the Chinese standard variety established in the PRC, since the major occasions in politics depicted in this chapter (such as the collapse of the Qing dynasty 1912 and the taking over of the CCP in 1949) have a strong influence on them as well.

3.3.2. Modern Written Chinese

By the time the Qing dynasty broke apart in the early 20th century, the Chinese were accustomed to a well-established binary writing system – Wényán for highly official functions and Baíhuà for less prestigious activities that required writing (see chapter 3.2), the former being very distinct from, the latter relatively close to spoken Chinese. Even prior to 1912 there were scholars advocating a written standard language that reflects the spoken language, such as Huang Zunxian, who summarized his proposition with 我手寫我口 (Wǒ shǒu xiě wǒ kǒu; literally *my hand writes my mouth*). People raised their voices to replace Wényán with Baíhuà completely, and small steps were taken towards this goal, such as the publishing of a dozen newspapers in Baíhuà by the turn of the century. Yet, the single most important event that supported this objective was the abolishment of the state examination in 1905, in which prospective bureaucrats were tested in their Wényán competencies (see also chapter 3.2). Through this change, students were less inclined to acquire Wényán, yet it continued to play a major role in everyday life, e.g. in trade, business, administration and the law sector. For decades there existed several different written languages (traditional Wényán, modern

Wényán, traditional Baíhuà and new-style Baíhuà) until the point was reached to question whether to revive Wényán fully or to enhance Baíhuà. This decision was made by the CCP when they defined Pǔtōnghuà through "looking to exemplary modern works in *baíhuà* 'vernacular literary language'" (Wang 1995, taken from Chen 1999: 24; emphasis in original; full quote in chapter 3.3.1), turning Baíhuà into China's written standard language. (Chen 1999: 70–82)

However, Baíhuà was not fully elaborated for playing such a role at that time, i.e. it did not provide for "*maximal variation in function*" (Haugen 1966: 931; emphasis in original), since it was primarily used for theater plays and other cultural activities. Wényán on the other hand provided a sheer endless pool of linguistic resources, due to its long usage over many centuries. For that reason, writers using the newly introduced written standard often turned to Wényán to express matters that Baíhuà could not provide for. That way, in a matter of fact, Wényán is still in use today, whereas Latin turned into a dead language once it was replaced by the various new Ausbau languages. Furthermore, Wényán has the ability to express matters very concisely due to its monosyllabic nature and is perceived as highly formal, therefore, Wényán can cover certain registers that Baíhuà could not at the beginning and offers a fruitful source for Modern written Chinese. (Chen 1999: 84–85)

However, Wényán was not the only influence on the newly defined standard Pǔtōnghuà. In its modernization efforts, Chinese was lacking expressions for many new concepts and things. To compensate that, Chinese borrowed these expressions from foreign languages, especially from Japanese and European languages. According to a study (Wang 1979, taken from Chen 1999: 85–86), over 50% of commonly used expressions in written Chinese today are borrowed in comparison to 1840. However, Pǔtōnghuà did not undergo big changes only terminology-wise, foreign languages also had a great impact on its grammatical norms. Especially in the first half of the 20th century, many Western and Japanese works were translated into Chinese by translators who were not qualified for the job – they adhered to the grammatical structures of the source language rigorously, creating uncommon expressions and structures. These new patterns of *translationese* – this is how this kind of Chinese was commonly referred to – found their way into the written standard language through translations of political writings that had to be consumed by all (literate) Chinese. An example of a grammatical norm that was introduced by translationese is the passive marker

被 (bèi), which initially was only used for the description of negative events. Over time, however, it adopted the general function of a passive marker, regardless of the meaning conveyed in the sentence. Another aspect that gained in popularity through Western influence is the usage of affix-like morphemes that usually match those used in European languages, especially English – e.g. the prefix 非 (fēi) expresses *non-*, the suffix 化 (huà) represents *-ize* for verbs or *-tion* for nouns. (Chen 1999: 85–97)

The least influential factor on the new standard written language, yet worth mentioning, is the Southern Fāngyán. Some grammatical features that were originally only typical in Southern Fāngyán were adopted in Pǔtōnghuà, such as an alternative pattern for yes-no questions in addition to the one typical in Mandarin. (Chen 1999: 91)

It has been pointed out throughout this Master's thesis that dictionaries play a crucial role in the creation of a standard language and the definition of a pluricentric variety (see e.g. chapter 1.1 or chapter 2.3.3). Klöter (2016: 64–67) takes a closer look at how the *Dictionary of Modern Chinese* (現代漢語詞典; Xiàndài Hànyǔ Cídiǎn) compiles its entries and shows how intertwined descriptivism and prescriptivism are in the creation and updating of a dictionary. Chinese society underwent major social changes, through which new terms were coined, such as words with a negative connotation to refer to singles around the age of 30 (剩女 shèngnǚ *leftover woman* and 剩男 shèngnán *leftover man*). These words did not make it into the 2012 edition of the dictionary because the editors did not want to include such a negative term in their dictionary out of respect to the people who are referred to with this word – even though shèngnǚ and shèngnán are in common usage. Another example for the editors' prescriptive approach is the word 同志 (tóngzhì) that used to denote *comrade* only but has undergone a semantic shift towards the meaning *homosexual*. However, in a comparison of the entries of tóngzhì in the edition of 1979 and 2012, Klöter shows that this semantic shift is not represented in the latter. In an interview, the chief editor of the dictionary states that "we won't put it into a standard dictionary because we don't want to encourage such things. We don't want to draw attention to such things" (CCTV News 2012, taken from Klöter 2016: 66–67).

Similar to the struggles in finding a standard in spoken Chinese (see chapter 3.3.1), written Chinese underwent big changes and was the cause for debates stretching over many

decades. When the decision was made to rely only on Baíhuà, it could not cover all functions of a written standard, hence especially Wényán and foreign languages provided a pool which Pǔtōnghuà could draw from. The elaborations on the compilation of dictionary entries show that ideological beliefs have a strong impact on what is conceived as correct and what is condemned as incorrect.

3.3.3. The Modern Chinese Writing System

The Chinese script has been perceived as a hindrance for development for a long time since the complex script was difficult to acquire and literacy rates were low. The debates revolving around script reform grew louder already in the second half of the 19th century, yet, by the time the Qing dynasty fell, the notion of full phonetization of the Chinese script was very popular. The discussions ebbed after roughly two decades due to difficulties in implementation in such a short period of time and it was worked on the simplification of the Chinese script instead. By 1935 there already existed a list containing 2400 characters in a simplified form, yet, the simplification of the Chinese script took on a new scale once the CCP took over in 1949. (Chen 1999: 150–153)

There are mainly two ways how the Chinese script could be simplified. On the one hand, the number of strokes can be reduced per character. This can be reached by either replacing a component of a character with another that has fewer strokes, replacing the character with one of its components, replacing the entire character with another, simpler one with the same pronunciation. On the other hand, the amount of characters in common use can be reduced. Due to the usage of characters over a long period of time and in a vast region, many characters developed that denote the same meaning, yet, are slightly different from one another. By choosing only one of these variants, the amount of characters could be drastically cut down. On this basis, the simplification was implemented according to the following principles: characters are replaced by the simplified version if the latter has been in common use already, characters are simplified by replacing them with (near) homophones that have fewer strokes, and characters are replaced by newly invented or uncommonly used ones with fewer strokes. (Chen 1999: 148, 156–157)

The first wave of script simplification was conducted in the years 1955 and 1956. 810 groups of characters were accumulated from which only one was chosen each (following the second method of simplification mentioned above) – that way over a thousand characters

were abolished from the standard language, making them non-standard variants. Additionally, several lists proposed over 500 simplified characters, another with over 50 simplified character components. After some trial usage, a list of over 2000 characters in total was published in 1964 that marked the end of the CCP's first simplification efforts. Decades later, character simplification was taken up again, however the proposed characters from 1977 were repealed by 1986 since they were harshly criticized. Instead, the official list of simplified characters was merely republished in the same year with some minor adaptations. (Chen 1999: 54–56)

One might suspect that a simplification of the script would meet with fierce criticism and would not be accepted (something that is required for introducing a new standard according to Haugen 1966 (see chapter 2.2.1)), however, roughly 90 percent of the simplified characters were already in common use before their introduction, hence, the simplification process merely offered official recognition to the writing habitus of the people. Nevertheless, the simplified characters do have their disadvantages – some characters became harder to differentiate (e.g. 儿 ér *son* and 几 jī *several*, in comparison to their traditional counterparts 兒 and 幾), some simplified characters convey less accurate phonetic information than the old characters and, in general, the simplification of the script made it harder to access writings in traditional script. But the advantages outweigh the disadvantages, since the simplified script is easier to acquire and recognize. (Chen 1999: 157–159)

3.3.4. Phonetization of Modern Chinese

Auxiliary code standardization, one of Nahir's linguistic goals of language policy (see chapter 1.1), is an essential part for the standardization of languages that have non-phonetic scripts, such as Chinese. Therefore, the phonetization of Chinese was discussed extensively. There are many ways how to develop a phonetic script for Chinese, Chen (1999: 167–178) elaborates on them in detail. First, the intended role of the phonetic script needs to be defined in regard to the Chinese characters – this can be auxiliary, supplementary, alternative, or superseding. Second, the script can take on various forms – shorthand-style scripts, Kana-style scripts (like in Japanese), or alphabets, such as the Roman one. Third, a question that is closely interlinked with the decision on Modern Spoken Chinese (see chapter 3.3.1) is whether the phonetic script shall follow a specific Fāngyán or be non-Fāngyán specific. Fourth, the syllabic structure can be represented in a syllabic, syllabo-phonemic or phonemic way. Lastly, there are several

ways how the Chinese tones can be represented in the script, e.g. through diacritics or numbers.

The first efforts for the phonetization of Chinese date from the 16th century by Western missionaries. Even though their phonetization schemes were never fully taken up by the Chinese, their attempts manifested the idea that phonetization is needed for promoting education, which led the Chinese to come up with their own ideas starting in the mid-19th century. The first major step for a phonetic script was the introduction of Zhùyīn Zìmǔ (later renamed to Zhùyīn Fúhào, see chapter 3.3.1) in 1918 that is still in use today in Taiwan. Zhùyīn Zìmǔ is an alphabet that bases its letters on simple strokes of the traditional script in a syllabic way. At the beginning it was based on the Beijing dialect with influences from other dialects before it shifted to a dialect-specific approach (focusing on only the Beijing dialect). The tones are represented by diacritics next to the letters, the first tone being unmarked. In contrast to earlier schemes, Zhùyīn Zìmǔ was never intended to take on a greater role than just as an auxiliary tool for indexing the pronunciation of characters. In the 1920s two other schemes were published, Guóyǔ Luómǎzì and Lādīnghuà Xīn Wénzì (拉丁化新文字; *Romanized New Script*), however, only Zhùyīn Zìmǔ stood a chance to the phonetization scheme introduced in the 1950s by the CCP – Hànyǔ Pīnyīn (漢語拼音; *Chinese transliteration*). (Chen 1999: 164–167, 180–182)

Hànyǔ Pīnyīn, introduced in 1958, took over the same role as Zhùyīn Zìmǔ, in that it provides an auxiliary tool for the phonetic annotation of Chinese characters. It is extensively used for acquisition and diffusion planning. Therefore, it is used not only by native speakers of Chinese (of both the Northern Fāngyán Mandarin and the many Southern Fāngyán) but also by foreign language learners to acquire the Chinese standard language through the aid of Hànyǔ Pīnyīn. Hànyǔ Pīnyīn consists of Roman letters, using diacritics to mark the tones. It represents the spoken standard – hence, it has a mono-dialectal basis. (Chen 1999: 187–188)

3.3.5. Standardization of Other Fāngyán

Referring to the great discussions revolving around the standardizing of Chinese in the first half of the 20th century, Ramsey (1987: 16) states that "[i]t would in any case be unrealistic to expect the practical men who had managed to consolidate control of the entire country to turn around and Balkanize China by recognizing the status of these dialects [=Fāngyán] as separate languages". As discussed in chapter 2.2, a major step for recognizing a variety as a

language is the governmental standardization of such. Yet, this never occurred in the case of Southern Fāngyán, even though the CCP was not an opponent of Fāngyán standardization before it became the party of government – its members in fact advocated for it. Yet, this quickly changed after the CCP gained power. A mono-dialectal standard variety was introduced on a massive scale and Fāngyán standardization was discouraged (see chapter 3.3.1 to 3.3.4) (Chen 1999: 121).

The standardization of Southern Fāngyán, firstly, is highly dependent on the specific Fāngyán and their particular socio-political situations, whether there exist any ambitions towards standardization. Secondly, even though there are standardization schemes for some Fāngyán, they fade in comparison with the degree of standardization that Pǔtōnghuà exhibits. Due to the continuous dominance of Mandarin, speakers of the Southern Fāngyán never got the chance or never saw the need to develop their own writing schemes and expand the functions of their varieties (see also chapter 3.1), therefore, the degree of standardization is generally low amongst the Southern Fāngyán. However, there are two Fāngyán that are more prominent in regard to standardization (not for the entire speech area, though) – namely Hokkien in Taiwan and Cantonese in Hong Kong, both of which will be shortly presented here.

As will be discussed in chapter 3.3.6, the Taiwanese had to abide to very strict language policies, first implemented by the Japanese colonial rule and after that by the ruling party Guómíndǎng.[8] Japanese and later Mandarin were promoted and the use of Hokkien was harshly restricted – until martial law was lifted in 1987, it was forbidden to write in Hokkien, hence, the accumulation of dictionaries and the use of Hokkien in schools was unthinkable. (Mair n.d.: 4–5, 20) However, "[i]n the changing political climate of the 1980s, official politics towards local languages shifted from oppression to toleration, and subsequently from toleration to active cultivation" (Klöter 2008: 12). Groups have formed for advocating Hokkien, dictionaries are compiled, and it is mandatory for school children to learn Hokkien (or Hakka) in school (Klöter 2008: 12–14; Scott & Tiun 2007: 60). These changes need to be taken with a grain of salt, though, since there are still fierce discussions led regarding Hokkien's written representation. A great hindrance is that roughly a quarter of commonly used morphemes, such as frequently used particles, do not have an adequate representation in the Chinese

[8] More information on the different language policy stages Taiwan went through, and on the impact it had on language practices can be found in Sandel (2003).

script (Chen 1999: 116) – and Hokkien continues to play a marginal role in comparison to Mandarin. However, Hokkien is also spoken in parts of Mainland China (especially in the Fujian province) and South-East Asia (e.g. Singapore), where its future seems less bright than in Taiwan.

As with Hokkien, Cantonese is not only spoken in Hong Kong, but also in the Guangdong province (and many communities overseas). However, the status in these regions could not be any more different: the use of Cantonese in Guangdong has drastically declined after the 1950s, especially in public domains, whereas Cantonese is growing in popularity in Hong Kong, even after returning under Chinese sovereignty in 1997. The use of spoken and written Cantonese in Hong Kong is frequent and widespread, covering a wide range of functions (e.g. advertisements, newspapers, government posters) – Cantonese can even cover the academic domain to some extent, as Bauer (2018) shows with his Cantonese abstract translation (2018: 104–106). The ambitious preparation of several dictionaries and creation of a transliteration scheme in the last two decades led Cantonese to be by far the most standardized Fāngyán after Mandarin, especially in its written form. Interestingly, there is one aspect, though, in which Taiwanese Hokkien is in a better position than Hongkongese Cantonese – that is in the institution school. Hongkongese school children are not taught Cantonese in school, yet they acquire the norms of Cantonese through informal ways. (Bauer 2018: 109, 116)

3.3.6. Developments of Mandarin Outside Mainland China

The standardization processes presented so far in chapter 3.3 are mainly concerned with those of Mainland China – with good reason, since the majority of Chinese speakers live in Mainland China and great influence was exerted from the ruling governments towards language policy that had an enormous impact on the language as a whole. However, Chinese also functions as at least one of several official languages in Hong Kong, Singapore and Taiwan. The developments and usage of standardized Mandarin in these regions will be discussed in this chapter.

Due to its 99 years under British rule and its rapid economic development in that time, Hong Kong experienced and continues to experience a great influence from the English language. Prior to 1974, English was the only official language of Hong Kong. The Official Languages Ordinance (1974: 3 (1)) introduced Chinese as the second official language, equal

to English. However, Chinese was not further defined and since the knowledge of Mandarin was generally low, this de jure policy allowed for a more central role of Cantonese (see also chapter 3.3.5). Mandarin gained importance since Hong Kong was returned to PRC, however the norms established for Pǔtōnghuà are not fully followed in Hong Kong. Especially on the lexical level, the English and Cantonese influence is visible, and the traditional script is still used in Hong Kong in comparison to Mainland China, where it was simplified (see chapter 3.3.3). Table 5 provides some examples of lexical differences in the four regions discussed in this chapter.

	Mainland China	**Taiwan**	**Hong Kong**	**Singapore**
'taxi'	Chūzū qìchē	Jìchéngchē	Díshì	Déshì
'petrol'	Qìyóu	Qìyóu	Diànyóu	Diànyóu
'folk-dance'	Mínjiānwǔ	Tǔfēngwǔ	Tǔfēngwǔ	Tǔfēngwǔ or mínjiānwǔ
'film'	Jiāojuǎn	Jiāojuǎn	Fēilín	Jiāojuǎn or fēilín
'motorcycle'	Mótuōchē	Jīchē	Diàndānchē	Mótuōxīkǎ or any other of the three
'disabled'	Cánjí	Cánzhàng	Shāngcán	Cánquē or any of the other three

Table 5: Examples for lexical variation in Mandarin (Chen 1999: 106)

Some similarities between the Mandarin used in Hong Kong and Singapore, as become apparent on the lexical level in Table 5, are remnants of a common past – Singapore was, too, under British rule until 1963. Singapore has four official languages: Malay, Mandarin Chinese, Tamil and English (Constitution of the Republic of Singapore 1965: 153A). Ethnic Chinese make up around 74% of the Singaporean population, hence Chinese is the most common language learned as a first language at home (roughly 36%) and the second most widely known language in Singapore after English (Tien 2012: 454). Singaporeans observed the language policy developments in Mainland China closely, which served as a model. Huáyǔ (華語; *the Chinese language*) – this is the term used to refer to the Mandarin Chinese standard variety in Singapore – has been taught in Chinese schools since the first half of the 20th century. The use of a Mandarin-based standard variety was welcomed, even though Mandarin does not have the majority of speakers. Up to six different Fāngyán are spoken as a home language by Singaporean Chinese, none of which with a great majority (unlike Cantonese in Hong Kong (see chapter 3.3.5) or Mandarin in Mainland China (see chapter 3.1)). When Singapore became an independent state in 1965, the promotion of Huáyǔ was intensified (ever since 1979, the

annual event *Speak Mandarin Campaign* is organized by the government) and Hànyǔ Pīnyīn (see chapter 3.3.4) was introduced as the standard annotation system, following the Mainland Chinese norms closely. The simplified script introduced in Mainland China was adopted, too. (Chen 1999: 32–33) Hence, the Singaporean standard variety of Mandarin follows the norms of Mainland China closely, however, differences still exist due to the influence of various languages and Fāngyán. These differences are not only on the lexical level (see Table 5) but also on e.g. the phonological level. Rhotacization (see chapter 3.3.4) is much less commonly used than in Pǔtōnghuà and many characters are pronounced with a tone that does not exist in the Mainland standard (sometimes referred to as the fifth tone, this pronunciation can be traced back to the rù tone in Medieval Chinese that is still used in many Southern Fāngyán today, but not in Mandarin) (Chen 1999: 49).

Singapore has some own norms, yet, it is strongly oriented towards the most dominant Mandarin standard variety – Pǔtōnghuà –, taking up a more exonormative orientation than the Taiwanese standard variety of Mandarin. Ever since the 18th century, the Taiwanese people did follow the norms established in the Chinese imperial court, however, the occupation of the island by the Japanese (1895–1945) broke this connection. The Japanese language was promoted vigorously. In 1903, a ban was imposed on Chinese publications, in 1920, the language of administration changed to Japanese, in 1922, Chinese was not taught obligatorily in schools anymore and by 1937, Chinese was cancelled from the school curriculum altogether, additionally forbidding pupils to speak any other language than Japanese. Before the end of Japanese rule, over 70% of the Taiwanese population was proficient in Japanese and many could not speak any Chinese at all. Due to this situation when Taiwan was returned to China, the re-installation of Chinese as the prominent language in Taiwan (i.e. as language of administration, language of instruction in schools etc.) was an urgent objective. At first Taiwan got support from the Chinese government, however, when the Republic of China was established on Taiwan in 1949 all ties were broken. The ruling party Guómíndǎng stuck to the name Guóyǔ for its standard variety – a term that was rejected on Mainland China (see chapter 3.3.1) – and they promoted Guóyǔ by harsh means, that are not unlike the ones implemented by the former colonial rulers. The usage of Fāngyán was strongly discouraged or even prohibited in some domains, and students in teacher education programs had to take Guóyǔ proficiency tests starting in 1958. Some of these harsh measures lasted for

a long time – the ban of using other languages/Fāngyán in school was lifted only in 1987. (Chen 1999: 30–32)

Until the mid-1980s there was barely any communication between Mainland China and Taiwan due to political reasons. This is the time in which the major changes took place in Mainland China regarding Pǔtōnghuà (see chapter 3.3.1 to 3.3.4), so naturally differences between the standardized Mandarin used in Mainland China and Taiwan developed. On the phonological level, rhotacization is much less common (similar to the Singaporean standard), and the pronunciation of several initials and finals merged, which are being differentiated clearly in the spoken standard of Pǔtōnghuà. This can be traced back to the predominant Fāngyán Hokkien, in which these differentiations are not made. The pronunciation of characters developed differently, too. According to a study conducted by Li (1992) over 20% of the character pronunciations of the 3500 most commonly used characters differs between Mainland China and Taiwan. (Chen 1999: 46–48) The Taiwanese government was first interested in the simplification of the Chinese script and took first steps towards the implementation of such. However, by 1956, when Mainland China introduced a simplified script, the Taiwanese Ministry of Education forbade the usage of simplified characters in order to protect the traditional script from what they perceived of as its destruction by the Communist government in Beijing. (Chen 1999: 162–163) This shows, how politically charged the discussion about the Chinese script is. However, it is not just the script that differs in Taiwan, it is also the annotation system. Instead of switching to Hànyǔ Pīnyīn, like Singapore did, Taiwan continues to use a modernized version of Zhùyīn Fúhào, which denotes the tones with diacritics, similar to that of Hànyǔ Pīnyīn (in earlier days, Zhùyīn Fúhào used a dot-system to indicate the tones). (Chen 1999: 189)

This chapter – chapter 3.3 on the standardization of Chinese – describes the developments of the many different aspects of standardizing Mandarin and the role of the Southern Fāngyán. Additionally, an overview was provided of the differences between the standard varieties in use in Mainland China, Hong Kong, Singapore and Taiwan. The following chapter (chapter 3.4) will go into depth on how these differences are perceived of academia in the context of pluricentricity.

3.4. Chinese Pluricentricity

This chapter finally focuses explicitly on Chinese pluricentricity. Firstly, indications of and elaborations on Chinese pluricentricity in the literature are presented and discussed (chapter 3.4.1). Secondly, Chinese pluricentricity is examined on the basis of the theories and conceptualizations presented in this Master's thesis (chapter 3.4.2 to 3.4.4). It will be assessed, whether the approaches to Chinese pluricentricity applied in the literature so far can withstand the arguments brought about by this thesis.

3.4.1. Indications of and Elaborations on Chinese Pluricentricity in the Literature

A first big hurdle for the application of pluricentricity to Chinese is the question regarding the definition of Chinese and whether to perceive of Fāngyán as independent languages. This issue has been discussed at length in chapter 3.1. This Master's thesis acknowledges that from a language-internal, i.e. Western, perspective, Fāngyán would be defined as languages, however, it refrains from following this approach due to the perception of Chinese speakers, who define Chinese as one entity. Hence, this Master's thesis deals with Chinese as one single language that consists of several Fāngyán.

Another hindrance for the interpretation of Chinese as pluricentric is legitimizing research on the differences between the standard varieties of e.g. Mandarin, which seem minor in comparison to the great distance between Fāngyán, and the distinction between standard and non-standard varieties. Sun (2006: 6), for instance, compares the standard varieties of Singapore and Taiwan with the words "[d]ifferent as huáyǔ 话语 and guóyǔ 国语 may appear, the standard is practically the same as pǔtōnghuà". Ramsey (1987: 15) expresses the same opinion by stating that ""what passes for Mandarin or Guoyu these days in Taiwan is virtually indistinguishable from the Putonghua spoken on the Mainland". Interestingly, both researchers use adverbs to soften their statements ("practically" and "virtually"). Kane (2006: 14) makes it even more explicit with the addition "but not entirely": "Mandarin spoken in Chinese communities in Southeast Asia, Hong Kong and Taiwan is virtually, but not entirely, identical with the standard language of Mainland China". However, Kane makes a rather confusing statement regarding Pǔtōnghuà only a few pages later: "Standard putonghua is based on the language of Beijing. People from all over China speak standard Chinese without any trace of local accent. However, one also hears putonghua spoken with a wide variety of regional accents, which might differ as much as British English from American English." (Kane

2006: 27) This statement is confusing in two ways. First of all, it is not at all clear what Kane means by “local accent” and “regional accent”. An assumption would be that people speak Pǔtōnghuà without any dialectal accents, yet, one can hear what their native Fāngyán is due to their regional accents – but this is all it is, an assumption. Secondly, he speaks of these regional variations as if they were pluricentric varieties, but only within China. This would ask for a pluriareal approach, excluding Taiwan and Singapore. DeFrancis (1984: 39) makes a similarly ambiguous statement decades earlier, namely: “Some two-thirds to three-quarters of the Chinese-speaking population speak what is loosely called Mandarin in English, or Putonghua in Chinese. Within this category there are differences roughly of the magnitude of the differences among the British, American, and Australian varieties of English.” In this Master’s thesis, Mandarin is not equated with Pǔtōnghuà (Mandarin is one of many Fāngyán, of which Pǔtōnghuà is one of several standard varieties, namely that of PRC), yet it is in this quote. Again, the differences within Mainland China, within Pǔtōnghuà, are compared to the differences between various pluricentric varieties of English. However, in Kane’s and DeFrancis’ defense, they could also refer to dialect varieties of English. This is not clear, though.

As can be seen from the short discussion of quotes above, there is no consistent opinion on Chinese pluricentricity. Major sources of misunderstanding are the Fāngyán and their influence on the spoken standard varieties on the one hand, and an inadequate differentiation between standard and non-standard in general on the other. Another issue for applying the pluricentric theory to Chinese is the political situation in East Asia, as discussed at the beginning of chapter 3, since pluricentricity is tightly interwoven with the concept of independent nation-states (see chapter 2.2.3). However, some researchers have dealt with the issue of Chinese pluricentricity, which will be presented in the following.

Bradley (1992) was the first to dedicate his attention to Chinese pluricentricity in his article *Chinese as a pluricentric language* in Clyne’s pivotal edited volume *Pluricentric Languages. Differing Norms in Different Nations*. Bradley recognizes three different standard varieties of Mandarin, namely Guóyǔ of Taiwan, Pǔtōnghuà in China and Huáyǔ in Singapore (1992: 306–307). Additionally, he points out that Cantonese has a highly developed standard, too, yet he does not incorporate it in his pluricentric scheme (1992: 307). The Mandarin varieties differ on various levels – Bradley (1992: 313–318) points out the differences in the lexicon, phonology, morphosyntax and communicative competence (i.e. pragmatics). He

ascribes these differences to the fact that Guóyǔ was formulated in the 1920s, whereas Pǔtōnghuà was codified in the 1950s and due to the opposing political systems, they developed further apart (1992: 313). In general, Bradley perceives of the Singaporean and Taiwanese standard as more similar than Pǔtōnghuà.

This publication on Chinese pluricentricity was a first step to cover the research gap, however, many aspects remain unsatisfying. Even though Bradley dedicated several pages to the description of Fāngyán, and making out that Cantonese is standardized, he does not incorporate them in any form in his scheme. He assumes that all Fāngyán share the same written form (Bradley 1992: 305) – an opinion that this Master's thesis rejects (see chapter 3.1). Additionally, he speaks of "Guangzhou Province" (1992: 308), which is wrong, plain and simple (Guangzhou is the capital city of the province Guangdong). The author does not elaborate on some other aspects in full extent, for instance, the usage of 同志 (tóngzhì; *comrade*) in the PRC, which rapidly declined – mentioned by Bradley (1992: 317) in the section on pragmatic differences between the Mandarin standard varieties. He does not provide a reason for this, however, Klöter (2016: 65–66) points out that there occurred a semantic shift from the meaning *comrade* to *homosexual*, hence, this could cause the decrease in usage (see also chapter 3.3.2). Generally speaking, Bradley covered several aspects that are not relevant for Chinese pluricentricity, at least not in a first attempt to address the topic, such as his long elaborations on Chinese communities overseas (1992: 305–308) and the Chinese spoken by non-Han groups (1992: 318–320). Therefore, his discussion of the pluricentricity of Mandarin is comparably short.

Tien (2016: 45–46), like Bradley (1992), defines Mandarin with three centers, namely Beijing (Pǔtōnghuà), Taibei (Guóyǔ) and Singapore (Huáyǔ). The author (2016: 47–48) reflects on the different approaches on how to define Chinese and comes to the conclusion that the perspective of Western linguists is a different one in comparison to Chinese linguists. The former equate Chinese with Mandarin, therefore, allowing for the application of pluricentricity only on Mandarin. From a Western standpoint, Tien suggests that concepts such as *dialect continuum* and *Sprachbund* may apply to Chinese. Chinese linguists, on the other hand, have a broader definition of Chinese, that also includes the Southern Fāngyán, and therefore, other Fāngyán can be interpreted as pluricentric. Due to the two different scripts in use today (see chapter 3.3.3), Tien (2016: 50) suggests, that the Chinese script is

pluricentric with its norm setting centers in Beijing for the simplified characters and Taibei for the traditional characters. The same applies to the phonetization systems (Tien 2016: 51) – Mainland China being the center for Hànyǔ Pīnyīn and Taiwan for Wade-Giles as well as Zhùyīn Fúhào. Tien (2016: 56–57) further includes domain-specific language use, socio- and generatiolects in his understanding of pluricentricity. In another article, Tien (2012) proposes that not only Mandarin is pluricentric, but also the Fāngyán Hokkien. He focuses on the Hokkien variety in Singapore that has only little codification. Singaporean Hokkien is the most prominent Chinese Fāngyán after Mandarin and especially culturally influential (2012: 455). Tien (2012: 457) makes out four Hokkien varieties worldwide – that of Singapore, the Fujian province in Mainland China, Taiwan and the Chinese communities (in e.g. Malaysia and Indonesia) –, with the Taiwanese variety as the strongest base. The most outstanding feature of the Singaporean variety, Tien (2012: 469) states, is its lexicon.

The assumption of Tien (2016) that the Chinese script and phonetization system are pluricentric can only be supported due to two reasons. First, there are indeed major differences that are administered by the centers Mainland China and Taiwan, and second, the written and spoken forms of Chinese have developed independently for several millennia, scrutinizing the various aspects (as presented in chapter 3.3.1 to 3.3.4) separate from each other is therefore legitimate. Additionally, Tien's (2012) application of the pluricentric theory to Hokkien is something that is welcomed, since until then, only Mandarin was perceived as pluricentric. There are some issues, though, with the low degree of standardization that Hokkien has in Singapore, amongst the other defined areas as centers for Hokkien. This will be addressed in more detail in chapter 3.4.4. However, there are some controversies in Tien's articles as well. His elaborations on the inclusion of domain-specific language uses, socio- and generatiolects in a scheme of Chinese pluricentricity uncover that there are major differences in the perception of pluricentricity between Tien and this Master's thesis because this Master's thesis does not include these varieties in the pluricentric conceptualization. There may be differences in e.g. the sociolect typical for a certain group in Mainland China and Taiwan – these differences, however, cannot be traced back to entirely their existence in different pluricentric centers. The same goes for generatiolects – people in their 80s may speak differently in Mainland China and Taiwan but this is not entirely due to the different pluricentric varieties, this may also be caused by their different life experiences and their different environments (e.g. the occupation of Taiwan by the Japanese, who had very strict

language policies promoting Japanese and diminishing the use of Chinese, no matter if Mandarin, Hokkien or another Fāngyán). These differences in the conceptualization of pluricentricity are unfortunately not made explicit by Tien, hence, no comparison can be conducted here. Counterarguments for the suggestion that the concepts *dialect continuum* and *Sprachbund* may be applicable to Chinese are easily found as well. Wherever there are non-standard varieties, there may be dialect continua, in which the non-standard varieties geographically close to each other are mutually intelligible, whereas the non-standard varieties that are geographically distant are not, even though they are attributed to the same language. As mentioned in chapter 3.1, there is a non-standard continuum between Pekingese Mandarin and Shanghainese Wu (Ramsey 1987: 7), but this does not hold true for all Chinese Fāngyán. The suggestion to apply the concept *Sprachbund* to Chinese does not withstand any criticism. Sprachbund describes languages that are typologically closer than their genetic relationship would imply (see footnote 3). The fact, that Fāngyán are perceived as subdivisions of Chinese points out clearly that all the Fāngyán share a common origin (as discussed in chapter 3.1). Hence, the concept of Sprachbund can be ruled out.

Muhr (2016) makes two major proposals for the description of Chinese pluricentricity in his article *The state of the art of research on pluricentric languages: Where we were and where we are now*, which were both scrutinized already in chapter 2.3.2 but shall be discussed in the context of Chinese here. The first proposal is the division between external and internal pluricentricity, in which the former describes the standard variation across nations and the latter within these nations. Muhr also includes diglossia in internal pluricentricity as well as "the existence of '*fangyans*' in Chinese and '*mother tongues*' in Hindi that make these languages pluricentric in several ways" (Muhr 2016: 20; emphasis in original). The second proposal is of interest in the context of Chinese pluricentricity as it creates an own type of pluricentricity for the linguistic situation of only Chinese and Hindi. As cited in chapter 2.3.2, type 8 in Muhr's (2016: 23; emphasis in original) categorization refers to "*PCLs* [pluricentric languages] *that act as a 'dachsprache' (roof language) for (a) many so-called 'mother tongues' and (b) as a PCL towards the other standard varieties*". He further elaborates on the Chinese situation by stating that "[a] similar case is 'Chinese' which can also be seen as a Dachsprache (Tien, 2016, Clyne/Kipp, 1999) as it is pluricentric in respect to a large number of mutually unintelligible *fangyan* varieties (Cantonese, Hokkien, Mandarin, etc.) and by *Mandarin*, the standard variety, that is pluricentric in respect to the different Chinese-speaking countries"

(Muhr 2016: 24; emphasis in original). As indicated in the quote already, Clyne and Kipp (1999: 5) propose a very similar differentiation between two pluricentric levels for Chinese: "1. according to *fangyan* (Mandarin, Cantonese, Hokkien, etc.); and 2. according to the national variety of Mandarin, which has different norms in different countries (PRC, Taiwan, Singapore)" (emphasis in original).

At first glance, a differentiation between external and internal pluricentricity makes sense, especially seen from a pluriareal perspective, in which the focus lies on standard differences within nations (and similarities across national borders – this is something that still would not be covered by this conceptualization). However, from a Chinese perspective, this approach shows several weaknesses. First of all, Muhr explicitly connects external pluricentricity to nations, which is a hindrance for the application to Chinese (see also the introduction of chapter 3). This Master's thesis generally moved away from a rigid understanding of nation-states in connection with pluricentricity, since it is only a great hurdle for its application but has no advantages. Instead, self-governing entities (such as Taiwan) are the entities relevant for Chinese pluricentricity. Additionally, a focus is put on language definition and standard language because these are aspects much more relevant for the Chinese case. Hence, the connection between external pluricentricity and nations is not an issue for this Master's thesis, it is understood as the pluricentricity of several self-governing entities, regardless of their official recognition. The greater issue lies with the concept of internal pluricentricity when one reflects about the examples given. The first example are regional standard varieties that are strongly connected to their Bundesländer/Kantone, i.e. their political entities they are used in within the German speaking countries Austria, Germany, and Switzerland. It can be assumed that the differences amongst these varieties are relatively small or at least do not surpass the differences found across the national varieties. The next example is diglossia, hence involving non-standard varieties that may differ greatly from the standard variety used alongside with it. Chapter 2.3.2 pointed out already that diglossia describes a different aspect of language usage in comparison to pluricentricity, the former including standard as well as non-standard varieties, the latter only concerned with standard varieties, hence the reason for including diglossia in a conceptualization of pluricentricity is unclear. With his last example for internal pluricentricity, Muhr expands the degree of variation even further because he addresses the Chinese Fāngyán. The variation between the standard varieties of Mandarin, as discussed by Bradley (1992), are relatively minor in

comparison to the differences amongst Fāngyán (see also chapter 3.1). Additionally, it is not clear, whether Muhr only refers to standard varieties of Fāngyán (in so far they exist, see chapter 3.3.5), or also includes non-standard varieties of Fāngyán, similar to including diglossia in this framework. Generally speaking, internal pluricentricity is a collection of different concepts and various degrees of variation, mixing standard and non-standard varieties and entirely unregulated in comparison to external pluricentricity, for which many aspects have been covered in the literature (see chapter 2.3).

The second proposal by Muhr (as well as Clyne and Kipp) needs to be criticized as well, in which he discusses a two-level pluricentricity of Chinese – the pluricentricity of Mandarin that functions as a Dachsprache to the many Fāngyán, which are pluricentric, too. As already elaborated on in chapter 2.3.2, the term *Dachsprache* does not add any value in a discussion on pluricentricity but only terminological confusion, since a Dachsprache is per definitionem a standard language. An interpretation of the term *Dachsprache* in this context could be linked to the roof metaphor it contains – standardized Mandarin (which has its different regional varieties Pǔtōnghuà, Guóyǔ and Huáyǔ) also provides a standard language to Southern Fāngyán speakers. However, what is forgotten, is that Southern Fāngyán can be standardized, too, just not all of them have, i.e. Southern Fāngyán may also have a spectrum between non-standard and standard language within themselves, excluding the Mandarin standard varieties. This Master's thesis advocates defining all Fāngyán as a part of Chinese, since this is the way most Chinese speakers perceive of their language situation (see chapter 3.1). However, at the same time, one may not forget the major linguistic variation that occurs between Fāngyán. Referring to the various Fāngyán of pluricentric varieties of Chinese (even though not all of them have a defined standard, neither top-down nor bottom-up) equates these Fāngyán with the pluricentric varieties of other languages, e.g. English. The variation between Fāngyán is much greater – oftentimes referred to as independent languages by Western linguists (see chapter 3.1) – than the variations within English standard varieties. This equation does not add up, hence the only plausible conclusion is to reject this conceptualization of Chinese pluricentricity.

As presented in this chapter, there is no consistent opinion on Chinese pluricentricity in the literature, which does not deal with this topic directly. Major sources for confusion and misunderstanding are, on the one hand, the Fāngyán influences on spoken standard varieties

and, on the other, the inadequate differentiation between standard and non-standard varieties. Bradley (1992) proposed the notion of three pluricentric centers of Mandarin, namely Huáyǔ, Pǔtōnghuà and Guóyǔ, which is generally conceived well in the literature. He points out that Cantonese has a standard variety as well, though he does not elaborate on it further in the context of Chinese pluricentricity. Tien (2012 & 2016) argues in support of Hokkien being a pluricentric Fāngyán as well, focusing on the Singaporean variety of Hokkien, but making out four possible centers altogether – namely Singapore, Taiwan, the Fujian province in the PRC and various communities overseas. Additionally, Tien proposes to perceive the Chinese script and phonetization system as pluricentric. Even though Muhr's (2016) and Clyne and Kipp's (1999) conceptualization of Chinese was rejected, it still shows that there is some awareness within the academic community, that Chinese pluricentricity needs to be dealt with and that does not end with merely defining Mandarin as pluricentric. Klöter (2016) and Bauer (2018: 108) are amongst the researchers who advocate the application of pluricentricity to Chinese, pointing out that "[w]hether one subscribes to this particular distinction [referring to Clyne and Kipp's (1999: 5) two-level pluricentricity] or whether one calls for a more refined definition is of secondary importance. The crucial point is that the notion of 'pluricentric languages' seems fully applicable to Chinese" (Klöter 2016: 64).

As can be seen in this chapter, there are many misconceptualizations regarding the Chinese language, and also regarding the theory of pluricentricity – hence, the detailed elaborations on language, standard language and standard language variation were necessary in chapter 2 and the general impact of language policy had to be made explicit in chapter 1. In the following chapters (chapters 3.4.2 to 3.4.4), Chinese is scrutinized on the basis of the framework created in this Master's thesis.

3.4.2. Step I: What Is *Chinese*?

The first step for perceiving Chinese within the framework of pluricentricity was to define what the term *Chinese* refers to and what it incorporates – this was accomplished in chapter 3.1. Due to a lack of further classification, the term *Chinese* is located on the same level as various language branches within the Sino-Tibetan language family, therefore, the term contains great linguistic diversity.

The seven major varieties within Chinese are often referred to as *dialects* in the literature, however, this Master's thesis refrains from doing so and uses the Chinese term

Fāngyán for these varieties instead. Other terms, namely *topolect, regionalect* and *regiolect*, were also rejected, since they all do not accurately represent Fāngyán. Fāngyán themselves can cover the whole spectrum between dialect to standard language, hence referring to them as dialects is misleading. Regiolects are varieties that can be located somewhere in between the outer points of the dialect-to-standard continuum. However, in the Chinese case this would indicate a variety, that can be placed in between local non-standard varieties (of any Fāngyán) and the standard varieties of Mandarin – these are varieties that are not on the same continuum, though. The standard varieties of Mandarin can be placed on a scale that reaches from local Mandarin non-standard varieties to the various Mandarin standard varieties. By contrast, the local non-standard varieties of Southern Fāngyán are the outer end on the Fāngyán spectrum between dialect to standard language (the degree of standardization varies greatly across Southern Fāngyán, though).

Due to the great variation within Chinese, some researchers go the other direction and refer to Fāngyán as separate languages, or varieties that would be languages if defined via language-internal/Western means. Basing the definition of languages solely on language-internal arguments was rejected in chapter 2.1, because it excludes the influence humans have on the language they use. Instead, an approach to language definition was proposed for which language attitudes are paramount (see chapter 2.1.3). The consequence of applying this approach is to define Chinese as one language, since perceiving Chinese that way is extremely common in the Chinese-speaking realm (see e.g. Bradley 1992: 305 and Groves 2010: 541–542). However, this does not automatically diminish the great linguistic variation within the Chinese language – this needs to be kept in mind. As Mühlhäusler (1996: 328; see also chapter 2.2.3) points out, the field of linguistics believes “in the possibility of using the same descriptive labels for all languages”. However, Chinese illustrates that the descriptive labels developed for Western languages are not fully compatible with it.

Chapter 2.1.2 is dedicated to the concepts *Ausbau* and *Einbau languages* and it was mentioned that Chinese may also be a case of einbauization (García 2015: 357). The rapid process of standardization that Mandarin underwent in Mainland China, described in chapter 3.3, clearly shows that the de jure language policies were not implemented in the sake of gaining independence from another variety, but for modernization purposes. Additionally, there were no great attempts of standardizing Southern Fāngyán, so the Ausbau concept does

not fit in this case either. Since the Southern Fāngyán did not undergo a standardization process simultaneous to that of Mandarin, no traces of einbauization can be found either. Southern Fāngyán were not forced to adapt Mandarin features or vice versa. So, neither ausbauization nor einbauization become apparent in overt language policies. Yet, García's proposal can still be endorsed, since einbauization is present in language attitudes, i.e., in covert language policies – in a matter of fact, to an extreme degree, since Chinese is generally perceived as a single language, regardless of its great linguistic diversity. Therefore, Chinese can be listed as an example of einbauization – it is crucial to emphasize on what language policy dimension, though. From a theoretical perspective, einbauization processes within language attitudes are represented by *polytypic languages* in Croft's understanding of language (see chapter 2.1.3). Chinese is a prime example for polytypic languages, since its speakers perceive of it as one language, even though one could identify several languages by structural means.

3.4.3. Step II: Which Fāngyán Have Explicit and/or Implicit Norms?

After the definition of Chinese, the second major step for applying the concept of pluricentricity is to discuss its standardization. As chapter 2.2 shows, defining standard is challenging, since it is closely intertwined with language ideologies (chapter 2.2.2) and a useful tool for nation-states (chapter 2.2.3). Nevertheless, the stages (of implementation) of standardization, proposed by Haugen (1966) and Milroy and Milroy (1985) (see chapter 2.2.1), offer a rough structure for orientation. The standardization situation of Mandarin is fairly simple to identify. A mono-dialectal basis was chosen for the basis of standardization, it was codified rigorously, its functionality was expanded, and the new standard language was (and is) diffused extensively (see chapter 3.3.1 to 3.3.4). The standard variety of Mainland China is further corroborated by the high prestige attributed to it. It can be stated with confidence, that Mandarin has at least one standard variety.

Dealing with the standardization of Southern Fāngyán is not that easy. In the case of Cantonese, there are no attempts of standardization from the regional governments, hence the selection of the norm basis is not explicit. Even though there are no explicit norms, there are well-established implicit norms and Cantonese is being used in a wide range of domains, indicating that Cantonese has an elaborated standard variety (see chapter 3.3.5). The outcome of a study on language attitudes of Chinese speakers (Groves 2010) illustrates that Cantonese

has in fact something resembling a standardized form, not only in its written form, as proposed by Bauer (2018), but also in its spoken form. In this study, three groups of Chinese speakers were presented with an open-ended questionnaire: 53 Cantonese-speaking people from Hong Kong, 18 Cantonese-speaking participants from Mainland China and 72 Pǔtōnghuà-speaking Mainlanders, all of which being university students in Hong Kong. Groves (2010) uncovers an interesting contradiction in her data. When the participants were asked directly, whether written Cantonese should be standardized and taught as a subject in school, the majority of participants negated the question (nearly three quarter of the Pǔtōnghuà speakers and roughly three-fifths of all Cantonese speakers) (Groves 2010: 538–539). It must be noted that about a fifth of the Hongkongese participants declined this question because they believe that "it was simply unnecessary as it was already happening anyway" (2010: 539), in reference to the young generation picking up the Cantonese implicit norms informally. In contrast to that, the answers of questions 4 ("Do you speak Cantonese with an accent?") and 5 ("Where is the best Cantonese spoken?") indicate that there are actually implicit norms. The majority of the Hongkongese participants (43 out of 53) assumes that they speak Cantonese with no accent – this is interpreted by the study author as "recognizing that they spoke a variety recognized as a standard" (Groves 2010: 543). The same participants assume that the Hongkongese variety of Cantonese is the best (roughly 85%). This opinion is not widespread amongst the Cantonese speakers of Mainland China (only about 6%), they state that the Cantonese spoken in the area of Guangzhou is the best (50%), or that the Cantonese spoken in Hong Kong and around Guangzhou are equally good (about 39%). "This", Groves (2010: 543) concludes, "acknowledges the two major competing varieties of Cantonese and reinforces the statements in the literature that the Hong Kong variety has recently taken over from Guangzhou ('Canton') Cantonese as the more prestigious variety". This shift is also discussed in e.g. Poon (2010: 50).

Thinking back to the stages (of implementation) of standardization (Haugen 1966; Milroy & Milroy 1985), Cantonese is an example for going through most of these stages without governmental implementation but with de facto language policies. The situations of Cantonese in Hong Kong and Guangdong are very different, though. Whereas Cantonese is only a home language in Mainland China and not taught and used in schools at all, Cantonese covers a wide range of domains in Hong Kong and, even though not taught as a subject,

Cantonese is used as a medium of instruction in Hongkongese schools. So, the two varieties differ especially regarding their diffusion, as well as their prestige.

The third Fāngyán emphasized in this Master's thesis is Hokkien. As mentioned in chapter 3.3.5, Hokkien holds its strongest position in Taiwan. The Taiwanese government is backing the use of Hokkien and works towards revitalizing it. These efforts include a broad range of corpus planning activities. On the website of the Taiwanese Ministry of Education (MOE 2012), one can find instructions on how to select characters for writing Hokkien, which characters are generally recommended for writing Hokkien, and a user manual for a standardized phonetization system for Hokkien. Additionally, the *Dictionary of Frequently-Used Taiwan Minnan* was made publicly available online in 2011 (MOE 2013). Hence, it can be assumed that Hokkien enjoys a certain degree of standardization in Taiwan. It is codified to some extent (the writing, as well as the phonetization system), and diffused, since Hokkien is taught in Taiwanese schools. In comparison to the Hongkongese standard of Cantonese, Taiwanese Hokkien has a weaker position, though. Hokkien has not been elaborated to cover a wide range of functions, and prestige-wise is also less esteemed. In Singapore, no governmental efforts are supporting the use of Hokkien, however, there is significant public interest in reviving the once widespread use of Hokkien, as an article of the New York Times points out (New York Times: 2017). As Tien (2012 & 2016) emphasizes, Hokkien is also a common home language in neighboring countries, such as Malaysia, in which the *Hokkien Language Association* organizes the *Speak Hokkien Campaign* (Persatuan Bahasa Hokkien Pulau Pinang 2017). Within Mainland China, more precisely in Fujian, no activities could be made out about attempts to standardize and/or revive Hokkien. One reason may be the rigorous diffusion of Pǔtōnghuà that replaces Hokkien. Another hindrance could be that the non-standard varieties in the Fujian province are highly diverse, Ramsey (1987: 108) even calls them "the most heterogenous in China". Generally speaking, there is too little research on Hokkien available to the author of this Master's thesis for drawing profound conclusions on its standardization situation – in Mainland China and elsewhere. Hence, additional analyses on Hokkien, especially studies on language attitudes of its speakers, are a desideratum for further research.

3.4.4. Step III: Which Fāngyán Are Pluricentric?

The third and last step for perceiving Chinese within the framework of pluricentricity is to scrutinize the above identified standard varieties under the theoretical deliberations presented in chapter 2.3 about standard language variation. The first Fāngyán discussed from this perspective will be the most dominant one, Mandarin. Cantonese and Hokkien follow accordingly.

The literature on Chinese pluricentricity generally agrees on the existence of three centers of Mandarin – Mainland China, Singapore and Taiwan. Additionally, there are some indications, that Hong Kong does not follow the Beijing norms entirely, but has some own implicit norms. One important aspect is that the simplified characters of Mainland China, attached to Pǔtōnghuà, are not used in Hong Kong, but the traditional script is used instead (see chapter 3.3.6). The pluricentric varieties are not geographically continuous and there are many different reasons for the pluricentricity of Mandarin. In the case of Mainland China and Taiwan, there is the *political division* alongside the Capitalist-Communist spectrum. Chinese, in general, made its way to Singapore through *immigration*, however, the implementation and diffusion of Mandarin was a political decision – this is something Clyne's (2004: 296) categories do not address. The Hongkongese situation could be attributed to the category *redrawing of borders* in its widest sense, since Hong Kong underwent a transfer of sovereignty twice (to the UK and back to PRC). The two criterion that need to be met in Muhr's (2016: 20–21) list of criteria for determining a pluricentric language can both be affirmed – Mandarin is spoken in all four regions (criterion 1) and they all have official status in these regions (criterion 2).

Determining the degree of the asymmetrical relation between pluricentric varieties is too big of a task for this Master's thesis, since it would require detailed analyses of model speakers' language usage. This is another desideratum for further research. This Master's thesis, hence, puts its focus on dictionaries, to at least uncover certain tendencies that could be validated in further studies.

Mainland China has several Chinese dictionaries that are published within the center – an example is the aforementioned *Dictionary of Modern Chinese*, explicitly using *Hànyǔ* in its title. Taiwanese publishing houses have also issued Chinese dictionaries, e.g. 康軒國語詞典 (Kāngxuān Guóyǔ Cídiǎn; *Kangxuan Chinese dictionary*), using *Guóyǔ* in its title. With the great dominant position Pǔtōnghuà occupies and the many published dictionaries, it can be

assumed that Pǔtōnghuà is a full center, i.e. fully endonormative. Guóyǔ does not enjoy such a powerful position, nevertheless, it has its own dictionaries, therefore implying that the Taiwanese variety of Mandarin is more endonormative than exonormative. No Singaporean and Hongkongese dictionaries for Mandarin could be found, suggesting that they are predominantly exonormative.

The single most important dictionary for Chinese pluricentricity, though, is the *Global Chinese Dictionary*, published in 2010. This dictionary contains a collection of regional variants, indicates where these variants are commonly used, and informs its readers about possible semasiological differences. The regions included in the dictionary are not only the ones covered in this Master's thesis (Mainland China, Hong Kong, Taiwan, and Singapore), but also, amongst others, Macao, Malaysia, Thailand, Indonesia, Japan, Australia, the USA, and Canada (see Li 2016: 282–283). The description[9] provided in an online store for the *Global Chinese Dictionary* (Chinabooks 2018) states that the dictionary is oriented towards the regional standards and tries to consider the habits of the various Chinese communities. The key in this statement is the reference to the regional varieties as 标准 (biāozhǔn; *standard*), which shows an explicit acknowledgement of Mandarin pluricentricity. This dictionary is an indication that the dominant variety Pǔtōnghuà recognizes that there are, in fact, other varieties of standardized Mandarin – something that is not always the case in pluricentric languages (see Clyne's (2004: 297) differences between dominant and non-dominant varieties in chapter 2.3.3). The dictionary has over ten thousand entries, of which more than 1400 entries are words that are used anywhere, over 5000 are used in the Mainland, roughly 4200 in Hong Kong and Macao, more than 3300 in Taiwan and another 1800 in Singapore, Malaysia, and Thailand (Li 2016: 283). This is interesting, as there are more entries for Hong Kong – a Mandarin variety not covered in the theoretical elaborations on Chinese pluricentricity yet – than for Taiwan – a regional variety that has much stronger recognition for being a pluricentric variety of Mandarin. This also speaks for Hongkongese Mandarin being its own regional variety. However, Li (2016: 284) emphasizes that the geographical affinity of entries in the dictionary are highly unbalanced.

[9] This description is probably taken from the dictionary itself, e.g. its preface, since it says "本词典" ("běn cídiǎn"; "our dictionary"). However, this is not made explicit in the online store.

The *Global Chinese Dictionary* is interpreted as a verification of Mandarin pluricentricity, confirming the existence of at least four centers – Mainland China, Taiwan, Singapore and Hong Kong. The examination of other Chinese communities (such as in Malaysia and Indonesia) would lead to a more comprehensive description of the regions not covered in this Master's thesis.

Cantonese – the second Fāngyán under scrutiny for pluricentricity in this Master's thesis – does not have any explicit norms compiled on a governmental level, such as Mainland Pǔtōnghuà has. However, implicit norms are in place, as the discussion in chapter 3.4.3 shows. These norms are not monocentric, though. Groves (2010: 543) confirms in her study that Cantonese has two centers – Hong Kong and the area around Guangzhou – that are competing in prestige. Poon (2010: 8) gives four reasons, why an own independent Cantonese standard developed in Hong Kong. First, Cantonese in Hong Kong has been exposed to English, causing loanwords to find their way into the Hongkongese variety of Cantonese. Second, it has become popular to use mixed code in Hong Kong (i.e. using English words/phrases in Cantonese). Third, Cantonese in Guangdong is under strong influence of Pǔtōnghuà, making it adapt phrases of Mandarin in their Cantonese variety. Last but not least, Cantonese in Hong Kong was able to expand to various domains and new expressions are coined, especially in the entertainment sector. Therefore, this Master's thesis assumes that Cantonese has two pluricentric varieties.

The cause for Cantonese being pluricentric is the same as the occurrence of an independent Hongkongese variety of Mandarin: the *redrawing of borders*, due to the two-time transfer of sovereignty. In contrast to Mandarin, the Cantonese pluricentric situation is a geographically contiguous one. The two first criteria of Muhr (2016: 20–21) are met by Cantonese, too. Cantonese occurs in at least two self-governing entities (Hong Kong and Mainland China). Regarding the official status, the second criteria is more challenging because Cantonese has an official status in Hong Kong, in that Chinese was introduced as the second official language of Hong Kong in 1974 and *Chinese* was interpreted as *Cantonese* (see chapter 3.3.6). Cantonese is not referred to by its name as an official language of Hong Kong, though. Nevertheless, Cantonese enjoys strong ethno-linguistic awareness in both Hong Kong and Guangdong, hence, the second criteria is met as well. Regarding the relation between the two varieties, this Master's thesis does not have the insight into Cantonese codices and model speakers to place the two varieties precisely on Ammon's (1989: 90) spectrum of full endo- to

full exonormativity. However, it appears in the study by Groves (2010) that the Hongkongese variety is more dominant and, hence, more endonormative, than the Mainland Chinese variety.

The last Fāngyán examined from a perspective of pluricentricity is Hokkien. There is one article that is dedicated to Hokkien pluricentricity. Tien (2012: 457) makes out four major Hokkien varieties – namely in Taiwan, Singapore, the Fujian province in Mainland China and overseas (encompassing the South-East Asian region, such as Malaysia and Indonesia), as mentioned already in chapter 3.4.1. Hokkien is very diverse in its original home region in the Fujian province in Mainland China. To incorporate it in a pluricentric understanding of Hokkien, Tien (2012: 458) points out that the variety used around the city of Xiamen is likely to be the most understood. The Hokkien variety enjoying the best de jure position is the Taiwanese one. There are efforts made by the Ministry of Education to codify and spread Hokkien, whereas Hokkien plays a non-dominant role in all other regions – dominated either by a pluricentric variety of Mandarin, or by other languages, such as English or Bahasa Indonesia (see chapter 3.4.3). The Singaporean variety of Hokkien gains its distinctiveness mainly through lexical and pragmatic features (Tien 2012: 461).

The reason for Hokkien's pluricentricity is *migration* – over the centuries, many people emigrated from Fujian/China to Taiwan and South-East Asia. Hence, Hokkien is not geographically continuous. Muhr's (2016: 20–21) first criteria is met, since Hokkien does occur in various regions, however, it only enjoys official status in Taiwan. In the other regions, Hokkien has some ethno-linguistic awareness, how much, indeed, would need to be uncovered in further studies. Since Hokkien is being codified in Taiwan, this suggests that Taiwanese Hokkien is a variety that follows its own norms, i.e. being endonormative. However, Hokkien's original speech area is Fujian, and references to the variety spoken there can be found even in an article reflecting on the creation of a Hokkien textbook for the Taiwanese variety (Fuehrer 2016). Therefore, the relation between the various Hokkien varieties begs for more in-depth research. In general, it can be assumed that Hokkien has at least three centers, all of which play a non-dominant role in the region they are spoken in: Fujian, Taiwan and Singapore. More research needs to be conducted in regard to the possible varieties in South-East Asian countries.

3.5. The Chinese Language in Retrospection

Chapter 3 has provided detailed insight into the Chinese language from the theoretical perspective this Master's thesis has presented in chapter 1 and 2. First of it was emphasized that this Master's thesis does not put its focus on the existence of officially recognized nation-states but to self-governing entities instead, hence allowing for a more realistic description of Chinese that is dedicated to pluricentricity. By both rejecting to call the Chinese Fāngyán *dialects* nor *languages*, the path was paved towards the application of pluricentricity, scrutinizing each Fāngyán independently. Beforehand, the history of the Chinese language was sketched and the development of the standardization process in Mainland China was depicted in detail, for it is highly relevant for the application of pluricentricity. Mandarin, Cantonese, and Hokkien were scrutinized in regard to their degree of standardization, on the basis of which potential pluricentric varieties were categorized and discussed.

This Master's thesis has uncovered various aspects regarding Chinese pluricentricity that are in dire need of further research (for concrete research desiderata see chapter 4). Nevertheless, this Master's thesis proposes that all the Fāngyán discussed here – namely Mandarin, Cantonese, and Hokkien – are pluricentric (see figure 5).

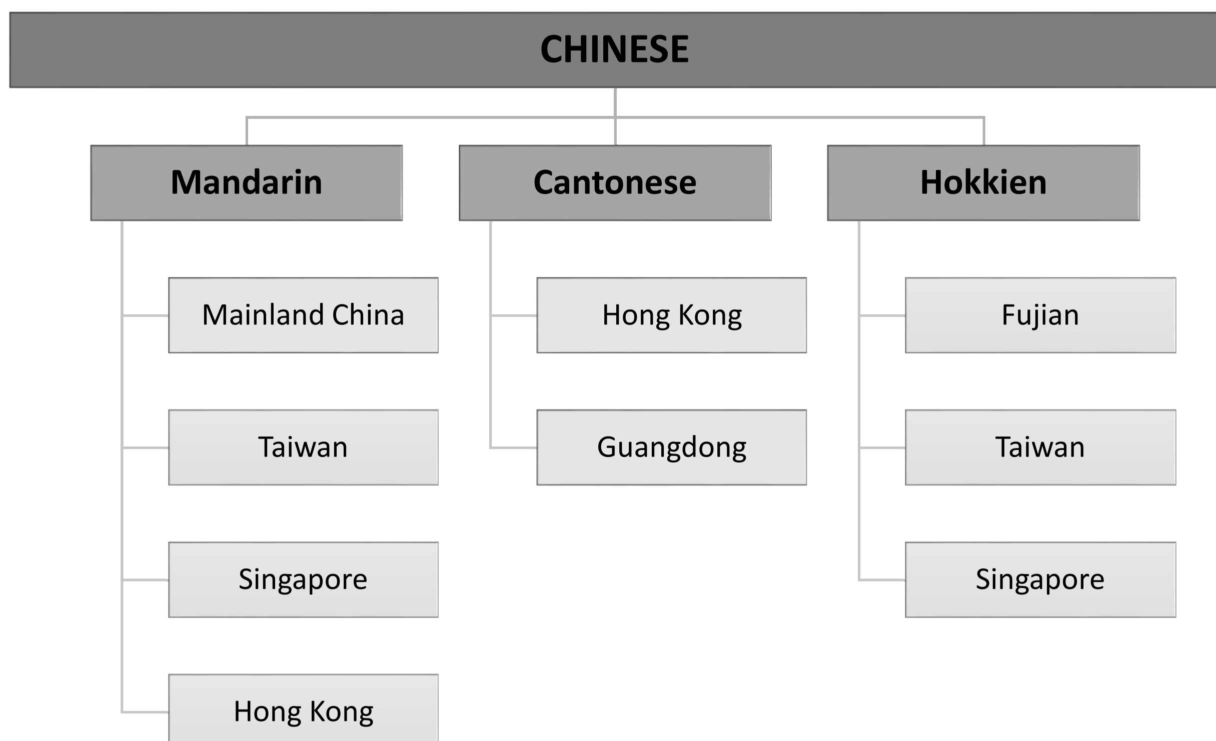

Figure 5: The pluricentric varieties of Chinese

This proposal is possibly still expandable and certain varieties need more in-depth research for verification. Furthermore, the order, in which the Fāngyán and their varieties are listed in figure 5, shall not indicate any role, rank order whatsoever, the same goes for figure 6.

In addition to the pluricentricity of the various Chinese Fāngyán, it is also legitimate to speak of the Chinese scripts and the different annotation systems as pluricentric, as Tien (2016: 50–51; see chapter 3.4.1) suggests. Figure 6 illustrates this including each variety's center in brackets to conclude the picture of Chinese pluricentricity.

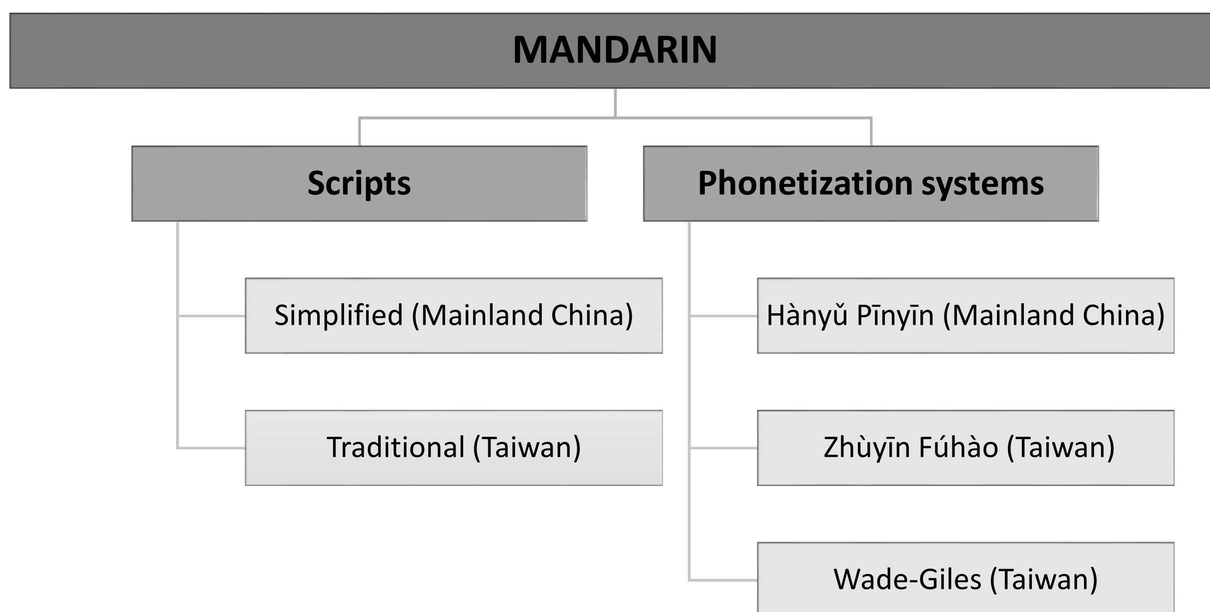

Figure 6: The pluricentricity of scripts and phonetization systems in Mandarin (Tien 2016: 50–51; visually edited by author)

An aspect, that is highly important, yet not made explicit by Tien (2016), is that the pluricentricity of scripts and phonetization systems refers only to the dominant Fāngyán Mandarin. Possible pluricentric varieties in script and phonetization systems may still be uncovered in the Southern Fāngyán – another desideratum for future research.

Chapter 4 will now conclude this Master's thesis and will offer an outlook to further possible research questions in the field of Chinese pluricentricity.

4. The Applicability of Pluricentricity to Chinese – Conclusion and Outlook

> [I]t seems paradoxical that the application of the notion of pluricentricity or 'pluricentric languages' to Chinese has not gained any more acceptance. (Klöter 2016: 64)

The same thought came to the author of this Master's thesis after having attended Mandarin language courses in both Mainland China and Taiwan and having travelled extensively throughout both regions. Detecting the various differences in the taught and used standard varieties is not a difficult task, even for a person who is not fluent in Mandarin. Hence, the research question for this Master's thesis emerged: *To what extent is the notion of pluricentricity applicable to the Chinese language?*

Many aspects had to be discussed before the full extent of Chinese pluricentricity could be grasped. Chapter 1 is dedicated to language policy, describing the shift in research focus from explicit language planning to implicit language beliefs and behavior. This chapter provided the tools for a systematic approach to Chinese pluricentricity, understanding its underlying dynamics.

Chapter 2 constructs the theoretical scaffold of this Master's thesis, covering the ways of defining language and analyzing the notion of standard language before going into depth on the concept of pluricentricity. Two ways of defining language could be identified. The language-internal approach defines languages based on their intrinsic components, whereas the language-external approach takes humankind into account as an influential factor, basing its language definition on language attitudes. Subsequently, the notion of standard language was identified as a product of language ideologies. Standard languages are a powerful tool for the creation of nation-states, in which the ideology of homogeneity in language is utilized as evidence for the homogeneity and, hence, shared identity of a people, through which nation-states legitimize their existence. Since standard language is a product of ideology, only indices can be collected for determining the degree of standardization. Furthermore, it is important to keep in mind, that only through the creation of (the belief in) a standard language and by defining which local variety/varieties is/are taken as the basis for this standard language, the notion of non-standard varieties emerges. Put more simply, if there is no standard language, there is no dialect. As a final step in the theoretical elaborations in this Master's thesis, the

concept of pluricentricity was taken under scrutiny. Nation-states play a crucial role in pluricentricity, since one of the criteria for a pluricentric language is that the language has to be used in several nation-states. However, the official recognition of certain regions in the East Asian context (e.g. Taiwan) poses a complex political issue. Therefore, this Master's thesis does not attribute too much importance to the official recognition of such controversial regions, but emphasizes the existence of self-governing entities instead, since they can have an influence on the standard variety used in such regions. Hence, this Master's thesis differentiates between Taiwan and China, without positioning itself in the ongoing debate regarding Taiwan's independence. Instead, this Master's thesis strives for an accurate depiction of the linguistic and social reality of the Chinese-speaking areas. Various aspects of pluricentricity were presented in chapter 2, such as the reasons for the existence of pluricentric varieties, possible approaches to categorizing pluricentric languages, and insights in the asymmetrical relation between the standard varieties of a language.

The centerpiece of this Master's thesis is chapter 3, in which the theoretical elaborations in chapter 1 and 2 were applied to the Chinese language. The first part is dedicated to contextualizing and conceptualizing Chinese – a very basic, yet crucial step for the application of pluricentricity to Chinese. The term *Chinese* is on the same level with various language groups in the Sino-Tibetan language family, making it a highly diverse, yet underclassified language. Many researchers advocate the further classification of Chinese, perceiving of the seven main varieties as independent languages. In stark contrast, others refer to these varieties as mere *dialects* of Chinese. Both approaches have their advantages, however, they both do not align with the linguistic reality found in East and South-East Asia. In accordance with the theory presented in the previous chapters, this Master's thesis rejects both approaches and proposes to call these varieties by their Chinese name, *Fāngyán*. Also, other terms, such as *topolect* and *regiolect*, were rejected, since they all only cast light on certain aspects of Chinese, but never fully illuminate the whole picture. As a next step, the developments of the Chinese language were sketched, and an emphasis was put on the standardization processes in Mainland China in the first half of the 20th century. This turbulent time was highly defining for the Chinese language as a whole and for the various Fāngyán in particular. Lastly, Chinese was discussed from a pluricentric standpoint. The indications of and elaborations on Chinese pluricentricity in the academic literature were compiled and presented at first, before using the theoretical scaffold of this Master's thesis to define to

what extent Chinese can be classified as a pluricentric language. Several suggestions from the literature were rejected, since they presented the Chinese language as if new categories must be formulated so that it could be included in the theory of pluricentricity. This Master's thesis strongly believes, that there is a way of incorporating Chinese in the list of pluricentric languages, without doing so.

By means of detailed discussion of the explicit and implicit norms, the socio-political situations and the speakers' language attitudes, various Chinese varieties (of Mandarin, as well as of Cantonese and Hokkien) could be identified as pluricentric. This Master's thesis has defined all three Fāngyán in focus as pluricentric. There are different standardized varieties of Mandarin used in Mainland China, Taiwan, Singapore, and Hong Kong. Cantonese has two different de facto standard varieties, one center being Hong Kong, the other the province of Guangdong in Mainland China. Last but not least, at least three different standard varieties of Hokkien could be identified, namely that of the Fujian province in Mainland China, Taiwan, and Singapore – all of which are in a non-dominant position in comparison to the Mandarin standard varieties used in these areas. Furthermore, Mandarin cannot only be classified as pluricentric in regard to the differences it reveals in the different centers, its scripts and phonetization systems can also be referred to as pluricentric. There are two Chinese scripts, the simplified one with its center in Mainland China, and the traditional one with Taiwan acting as its center. There are three common phonetization systems for Mandarin. Mainland China exerts strong influence on Hànyǔ Pīnyīn, whereas Taiwan has the greatest impact on Zhùyīn Fúhào and Wade-Giles.

On the basis of the extensive discussion presented in this Master's thesis, the research question can be answered confidently: The notion of pluricentricity is fully applicable to the Chinese language and it does not require the introduction of new categories or the like. If Chinese is defined with a language-external approach, respecting the way Chinese is perceived by its speakers, then nothing stands in the path of applying the notion of pluricentricity to Chinese in a way that aligns with the linguistic and social reality.

This Master's thesis offers the first application of pluricentricity to Chinese that encompasses the entirety of linguistic variation of the Chinese language, without creating new categories and conceptualizations for doing so. This is a crucial accomplishment for research on pluricentricity, since it makes Chinese comparable to other pluricentric languages. Yet, it

still only scratches the surface of Chinese pluricentricity. In the course of this Master's thesis, many research gaps could be identified that may serve as a point of departure for future analyses. As to the limited space and resources, this Master's thesis could not deal with the language usage of model speakers, which is highly relevant for determining the norm orientation of the pluricentric varieties. Qualitative studies on language attitudes would give more insight e.g. into the existence of implicit norms. In general, more research is required on the pluricentricity of Hokkien, Cantonese and the Mandarin variety of Hong Kong – the latter two of which are characterized as pluricentric for the first time altogether in this Master's thesis. This includes possible additional centers not addressed in this Master's thesis and the degree of ethno-linguistic awareness in the discussed pluricentric centers. Last but not least, the pluricentric perspective assumed for the scripts and phonetization systems of Mandarin may also be applicable to the script and phonetization system of Cantonese and Hokkien.

Bibliography

Aikhenvald, Alexandra Y. & Dixon, R. M. W. 2006. Introduction. In: Aikhenvald, Alexandra Y. & Dixon, R. M. W. (eds.). *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press, 1–26.

Ammon, Ulrich & Bickel, Hans & Ebner, Jakob & Esterhammer, Ruth & Gasser, Markus & Hofer, Lorenz & Kellermeier-Rehbein, Birte & Löffler, Heinrich & Mangott, Doris & Moser, Hans & Schläpfer, Robert & Schlossmacher, Michael & Schmidlin, Regula & Vallaster, Günter (eds.). 2004. Die nationalen Voll- und Halbzentren des Deutschen. In: *Variantenwörterbuch des Deutschen*. Berlin/New York: de Gruyter, XXXI – LXXV.

Ammon, Ulrich. 1989. Towards a Descriptive Framework for the Status/Function (Social Position) of a Language within a Country. In: Ammon, Ulrich (ed.). *Status and Function of Languages and Language Varieties*. Berlin/New York: Walter de Gruyter, 21–106.

Ammon, Ulrich. 2005a. Pluricentric and Divided Languages. Plurizentrische und geteilte Sprachen. In: *Sociolinguistics: An International Handbook of the Science of Language and Society.* 2nd edition. Berlin: de Gruyter, 1536–1542.

Ammon, Ulrich. 2005b. Standard und Variation: Norm, Autorität, Legitimation. In: Eichinger, Ludwig M. & Kallmeyer, Werner (eds.). *Standardvariation: Wie viel Variation verträgt die deutsche Sprache?* Berlin/New York: de Gruyter, 28–40.

Anderson, Benedict. 2006. *Imagined Communities*. London: Verso.

Baldauf, Richard B. 1994. ‚Unplanned' language planning and policy. In: Grabe, William (ed.). *Annual Review of Applied Linguistics*. Cambridge: CUP, 82–89.

Bauer, Robert S. 2018. Cantonese as written language in Hong Kong. In: *Global Chinese*, 4/1, 103–142.

Bradley, David. 1992. Chinese as a pluricentric language. In: Clyne, Michael (ed.). *Pluricentric Languages. Differing Norms in Different Nations*. Berlin/New York: Mouton de Gruyter, 305–323.

CCTV News. 2012. *Shè kē yuàn yuán fù yuàn chángjiāng lán shēng: "Shèng nán" "shèngnǚ" duì mǒu xiē wèihūn nánnǚ yǒu shī zūnjìng* [Jiang Lansheng, former vice president of the Chinese Academy of Social Sciences: "leftover males" and "leftover women" is disrespectful towards some unmarried men and women]. Retrieved from http://news.cntv.cn/20120719/108538.shtml, 15th August 2018.

Chambers, J. K. & Trudgill, Peter. 1998. *Dialectology*. Cambridge: Cambridge University Press.

Chen, Ping. 1999. *Modern Chinese: History and Sociolinguistics*. Cambridge: Cambridge University Press.

Cheng, Robert & Zheng Liangwei. 1987. Borrowing and Internal Development in Lexical Change – A Comparison of Taiwanese Words and Their Mandarin Equivalents. In: *Journal of Chinese Linguistics*, 15/1, 105–131.

Chinabooks. 2018. *Quanqiu huayu cidian*. Retrieved from http://www.chinabooks.ch/catalog/product_info.php?products_id=8176, 15th August 2018.

Clyne, Michael & Kipp, Sandra. 1999. *Pluricentric Languages in an Immigrant Context. Spanish, Arabic and Chinese*. Berlin/New York: De Gruyter.

Clyne, Michael. 1992. *Pluricentric Languages: Differing Norms in Different* Nations. Berlin et. al.: Mouton De Gruyter.

Clyne, Michael. 2004. Pluricentric Language / Plurizentrische Sprache. In: Ulrich Ammon (ed.). *Sociolinguistics: an international handbook of the science of language and society.* 2nd edition. Berlin: Mouton de Gruyter, 296–300.

Constitution of the Republic of Singapore. 1965. Retrieved from https://sso.agc.gov.sg/Act/CONS1963?ValidDate=20170401&ProvIds=P1XIII-, 15th August 2018.

Cooper, Robert. 1989. *Language Planning and Social Change*. Cambridge: University Press.

Croft, William. 2000. *Explaining Language Change. An Evolutionary Approach*. Harlow: Longman.

Crystal, David. 2008. *A Dictionary of Linguistics and Phonetics*. 6th edition. Oxford: Blackwell.

Crystal, David. 2010. *The Cambridge Encyclopedia of Language*. 3rd edition. Cambridge: Cambridge University Press.

DeFrancis, John. 1984. *The Chinese Language. Fact and Fantasy.* Honolulu: University of Hawaii Press.

Deumert, Ana. 2003. Standard Languages As Civic Rituals – Theory and Examples. In: *Sociolinguistica*, 17, 31–51.

Encyclopedia of China [Zhōngguó dà bǎikē quánshū]. 1988. *Yǔyán* [Language]. Beijing: Encyclopedia of China Publishing House.

Ferguson, Charles A. 1959. Diglossia. In: *Word*, 15, 325–340.

Fishman, Joshua A. 1983. Modeling rationales in corpus planning: modernity and tradition in images of the good corpus. In: Cobarrubias, Juan & Fishman, Joshua A. (eds.). *Progress in language planning: International perspectives*. Berlin: Mouton, 107–118.

Fishman, Joshua A. 2000. The Status Agenda in Corpus Planning. In: Lambert, Richard D. & Shohamy, Elana (eds.). *Language Policy and Pedagogy. Essays in Honor of A. Ronald Walton*. Philadelphia/Amsterdam: John Benjamins, 43–51.

Fishman, Joshua A. 2008. Rethinking the *Ausbau-Abstand* dichotomy into a continuous and multivariate system. In: *International Journal of the Sociology of Language*, 191, 17–26.

Fishman, Joshua A., Ferguson, Charles A. & Das Gupta, Jyotirindra (eds.). 1968. *Language Problems of Developing Nations*. New York: John Wiley & Sons.

Fuehrer, Bernhard. 2016. *Southern Hokkien: An Introduction*: What We Did and Why We Did What We Did. In: *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, 166/2, 425–441.

Gal, Susan & Irvine, Judith T. 1995. The Boundaries of Languages and Disciplines: How Ideologies Construct Difference. In: *Social Research*, 62/4, 967–1001.

García, Ofelia. 2015. Language Policy. In: *International Encyclopedia of the Social & Behavioral Sciences*, 2/13, 353–359.

Glauninger, Manfred. 2013. Deutsch im 21. Jahrhundert: „pluri“-, „supra“- oder „postnational“? In: Fiala-Fürst, Ingeborg & Joachimsthaler, Jürgen & Schmitz, Walter (eds.). *Mitteleuropa: Kontakte und Kontroversen. Dokumentation des II. Kongresses des Mitteleuropäischen Germanistenverbandes (MGV) in Olomouc/Olmütz*. Dresden: Thelem, 459–468.

Goebl, Hans. 1975. “Le *Rey* est mort, vive le *Roy*.” Nouveaux regards sur la scriptologie. In: *Travaux de linguistique et de littérature* XIII, 1, 145–210.

Groves, Julie May. 2010. Language or dialect, topolect or regiolect? A comparative study of language attitudes towards the status of Cantonese in Hong Kong. In: *Journal of Multilingual and Multicultural Development*, 31/6, 531–551.

Haugen, Einar. 1959. Planning a Standard Language in Modern Norway. In: *Anthropological Linguistics,* 1/3, 8–18.

Haugen, Einar. 1966. Dialect, language, nation. In*: American Anthropologist*, 68, 922–935.

Haugen, Einar. 1983. The implementation of corpus planning: theory and practice. In: Cobarrubias, Juan & Fishman, Joshua A. (eds.). *Progress in language planning: International perspectives*. Berlin: Mouton, 269–290.

Hornberger, Nancy H. 2006. Frameworks and Models in Language Policy and Planning. In: Ricento, Thomas (ed.). *An Introduction to Language Policy. Theory and Method*. Malden, MA: Blackwell, 24–41.

Hull, David L. 1988. *Science as a process: an evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press.

Ising, Erika. 1987. Nationalsprache/Nationalitätensprache. In: Ammon, Ulrich & Dittmar, Norber & Mattheier, Klaus J. (eds.). *Sociolinguistics/Soziolinguistik*. Berlin/New York: de Gruyter, 335–344.

Jernudd, Björn H. & Neustupný, Jiří V. 1987. Language planning: for whom? In: Laforge, Lorne (ed.). *Proceedings of the International Colloquium on Language Planning*. Québec: Les Press de L'Université Laval, 69–84.

Joseph, John E. & Taylor, Talbot J. 1990. *Ideologies of Language*. London: Routledge.

Kane, Daniel. 2006. *The Chinese Language. Its History and Current Usage.* Tokyo/Rutland (Vermont)/Singapore: Tuttle Publishing.

Kangxuan Chinese dictionary [Kāngxuān Guóyǔ Cídiǎn]. 2010. New Taibei: Kang Hsuan Educational Publishing Group.

Kloss, Heinz. 1967. ‚Abstand Languages' and ‚Ausbau Languages'. In: *Anthropological Linguistics*, 9/7, 29–41.

Kloss, Heinz. 1969. *Research Possibilities on Group Bilingualism: a report*. Quebec: International Center for Research on Bilingualism.

Kloss, Heinz. 1978. *Die Entwicklung neuer germanischer Kultursprachen seit 1800.* 2nd edition. Düsseldorf: Schwann.

Klöter, Henning. 2008. Re-Writing language in Taiwan. In: Shih, Fang-Long & Thompson, Stuart & Tremlett, Paul (eds.). *Re-writing Culture in Taiwan*. London: Routledge, 102–122.

Klöter, Henning. 2016. "What is Correct Chinese?" Revisited. In: Tieken Boon van Ostade, Ingrid & Percy, Carol (eds.). *Prescription and Tradition in Language. Establishing Standards across Time and Space*. Multilingual Matters: Bristol, 57–70.

Kurpaska, Maria. 2010. *Chinese Language(s): A Look Through the Prism of The Great Dictionary of Modern Chinese Dialects*. Berlin: De Gruyter Mouton.

Lenz, Alexandra N. 2009. Emergence of varieties through restructuring and reevaluation. In: Auer, Peter & Schmidt, Jürgen Erich (eds.). *Theories and Methods*. Handbooks of Linguistics and Communication Science. Berlin/New York: Walter de Gruyter.

Li, Qingmei. 1992. Hǎixiá liǎng'àn zìyīn bǐjiào [Phonetic comparison of Mainland China and Taiwan]. In: *Yǔyán Yánjiū yǔ Yìngyòng* 3, 42–48.

Li, Rui. 2016. Review of A Dictionary of Global Huayu. In: *Advances in Language and Literary Studies*, 7/3, 282–285.

Li, Rulong. 1995. Mǐnnán fāngyán dìqū de yǔyán shēnghuó [Life in the Hokkien Fāngyán area]. In: *Yǔwén Yánjiū,* 2, 34–37.

Li, Yuming (ed.). 2010. *Quánqiú Huáyǔ Cídiǎn* [Global Chinese Dictionary]. Beijing: Commercial Press.

Mair, Victor H. 1991. What Is a Chinese "Dialect/Topolect"? Reflections on Some Key Sino-English Terms. In: *Sino-Platonic Papers*, 29, 1–31.

Mair, Victor H. n. d. H*ow to Forget Your Mother Tongue and Remember Your National Language*. Retrieved from http://pinyin.info/readings/mair/taiwanese.html, 15th August 2018.

Milroy, James & Milroy, Lesley. 1985. *Authority in language. Investigating language prescription and standardization*. London/Boston/Henley: Routledge & Kegan Paul.

Milroy, James. 2001. Language ideologies and the consequences of standardization. In: *Journal of Sociolinguistics*, 5/4, 530–555.

MOE [Ministry of Education, Taiwan]. 2012. *User Manual for Romanizing the Minnan Language*. Retrieved from https://english.moe.gov.tw/cp-34-14780-4DA52-1.html, 15th August 2018.

MOE [Ministry of Education, Taiwan]. 2013. *Dictionary of Frequently-Used Taiwan Minnan*. Retrieved from https://english.moe.gov.tw/cp-34-14785-E53FD-1.html, 15th August 2018.

Mühlhäusler, Peter. 1996. *Linguistic Ecology*. Oxford: Blackwell.

Muhr, Rudolf. 1997. Zur Terminologie und Methode der Beschreibung plurizentrischer Sprachen und deren Varietäten am Beispiel des Deutschen. In: Muhr, Rudolf & Schrodt, Richard (eds.). *Österreichisches Deutsch und andere nationale Varietäten plurizentrischer Sprachen in Europa*. Wien: Hölder-Pichler-Tempsky, 40–66.

Muhr, Rudolf. 2012. Linguistic dominance and non-dominance in pluricentric languages: A typology. In: Muhr, Rudolf (ed.). *Non-dominant Varieties of pluricentric Languages. Getting the Picture. In memory of Michael Clyne*. Wien: Peter Lang, 23–48.

Muhr, Rudolf. 2016. The state of the art of research on pluricentric languages: Where we were and where we are now. In: Muhr, Rudolf (ed.). *Pluricentric Languages and Non-Dominant Varieties Worldwide. Part I: Pluricentric Languages across Continents. Features and Usage.* Frankfurt/Main: Peter Lang, 13–38.

Muljačić , Žarko. 1989. Über den Begriff *Dachsprache*. In: Ammon, Ulrich (ed.). *Status and Function of Languages and Language Varieties*. Berlin/New York: de Gruyter, 256–277.

Nahir, Moshe. 1984. Language planning goals: A classification. In: *Language Problems and Language Planning*, 8, 294–327.

Nekvapil, Jiří. 2006. From Language Planning to Language Management. In: *Sociolinguistica*, 20, 92–104.

New York Times. 2017. *In Singapore, Chinese Dialects Revive After Decades of Restrictions*. Retrieved from https://www.nytimes.com/2017/08/26/world/asia/singapore-language-hokkien-mandarin.html, 15th August 2018.

Official Languages Ordinance. 1974. Retrieved from https://www.elegislation.gov.hk/hk/cap5!en@2017-02-15T00:00:00?INDEX_CS=N&keyword.PIT_DATE=&keyword.WITHIN_ALL=&keyword.LEG_TYPE=3&keyword.LEG_TYPE=1&keyword.LEG_TYPE=2&keyword.NEAR_PHRASE=&keyword.PART_NO=&keyword.NEAR_IN_ORDER=&keyword.WITHIN_TEXT=Y&keyword.SEARCH_KEYWORD=Official%20Languages%20Ordinance&keyword.NEAR_DIST=&keyword.ANY_WORDS=&keyword.ENABLE_WORD_STEMMING=&keyword.SCHEDULE_APPENDIX_ANNEX_NO=&keyword.EXACT_PHRASE=&keyword.SUBDIVISION_NO=&keyword.SEARCH_MODE=P&keyword.WITHIN_TAGS=&keyword.NO_WORDS=&keyword.SEARCH_FIELD=E&keyword.ALL_WORDS=&keyword.PIT_TIME=&keyword.CASE_SENSITIVE=&keyword.ADVANCED_SEARCH=&keyword.CHAPTER_TITLE=&keyword.SECTION_RULE_NO=&keyword.CHAPTER_NO=&keyword.DIVISION_NO=&keyword.PIT_TYPE=C&keyword.BILINGUAL=, 15th August 2018.

Persatuan Bahasa Hokkien Pulau Pinang. 2017. *Speak Hokkien Campaign*. Retrieved from https://www.speakhokkien.org/english, 15th August 2018.

Pohl, Heinz Dieter. 1997. Gedanken zum Österreichischen Deutsch (als Teil der „pluriarealen“ deutschen Sprache). In: Muhr, Rudolf & Schrodt, Richard (eds.). *Österreichisches Deutsch und andere nationale Varietäten plurizentrischer Sprachen in Europa*. Wien: Hölder-Pichler-Tempsky, 40–66.

Poon, Anita Y.K. 2010. Language use, and language policy and planning in Hong Kong. In: *Current Issues in Language Planning*, 11/1, 1–66.

Ramsey, Robert S. 1987. *The Languages of China*. Princeton/ New Jersey: Princeton University Press.

Ricento, Thomas (ed.). 2006. *An Introduction to Language Policy. Theory and Method*. Malden, MA: Blackwell.

Ricento, Thomas. 2000. Historical and theoretical perspectives in language policy and planning. In: *Journal of Sociolinguistics*, 4/2, 196–213.

Riesel, Elise. 1953. K voprosu o nacional'nom jazyke v Avstrii [Regarding the national language in Austria]. In: *Charkow*, 5, 157–171.

Sandel, Todd L. 2003. Linguistic Capital in Taiwan: The KMT's Mandarin Language Policy and Its Perceived Impact on Language Practices of Bilingual Mandarin and Tai-gi Speakers. In: *Language in Society*, 32/4, 523–551.

Schiffman, Harold F. 1996. *Linguistic Culture and Language Policy*. New York: Routledge.

Scott, Mandy & Tiun, Hak-Khiam. 2007. Mandarin-Only to Mandarin-Plus: Taiwan. In: *Language Policy*, 6, 53–72.

Shohamy, Elana. 2006. *Language Policy: Hidden agendas and new approaches*. New York: Routledge.

Smakman, Dick. 2012. The definition of the standard language: a survey in seven countries. In: *International Journal of the Sociology of Language*, 218, 25–58.

Spolsky, Bernard & Shohamy, Elana. 2000. Language Practice, Language Ideology, and Language Policy. In: Lambert, Richard D. & Shohamy, Elana (eds.). *Language Policy and Pedagogy. Essays in Honor of A. Ronald Walton*. Philadelphia/Amsterdam: John Benjamins, 1–41.

Spolsky, Bernard. 2004. *Language Policy*. Cambridge: Cambridge University Press.

Stepanov, Georg V. 1957. Problemy izucenija ispanskogo v jazykac Latinskoj Amerike [Problems of Researching the Spanish language in Latin America]. In: *Vosprosy jazykoznanija,* 4, 16–25.

Stewart, William A. 1968. A Sociolinguistic Typology for Describing National Multilingualism. In: Fishman, Joshua A (ed.). *Readings in the Sociology of Language*. The Hague/Paris: Mouton, 531–545.

Sun, Chaofen. 2006. *Chinese. A Linguistic Introduction*. New York: Cambridge University Press.

Thomason, Sarah & Kaufman, Terrence. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.

Tien, Adrian. 2012. Chinese Hokkien and its lexicon in Singapore: Evidence for an indigenised Singapore culture. In: Muhr, Rudolf (ed.). *Non-dominant varieties of pluricentric languages. Getting the Picture. In memory of Michael Clyne*. Wien et. al.: Peter Lang, 453–472.

Tien, Adrian. 2016. Perspectives on „Chinese“ pluricentricity in China, Greater China and beyond. In: Muhr, Rudolf (ed.). *Pluricentric languages and non-dominant varieties worldwide: part I: pluricentric languages across continents: features and usage*. Frankfurt am Main: Peter Lang, 41–60.

Tollefson, James W. 1991. *Planning language, planning inequality. Language policy in the community*. London/New York: Longman.

Tosco, Mauro. 2008. Introduction: *Ausbau* is everywhere! In: *International Journal of the Sociology of Language*, 191, 1–16.

Wang, Hui & Yuan, Zhongrui. 2013. The promotion of Putonghua (Mandarin Chinese): An overview. In: Li, Yuming & Li, Wei (eds.). *The Language Situation in China*, Volume 1, 2006–2007. Berlin/Boston: de Gruyter, 27–39.

Wang, Jun. 1995. *Dāngdài Zhōngguó de wénzì gǎigé* [The Chinese character reform of China]. Beijing: Dāngdài Zhōngguó Chūbǎnshè.

Wang, Li. 1979. Báihuàwén yùndòng de yìyì [The significance of the movement of Báihuà writing]. In: *Zhōngguó Yǔwén*, 3, 161–162.

Weber, Jean-Jacques & Horner, Kristine. 2012. *Introducing Multilingualism. A social approach*. New York: Routledge.

Wikimedia Commons. 2017. *Yellow River map*. Retrieved from https://en.wikipedia.org/wiki/Yellow_River#/media/File:Yellowrivermap.jpg, 15th August 2018.

World Population Review. n.d. *2018 World Population by County (live)*. Retrieved from http://worldpopulationreview.com/, 15th August 2018.

Wright, Sue. 2004. *Language Policy and Language Planning. From Nationalism to Globalisation.* New York: Palgrave.

Xu, Shirong. 1982. ’Guójiā tuīxíng quánguó tōngyòng de Pǔtōnghuà’ – dù ‘Xiànfǎ xiūgǎi cáo’ān’ dì èrshí tiáo [‘The state promotes Putonghua which is universally used throughout China’ – On reading Article 20 of the ‘Draft of the Revised Constitution’]. In: *Wénzì Gǎigé,* 2, 15–16.

Appendix

English Abstract

This Master's thesis examines to what extent the notion of pluricentricity is applicable to the Chinese language. There is little consensus in the literature about how to deal with Chinese pluricentricity. The concept of pluricentricity is either not applied to Chinese as a whole (but only to Mandarin), or entire new categories and conceptualizations are developed that describe only the phenomenon of Chinese pluricentricity, making it difficult to compare Chinese with other pluricentric languages. Additionally, the definition of Chinese itself is highly controversial, some arguing to perceive of the seven major varieties (方言, Fāngyán) as dialects, others advocating for dealing with Fāngyán as independent languages.

Therefore, this Master's thesis is devoted to the definition of language and the concept of standard language, before Chinese is scrutinized through a pluricentric lens. This Master's thesis assumes that Fāngyán can cover the whole spectrum between dialect and standard, yet it does not perceive of them as independent languages, since this would not align with the language attitudes of Chinese speakers. On the basis of this assumption, this Master's thesis could identify at least three Fāngyán that have several standard varieties: Mandarin (with its centers in Mainland China, Taiwan, Singapore, and Hong Kong), Cantonese (in Hong Kong and Guangdong), and Hokkien (in Fujian, Taiwan, and Singapore). Furthermore, Mandarin cannot only be classified as pluricentric in regard to the differences it reveals in the different centers, its scripts – simplified (Mainland China) and traditional (Taiwan) – and phonetization systems – Hànyǔ Pīnyīn (Mainland China) and Zhùyīn Fúhào and Wade-Giles (Taiwan) – can also be perceived as pluricentric.

German Abstract

Diese Masterarbeit untersucht, inwieweit das Konzept der Plurizentrik auf die chinesische Sprache anwendbar ist. In der Literatur gibt es nur wenig Konsens darüber, wie man mit chinesischer Plurizentrik umgehen soll. Das Konzept der Plurizentrik wird entweder nicht auf die Gesamtheit des Chinesischen angewandt (sondern nur auf Mandarin), oder komplett neue Kategorien und Konzeptualisierungen werden entwickelt, die nur das Phänomen der Plurizentrik des Chinesischen beschreiben und somit Chinesisch nur schwer mit anderen plurizentrischen Sprachen verglichen werden kann. Außerdem ist die Definition der chinesischen Sprache an sich sehr umstritten, manche setzen sich dafür ein, dass man die sieben Hauptvarietäten des Chinesischen (方言, Fāngyán) als Dialekte ansieht, andere argumentieren für die Handhabung von Fāngyán als unabhängige Sprachen.

Daher widmet sich diese Masterarbeit zuerst der Definition von Sprachen und dem Konzept von Standardsprache, bevor Chinesisch durch eine plurizentrische Linse betrachtet wird. Diese Masterarbeit nimmt an, dass Fāngyán das ganze Spektrum von Dialekt bis Standard bedienen können, trotzdem werden sie nicht als unabhängige Sprachen bezeichnet, da dies nicht mit den Spracheinstellungen von Chinesischsprecher*innen übereinstimmen würde. Auf Basis dieser Annahme konnte diese Masterarbeit mindestens drei Fāngyán ausmachen, die mehrere Standardvarietäten aufweisen: Mandarin (mit den Zentren in Festland-China, Taiwan, Singapur und Hong Kong), Kantonesisch (in Hong Kong und Guangdong) und Hokkien (in Fujian, Taiwan und Singapur). Des Weiteren kann Mandarin nicht nur als plurizentrisch hinsichtlich der Unterschiede zwischen den Zentren angesehen werden, sondern auch die Schriften – Kurzzeichen (Festland-China) und Langzeichen (Taiwan) – und die Phonetisierungssysteme – Hànyǔ Pīnyīn (Festland-China) und Zhùyīn Fúhào und Wade-Giles (Taiwan) – können als plurizentrisch bezeichnet werden.