Mutual intelligibility of Chinese dialects : an experimental approach
Tang, C.

Mutual intelligibility of Chinese dialects

An experimental approach

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6406
e-mail: lot@let.uu.nl
http://www.lotschool.nl

Cover illustration:
Map of mainland China with the locations of the target dialects of this study indicated.

ISBN: 978-94-6093-001-0

NUR 632

MUTUAL INTELLIGIBILITY OF CHINESE DIALECTS

AN EXPERIMENTAL APPROACH

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 8 september 2008
klokke 13.15 uur

door

CHAOJU TANG

geboren te Chongqing, China
in 1968

Promotiecommissie

Promotor: Prof. dr. Vincent J. van Heuven

Overige leden: Prof. dr. Willem F.H. Adelaar
Dr. Yiya Chen
Dr. Charlotte S. Gooskens-Christiansen (Rijksuniversiteit Groningen)
Prof. dr. ir. John Nerbonne (Rijksuniversiteit Groningen)

Contents

Acknowledgments

My first word of thanks goes to Jos Pacilly, engineer and technician in the LUCL Phonetics Laboratory. It is his extraordinary patience and tolerance that led me overcome my fear of machinery. His motto 'solve problems one by one' helped me a lot.

I thank Dr. Wilbert Heeringa at the Meertens Institute, Amsterdam and Drs Peter Kleiweg in Groningen University. Without their help, it would not have been possible to compute the various Levenshtein distances with the LO4 software.

A special word of thanks is due to professor Liang Jie, who not only encouraged me to strive for a PhD position in my first year in Leiden, but also offered me practical help with literature, and introduced me to professional experts on my research topic in China.

Next, I would like to thank my teachers and fellow students in the phonetics laboratory of LUCL. Whenever I met with a problem, I could always turn to them and ask for help. I thoroughly enjoyed the conversations with Vincent, Maarten, Jos, Gijs, Elisabeth, Ellen, Jurgen, Jurriaan, Rob and Yiya in the phonetics 'bibliotheek' during coffee and tea breaks; they gave me a true introduction to Dutch culture. Towards the end of my stay in Leiden, I received much encouragement and help from Ezzeldin, Willemijn and Franziska.

I also wish to acknowledge the help from the librarians at Leiden University, both in the Sinology institute and in the main library. I am most grateful for having been given access to the scanning equipment in order to convert numerous pages of *the Atlas of Chinese Languages* to PDF. Also, the 'Stack Permission' card issued by Hanno Lecher, the head librarian of the Sinology library was of invaluable help.

I feel very much indebted to the CSC (China Scholarship Council), which organization gave me financial support for one year of tuition at the M. Phil level and for 48 months of subsistence. The same gratitude goes to my home university (Chongqing Jiaotong University), which also assisted me financially (and spiritually). I will never forget the encouragement and practical help from the administration, management and my dear colleagues there. They never refused me their help when I needed it. I also benefited enormously from subsidies granted to me by the Leiden University Fund, which allowed me to attend international conferences in Spain and Germany.

I thank the experts in the Department of Linguistics of the Chinese Academy of Social Sciences (CASS) for making recordings and digital databases available to me. I am also very much indebted to my areal contact persons in China, and my experimental subjects who acted as dialect speakers and listeners. Similarly, I thank my fellow students and researchers in LUCL who served as my subjects in the sound-quality judgment test.

I am greatly indebted to my family members. I owe so much to my only son Jinhong, who has always been my spiritual support. The love between us never goes away, wherever I am or whatever difficulties I meet.

Last but not least, my gratitude goes to my beloved father, who is so devoted to his family. Without his care, I could never have stayed in the Netherlands. He sacrificed his physical health in order to support me in my attempts to gain the doctoral degree. I am heart-broken now that he is suffering from disease and can only hope that my doctor's degree is the cure he needs.

Chapter One

Introduction

1.1 Questions

When we do research on language variety, very often we encounter questions such as these: (1) How should we distinguish a 'dialect' from a 'language'? (2) How much do two language varieties resemble one another, or how different are they? The answers to these two questions are concerned with the same problem: measuring the linguistic distance between language varieties.

1.1.1 Dialect versus language

It is not easy to distinguish 'dialect' from 'language'. The concepts of dialect and language involves non-linguistic as well as linguistic factors. Some speech varieties are very similar to each other but they are defined as different languages (e.g., German versus Dutch), while some speech varieties are quite different but are defined as dialects of the same language (e.g., Mandarin versus Cantonese).

A linguistic view defines a dialect as a speech variety or subdivision of a language which is characteristic of a particular group of speech speakers who are set off from others. This variety is distinguished from other varieties of the same language by features of the phonology (phonetics and pronunciation), grammar, and usage of vocabulary (cf. Oxford English Dictionary, online links: http://dictionary.oed.com/ and http://dictionary.reference.com/).

Based on this definition, the criterion for the dialect versus language distinction is determined by the (dis)similarities of structural features between two language varieties. The more two language varieties are structurally like each other, the more closely they are related to, or genealogically connected with, each other; that is, they are probably dialects of the same language. Otherwise, they are distant languages evolved from different proto-language families or phyla.

1.1.2 Resemblance versus difference

When we know that language varieties are dialects of some parent language, we further want to know how large their resemblance or difference is. This determines the affinity classification of dialects. If two language varieties are more alike each other, they should be closely grouped together to form a sub-division of a language phylum. Otherwise, they will be classified at different hierarchical levels of the language cladistic structure when we interpret the affinity relationship between language varieties into a tree structure.

1.1.3 Complex versus simplex

Determining the resemblance or difference between language varieties is a matter of measuring linguistic distance. There are various means to measure the linguistic distance between language varieties. Language varieties differ from each other not in just one dimension but in a great many respects: in their lexicon, in phonetics, in phonology, in morphology, in syntax, and so on. And at each of these linguistic levels, the ways in which language varieties may vary are further subdivided along many different parameters. Phonologically, they may differ in their sound inventories, in the details of the sounds in the inventory, as well as in their stress, tone and intonation systems. In order to express the distance between two language varieties, one would have to come up with a weighted average of the component distances along each of the dimensions identified (and probably many more). So, measuring linguistic distance is a multidimensional problem and we have no *a priori* way of weighing the dimensions.

Ideally, however, we would want to express the linguistic distance between language varieties in a single number on a one-dimensional scale rather than as a distance between points in some multi-dimensional hyperspace.

1.1.4 Intelligibility versus Mutual Intelligibility

A way-out would be to use intelligibility as a criterion for weighing the structural dimensions. Intelligibility can be interpreted as 'voice communication', or as 'the capability of being understood – the quality of language that makes it comprehensible.' The measuring index for intelligibility refers to the degree of accuracy to which speech can be understood. With specific reference to the speech communication system, intelligibility denotes the extent to which language listeners can identify words or phrases that are produced by speakers and transmitted to listeners via the communication system (cf. http://en.wikipedia.org/wiki/Intelligibility_(communication))

Intelligibility testing is a helpful approach, proposed by linguists, to integrate various linguistic distance measures. Intelligibility can be tested at several levels of the linguistic hierarchy, e.g. at the level of meaningless units (sounds or phonemes), at the level of meaningful units such as morphemes and words, or at the level of continuous sequences of sentences and spoken texts. Typically, intelligibility tests are composed of a test battery that addresses sounds, words and sentences separately. When we want to apply speech intelligibility tests to the problem of establishing the success of communication between speaker and hearer of related language varieties, we are not so much interested in the success with which listeners identify individual sounds. Rather, we are interested in the percentage of words that they get right. Therefore, word recognition is the key to speech understanding. The implication is that the measure of intelligibility is the percentage of correctly recognized words. The degree of intelligibility is best viewed as a scalar variable that expresses how well listener A understands speaker B, for instance on a scale from 0 (no understanding at all) to 100 (perfect comprehension). Therefore, intelligibility testing measures how well a listener of variety B understands or comprehends a speaker of variety A. The testing result can be expressed as a single number. For example, if listener B does not understand speaker A at all, the number should be zero. If the listener B gets every detail of speaker A's intentions (completely prefect comprehension), the score should be maximal. A convenient range between minimum and maximum understanding (or 'comprehension') could be between the percentage of 0 and 100.

American structuralists Voegelin & Harris took the initiative to test intelligibility in order to distinguish between language and dialect. Voegelin & Harris (1951) developed two techniques to assess the dialect intelligibility. One approach was called '*asking* the informants' about perceived dialect (dis)similarity, the other was called '*testing* the informants' comprehension' of the dialects in question based on the proportion of correctly translated words in the dialects at issue. Hickerson, Turner & Hickerson (1952) applied 'the testing-the-informants'-comprehension' approach in order to determine the relationship between seven Iroquois dialects.[1] A similar study of intelligibility testing was done by Bruce Biggs (1957) for Yuman languages.[2]

Linguists realized that the intelligibility between dialects is not necessarily reciprocal. The intelligibility between two language varieties is asymmetrical rather than symmetrical (or 'reciprocal') when the percentage of correctly recognized linguistic units by the listeners of language variety B is not equal to that by the listeners of language variety A. Typically, when language A makes a distinction between categories that is neutralized in language B, speakers of A are more difficult to understand for listeners of B than *vice versa*.

It is always the case that the intelligibility for language testing involves two-way communication. The non-reciprocal intelligibility between two California Indian languages – Achumawi and Atsugewi – was reported early on. Achumawi and Atsugewi are genealogically related languages of the Shanstan branch of Hokan. Achumawi was better understood by Atsugewi speakers than the other way around (Merriam 1926, Voegelin 1946). Olmsted (1954) definitively ascertained the asymmetry between these two California Indian languages. Some improvements were suggested on the intelligibility testing approach, addressing especially the problem of 'non-reciprocal intelligibility' between language varieties. As a case in point, Pierce (1952, 1954) adapted the Hickerson-Turner method by calculating the arithmetic mean of the two single intelligibility scores, i.e. the intelligibility from speaker A to listener B and *vice versa*. The

---

[1] Iroquois dialects belong to the family of North American Indian languages spoken by the Iroquois (the race of people living in America when Europeans arrived).

[2] Yuman languages are a group of languages of the Hokan family in Arizona, California and Mexico.

scores were collected from speakers of a set of Algonquian languages.[3] In later developments, intelligibility testing involved more refined materials, and devised methods and accurate computations (Wolff, 1964). In the 1960s, a team of researchers from the Summer Institute of Linguistics (SIL) did groundbreaking work on intelligibility testing of dialects in Mexico, on, for example, Mixe (Crawford 1967), Mixtec (Bradley 1967), Tzotzil (Stoltzfus 1967), Choapan (Casad 1969), and Mazatec (Kirk 1970). All of these dialectal studies are examples of further applications and modifications of techniques to be employed for intelligibility testing of multiple language varieties (Casad 1974, 1987). Later research confirmed the asymmetrical intelligibility between more pairs of (related) language varieties, also for Western languages. It has been shown that Portuguese listeners understand Spanish better than Spanish listeners understand Portuguese (Jensen 1989). Similarly, it is clear that Danes understand Swedes quite well but not vice versa (Delsing & Lundin-Åkesson 2005, Gooskens, Van Heuven & Van Bezooijen 2008).

To be more accurate, the notion of 'mutual intelligibility' is used to express the asymmetrical comprehension between language varieties. Mutual intelligibility is best defined as the *average (mean)* of the intelligibility of speakers of language variety A for listeners of language variety B and vice versa (Pierce 1952, 1954). In other words, mutual intelligibility is actually the (gradient) ease/difficulty of two-way communication between speakers/hearers of different language varieties. When speakers of language (variety) A can naturally readily understand speakers of language (variety) B and *vice versa* without prior exposure, intentional study or extraordinary effort, we say these language varieties are mutually intelligible and there exists some degree of mutual intelligibility between these two languages: A and B.

By definition, mutual intelligibility is an overall criterion that may tell us in a psychologically relevant way whether two languages are similar/close to each other. Theoretically, by comparing a large number of languages differing along many dimensions we may establish the relative importance of the various dimensions using mutual intelligibility as the overall criterion variable. When two language varieties are mutually intelligible, beyond some threshold level, the varieties should not be considered distinct languages, they are probably dialects of the same language. Conversely, for varieties to belong to different languages they should not be very mutually intelligible. This, then, would provide us with a solid, experimentally grounded, foundation for traditional claims about genealogical relatedness among language varieties as proposed by linguists.

Mutual intelligibility (instead of intelligibility alone) is, therefore, used as a reasonable criterion to measure the (dis)similarities between two language varieties. If the mutual intelligibility between two language varieties is sufficiently high, these two varieties are supposed to be regarded as the dialects from the same parent language, otherwise, they belong to different languages. Contrary to inherently multi-dimensional structural distance measures, mutual intelligibility is a single criterion.

[3] Algonqian languages are languages belong to a subfamily of native American languages that includes most of the languages in the Algic language family.

1.2 (Mutual) intelligibility tested experimentally

The research on testing intelligibility of dialects (from non-reciprocal to mutual intelligibility) has received considerable attention for a long time. Taken the cue of American structuralists' techniques, (mutual) intelligibility can be experimentally tested through functional and judgement approaches. A functional approach is the '*testing* the informants' technique; the opinion/judgment approach is the '*asking* the informants' technique as identified by Voegelin & Harris (1951)

1.2.1 Functional testing method

The '*testing* the informants' technique measures to what extent a listener actually recognizes linguistic units (words) in spoken stimuli. This functional intelligibility testing approach tests the (mutual)comprehension of the dialects in question based on the proportion of correctly translation of words in the dialects at issue: how well does listener A ***actually*** understand speaker B (and vice versa). The typical metric is to count the average percentage of correctly recognized or translated words from language variety A to language variety B (and vice versa).

In word recognition tasks, which are often part of functional intelligibility tests, words that were successfully recognized in an earlier part of the test will linger in the listener's mind and will be recognized with little effort the next time they occur. This so-called 'repetition priming' results in ceiling effects. In order to avoid priming effects, word recognition experiments take the precaution to block the different versions of stimulus words over different listeners such that a listener hears only one version of each stimulus word.

1.2.2 Opinion testing method

The '*asking* the informants' technique solicits judgments or opinions about perceived dialect distance or (dis)similarity. This testing approach is an alternative to functional testing methods. In opinion testing, listeners are asked how well they *think* they would understand a speech sample presented to them. The same sample can be presented to the same listener in several different versions, for instance, synthesized by several competing brands of reading machines and by a human control speaker (Pisoni et al. 1979). The listener is familiarized with the contents of the speech sample before it is presented so that recognition does not play a role in the process. All the listener has to do is to imagine that s/he has not heard the sample before and to estimate how much of its contents s/he thinks s/he would grasp. The response is an intelligibility judgment, expressed as a position on an intelligibility scale between a minimum and a maximum score, for instance 0 for 'I think I would not get a single word of what this speaker says' to 10 for 'I would understand this speaker perfectly, I would not miss a single word.'

1.2.3 The application of functional testing and judgment/opinion testing

Outside the area of linguistic fieldwork, intelligibility testing has been a topic of considerable importance in the areas of audiology, speech technology and in foreign language testing. In the literature on quality assessment of speech synthesis a division is often made between functional intelligibility testing and opinion testing. In the field of audiology, intelligibility tests were developed that measure intelligibility as function of the patient's hearing loss at the level of individual sounds, of words and of sentences (see, for instance, Kalikow, Stevens & Elliott 1977). More recently, similar techniques were adopted and extended in order to test the intelligibility of, and diagnose problems with, talking computers (see, for example, Van Bezooijen & Van Heuven 1997 and references therein). The same techniques were also fruitfully applied to the intelligibility testing of foreign-accented speech (e.g. Wang & Van Heuven 2007, Wang 2007 and references therein).

Although the methods for intelligibility testing have been well established, efforts spent on establishing testing mutual intelligibility among languages and language varieties have been disappointingly poor.

As mentioned above, early attempts at functional testing were made by American structuralists around 1950, trying to establish mutual intelligibility among related Amerindian languages based on listeners' comprehension of the material tested (Voegelin & Harris 1951, Hickerson, Turner & Hickerson 1951, Pierce 1952). The method was generalized and is still often used in the context of literacy programs, where a single orthography has to be developed that serves multiple closely related language varieties (Casad 1974, Brye & Brye 2002, Anderson 2005). The method works as long as the number of language varieties targeted is small. For instance, Van Bezooijen & Van den Berg (1999) studied the intelligibility of four Dutch and one Frisian varieties to Standard Dutch listeners; Gooskens (2007) determined mutual intelligibility among three West-Germanic languages (Frisian, Dutch, Afrikaans). In these methods listeners either summarize, or answer questions about, the contents of a speech sample they just heard.

A major problem with this method is that it is very difficult, if not impossible, to come up with speech samples and questions of equal difficulty in each of a set of language varieties, so that reproducibility of the results is compromised. Some attempts were made to determine mutual intelligibility for even small sets of related languages but came up with unsatisfactory results, mainly due to the fact that unsuitable materials or tasks were employed. As a case in point, one study (Delsing & Lundin-Åkesson 2005) tried to determine mutual intelligibility among Scandinavian languages Danish, Norwegian and Swedish using a *comprehension* test with just five open questions. As a consequence, these attempts were compromised by practical problems and by infelicitous choice of tasks and materials.

The practical problems are prohibitive when mutual intelligibility has to be established for, say, all pairs of varieties in a set of 15 dialects (yielding 225 pairs of language varieties). An alternative solution to this problem is to use judgment or opinion testing, which simply ask listeners how much the speech in language B differs from their own language A. This is called 'the perception of degrees of difference between a local variety and surrounding varieties' by Preston (1987: 4). Subjects listen to a recorded speech sample of a variety B and are asked to judge how different the variety is from their own variety A on some continuous rating scale. The assumption is that listeners are able to judge the (dis)similarity of the sample dialect to their own dialect based on the intelligibility testing. This is actually the measure of 'perceived linguistic distance' or 'estimated linguistic distance'.[4] The first study using this methodology was done, in the Netherlands, by Van Hout & Münstermann (1981), who asked listeners to rate the distance between recorded samples of nine different regional varieties of Dutch from the standard language on a 7-point scale. More recently, the same approach was used by Gooskens & Heeringa (2004), who played speech samples in 15 Norwegian dialects to groups of listeners from the same 15 dialect areas and asked the listeners to judge how much the samples differed from their own dialect. Listeners appear to have reliable (i.e. reproducible) ideas about how much language B differs from their own, even if they know the stimulus language from past exposure, and even if the recording quality of the speech samples may differ substantially.

1.3 Statement of the problem

1.3.1 The choice between functional and opinion testing

Functional testing and opinion testing have their own respective advantages and disadvantages. The earlier applications of functional and opinion testing leave us some room to do the mutual intelligibility measuring for related language varieties on several aspects. Firstly, functional testing has only been applied to small sets of related language varieties. No-one has yet attempted a large-scale comparison of 15 language varieties (yielding 225 pairs). Secondly, we have insufficient ground to decide which mutual intelligibility testing approach (functional approach or opinion approach) is a better choice. No reports exist about the correlation between functional and opinion tests. We need to (i) correlate the functional tests with the opinion tests; (ii) correlate both mutual intelligibility testing (functional and opinion methods) with objective structural measures; (iii) validate the correlations with traditional dialect taxonomy. Solid evidence (such as better correspondence with the traditional language/dialect taxonomy) is still needed to determine whether opinion tests are really a shortcut or an ideal substitute for functional tests.

Earlier work on predicting mutual intelligibility between language varieties from the structural measures can be found in Pierce (1954) on Crow and Hidatsa languages, which are two linguistically closely related varieties of the Crow-Hidatsa language family, belonging to the Siouan stock, e.g. testing the degree of overlap between mutual intelligibility and glottochronological estimates of linguistic distance. [5] Biggs (Casad

[4] Alternatively, subjects are asked to rate the distance between A and B without auditory samples but relying purely on preconceived ideas triggered by geographic names. (Gooskens 2009)

[5] Crow is a Missouri Valley Siouan language variety spoken primarily by the Crow Nation in present-day south-eastern Montana. It has one of the largest populations of American Indian languages with 4,280 speakers according to the 1990 US Census; Hidatsa is a language variety

1974, 1987) also studied the relationship between mutual intelligibility and the number of shared cognates. More recently, work was done by Gooskens & Heeringa (2004) on 15 Norwegian dialects correlating perceived linguistic distance and computed Levenshtein distance.[6] The work on correlating the results of functional intelligibility tests with structural distance measures was also done by colleagues in Groningen (Gooskens 2007, Beijering, Gooskens & Heeringa 2008)

This dissertation aims to (i) establish the mutual intelligibility between 15 Sinitic speech varieties (yielding 225 pairs of varieties to be compared) by running experiments both via functional and opinion methods; (ii) correlate functional methods with opinion methods to see to what extent the latter can be used as a substitute of the former; (iii) use more structural measures (e.g., lexical similarity, phonological correspondence, segment inventories and lexical frequencies of the vowels and consonants in the inventories, and Levenshtein distance) as predictors to validate the mutual intelligibility tests; (iv) determine through multiple regression techniques which structural measures afford better prediction of (mutual) intelligibility; (v) cross-validate mutual intelligibility testing methods by comparing the test results with traditional language taxonomy.

1.3.2 Problems in this research

1.3.2.1 The classification issue of Sinitic varieties

There is a basic agreement that Sinitic varieties have a primary split between the Mandarin and the non-Mandarin (or Southern) branches, whose dichotomy is essentially based on the phonological characteristics and tone evolution from Middle Chinese (for more details, see Chapter Two).

In a broad sense, language varieties in the Sinitic stock are often called Han Chinese, which is a sub-phylum of Sino-Tibetan.[7] This sub-phylum is one of the few language stocks, outside the Indo-European phylum that has a long tradition of linguistic

---

spoken by the Hidatsa tribe of the Dakotas. Crow and Hidatsa are closely related to each other. The ancestor of Crow-Hidatsa may have constituted the initial split from Proto-Siouan. The Crow and Hidatsa language varieties are classified as a subfamily in the Siouan language family. Crow and Hidatsa are not mutually intelligible, however the two languages share many phonological features, cognates and have similar morphologies and syntax. (cf. http://en.Wikipedia.org/wiki/Crow_language).

[6] Levenshtein distance, also called string edit distance, is named after the Russian scientist Vladimir Levenshtein, who devised the algorithm in 1965. It is a metric for measuring the amount of difference between two sequences (a string distance measure) that is based on the minimum number of string operations (insertion, deletion, substitution) needed to transform one string into the other. It is often used in applications that need to determine how similar, or different, two strings are, such as converting the phonetic transcription of a word in language A to its counterpart in language B (or vice versa). (for more details, I refer to Gooskens & Heeringa 2004; also the websites: http://en.wikipedia.org/wiki/Levenshtein_distance; http://www.merriampark.com/ld.htm).

[7] Han Chinese, (also Hanyu in Pinyin), means the native languages spoken by Han people (the majority people among the 56 peoples in China).

scholarship of its own. Varieties in this sub-phylum are traditionally split into Mandarin and Southern branches. Each branch comprises several different families respectively (details are in Chapter Two). However, the affinity between these varieties (i.e. how close or distant these varieties are) has been elusive. The classification of Sinitic language varieties is still controversial and has not been settled, i.e., the question whether individual varieties should be classified as either the primary division of Mandarin or non-Mandarin (Southern) is an issue of debate. Also, the internal structure within the main branches is debated a lot. A case in point is the grouping of Jin varieties (having Taiyuan as their representative). Traditionally, Jin varieties are classified into the Mandarin branch (see the linguistic map from the website: http://www.chinadata.ru/linguistic_group_map.htm). However, some linguists have recently branched Jin varieties off from the Mandarin split, arguing that Jin varieties have kept the Ru tone, which is one of the typical characteristics of non-Mandarin(Southern) varieties (see the *Language Atlas of China*, Wurm, T'sou, Bradley, Li, Xiong, Zhang, Fu, Wang & Dob 1987). This dissertation will decide the position of the Taiyuan variety (representing the Jin varieties) through validating the results from *mutual* intelligibility testing to the traditional dialect taxonomy.

1.3.2.2 Asymmetrical mutual intelligibility between Mandarin and non-Mandarin varieties

The mutual intelligibility between these Sinitic varieties maintains debated as well. The impressionistic claims are: (i) Mutual intelligibility between the Mandarin branch and the Southern branch is rather poor; (ii) Mandarin varieties are more intelligible to Southern varieties than vice versa; (iii) Language varieties within the Mandarin branch are more intelligible to each other than that within the Southern branch. (Duanmu 2000:2, Yan 2006:2)

This dissertation will pinpoint the issues mentioned above and try to validate the traditional split of the Mandarin and Southern branches by establishing the methods of mutual intelligibility testing. Further efforts will be made to test the impressionistic claims concerning the asymmetry of intelligibility between the Mandarin and Southern varieties and finally offer a solution to the debated Jin varieties via testing the mutual intelligibility between Taiyuan and other varieties based on experimental data.

1.3.3 Predicting mutual intelligibility from structural distance measures

As I expressed in § 1.1.3, language varieties may differ in various structural dimensions. Structural distance is by nature a symmetrical notion. That is to say, the distance from language variety A to language variety B is exactly the same as the distance from language variety B to language variety A (just as the distance from city A to city B is identical to that from city B to city A). Indeed, many popular linguistic distance measures reflect this property of symmetry. An example is the measure of lexical affinity between two language varieties. Lexical affinity is commonly defined as the proportion of cognate words shared between two related language varieties A and B. In order to compute this proportion, we first count the number of lexical items in the union of the vocabularies of A and B. We then divide this number into the number of words that are cognates in A and B. Obviously, the number of cognates is the same between A and B as in B and A, so that the lexical distance between A and B and between B and A is identical. A similar principle applies to the highly popular string edit distance measures (also called 'Levenshtein distances') between language varieties.

We argue that mutual intelligibility can be predicted from the various structural measures to some extent. Once we establish the mutual intelligibility between language varieties, we can correlate it with various structural distance measures through multiple regressions in order to find out how much of the mutual intelligibility can be predicted from the structural distance measures.

1.3.3.1 Structural measures for European language varieties

With the development of measurement methodologies in linguistics, measures on linguistic differences/similarities between languages were proposed. Various structural measures on European speech varieties (mostly non-tonal languages) originated in the 1930s. For example, a correlation method was used for language classification for Indo–European (Kroeber & Chretien 1937, 1939) and Middle English (Ogura 1990). Glotto-chronological methods were applied to American English in the 1950s (Swadesh 1950, Reed & Spicer 1952). Other distance measure methods for language classification were proposed by Hsieh (1973), Krishnamurti, Moses & Danforth (1983), and by Cavalli-Sforza & Wang (1986).

Further work on structural measures of difference between non-tonal languages has been done, for instance, at Stanford University (for Gaelic Irish dialects, Kessler 1995), and at the University of Groningen for Dutch (Nerbonne et al. 1996) and Sardinian (Bolognesi & Heeringa 2002) dialects. Recently, such methods for measuring structural difference were applied to tonal languages as well. The first attempt was done on Norwegian dialects, with a binary tone contrast at the word level, using the Levenshtein distance algorithm based on phonetic transcriptions, where all transcription segments for each word against its cognate were aligned for algorithmic comparison (Gooskens & Heeringa 2004). In the computation of phonetic distance between word pairs, the tone symbol was counted as if it was just another phoneme. The results of this objective measurement were then used to build a tree structure (through hierarchical cluster analysis via average linkage method) and the tree is used to validate the language family/affinity tree as constructed by linguists (Gooskens & Heeringa 2004).[8]

[8] The cluster analysis first establishes a group by finding the pair of dialects having the minimum distance. Then the next minimally distant pair is found, then the average distance between the two pairs is calculated and will be linked with next minimally distant pair and so on and so forth. Fortunately, we do not have to do this work by hand; computer software such as SPSS (Statistical Package for the Social Sciences) is able to do that for us automatically.

1.3.3.2 Structural measures on Chinese language varieties

Since the 1960s, the measurement methodology such as the lexicostatistical method began to be applied to determining linguistic relationships between Chinese dialects (Wang 1960). Extensive investigations of affinity among Chinese dialects were carried out between 1970 and 1990, aided by the development of computer technology (Cheng 1973, 1982, 1986, 1987, 1988, 1991, 1993, 1997; Wang 1987).

Instead of using the Levenshtein distance algorithm, Chin-Chuan Cheng (henceforth Cheng) computed structural distances between pairs of Chinese dialects along many different dimensions.[9] Since the 1970s, Cheng aimed at measuring dialectal differences in terms of tone height with respect to the Yin and Yang split in the tone systems between pairs of 17 Chinese dialects (Cheng 1973, 1991).[10] In the late 1970s till 1990s, Cheng did work on calculating the lexical correlation based on the *Hanyu Fangyan Cihui [Chinese dialect word list]* (Beijing University, 1962, 1964) converted to a computer database with 6,454 cognate variants for 905 words shared by 18 Chinese dialects (Cheng 1982, 1991,1993, 1997).[11] Employing the computer-based data file of *Hanyu Fangyan Zihui [Chinese dialect character pronunciation list] (*Beijing University, 1962, 1964*)*, Cheng also did measures on the genealogical relationship among 17 Chinese dialects correlating their phonological correspondence (the complexity of the rule system needed to convert phonological forms in one dialect to their cognates in the other dialect) of Modern-MC (Middle Chinese) reflexes in terms of initials (syllable onsets), finals (syllable rhymes) and tones and their combinations cross the 2,700 words (Cheng 1991, 1993, 1997).[12]

It is commonly held that Chinese, as an isolating language, has little or no grammar in terms of inflections of person, case, number, tense, voice and the like.

> 'When any of the Chinese dialects, including Mandarin, is compared to nearly any other language, one of the most obvious features to emerge is the relative simplicity of the words of Chinese … It is clear that Mandarin is quite striking in its general lack of complexity in word formation.' (Li & Thompson 1981: 10)

In this sense, most structural research on Chinese focuses on lexical entries and phonological (including tonal) features. That is, the genealogical relations among language varieties are usually determined by phonological correspondences and the

---

[9] Chin-Chuan Cheng, is an Academician and a linguist in the Institute of Linguistics at the Academia Sinica (Taipei, Taiwan)

[10] The 17 dialects on which tonal difference based are: Beijing, Jinan, Xi'an, Taiyuan, Hankou, Chengdu, Yangzhou, Suzhou, Wenzhou, Changsha, Shuangfeng, Nanchang, Meixian, Guangzhou, Xiamen, Chaozhou and Fuzhou.

[11] The 18 dialects are Beijing, Jinan, Shengyang, Xi'an, Chengdu, Kunming, Hefei, Yangzhou, Suzhou, Wenzhou, Changsha, Nanchang, Meixian, Guangzhou, Yangjiang, Xiamen, Chaozhou and Fuzhou. This is not the super-set of the previous 17 dialects.

[12] *Hanyu fangyan cihui*, see § 5.2.1; *Hanyu fangyan zihui*, see Note 63. This set of 17 dialects is not a subset of the 18 dialects for lexical correlations but they share many common dialects.

incidence of lexical cognates. The relative importance of these linguistic entities is still at issue.

1.3.3.3 Predicting mutual intelligibility of Sinitic varieties

Although methods of structural measures of linguistic similarity and difference between Sinitic varieties are as well established as those for European language varieties, less work on mutual intelligibility testing has been done on Sinitic varieties.

Mutual intelligibility tests (e.g. through functional testing and judgment testing) were already applied to many language varieties (e.g Amerindian, Dutch, Norwegian, and African language varieties). However, little such work is done about how to establish mutual intelligibility among Sinitic varieties experimentally, as Cheng (1992) stated as follows:

> In this paper, however, I have proposed a different measurement that takes into consideration the weights of signal and noise in inter-dialectal communication. The calculated intelligibility is called systemic intelligibility since it is based on dialects as linguistic systems and not on speakers' experience. It is hoped that systemic intelligibility will provide a basis for exploring the questions how individuals as language users understand the speech of other dialects. But questions such as those concerning how 'participant intelligibility' is to be calculated are yet to be answered. (Cheng 1992: 167)

One question is whether we can predict the mutual intelligibility between Sinitic language varieties from various structural distances and, if so, to what extent.

Practically, once the distance measures on the linguistic structures and the mutual intelligibility scores from the experiments are available, their correlation coefficients can be obtained. Similar work has recently been done by colleagues in Groningen University. Gooskens & Heeringa (2004) obtained linguistic distance judgments for 15 Norwegian speech samples based on melodic and monotonized readings of the fable *The North Wind and the Sun*. They then correlated the judgment scores with objective Levenshtein distance scores. The results showed that subjectively judged similarity/ distance between sample dialects and the listener's own dialect correlated substantially with the objective Levenshtein distance (r = .62 without melody and r = .67 with melody, $p<0.001$ (excluding distance judgments by listeners on their own dialects). Gooskens (2007) correlated lexical and phonetic distances with mutual intelligibility scores for three Mainland Scandinavian Standard languages (Danish, Norwegian and Swedish). The results showed a high correlation between intelligibility scores and phonetic distances ($r = -.80$, $p < .01$) but not significantly high with lexical distance($r = -.42$, $p = 0.11$). Beijering, Gooskens & Heeringa (2008) collected mutual intelligibility scores for 18 Scandinavian language varieties assessed by young Danes from Copenhagen. They then correlated these judgment scores with the linguistic distances between Standard Danish and each of the 18 varieties at the lexical level and at several phonetic levels. The results showed that both correlations are significant at the .01 level, but the correlation with phonetic distances is almost significantly higher than with lexical distances ($r = -.86$ versus $r = -.64$, $p = .08$). In particular, consonant substitutions, vowel insertions and vowel shortenings contribute significantly to the successful prediction of intelligibility.

In this manner, subjective intelligibility judgments were used to validate an objective linguistic distance measure, i.e. the Levenshtein distance. Tang & Van Heuven applied this judgment testing method to Chinese dialects and claimed the relative importance of structural dimensions can then be found through some form of statistical optimization (multiple regression techniques). Furthermore, we can decide which mutual intelligibility testing approach can be better predicted from structural measures when we validate the testing results with the traditional language taxonomy proposed by linguists (Tang & Van Heuven 2007, 2008, 2009).

This dissertation is a first try on tackling Cheng's question about how to establish the mutual intelligibility based on participants of Chinese dialects, by running both opinion-judgment experiments and functional experiments. The test results will be compared with Cheng's objective structural measures, using the latter as predictors of experimentally established mutual intelligibility between Sinitic language varieties. I will also compute other objective distance measures, such as Levenshtein distance measures based on the 764 Chinese words in the database compiled by linguists at the Institute of Linguistics of the Chinese Academy of Social Sciences (CASS), and see how well the mutual intelligibility between Sinitic language varieties correlates with various structural distance measures. Finally, I will relate all the measures, both objective counts on corpora and subjective data obtained with human subjects, with traditional dialect taxonomies proposed by Chinese linguists to see how well the mutual intelligibility between Sinitic language varieties can be predicted from the structural measures.

1.4 Determining the power of functional testing against opinion testing

The work done by Gooskens & Heeringa represents a complication relative to earlier work (for example, on Gaelic and Dutch varieties) in that their Norwegian dialects are tone languages whilst the Gaelic Irish and Dutch dialects are not. Since it is unclear how tonal differences should be weighed in this distance measure, Gooskens & Heeringa (2004) collected distance judgments for the same reading passages resynthesized with and without pitch variations.

They recorded 15 Norwegian speech samples from 15 different dialect speakers who read the same text, i.e. the fable *The North Wind and the Sun*, in their own dialects. They found 15 groups of listeners, one group from each of the locations where the 15 dialects are spoken. These subjects listened to the recordings and judged each dialect on a scale from 1 (similar to own dialect) to 10 (distant from own dialect) according to their own subjective opinions. Because dialect A is not necessarily as intelligible to the listener of dialect B as in the reverse case, two asymmetrical scores reflecting the dialect (dis)similarity/distance were obtained for each pair of the dialects. One is the mean of the judgment scores from listeners of dialect A to dialect B, the other is that from the listeners of dialect B to dialect A (Gooskens & Heeringa 2004). They then correlated the mean value of the two asymmetrical scores from both the full matrix, and from the matrix with only the off-diagonal scores, with the Levenshtein distance (Levenshtein distance is perfectly symmetrical because the distance from the string X to string Y is exactly the same as the distance from Y to X) based on the (both cognate and non-cognate) word pairs in the fable.

The difference in judged distance between the pairs of versions (with and without pitch) would then be an estimate of the weight of the tonal information. Norwegian, however, is a language with just a binary tone contrast. I will extend the research to a set of fully-fledged tone languages, viz. Chinese, a language (family) with much richer tone inventories varying from four (Mandarin) to as many as nine (Cantonese). Taking a cue from Gooskens & Heeringa's work, I want to apply their methodology and predict the mutual intelligibility between Sinitic/Chinese language varieties not only through judgment/opinion tests but also through functional tests, using not merely Levenshtein distance measures but also various structural measures published by Cheng or collected by myself. I will correlate the two types of experimental results with one another to find out to what extent opinion testing may serve as a feasible alternative to functional intelligibility testing in the area of language variation studies.

I believe that Sinitic languages offer a promising testing ground for mutual intelligibility studies as the dimensionality of the comparison is somewhat reduced. Sinitic languages are characterized by the absence of morphology, and they differ relatively little in terms of their syntax. As a result, differences in mutual intelligibility are primarily related to lexicon and phonology (including tone). It is also a fortunate circumstance that Chinese linguists have established an impressive body of digital resources that can be used to study objective structural similarities and differences among the many dialects/ languages spoken in China.

1.5 Goal of this research

If a procedure could be developed by which mutual intelligibility between any two languages could be established, we would have a powerful instrument, a communicatively meaningful way of arguing about linguistic distance. One important aim of the dissertation is to address this issue. This dissertation will: (i) aim to determine the mutual intelligibility between Sinitic varieties and will also (ii) find out the prediction power of various structural distance measures on Sinitic varieties for the mutual intelligibility testing, (iii) ultimately offer the contributions to establishing a measure of affinity among the members of the Sinitic language varieties.

Following western methods, as a first try, I will compute the Levenshtein distance between the cognates shared by the pairs of the Sinitic languages. I will see to what extent the structural measures and mutual intelligibility testing results converge with the traditional Chinese classification/ taxonomy respectively. Then I will correlate all the objective distance measures (obtained from the literature and computed by ourselves) with the subjective measures to see how well they correlate with one another, how well we can predict the mutual intelligibility between pairs of Sinitic languages from the objective structural measures. Finally, I will validate results from all these objective and subjective measurement with the traditional language taxonomy postulated by Chinese linguists, to see to what extent these subjective and objective distance measures reflect the classification of Chinese languages.

1.6 Summary of research questions

Specifically, in this dissertation I will aim to find answers to the following questions:

i) What is the correlation between judged (mutual) intelligibility and judged similarity in pairs of 15 target Sinitic dialects?
ii) Do the opinion-test scores confirm *a priori* expectations/claims with respect to mutual intelligibility between pairs of Chinese dialects?
iii) To what extent are dendrograms (affinity trees) based on our judgment scores compatible with traditional Chinese dialect taxonomies?
iv) What is the correlation between word-intelligibility and sentence-intelligibility obtained through functional testing on pairs of our 15 target Sinitic dialects?
v) Do the results obtained from functional testing confirm *a priori* expectations/claims with respect to mutual intelligibility between pairs of Chinese dialects?
vi) To what extent are dendrograms (affinity trees) based on functional test scores compatible with traditional Chinese dialect taxonomies?
vii) To what extent are the experimental results in accordance with observations on the characteristics of Chinese dialects?
viii) What is the Levenshtein distance between all pairs of the 15 Chinese dialects based on the cognates in the CASS database?
ix) How can we optimally predict the subjective measures (obtained from both opinion scores and functional scores) from (some combination of) objective measures (whether collected from the literature or computed by ourselves)?
x) Which of the subjective test measures (opinion tests and functional tests) can be predicted better from objective measures?
xi) To what extent do the objective measures reflect the traditional dialect classifications?
xii) To what extent can methodologies developed on European languages/dialects be applied to Chinese tonal languages/dialects?
xiii) Can we extend existing methodologies so as to enable mutual intelligibility testing between languages with complex lexical tone systems?

1.7 Research design and plan

Following Gooskens & Heeringa's methodology, I will run experiments using judgment/opinion testing and augment these with functional tests to determine the *mutual* intelligibility of Chinese dialects. I will target 15 Chinese dialects (a subset of Cheng's 17 dialects). These dialects are Beijing, Chengdu, Jinan, Xián, Taiyuan, Hankou, (Mandarin dialects), Suzhou, Wenzhou (Wu dialects), Nanchang (Gan dialect), Meixian (Hakka dialect), Xiamen, Fuzhou, Chaozhou (Min dialects),[13] Changsha (Xiang dialect), Guangzhou (Yue dialect).[14] Only the dialects of Yangzhou and Shuangfeng are excluded from Cheng's dialect set. In the following sections I will briefly describe the experimental and lexico-statistical datasets that I collected in the course of the present study.

1.7.1 Judgment/opinion tests

The purpose of this experiment is two-fold. First, I aim to measure the judged distance between language variety X and Y, that is, how much does language variety X differ overall from language variety Y (by listeners' judgments on a rating scale). Second, we will test the *mutual* intelligibility between speech varieties X and Y as judged by the same listeners. Here we asked listeners of variety X how well they *think* they understand speakers of variety Y (and vice versa). For both tasks we used existing recordings of the fable *The North Wind and the Sun* spoken by a native speaker for each of 15 target Sinitic dialects. Chapter Three reports on this experiment in details.

1.7.2 Functional tests

This experiment tests how well listener A *actually* understands speaker B (and vice versa). In order to obtain experimental data, I designed two tests: one at the level of isolated words, the other at the sentence level. The test scores reflect the number of words correctly recognized (in the word-level test) or translated (in the sentence-level test).

In the word-intelligibility test target word recognition is tested through semantic multiple-choice categorization. Listeners indicated to which of ten pre-given semantic categories a spoken word belongs. For instance, if the listener heard the word for 'apple', s/he should categorize it as a member of the category 'fruit'. Here, the assumption is that correct categorization can only be achieved if the listener correctly recognized the target words.

Word recognition in sentence context was tested by a Chinese version of the SPIN ('Speech Perception in Noise') test, which was originally developed for English by Kalikow, Stevens & Elliott (1977). In the SPIN test the listener has to write down only the last word in a number of short spoken sentences. In the materials I used, the identity of the final word was largely predictable from the earlier words in the sentence, so that this test addresses the efficient interaction of bottom-up (information from the speech signal) and top-down (expectations derived from earlier context) processes in

---

[13] In more details, there are many clusters in Min subgroup, actually, Xiamen dialect is the representative of South Min, Fuzhou represents East Min, Chaozhou represents Chao-Shan group.

[14] In *the Language Atlas of China*, Taiyuan is separated from the Mandarin branch, and belongs to a new non-Mandarin branch: Jin group.

continuous speech recognition. Earlier work has shown that this type of test is highly sensitive to differences in intelligibility due to different language backgrounds of speakers and listeners (Wang 2007).

One additional question that we hope to answer on the basis of the present research, is to what extent the recognition of isolated words (bottom-up information only) and of words in context (interaction of bottom-up and top-down information) are predictable from each other. If recognition of words in context is largely predictable from isolated-word recognition scores, the latter type of test will suffice for future work on functional mutual intelligibility testing in the Chinese language area. Chapter Four is about this experiment.

1.7.3 Levenshtein distance measure

Levenshtein distance measures were computed on a set of 764 words commonly shared in our 15 target dialects extracted from the database established by linguists in the Institute of Linguistics of Chinese Academy of Social Science (CASS). For further details, I refer to Chapter Five.

1.7.4 Other distance measures

The other distance measures mainly concern lexical affinity and especially phonological affinity between all pairs of our 15 dialect sample. The computations are based on different sources: they were either copied from existing literature (Cheng 1997), or derived from published sound inventories of Sinitic languages (Yan 2006), from Campbell's website on Sinitic dialects, and from the CASS database. Chapter Five explains these measures in detail.

1.8 Outline of the dissertation

In the next chapter (Chapter Two) I will introduce the language situation in China and its historical development, in so far as relevant to the 15 dialects that constitute my sample of Sinitic varieties. Chapter Three will focus on the collection and analysis of subjective distance and intelligibility measures through judgment (opinion) tests. In Chapter Four I will describe the functional intelligibility tests I carried out, and analyse the results. In Chapter Five I will collect a large number of objective distance measures between the 15 dialects, and consider to what extent these reflect the traditional dialect taxonomy for Sinitic languages. Chapter Six is about the correlations between all kinds of subjective and objective measures. All the results will be validated against the traditional Chinese dialect taxonomy. Chapter Seven presents a summary of main findings and draws conclusions with respect to the questions that were raised above in this introduction.

Chapter Two

The Chinese language situation

2.1 Introduction

In this chapter, I will introduce the language situation in the People's Republic of China (henceforth China). The language situation in China is complex. For the sake of simplicity, I will only present a general survey of traditional language taxonomies of the Sinitic stock. These taxonomies have been proposed by Chinese linguists. I select two of these taxonomies as references for comparison with various results of linguistic distance measures (to be computed in Chapter Five). One taxonomy is published as a map called 'Chinese linguistic groups' which is available from the internet link http://www.chinadata.ru/linguistic_ group_map.htm. The other taxonomy has been proposed in the *Language Atlas of China* (Wurm, T'sou, Bradley, Li, Xiong, Zhang, Fu, Wang & Dob 1987). These two taxonomies are shown in Figures 2.1 and 2.2. § 2.1 summarizes the language situation of Chinese dialects and introduces the Sinitic stock. § 2.2 explains the traditional primary split between the Mandarin and non-Mandarin dialects according to their phonological characteristics and tone evolution. § 2.3 deals with the various groupings/classifications of Chinese dialects proposed in the literature sources. Measures of linguistic distance between Chinese dialects are described in § 2.4. §§ 2.5 and 2.6 deal with the mutual intelligibility between Chinese dialects and how to determine the degree of the mutual intelligibility between Chinese dialects.

2.2 Taxonomy of Chinese language varieties

The linguistic wealth of China is a rich diversity of language varieties spoken today. Research on these language varieties has produced significant results that have greatly expanded our knowledge of the origin, the evolution, and the diversity of Chinese languages and their dialects.

China consists of 56 ethnic groups and each ethnic group has its own unique language variety. Han people have the largest population in the country. The language Han people speak is called Han Chinese. The other 55 peoples are the smaller part of the population and are called ethnic minority peoples. These ethnic peoples speak their own languages and their languages are often indicated by the same name, for instance, the ethnic language of Zhuang people is called the Zhuang language, and the language spoken by inner Mongol people is called Mongolian, etc. Typologically, languages spoken within China can be classified into several phyla, i.e. the Sino-Tibetan phylum, the Austro-Tai phylum, the Austro-Asiatic phylum, the Altaic phylum, Indo-European phylum and the Austronesian (sub)phylum (Lee 1987, A-1).[15] My study in this dissertation just focuses on the Sinitic stock.

Han Chinese (or Hanyu, the native language of Han majority people, henceforth Chinese) forms the Sinitic stock, which is a sub-phylum of Sino-Tibetan phylum. This stock is one of the few language stocks, outside the language stocks of the Indo-European phylum that has a long tradition of linguistic scholarship of its own. According to the tradition, the Sinitic stock comprises seven (super)groups and some unclassified language varieties: the (super) groups of Mandarin, Min, Yue, Wu, Hakka, Gan, Xiang. Geographically, the Mandarin (super)group is found mainly in the northern part of China, while all the other non-Mandarin (super)groups are distributed in the southern part of China, mostly along the coastal line. Accordingly, they are often called Southern groups. Recently, a more detailed classification was proposed. For instance, Li (1987: A-1) separated a Jin group from the Mandarin supergroup (later, I will deal more extensively with the new non-Mandarin sub-group — Jin). Li also added two new groups — Hui and Pinghua — to the non-Mandarin groups of this Sinitic stock (cf. Li 1987, A-1). In most cases, I will use the terms 'non-Mandarin group(s)' and 'Southern groups' exchangeably.

In the Sinitic stock, there are approximately 1,500 recorded language varieties (Campbell, see http://www.glossika.com/en/dict/faq.php#1). For practical purposes, members in this stock can be treated as separate languages. However, through history, the speakers of these languages were united by common political, ethnic, and cultural ties. They share the same literary heritage which is actually the character-based orthographic system called 'hanzi' (Han Characters). Three common characteristics can be found in speech varieties within the Sinitic stock: (i) the common phenomenon to use monosyllabic meaningful units, (ii) a shared system of tones originally developed or evolved from earlier consonantal features, (iii) the tendency to devoice the earlier voiced initial consonants. In sum, they share the uniform Chinese character-based writing system and are based on largely the same grammar rules: e.g. the similarities or correspondences in word forms, in grammatical elements such as prefixes, suffixes, vocalic and consonantal alternations and in general literarily syntactic structure (although differences in word order are widespread, despite the colloquial usage of some dialectal grammar). It is customary to call the speech varieties of the Sinitic stock 'dialects of Chinese'.

However, the classification of these dialects is tentative and still controversial. According to the consensus of Chinese dialectologists, dialects of the Sinitic stock are primarily bifurcated into the super-groups of Mandarin and Southern (non-Mandarin). The (sub)groupings for some members of these dialects are consistently agreed upon by linguists, but for certain members and some internal structures (subgroupings and clusters), sources do not agree which dialects should be assigned to which of the two primary branches. For instance, even in the authoritative *Language Atlas of China*, there are discrepancies in the classification within the Sinitic stock. Lee (1987: A-1) agrees

[15] In *The Language Atlas of China* (Wurm, T'sou, Bradley, Li, Xiong, Zhang, Fu, Wang & Dob 1987), the Austronesian phylum bifurcates into phyla of Austro-Tai and Austronesian. This genealogy is still controversial.

there are seven (super)groups and two unclassified groups in this stock: the Mandarin Supergroup, the Min Supergroup, the Yue group, the Wu group, the Hakka group, the Jin group, the Xiang group and two as yet classified varieties: Tuhua and Xianghua. No sub-groups or clusters are further classified by Lee. Lee explicitly claimed that this classification is commonly accepted outside China but does not agree with those put forward by Chinese linguists (1987: A-1). Li (1987: A-1) proposed the following classification: the Mandarin Supergroup comprising several subgroups (the Northeastern group, the Beijing group, the Beifang (Jilu) group, the Jiaoliao group, the Zhongyuan group, the Lanyin group, the Southwestern group, the Jianghuai group, and not-yet-grouped-Mandarin), the Jin group, the Wu group, the Hui group, the Gan group, the Xiang group, the Min Supergroup, the Yue group, the Pinghua group, the Hakka group, and not-yet-grouped-non-Mandarin.[16]

In this dissertation, I will adopt two dialect taxonomies. One taxonomy of Chinese dialects is proposed by the *Language Atlas of China* (Wurm, T'sou, Bradley, Li, Xiong, Zhang, Fu, Wang & Dob 1987), the other is a simplified version that was published as 'Chinese linguistic groups' on the internet at http://www.chinadata.ru/linguistic_ group_map.htm (cf. Tang & Van Heuven 2007). In order to directly present the genealogical relationships among my selected 15 dialects, I graphically interpreted the two dialect taxonomies with tree structures (see Figures 2.1 and 2.2) and omit irrelevant dialects (which would lead to differences between the trees at the leaf-level of the trees only.[17]

[16] In the *Language Atlas of China* (Wurm et al. 1987), the subgroup name 'Beifang' is later called 'Jilu'. I will use the latter name in my dissertation. Hopefully, this is not confusing.

[17] In an earlier publication (Tang & van Heuven 2007), we adopted the sub-branches of the Mandarin branch from a map of 'Chinese Linguistic Groups' (http://www.chinadata.ru/ linguistic_ group_map.htm), in which there is no detailed sub-division of the secondary split within the Mandarin branch. Mandarin dialects are roughly classified into three sub-branches called Northern, Eastern and South-Western (see Figure 2.3 in this chapter). One of our target dialects, Jinan was provisionally classified into the Eastern sub-branch in Tang & van Heuven (2007), but it actually should go to the Northern Mandarin subgroup. (see Figures 2.1 and 2.2).

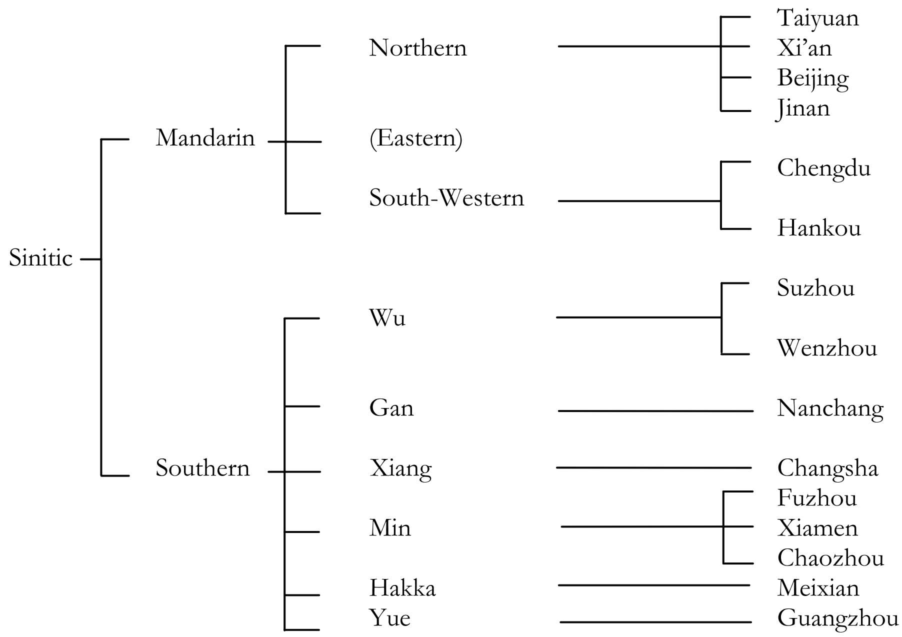

Figure 2.1 Dialect taxonomy based on 'Chinese Linguistic Groups'.[18] The Eastern sub-branch (in parentheses) is not represented in my 15-dialect sample. The Mandarin super-group comprises the groups of the Northeastern, the Eastern and the Southeastern varieties. The Southern (non-Mandarin) branch has groups of Wu, Yue, Gan, Kejia (Hakka), Min and Xiang. [19]

[18] This division follows the Chinese Linguistic Groups (see Figure 2.3, cf. http://www.china-travel.com/china-travel-guides/china-maps) In the *Language Atlas of China*, the sub-groups of the Mandarin branch are: Northeastern, Beijing, Jilu (Beifang), Jiaoliao, Zhongyuan, Lanyin, South-western, Jianghuai. (see Figure 2.2) Furthermore, Jin is branched off the Mandarin branch in the *Language Atlas of China* (Wurm et al. 1987).

[19] In the *Language Atlas of China* (Wurm et al. 1987), there are other non-Mandarin sub-branches: Pinghua and Huiyu.

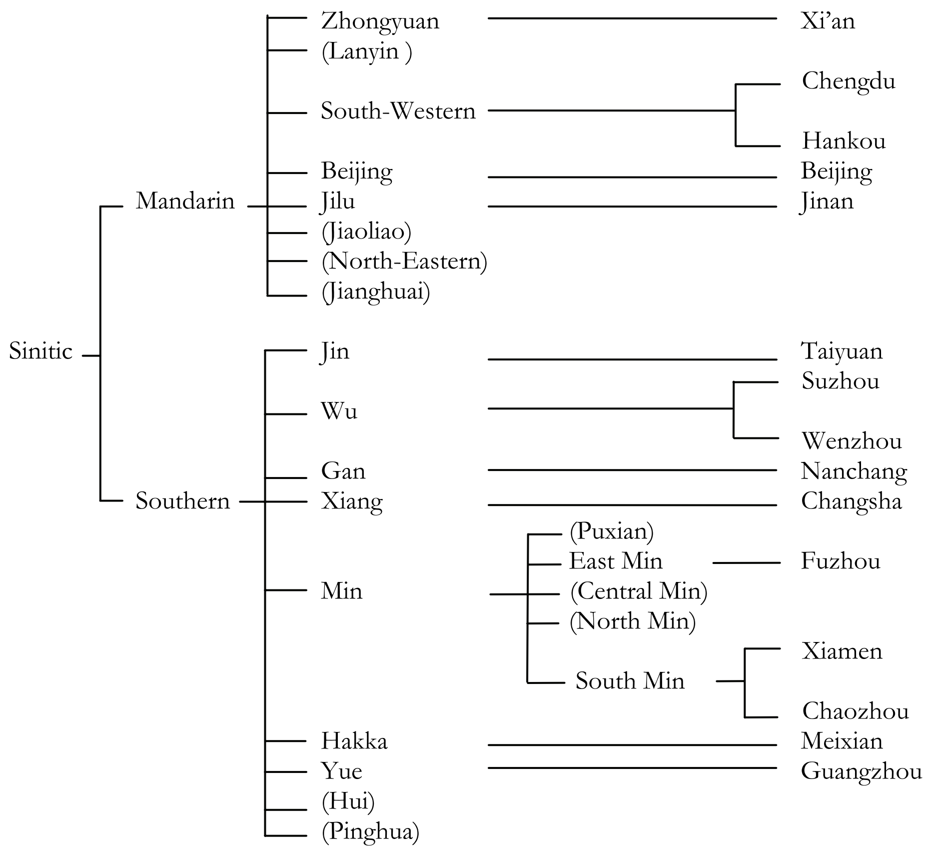

Figure 2.2 Dialect classification based on the *Language Atlas of China* (Wurm et al. 1987). (Sub)branches in parentheses are not represented in my 15-dialect sample.

Observing both linguistic maps of Chinese, we find that the basic division line between the branches of Mandarin and non-Mandarin (Southern) is through the middle from the North to the South part of China. As a result, most non-Mandarin (Southern) dialects are geographically distributed along the coastal line (cf. Figure 2.3, Tang & Van Heuven 2007, and Li 1987).

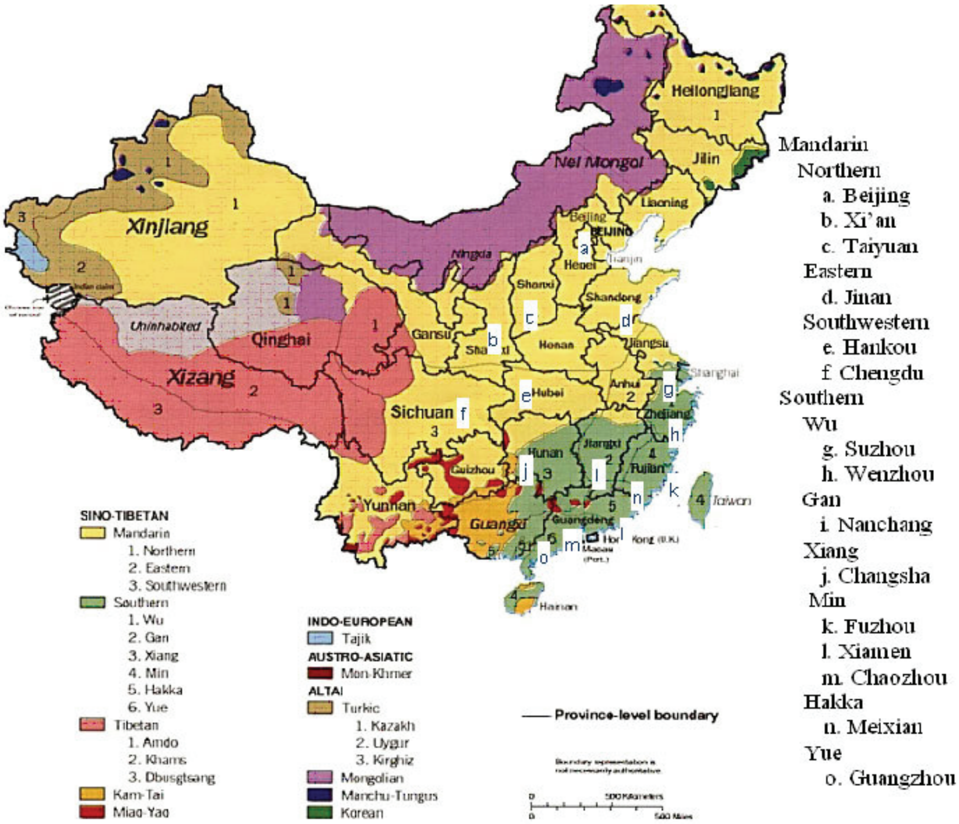

Figure 2.3 Language distribution in the P. R. China (downloaded from http://www.chinatravel.com/chinatravel guides/china-maps). The 15 target dialects (a-o) of my study are identified on the map and listed hierarchically according to dialect (sub)group.

By and large, the popular dialect map published on the internet is a simplified version of the more detailed proposal by the *Language Atlas of China*. For instance, the internet map groups Xi'an, Beijing and Jinan together as Northern Mandarin dialects, whilst Chinese dialectologists, notably Li (1987), consider each of these dialects to be instances of separate branches within the Mandarin branch, i.e. Zhongyuan, Beijing and Jilu, respectively.[20] Similarly, Li et al. (1987) set up a number of subgroups within the Min Supergroup, grouping Xiamen and Chaozhou as dialects of South Min, and Fuzhou as an instance of East Min. Such subdivisions within the Min group are not made in the internet map (1987: A-1).

There is, however, one major discrepancy between the two taxonomies. It concerns the status of Taiyuan. In the internet map it is grouped with the Northern Mandarin

[20] In the *Language Atlas of China* (Wurm et al. 1987), Jilu was first called 'Beifang'.

dialects, together with Xi'an, Beijing and Jinan. In Li (1987: A-1), Taiyuan is set up as a language branched off from the Mandarin super-group, more specifically, as an instance of a Jin super-group which falls into the non-Mandarin (Southern) branch. There are some differences that separate Jin dialects from the Mandarin branch. According to Yan (2006: 71), Taiyuan (the representative of Jin group) is different from the other neighboring northern dialects because it did not realize the Middle Chinese (MC) /g-/ as an unaspirated /k-/ (whilst dialects in the Mandarin branch did) but rather as an aspirated /k'/ in non-level tones. This is a feature that is typical of Hakka and Gan dialects. The main reason, however, for Li (and others) to consider Taiyuan to be a non-Mandarin (Southern) dialect is that it kept the *Ru* tone, which is something Mandarin dialects do not normally do. Yet, Eastern Mandarin and some South-Western Mandarin dialects (which are not included in our sample) also kept the *Ru* tone, so that the classification of Taiyuan as a non-Mandarin (Southern) dialect is not straightforward. One of the purposes of this dissertation is to check the classification of the controversial Jin group (represented by Taiyuan). For the moment, I will leave the status of Taiyuan undecided; later, I will have occasion to settle the issue on the basis of my own experimental and objective lexico-statistical data. I aim to find out how distant the Taiyuan variety is from the Mandarin members and then decide on its grouping.

2.3 Primary split between Mandarin and non-Mandarin branches

The split between the Mandarin and non-Mandarin (Southern) branches of Sinitic Chinese can be traced back to an evolution of phonological features, most notably in the changes of initial consonants, of final consonants and the emergence of a split in the tone system. The changes of initial consonants and of final consonants will be explained in the following passages, the tone evolution from the Middle Chinese (MC) will be illustrated in Tables 2.1 to 2.3 and described in the text.[21] Table 2.1 shows the tone split that took place in Middle Chinese, Table 2.2 illustrates the tone system of the non-Mandarin (Southern) branch of Sinitic Chinese, and Table 2.3 explains the tone system of the Mandarin branch of Sinitic Chinese.

Historically, according to Li (1973), initial consonant clusters such as /gl-, bl-, ml-/, and many final consonants such as /-b, -d, -g, -p, -t, -k, -m, -n, -ŋ / existed in the syllable structure of Archaic Chinese.[22] With a gradual change until 600 A.D., the evolution of Chinese was characterized as a continuous process of merging and simplification of syllable types. The nasal and stop endings underwent varying degrees of neutralization, weakening, and loss (Chen 1973). After 600 A.D. Sinitic Chinese saw a series of evolutions. The complex initial clusters were simplified and the /-b, -g, -d/ finals disappeared. In the same period, the contrast of voiced-voiceless initial consonants appeared, although some dialects retained the old sonant initials /b'-, d'-, g'-/, whilst

[21] Middle Chinese (MC) was the language used during the Sui, Tang, and Song dynasties (6th through 10th centuries A.D.). It can be divided into an early period, reflected by the *Qieyun* rhyme table (601 A.D.), and a late period in the 10th century, reflected by the *Guangyun* rhyme table.
[22] Archaic Chinese refers to the Chinese spoken from the Zhou Dynasty (1027-256 B.C.) well into the Former Han Dynasty (206 B.C. to 9 A.D.). Archaic Chinese was followed by Middle Chinese and – more recently – by Modern Chinese.

other dialects changed these into voiced /p-, t-, k-/ and voiceless /p'-, t'-, k'-/.[23] All language varieties which underwent the unvoicing of the initial consonants, the merging of Middle Chinese (MC) nasal endings, the dropping of Middle Chinese (MC) final stops /-p, -t, -k/ and the simplification of rhymes, form the Mandarin branch.[24] Basically, the Mandarin branch is identified by the ending of the syllable, i.e., in the Mandarin branch most codas of Middle Chinese (MC) were reduced or disappeared altogether; only /-n, -ŋ/ were kept from the list of /-p, -t, -k, -m, -n, -ŋ/ whilst dialects in the non-Mandarin (Southern) branch kept the voiced initial consonants and the codas /-p, -t, -k/.

Secondly, a tone split, which Middle Chinese (MC) underwent, contributed to the Mandarin and non-Mandarin (Southern) division of Chinese dialects. According to the *Language Atlas of China* (Wurm, T'sou, Bradley, Li, Xiong, Zhang, Fu, Wang & Dob 1987), the basic criterion for the division into Mandarin and non-Mandarin (Southern) lies in the tonal changes, viz, the evolution of the so-called *Ru* tone (an abrupt tone) from voiceless initial consonants in Middle Chinese (MC).

More specifically, there were originally four tone categories in MC, i.e. the tone melodies of *Ping Sheng* (level), *Shang Sheng* (rising), *Qu Sheng* (departing), and *Ru Sheng* (entering). These four tone melodies then each split into two registers around late MC: *Yin* (upper register) and *Yang* (lower register).[25] Theoretically, this yields eight tones with two registers of four tone melodic types: *Yin Ping*, *Yang Ping*, *Yin Shang*, *Yang Shang*, *Yin Qu*, *Yang Qu*, *Yin Ru* and *Yang Ru*, as indicated in Table 2.1:

Table 2.1. Traditional names and organization of word melodies in late Middle Chinese in terms of tones and register.

Register	Tone (Sheng)			
	Level (Ping)	Rising (Shang)	Departing (Qu)	Entering (Ru)
Upper (Yin)	Yin Ping	Yin Shang	Yin Qu	Yin Ru
Lower (Yang)	Yang Ping	Yang Shang	Yang Qu	Yang Ru

In the course of the above evolution, some dialects simplified their system by merging certain lexical tones. For instance, Mandarin dialects have lost some tones and kept only four or even as few as three tones (for example, in the Lanyin subgroup): *Yin-Ping*, *Yang-Ping*, and *Qu* (or sometimes *Ping*, *Shang* and *Qu*). Other dialects (typically the Southern/non-Mandarin dialects), however, may not have merged any tones or even underwent further tone splits (e.g. Cantonese had a further split of the *Yin Ru* tone, and now has nine tones). As a result, most Mandarin dialects have no *Ru* (Entering) tone.

[23] The diacritic ' represents aspiration in the Chinese dialectological tradition.
[24] For Mandarin dialects, they experienced the procedure of the final stops' disappearance: from /-p, -t, -k/ to /-k/, /-ʔ/ and to /0/ (no such final stops at all), whilst some non-Mandarin (Southern) dialects kept either complete final stops (Yue dialects) or partial final stops (Min dialects, Wu dialects etc.).
[25] Tones of syllables with Middle Chinese (MC) voiceless initials are called Yin, whilst tones of syllables with Middle Chinese (MC) voiced initials are called Yang (Cheng 1973: 95).

Furthermore, only the *Ping* (Level) tone split into *Yin* and *Yang* registers (cf. Table 2.1), and the MC entering tone merged into other tones, yielding the total of four tone melodies. Retroflexion, rhoticization, and tone neutralization are commonly found in most Mandarin dialects. Dialects within the Mandarin branch share most of these common phonological features and have consequently been claimed to constitute a homogenous set.

Dialects in the non-Mandarin (Southern) branch experienced less merging of voicing features but more tone splits, so that most non-Mandarin (Southern) dialects have the voiced-voiceless initial contrast as well as complex tone inventories typically in excess of five tones. Dialects in this branch vary so much that a traveler often has the feeling of encountering a new language at every two or three miles. The inhabitants of neighboring villages have a hard time understanding each other, especially along the Southern coast of China. Dialects in these regions have remained archaic and are heterogeneous.

2.3.1 The non-Mandarin branch

In the non-Mandarin (Southern) branch, dialects generally have characteristics of keeping *Ru* tones and the corresponding stop endings. As for the changes of the initial consonants, there are various cases in the situation of each group. The following passages will explain the overall phonological characteristics of these groups in the non-Mandarin (Southern) branch. For some groups, I will use my selected dialects as examples to show the characteristic features in their own groups.

Dialects in the Wu group have a distinctive voiceless-voiced contrast. According to Yan (2006), the Wu dialects are characterized by having kept the MC slack voiced obstruent (plosives and fricatives) initials, such as /b̥-, d̥-, g̊-, z̥-, v̥-/, etc. as voiced in contrast with their voiceless counterparts which have remained voiceless in their modern reflexes, thus maintaining the three-way contrast of MC initial stops /p-, p'-, b-, t-, t'-, d-, k-, k'-, g-/ and affricates /ts-, ts'-, dz-/z-; tɕ-,tɕ'-,dʑ-/, etc. Some MC initials underwent splitting and have a literary-colloquial contrast, e. g. MC /mj-/ split into literary /v/ and colloquial /m/ (in Suzhou and Wenzhou dialects). Most MC finals (such as /-am, -em, -an, -uan, -en, -uen/) lost their nasal endings and became open or nasalized (e.g. in Suzhou and Wenzhou dialects). The three nasal codas /-m, -n, -ŋ/ merged into /-ng/, and the three finals stops /-p, -t, -k/ merged into /-ʔ/ with *Ru* tone. Most diphthongs became monophthongs, e.g. MC finals /-ai, -uai, -ei, and -uei/ remained open but lost the /-i/ ending.

Each MC tone split into *Yin* (high) and *Yang* (low) except the *Shang* tone in some dialects (e.g. in Suzhou dialect, *Yang-Shang* merged into *Yang-Qu*) according to the voicing and tenseness of the initials. High tones occur with voiceless, tense nasal and lateral initials, while low tones occur with voiced initials (including voiced nasal and liquid). Normally dialects in this group have seven or eight tones (see also Table 2.2).

The Min (super)group is claimed to have branched off from MC earlier than the other dialects. As a consequence, a greater amount of variation developed and this group is usually regarded as the farthest removed from the Mandarin branch. Min dialects are characterized by keeping the MC codas (e.g. /-m, -ŋ, -p, -k, -ʔ/ occur in nearly all of the Min dialects and by the change of the original voiced stops into unaspirated voiceless sounds, even in *Ping-Sheng* where the aspirated pronunciation is the prevalent one.

The Min (super)group further split into subgroups according to their internal differences: (1) the South Min subgroup comprises the Quanzhang cluster, represented by Xiamen (viz. Amoy) dialect, the Datian cluster and the Chao-Shan cluster (in the Chaozhou and Shantou dialect area), (2) the North Min subgroup is represented by Jian'ou dialect), (3) the East Min subgroup includes Houguan cluster (represented by the Fuzhou dialect) and Funing cluster, (4) the Central Min subgroup (represented by the Yong'an dialect), and (5) the Puxian subgroup (in the Putian-Xianyou area).[26]

South Min normally has seven codas: /-m, -n, -ŋ, -p, -t, -k, -ʔ/ but Chao-Shan dialect has no /-n/. East Min has a distinctive feature that sets it apart from the other Min dialects: the influence of tones on the nature of the vowels in finals. For instance, in Fuzhou, the *Ying-Ping*, *Yang-Ping*, *Shang* and *Yang-Qu* tones tend to co-occur with tense finals containing higher and more fronted vowels than in the case with *Yin-Qu*, *Yang-Qu* and *Yin-Ru* tones which show a tendency to co-occur with lax finals containing vowels that are lower and less fronted. Generally, most Min dialects have six to eight tones (e.g. Xiamen has seven tones besides the neutral tone).

[26] In some publications, more subgroups of Min are distinguished, such as Leizhou Min, Shaojiang Min and Qiongwen Min (see *the Language Atlas of China* and http://www.glossika.com/en/dict/ classification/min/index.php).

Table 2.2. Number of present-day tones and historical development of tone system for six non-Mandarin (Southern) Chinese dialect groups.

Dialect Family	# tones	Tone Split									explanations
		Level *Ping Sheng*		Rising *Shang Sheng*		Departing *Qu Sheng*		Entering *Ru Sheng*			
		Yin	*Yang*	*Yin*	*Yang*	*Yin*	*Yang*	*Yin*	*Yang*	neutral	
Wu	7/8	+	+	+	+*	+	+	+	+		*For some dialects (Suzhou, *Yang-Shang* merged into *Yang-Qu*)
Min	6-8	+	+	+	+	+	+*	+	+		*SM (Chaozhou) has 8, NM has 6, no *Yin-Yang* split for *Ping* and *Shang*, CM has 6, no *Yin-Yang* split for *Qu* and *Ru*, the others (Xiamen & Fuzhou) have 7.
Yue	9	+	+	+	+	+	+	+	+	+	
Hakka	6	+	+*			+	+	+	+		MC *Shang* tone became *Yin-Ping*
Xiang	6	+	+	+		+	+	+			
Gan	6/7	+	+	+*		+	+	+	+		*For some dialects, *Shang-Sheng* tone merged into *Yang-Qu*.

Note: SM=South Min, NM=North Min, CM=Central Min.

The Yue group (represented by Guangzhou, viz. Cantonese) has the most complex tone system among all Chinese dialects. It has a system of eight, nine, or more tones. This dialect group has kept most features of MC. Yue dialects are characterized by the preservation of all MC final consonants. A distinction between long and short vowels, as in Cantonese (i.e. Guangzhou), is also a special feature. Certain tone distinctions depend on the length of the vowel.

The Xiang group comprises Old Xiang and New Xiang. The former is represented by Shuangfeng dialect, which kept the voiced initials (stops and affricates) as in MC (i. e. /b'-, d'-, g'-/); this makes Old Xiang rather like Wu dialects (the latter has Changsha as a representative). New Xiang is closer to the Mandarin branch as the voiced obstruent initials of MC became voiceless (unaspirated) consonants (/p-, t-, k-/). Xiang dialects kept the ancient voiced stops as truly voiced consonants (except Changsha dialect). Final /-p, -t, -k/ are usually lost but the *Ru* tone is preserved in distinct tone classes.

The Hakka (Kejiahua, literally 'guest languages') group, which is spoken by the Hakka people mostly in Southern China, i.e. the eastern and northern parts of Guangdong province, is widely distributed in over 200 cities and counties. Among the Chinese dialect groups, Hakka is the only group which is not named geographically. Meixian (viz. Moi-yen or Moin-yan) in Guangdong province is the representative dialect of the Hakka group. Despite being a non-Mandarin (Southern) dialect, Hakka was actually the result of northern emigration. The Hakka group was formed in the course of migration, at some stage in the history, of some Northern inhabitants (in central plains) to the South, and later to the South-west. Hence Hakka shares many common features with the Mandarin branch, though arguably not enough to make them mutually comprehensible.

The Hakka emigrants made a conscious effort to preserve their own speech after they settled down in their new residence. As a result, Hakka is distinguishable from the local vernaculars spoken in the area the Hakka people migrated to. Hakka dialects have retained a high degree of internal uniformity and are internally mutually intelligible to one another despite the influence of neighboring dialects. Several phonological features of MC are preserved in Hakka. The fully voiced initials of MC have become aspirated unvoiced initials (stops and affricates).[27] MC final consonants /m, -n, -ŋ, -p, -t, -k, -ʔ/ survived in one of three alternative ways: (1) all of these endings are still present, (2) only /-n, -ŋ, -Ø (elision of nasal compensated by vowel nasalization)/ and /-t, -ʔ/ survived, or (3) only /-ŋ, -Ø / are still present. A noticeable phonetic characteristic of Hakka is that the MC *Shang* tone has become *Yin-Ping*. Only two of the four MC tones in Meixian exhibit a *Yin-Yang* split (*Ping* and *Ru* tone), yielding six tones.

The Gan group is represented by Nanchang. This dialect group has the smallest number of speakers and the smallest geographic distribution compared to the other five non-Mandarin (Southern) (super)groups. Geographically, Gan dialects are surrounded by Wu and Min dialects to the east, Xiang dialects to the west, Yue and Hakka dialects to the south, and Mandarin dialects to the north. As a result, the Gan group shares

[27] This feature is shared by dialects in the Gan group. This is supported by the Meixian (Hakka)-Nanchang (Gan) cluster in Figure 4.2 (sentence-intelligibility tree) in Chapter Four.

some common features with all these dialect groups, especially with Hakka (Kejia). For instance, it is common for all Gan dialects that the MC voiced initials (stops and affricates) have become voiceless aspirates. This is similar to Hakka so that some scholars have proposed an integrated family of Gan-Hakka or Ke-Gan. However, the Gan group – unlike Hakka – has the tendency to voice all aspirates in connected speech, whilst the Hakka group preserved the final consonants such as /-m, -p, -k/ much better than the Gan group. Dialects in Gan-Hakka are characterized by the change of the ancient voiced stops into aspirated voiceless sounds in all four original tone classes (aspirated in *Ping-Sheng* only in the three Mandarin groups). Finals /-p, -t, -k/ have been maintained to varying degrees depending on the specific dialect, for example, for Nanchang dialect, the MC stop ending /-p/ neutralized to /-t/. The *Ru* tone is preserved and there are often six or seven tones.

2.3.2 The Mandarin branch

The difference between subgroups within the Mandarin branch also lies in phonological features, typically in the coda, the tone registers, and the pitch change. I will now briefly characterize the differences between the subgroups in the Mandarin branch of Sinitic Chinese (cf. Li 1973).

The Northern Mandarin group occupies a large area in the north of China, from Manchuria in the north to Hubei in the south, from Xinjiang in the west to Jiangsu in the east. Dialects in this group represent most of the Mandarin characteristics: the unvoicing of the MC voiced obstruents (stops, affricates and fricatives), the *Yin-Yang* split of the *Ping Sheng* (level) tone and the disappearance of the 'entering tone' (*Ru Sheng*), so that as a rule there are only four tones: i.e. *Ying-Ping*, *Yang-Ping*, *Shang*, *Qu*. There are further subgroups in this group (see Table 2.3).

The Eastern Mandarin group is spoken along the lower Yangtze River in the provinces of Anhui and Jiangsu. These dialects differ from the other Mandarin groups by the survival of the *Ru* tone as a separate short tone; dialects in this group, therefore, have five tones. Also, the original final consonants /-p, -t, -k/, which accompanied the *Ru* tone, were replaced by a glottal stop.

The South-Western Mandarin group is a fairly uniform type of language spoken in Sichuan, Yunnan, Guizhou, and parts of Hubei and Guangxi. These dialects have no *Ru* tone. Although some dialect localities preserved it as a special (fifth) tone, the final consonants have completely dropped (not even a glottal stop was left). This group can also be sub-divided.

In sum, dialects within the Mandarin branch share many internal similarities. All of these Chinese dialects share more or less the same vocabulary (more homophones depend on tones to be distinguishable), and they share similar syllable and tonal structures as well.

2.4 The traditional (sub)grouping of Chinese language varieties

We will now review relevant literature on the grouping of Sinitic languages.

Based on the phonological features and tonal evolutions from Middle Chinese (MC), consensus has been reached that the Sinitic stock primarily splits into the Mandarin and the non-Mandarin (Southern) branches. However, the (sub)grouping of members in each branch is still controversial. It is difficult to count the number of Chinese dialects because the answer depends on the different criteria to classify the Chinese dialects.

As Lü pointed out:

> If we count as distinct dialects whenever the speech is slightly different, then such dialects are too numerous to count…. If our criterion is different in the phonological system, then such dialects number is the hundreds, perhaps even a thousand or two (1980: 85, translated by Wang 1996: 236, cf. Yan 2006: 8).

Traditional dialectologists differentiate language varieties according to different criteria. Research on Chinese dialects, i.e. the classification and or (sub)grouping of Chinese dialects, mostly focused on the descriptive qualities according to aspects of language phenomena. According to the literature, earlier Chinese dialects groupings were based on non-linguistic criteria, mostly on the geographical distribution. In the earlier period of the 20th century, there was no clear differentiation between Mandarin and non-Mandarin dialects (Zhang 1900). Zhou Zhenhe (1991:48) claimed that the major southward migrations between the pre-Zhou and West Jin dynasties gave birth to the various non-Mandarin (Southern) dialects and the later migrations after different periods of West Jin, Yuan, Qing dynasties shaped the fundamental pattern of their dialectal geography. Since the 1930s, there were basically two criteria for Chinese dialects groupings. One was the geographical distribution, for example:

1. Li Jinxi (prior to 1930) classified Chinese dialects into 12 groups according to their geographical names, mainly using the Yangtze River ('jiang' in Pinyin) and the Yellow River ('he' in Pinyin) as reference points: Hebei Group (literally, dialects in the north of the Yellow River), He'nan Group(dialects in the south of the Yangtze River), Hexi Group (dialects in the west of the Yellow River), Jianghuai Group, Jianghan Group, Jinhu Group, Zhongyuan Group, Jinsha Group, Taihu Group, Yuehai Group, Minhai Group, Ouhai Group. (cf. He Jiuying 1995:414-415).
2. Chao Yuen-ren (1934) firstly distinguished 'Mandarin' (including Jin, Xiang and Gan) from the non-Mandarin dialects, but roughly using the term 'North-east Mandarin' and 'South-east Mandarin'.[28]
3. Chao (1939) then reclassified the Chinese dialects into nine groups including subgroups: Northern (Shangjiang/Upper Yangtze River) and Southern (Xiajiang/ Lower Yangtze River) Mandarin, and separated Xiang and Gan from Mandarin groups.

[28] Chao's dialect classification was used in the Zhongguo Fensheng Ditu (Maps of China's Provinces), in 1934, with no texts. Its second edition with texts was published in 1939. The reclassification was found in the Mandarin Primer (1948). (From Yan 2006: 9)

Table 2.3. Number of tones in inventory and evolution of the *Ru* tone in the Mandarin subgroups.[29]

Mandarin subgrouping according to		Tone evolution		Representative dialect
		# tones	What happened to *Ru*	
Linguistic Map	Atlas			
Northern	Lanyin	3	merged into *Ping*, *Qu*	Lanzhou
	Jiaoliao	3 / 4	merged into *Ping*, *Shang*, *Qu*	Qindao
	Zhongyuan	3 / 4	merged into *Ping*	Xi'an
	North-Eastern	4	merged into *Ping*, *Shang*, *Qu* (mostly into *Shang*)	Shengyang
	Beijing	4	evenly merged into *Ping*, *Shang*, *Qu*	Beijing
	Jilu[30]	4	merged into *Ping*, *Qu*	Jinan
Eastern	Jianghuai	5	was kept	Yangzhou
South-Western	South-Western	4 / 5	normally merged into *Yang-Ping*; some dialects kept the *Ru* tone	Chengdu, Wuhan

(cf. http://zh.wikipedia.org/wiki/%E5%AE%98%E8%AF%9D, Chinese version)

[29] The corresponding of *the Language Atlas of China* subgroups with the sub-branches of the Linguistic Map is roughly based on their geographical distribution. So there is no absolutely clear division line and some discrepancies might not be avoided.
[30] In the text of *the Language Atlas of China*, Jilu is called Beifang.

4. In his later literature, Chao (1948) specifically classified the Chinese dialects into eleven groups, and furtherly distinguished 'Mandarin' varieties including: Beifang/ Northern Mandarin Region, Shangjiang /Upper Yangtze River Mandarin Region and Xiajiang/Lower Yangtze River Region from the other non-Mandarin varieties (Wu, Wan, Min, Chaoxian, Kejia/Hakka, Yue, Xiang and Gan) based on the Yangtze River. (Yan 2006:8-9).[31]

The other criterion was based on the linguistic features of these dialects. Most scholars classified dialects mostly based on historical sound changes, i.e. the phonological characteristics. The dialect groupings and dialect identifications always involved the phonological features of MC as the reference point.

1. Wang Li (in the 1930s) was the first to use the major phonetic characteristics for Chinese dialectal classification. Based on this principle, he classified the Chinese dialects into five groups, see Table 2.4 (cf. Wang 1996:249, Yan 2006).
2. Li Fanggui (1937: 1-13) then proposed to classify Chinese dialects into eight major groups based on some phonological features of MC.[32]
3. Yuan Jiahua (1960) merged all Mandarin dialects into Beifanghua but distinguished South Min from North Min according to their common or different features.
4. Zhan Bohui (1981, 1991) subgrouped the Northern dialects into Huabei, Xibei, Xi'nan, Jianghuai; bifurcated Xiang into Old Xiang and New Xiang; split Min into North Min, East Min and South Min.
5. Ting Pang-hsin (Ding Bangxin) (1982: 257-258) classified Chinese dialects into seven major groups according to 17 different evolutionary features of MC: 16 features concerning the development of initials, finals, tones and 1 principle in terms of historical sound changes under 6 universal conditions (cf. Wang 1996: 256, Yan 2006: 14-16).
6. Huang Jinhu (1987) subgrouped Mandarin dialects into Northern Mandarin, Beijing Mandarin, Jianghuai Mandarin, Lanyin Mandarin, Zhongyuan Mandarin, South-western Mandarin. He also furtherly split Min group into East Min, South Min, Puxian, Central Min and North Min.
7. Ramsey (1987) classified Chinese dialects into Mandarin group and non-Mandarin group including Wu, Xiang, Gan, Kejia, Yue and Min.
8. Norman (1988: 182) firstly proposed diagnostic features for the dialect classification and he grouped Chinese dialects into three major branches according to ten phonological, grammatical and lexical features: the Northern (Beifang) group

---

[31] The Mandarin dialects in the north part of Yangtze River were Northern Mandarin, in the upper Yangtze River were Shangjiang Mandarin, in the lower Yangtze River were Xiajiang Mandarin. Dialects geographically crossed southward of Yangtze River were Southern dialects or non-Mandarin dialects including Chaoxian, Xiang, Wu, Gan, Kejia (Hakka), Yue, Min and Wan. (see Table 2.4).

[32] His criteria accepted and developed by some scholars (i.e. Forrest 1948, Tung T'ung-ho 1953, Yuan 1960, Zhan 1981, 1991, Ting 1982, Huang 1987, Ramsey 1987, Norman 1988 and Lau 2002). Among them, Ting (1982: 258) argues to assign different weights to the historical and synchronic features in the decision to classify varieties into major dialect groups, secondary dialects, sub-groups etc. In a set of *Language Atlas of China* (Volume I and II) (Wurm et al. 1987, 1990), the hierarchy of dialect classification from macro- to micro-perspective is adapted reflecting a structure in terms of supergroup, group, subgroup, cluster and local dialect.

included Northern Mandarin, Southern Mandarin, North-west Mandarin and South-west Mandarin; the Central Group comprised Xiang, Wu and Gan; the Southern Group braced Kejia (Hakka), Yue and Min.

9. Based on Ting's criteria of historical sound changes and the tonal evolution feature plus the migration history, Lau (2002) proposed a new dialect classification leading to just four groups: Northern group (Beifanghua area), Wu, Min and Gan-Yue group.
10. The *Language atlas of China* proposed ten (super)groups (Mandarin Supergroup, Jin group, Wu group, Xiang group, Gan group, Min Supergroup, Yue group, Hakka/Kejia group, Hui group and Pinghua group) and detailed subgroups based on the notions of descending hierarchy: supergroup, group, subgrouping, cluster and local dialects (Wurm et al. 1987).

The various traditional dialect groups and sub-groups are listed in Table 2.4. (abbreviations: Note 33)[34]

Obviously, Table 2.4 tells us that the classification of Chinese dialects is discrepant at different stages according to the individual researchers or scholars. There is consensus that there are major branches of Mandarin and non-Mandarin (Southern) for the Sinitic language family but the detailed sub-groupings of these dialects within each primary split are rather diverse.

The literature and Table 2.4 show us that dialects sharing the common/similar phonological characteristics go together, while dialects in different groups and/or different subgroups are different in terms of phonological features. The classification did not tell us how large the similarity (or the difference) between dialects is. With further research of these dialects and more dialect data available, measures on the distances/similarities between dialects are addressed. The results of these measures tell us to what extent the dialects differ or are similar to one another in terms of quantitative scales. Conversely, the results can be used to validate the traditional dialect classification.

[33] HB: Hebei, HuaB: Hua Bei, HN: Henan, HX: Hexi, JH: Jianghuai, JHa= Jianghan, JHu: Jinhu, ZY: Zhongyuan, TH: Taihu, JS: Jinsha, YH: Yuehai, MH: Minhai, OH: Ouhai, N: Northern, SJ: Shangjiang, XJ: Xiajiang, UY: Upper Yangtze River, LoY: Lower Yangtze River, SW: South-Western, XB=Xibei, XN=Xi'nan, N: North, BJ: Beijing, LY: Lanyin, N: North, NW: North-West, S: South, SW: South-West, NE: North-East, JLu: Jilu, JL: Jiaoliao, SE: South-East, EG: Eastern Gan, WMK: Western Min-Ke, Ke: Kejia (Hakka), HL: Huizhou local dialect, NX: New Xiang, OX: Old Xiang, P: PuXian, (Min)S: South Min, (Min)E: East Min, (Min)C: Central Min, (Min)N: North Min, CX: Chaoxian, Hui: Huiyu, Ping: Pinghua..

[34] The classification in different periods does not match.

Table 2.4. Traditional Chinese Dialect groups and sub-groups (abbreviations are explained in Note 33).

Scholar (down)	Number (across) 1	2	3	4	5	6	7	8	9	10
Zhang 1900	No clear differentiation between Mandarin and non-Mandarin (Southern) branches. No Mandarin features.									
Li 1930	HB, HN, HX, JH, JHa, JHu, ZY		OH	TH	JS		YH	MH		
Chao 1934	North-east Mandarin	South-east Mandarin including Jin, Xiang and Gan				non-Mandarin				
Chao 1937-1948	Northern/Shangjiang, Southern/Xiajiang	Jin	Xiang		Gan			Jin included in Mandarin		
Chao 1939	Mandarin: N, SJ, XJ	CX		Wu		Kejia	Yue	Min	Wan	
Chao 1948	N, UY, LoY	CX	Xiang	Wu	Gan	Kejia	Yue	Min	Wan	
Wang Li 1930's	Mandarin phonetic group, including Xiang			Wu		Kejia	Yue	Min	other	
Li 1937	N, XJ, SW			Wu	Gan-Ke		Yue	Min		
Yuan 1960	Mandarin/BeifFangHua		Xiang	Wu	Gan	Kejia	Yue	Min: N, S		
Zhan 1981, 1991	Mandarin (Northern dialects): HuaB, XB, XN, JH		Xiang: OX, NX	Wu	Gan	Kejia	Yue	Min: N, S, E		

Table 2.4 (continued)

<table>
<tr><th rowspan="2">Scholar (down)</th><th colspan="10">Number (across)</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th></tr>
<tr><td>Ting 1982</td><td>Mandarin</td><td></td><td>Xiang</td><td>Wu</td><td>Gan</td><td>Kejia</td><td>Yue</td><td>Min</td><td></td><td></td></tr>
<tr><td rowspan="2">Huang 1987</td><td>Mandarin</td><td rowspan="2"></td><td rowspan="2">Xiang</td><td rowspan="2">Wu</td><td rowspan="2">Gan</td><td rowspan="2">Kejia</td><td rowspan="2">Yue</td><td>Min</td><td></td><td></td></tr>
<tr><td>N | BJ | JH | | LY | ZY | SW</td><td>N | C | S | E | PX</td><td></td><td></td></tr>
<tr><td rowspan="2">Norman 1988</td><td>Northern Group</td><td colspan="4">Central Group</td><td colspan="3">Southern Group</td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>N | NW | S | SW</td><td></td><td>Xiang</td><td>Wu</td><td>Gan</td><td>Kejia</td><td>Yue</td><td>Min</td></tr>
<tr><td>Ramsey 1987</td><td>Mandarin</td><td></td><td>Xiang</td><td>Wu</td><td>Gan</td><td>Kejia</td><td>Yue</td><td>Min</td><td></td><td></td></tr>
<tr><td rowspan="2">Lau 2002</td><td rowspan="2">Northern dialects</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2">Wu</td><td colspan="3">Gan-Yue</td><td>Min</td><td></td><td></td></tr>
<tr><td>EG | WNK | HL</td><td>Kejia</td><td>Yue</td><td></td><td></td><td></td></tr>
<tr><td>Atlas 1987</td><td>Mandarin
NE | BJ | JLu | JL | SE | LY | ZY | SW</td><td>Jin</td><td>Xiang</td><td>Wu</td><td>Gan</td><td>Kejia</td><td>Yue</td><td>Min</td><td>Hui</td><td>Ping</td></tr>
</table>

(cf.: Yan 2006:17-18)

Dialects may differ in various linguistic structures (phonetic sounds, lexical vocabularies, phonological features, morphological and syntactic differences etc.). The distance measures between pairs of dialects can be used as predictor parameters to determine the *mutual* intelligibility between two pairs of dialects. The more distant the two dialects are, the lower degree of *mutual* intelligibility is; and the closer the two dialects are, the higher degree of *mutual* intelligibility between them is (cf. Chapter One).

The following sections will briefly talk about the distance measures and the *mutual* intelligibility testing on Chinese dialects. § 2.5 reviews the structural measures on Sinitic varieties done by linguists. § 2.6 mentions the arguments about the *mutual* intelligibility between Sinitic varieties. § 2.7 introduces the status of Sinitic varieties spoken in China.

2.5 Structural distance measures on Sinitic language varieties

Traditionally, as mentioned in § 2.4, Chinese dialects were classified in terms of qualitative differences of language characteristics: isoglosses, synchronic or diachronic rules, the presence or absence of certain features. The final overall (sub)grouping or classification was based on the relatively more important features which could be determined by personal qualitative judgments. Despite the long tradition of linguistic and dialectological research in China, the issue of the (sub)grouping of Sinitic language varieties are still controversial, as Yan (2006: 238) said:

> Looking back at the development of Chinese dialectology since the 1930s, we can see that the issue of classification of Chinese dialects is still not settled.

Even the authoritative classification proposed by the *Language Atlas of China* (Wurm et al. 1987) has been questioned whenever different criteria or characteristics are used. For example, in the *Language Atlas of China* linguists/dialectologists (e.g. Li Rong) proposed that the divisions of Chinese might go even further to ten groups instead of seven (in tradition). To this effect, a new group (Jin) was separated from the Mandarin branch, as well as Pinghua from Cantonese (Yue group), and Huiyu from the Wu group. However, this classification is claimed to be challenged by Chinese linguists when various approaches have been adopted (Wang 1996: 235-267, Yan 2006: 238). Furthermore, some dialect continua at dialect contact/transitional areas or boundary areas are still debated on their division or classification, viz. Changsha (Xiang group), Nanchang (Gan group) and Meixian (Hakka group).[35] Therefore, Chinese language varieties are related to one another in a broad diachronic and synchronic sense, covering both genetic and contact relationships.

As Cheng (1993b, 1997) explained, when we ask the question: how different are Chinese dialects, we are just referred to their classifications. But then again, on what are these classifications based? The answer can be: on the differences between dialects. We seem to end up in circularity. Cheng (1997) summarized the problem by stating that

[35] Normally, the dialect is named regionally, e.g. the dialect is named after the city or region where it is spoken, for example, Changsha is the capital city of Hunan province, and then we call the dialect spoken there the Changsha dialect.

Chinese dialects are traditionally classified in terms of qualitative differences of linguistic characteristics, which may be represented by isoglosses (which may be stated as rules either synchronic or diachronic according to the presence or absence of some phonological features). However, the same dialect might be classified into different groups based on the initials or consonantal ending of the syllable, respectively. Even though we considered the overall differences by weighing some relatively important features, these weights were likely to be chosen on the basis of personal qualitative judgments, which need to be tested objectively:

> 'Chinese linguists need to agree on a set of objective diagnostic criteria including phonological, lexical and syntactical features for the classification of Chinese dialects. Of course, this task and settling on the number of criteria and the weight of each criterion for the classification of the major grouping and sub-grouping could be a great challenge to all the Chinese dialectologists.' (Yan 2006: 238)

Cheng (1997) was well aware that we should know whether two dialects are *mutually* intelligible or not when we are talking about their difference. However, Cheng believed that a *mutual* intelligibility score based on the subjectively counted percentage of correctly understood sentences in recorded passages (as done by Voegelin & Harris 1951, Hickerson, Turner & Hickerson 1952, Pierce 1952) was insufficient as a language affinity measure although he, too, agreed that *mutual* intelligibility plays an important role in determining how language varieties were related to each other. Therefore, he turned to calculating various linguistic structure measures on pairs of Chinese dialects.[36]

With the development of measurement methodologies, various methods were applied to measuring similarities and differences between related language varieties. Since 1930s, various structural measures have been used for linguistic research. For example, a correlation method was used for language classification for Indo-European (Kroeber & Chretien 1937, 1939), American English (Reed & Spicer 1952) and Middle English (Ogura 1990). Other proposals for language classification were also found, e.g. in Hsieh (1973), Krishnamurti, Moses & Danforth (1983), Cavalli-Sforza & Wang (1986) and Wang (1987).

Since the 1960s, the lexicostatistical methods were applied to determining linguistic relationships between Chinese dialects (Wang 1960). More extensive investigation into affinity of Chinese dialects became possible and feasible since 1970s with the arrival of computer technology (Cheng 1982, 1986, 1987, 1988, 1991; Wang 1987; Ogura 1990).

Structural similarity/difference comparison can be based on the linguistic aspects such as phonological correspondence, cognate occurrence, syntactic structure, Levenshtein distance between the transcriptions for each syllable or some meaningful combination of all these elements. Some Chinese researchers (notably Cheng) did much work on

[36] Confusingly, in his approach he called the phonological correspondence between the cognate word sets in dialects as mutual intelligibility (Cheng 1993b, 1997), which is not conceptually correct.

quantitatively measuring linguistic structure. He determined the similarities/differences between Chinese dialects on several aspects: Lexical similarity, phonological correspondence, differences of lexical tones between dialectal varieties, morpheme correspondence based on syllables in cognates of dialect pairs, morpho-syntactic structure comparison between dialects, etc. His various measurements were used to validate the Chinese dialect taxonomy proposed by Chinese dialectologists based on the evolution rules and phonological features developed from the Middle Chinese (Cheng 1973, 1986, 1991, 1993a, 1996, 1997).

In Chapter Five, I will focus on such structural measures. I will introduce Cheng's measures such as lexical affinity, phonological affinity, phonological correspondence for 17 Chinese dialects. I am in a fortunate situation that Cheng published most of his measures (lexical and phonological measures) so that I could reuse his measures to generate affinity trees according to my 15 target dialects. I will also offer my structural measures on my selected 15 Chinese dialects (all are the subset of Cheng's selection) in details, i.e. inventory affinity based on data from Campbell's website, phonological frequency affinity and Levenshtein distance based on the database done by the linguists in CASS (Chinese Academy of Social Sciences) for my 15 Chinese dialects. Following Cheng's cue, I will generate agglomeration trees from each of the measures. These hierarchical structures will be used to compare with the traditional Chinese dialectal taxonomy. Any tree which correctly reflects the primary split of the Mandarin and non-Mandarin (Southern) branches will be used in Chapter Six to predict my own subjective measures based on *mutual* intelligibility tests from Chapters Three and Four. We will decide on the classification issue of Jin dialects and determine the preferred mutual intelligibility testing approach for Chinese dialects after correlating the experimental results with traditional Chinese dialect taxonomy. Conclusions will be drawn in Chapter Seven.

2.6 *Mutual* Intelligibility between Chinese language varieties

Chinese dialects are characterized as the uniformity and divergence. Dialects in the Mandarin branch are generally homogeneous and are acknowledged to be relatively *mutually* intelligible to one another whilst dialects in the non-Mandarin (Southern) branch are more or less divergent and are claimed to be *mutually* unintelligible both with the Mandarin dialects and with their own non-Mandarin (Southern) dialects. This can be seen, for instance, in the following quotations (my italics, TC):

> It is generally accepted that there are eight major dialects groups in Chinese that *are mutually unintelligibile*. (Bao 1999: 8-9)

> A striking aspect of Chinese is *the lack of intelligibility* across dialect families, namely, speakers from different dialect families often *cannot understand each other*. (Duanmu 2000:2)

> It is traditional to speak of the different varieties of Chinese as 'dialects', even though they may be different from one another to the point of being *mutually unintelligible*. It is often pointed out, for example, that Cantonese and Mandarin

> differ from each other roughly as the Romance 'language' Portuguese and Rumanian do. … The greatest variations in terms of phonology, syntax, and vocabulary occur in the southern region of the country, The dialects of the Mandarin group …not only can claim the largest percentage of China's population, but also have a *higher degree of mutual intelligibility.* (Li & Thompson 1981:2)

> These dialects differ greatly in certain aspects of phonology, lexicon, and syntax and *are mutually unintelligible.* (Yan 2006:2)

> According to lexical statistical data of any two languages within the Sinitic branch (for example Wu and Mandarin), the data will always reveal that there is *less intelligibility between them than any two Romance languages* in Europe. For example, French has lexical similarity of about 75% to several other Romance languages. In comparison, Mandarin has 31% lexical similarity with Wu (Shanghainese) and 19% with Yue (Cantonese). …. Of all the dialects within the various languages probably the most uniform group, … is Mandarin. I know based on experience that South-western and Zhongyuan Mandarin are not difficult to understand with Beijing and the other northern varieties. (Campbell, 2009: http://www.glossika.com/en/dict/faq.php#1)

According to Chinese dialectologists, Chinese dialects differ from one another because of the phonological and tonal differences (and/or different grammatical usage). The differences among these dialects are as great as any two languages within the Indo-European language family (Yan 2006). However, linguistic research on dialect similarities and differences usually focuses on qualitative aspects of language phenomena: listings of phonological and syntactic rules and descriptions of certain shared linguistic characteristics among dialects. The classification and subgroups of Chinese dialects are based on the comparisons of phonological features and syntactic rules. These descriptions cannot capture the degree of similarity (or distance) between two varieties in a single number and do not afford a clear prediction of their degree of mutual intelligibility.

As mentioned above, dialects within the Mandarin branch are often claimed to be intelligible to each other to some extent, but are not mutually intelligible to dialects in the non-Mandarin (Southern) branch (despite the recent influence of Standard Mandarin).

The dialect members within their own branch do not only share a large number of cognates (i.e. words that are historically related in the two varieties) but also there exist some regular correspondence between the varieties in the phonological shapes of the words. In the Mandarin branch, for example, Beijing and Chengdu dialects mostly share the same phonemes in a syllable and differ in the tones only, whereby the low and falling tones in Beijing show up as falling and low tones in Chengdu, i.e. the tones have been switched between the two dialects. Therefore, dialects in the Mandarin branch are claimed to have a higher degree of mutual intelligibility to each other (Li & Thompson 1981: 2, Campbell 2009). In some cases, even dialects belonging to different branches share some correspondence rules, i.e., the sounds and tones of one dialect can be related to those of another through systematic rules. For instance, the diphthong [ai] in Chengdu corresponds with monophthong [e] in Shanghai dialect (Duanmu 2000: 2).

Most dialects in the non-Mandarin (Southern) branch are mutually unintelligible either to each other or to the Mandarin dialects. Moreover, dialects in the Mandarin branch and dialects in the non-Mandarin (Southern) branch are claimed mutually unintelligible across the border that separates these dialects into two primary branches (Duanmu 2000: 2, Yan 2006: 2). That is, listeners from Mandarin branch cannot well understand the speakers from the non-Mandarin (Southern) branch, whereas dialects within the Mandarin branch are sufficiently mutual intelligible. Generally speaking, within the Southern branch, dialects in the Min, Yue and Wu groups are least mutually intelligible to Mandarin dialects, followed by Hakka, Gan and Xiang. However, some cross-group intelligibility has been claimed in exceptional cases. For instance, Xiang dialects (belonging to the non-Mandarin/Southern branch) may share common terms and some degree of intelligibility with South-Western Mandarin dialects.

Many reasons may explain the mutual (un)intelligibility between Chinese dialects. For instance, the Northern part of China is situated on the plains, affording easy travel, whilst the Southern part is very mountainous and difficult to travel through. Accordingly, there may have been less language contact between non-Mandarin (Southern) dialects, which circumstance does not foster mutual intelligibility. It is, however, not the aim of the present dissertation to explain the reasons why the various dialects spoken in China grew apart to different degrees. What I am interested in, is: (i) to test the impressionistic claims of the asymmetry between Mandarin and non-Mandarin dialects, (ii) establish their degree of mutual intelligibility between these dialect pairs, and (iii) to see whether and to what extent I can predict the degree of mutual intelligibility from objective, structural differences between the dialects (such as the number of cognates shared by two dialects, and the transparency of the phonological differences between two dialects), (iv) to settle the status of controversial dialectal groups such as the Jin group and (v) decide which mutual intelligibility testing method is better regarding to the traditional Chinese dialects taxonomy. Specifically, in my study, the claims of mutual (un)intelligibility will be substantiated experimentally. I will determine the degree of mutual intelligibility between Chinese dialects experimentally by opinion tests and functional tests. 'Later studies of intelligibility all used tests with speakers, involving phonological perception, structural and contextual criteria, translation, or sociolinguistic factors' (Cheng 1992: 147). In this way, we are trying to answer Cheng's question how 'participant intelligibility' is to be calculated (Cheng 1992: 167). I will validate my experimental results with Cheng's structural measures and regress all these results to traditional Chinese taxonomy. In Chapter Three, I will describe my experimental methods used to obtain judgments of similarity and judgments of mutual intelligibility between pairs of my target 15 Sinitic dialects. The collected judgment scores will be processed by SPSS and then be converted to tree structures showing the distance/affinity relationships between pairs of these dialects. In Chapter Four, I will functionally test the mutual intelligibility at the levels of isolated words and of words in context (sentence level). The resulting tree structures will be compared with those obtained from the judgment tests. The relatively better method will be determined through validation of these experimental results against the traditional dialect taxonomy.

2.7 The popularity of Chinese dialects

Because of the divergence and the *mutual* unintelligibility among Chinese dialects, a lingua franca is needed for communication purposes. Therefore, a so-called 'Standard Mandarin' (or Putonghua – literally 'the common language') which is based on Beijing dialect) gained a prestige position and plays the dominant role in national and international communication. This is the variety that is officially prescribed by the Chinese national government.[37] However, speakers (most of them are bilingual – native dialect and Standard Mandarin) of each dialect tend to maintain their local dialects. Despite the fact that Putonghua is used as the language of education from kindergarten to university and is propagated through mass communication via TV broadcasting and the internet, various dialects (variants) (sub-standard or non-standard, social or geographical, prestigious or downtrodden) co-exist and are popularly spoken. In reality, dialects are actively used and spoken by most people, especially in the older generation, in daily conversation. People prefer to communicate with their fellow dialect speakers in their own dialect/vernacular. In addition to governmental media programmes which are broadcasted in Putonghua (the Standard Mandarin), local programmes (talk shows, television plays) in local dialects are also televised.

Moreover, following the long tradition of Chinese dialectal research, dialectology is part of curriculum at linguistic institutes or departments in many universities, and the native/local dialect still has the position of the students' mother tongue. As a consequence, people in the same dialectal area or transitional/neighbouring area may be mutually comprehensible, but communication may break down (completely) between speakers hailing from more distant areas (unless both parties use Standard Mandarin).

The popularity position of local dialect addresses various interesting research topics about Chinese dialects; the mutual intelligibility between these dialects is one of them. The wealth of collected Chinese dialectal data and the large number of live dialect speakers and listeners make our experimental research (Chapters Three and Four) possible and feasible. The previous structural measures developed for Chinese dialects by Chinese linguists (notably Cheng, see Chapter Five) afford a comparison between the objective measure results and our experimental results (in Chapters Three and Four). All the measures will be used to validate the traditional Chinese dialect taxonomy established by Chinese linguists and dialectologists.

[37] Beijing dialect is agreed as the dialectal norm for Mandarin. In this sense, Beijing dialect and Standard Mandarin are very similar to each other and they are phonologically identical except that there are a few differences: Beijing dialect prominently has more rhotic vowels (suffixed by –r, so-called er'hua), has more phonetic lenition and more slang words (http://en.wikipedia.org/wiki/Beijing_dialect).

Chapter Three

Mutual intelligibility of Chinese dialects: Opinion tests

3.1 Introduction

As explained in Chapter One, mutual intelligibility is one of criteria for distance measures on language varieties. It is often used to argue about the genealogical relationship between pairs of language varieties. The more two languages are like each other, the more closely they are related. Language varieties that are very close, are often called dialects of the same language. In order to determine the difference between language varieties, we need to measure the 'distance' between them. Distance between languages is used as a criterion when arguing about genealogical relationships between languages. The more the languages resemble each other, the more likely they are derived from the same parent language, i.e., belong to the same language family. However, it is difficult to quantify the distance between languages one-dimensionally since languages differ along many structural dimensions (e.g. phonetics, phonology, morphology, syntax).

Useful work on structural measures of difference between non-tonal languages has been done, for instance, at Stanford University (for Gaelic Irish dialects, Kessler 1995) and at the University of Groningen (for Dutch and Norwegian dialects, Heeringa 2004, Gooskens & Heeringa 2004), using the Levenshtein distance, a similarity metric that computes the mean number of string operations needed to convert a word in one language to its (cognate) counterpart in the other language. This objective measure was then used to build a tree structure (through hierarchical cluster analysis) which matched the language family tree as constructed by linguists.

It is unclear how various dimensions of language difference should be weighed against each other. That is, we do not know which structural correspondences are more or less important when constructing a difference/similarity measure. Obviously, the problem gets even more complex when we apply such distance measures to tonal languages. Ideally, we want to express the difference/similarity in a single number on a one-dimensional scale rather than as a distance between points in some multi-dimensional hyperspace. Therefore, we select a single criterion — mutual intelligibility. Mutual intelligibility exists between two languages A and B when speakers of language (variety) A can readily understand speakers of language (variety) B (and vice versa) without prior exposure, intentional study or extraordinary effort. By definition, mutual intelligibility is an overall criterion that may tell us in a psychologically relevant way whether two languages are similar/close to each other. When two language varieties are mutually intelligible, they are probably dialects of the same language; when their mutual intelligibility drops below some threshold measure, the varieties belong to different languages.

Although methods for determining intelligibility are well-established, for instance, in the fields of speech technology and audiology (e.g. Van Bezooijen & Van Heuven 1997), the practical problems are prohibitive when mutual intelligibility has to be established for, say, all pairs of varieties in a set of 15 dialects (yielding 225 pairs). Rather than measuring intelligibility by functional tests, opinion testing has been advanced as a shortcut (Van Bezooijen & Van Heuven 1997). That is, the indices of the measurements of mutual intelligibility between languages are generated from listeners' judgment scores.

The question can be raised to what extent the judged distance between a stimulus dialect and the listener's own dialect correlates with (judged) intelligibility. It should be realized in this context that perceived distance between some dialect and one's own is not necessarily the same as an intelligibility judgment. One of the aims of the present chapter is to test to what extent judged distance and judged intelligibility actually measure the same property.

The work done by Gooskens and Heeringa represents a complication relative to earlier work in that their Norwegian dialects are tone languages whilst the Gaelic Irish and Dutch dialects are not. Since it is unclear how tonal differences should be weighed in the distance measure, Gooskens and Heeringa collected distance judgments for the same reading passages resynthesized with and without pitch variations. The difference in judged distance between the pairs of versions (with and without pitch) would then be an estimate of the weight of the tonal information. Norwegian, however, is a language with a binary tone contrast. We want to test Gooskens and Heeringa's method on a full-fledged tone language, viz. Chinese, a language (family) with much richer tone inventories varying from four (Mandarin, e.g. Beijing, Chengdu) to as many as nine (Yue/Cantonese, e.g. Guangzhou).

The Norwegian language situation is rather unique in that Norwegian arguably has no standard language: Norwegians only use local dialects. This is a felicitous condition when trying to predict mutual intelligibility from structural differences between dialects. In the Sinitic (e. g. Chinese) language situation (as in most other countries) one dialect has the status of national language or standard language, so that it is widely used for the purposes of communication in education and mass media. I will test the hypothesis that mutual intelligibility can be predicted from structural differences more adequately when the standard language is excluded from the set of dialects in the study (more details in Chapter Five).

In the present and next chapters, I will describe the experimental procedures we followed to establish mutual intelligibility among a set of 15 Sinitic languages. The present chapter introduces the opinion experiment and Chapter Four is about functional tests.

In this chapter, I will first determine the judged mutual intelligibility and judged similarity between pairs of dialects through opinion tests by presenting recordings of the same spoken passage in 15 Chinese dialects to naïve listeners of the same set of dialects and asking them to rate the dialects along both subjective dimensions (similarity judgment and intelligibility judgment). Dendrograms (tree structures) can be generated from the judgement scores collected by the naïve listeners. I will then compare these trees with the traditional dialect taxonomy proposed by Chinese linguists. I will also test the impressionistic claims that non-Mandarin (Southern) dialects are less mutually intelligible than Mandarin dialects and the Mandarin dialects speakers are more intelligible to non-Mandarin (Southern) dialects than vice versa. I will finally determine to what extent the two subjective measures (judgment similarity and judgment intelligibility) correlate with each other.

Our opinion tests used sound files of the well-known fable *The North Wind and the Sun* as the uniform material for all the 15 dialects, spoken by native speakers of these 15 dialects, we technically processed these files into one consistent gender voice, and produced two versions of each dialect file, one is normal speech with melody, and the other is monotonous speech without melody.

For each dialect of the 15 dialect speeches in two versions, we found 24 listeners of one native dialect to listen to it and rate their scores of intelligibility and distance similarity scaled from 0 (completely unintelligible and different) and 10 (completely intelligible and exactly the same) from or as the listener's native dialect. All our comparisons will be based on these outputs produced by SPSS databank of the two scores (intelligibility and similarity). And our analysis will come from these outputs.

3. 2 Method

In this section I will first describe how I collected subjective estimations of intelligibility and similarity for all 225 pairs of 15 Chinese dialects (in § 3.2), and then correlate the collected judgment intelligibility with judgment similarity(in § 3.3). In § 3.4, I will compare the mutual intelligibility between and within the Mandarin and non-Mandarin groups. Conclusions are summarized in § 3.5, and I try to answer the questions raised in the introductory section § 3.1 concerning the asymmetry of the mutual intelligibility between and within (non-)Mandarin groups, the convergence and divergence between the experimental results and the traditional dialect taxonomy, the correlation of judgment similarity and mutual intelligibility. In § 3.6, I introduce a control experiment in order to test the possible artifacts of sound quality in terms of artificially processed sound manipulation in the main experiment in § 3.2.

3.2.1 Materials

The Chinese dialects we targeted are the following 15 (a proper subset from Cheng 1997): Beijing, Chengdu, Jinan, Xi'an, Taiyuan, Hankou (Mandarin dialects), Suzhou, Wenzhou (Wu dialects), Nanchang (Gan dialect), Meixian (Hakka dialect), Xiamen, Fuzhou, Chaozhou (Min dialects), Changsha (Xiang dialect), and Guangzhou (Yue dialect)[38]. For their geographic location see Figure 3.1.

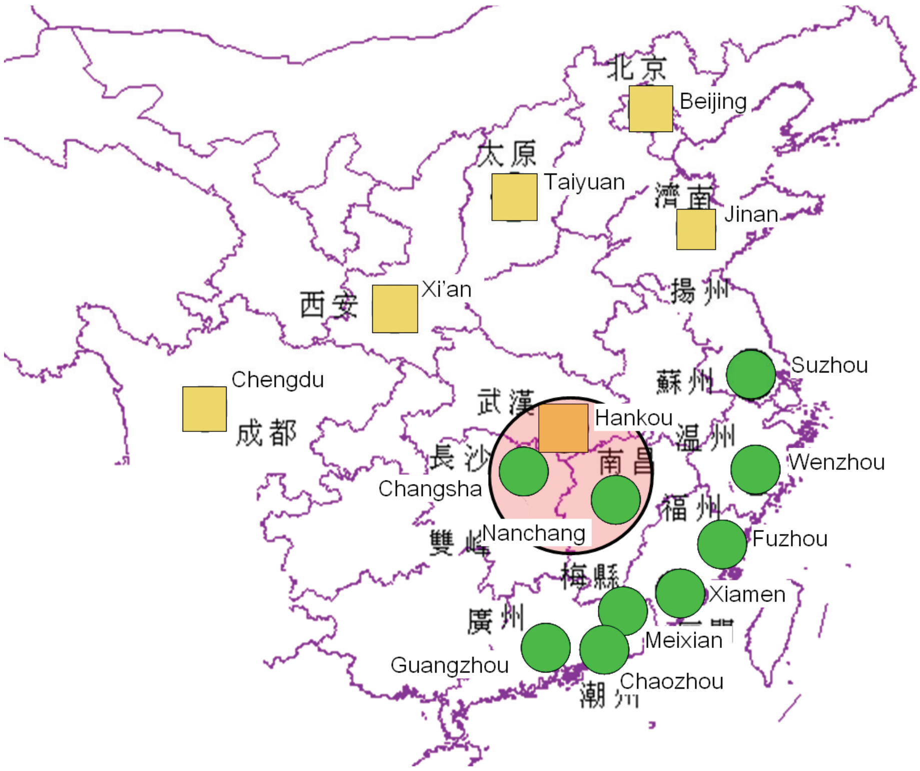

Figure 3.1 The geographical distribution of 15 selected Chinese dialects. Mandarin dialects (in the north) are represented by squares, non-Mandarin dialects (mainly along the coast) by circles. The three dialects in the central part (a transitional area) are enclosed in a bigger circle.

I used existing recordings of the fable 'The North Wind and the Sun' (supplied by the Institute of Linguistics of Chinese Academy of Social Science). Since each fable had been read by a different speaker (11 males and 4 females) with different speech habits, we processed the recordings (using Praat software, Boersma & Weenink 1996) such that all speakers sounded like males, all had roughly the same articulation rate and speech-pause ratio, and the same mean pitch. Also, each reading of the fable was pro-

[38] As I explained in the Introduction of Chapter One: the status of Taiyuan is undecided yet. I would like to provisionally treat it as a Mandarin dialect as it was traditionally treated. Later, Taiyuan will be reconsidered when all measures are collected.

duced in two melodic versions, i.e., one with the original pitch intervals kept intact, and one with all pitch movements replaced by a constant pitch (monotone), which was the same as the mean pitch of the fragment with melody (and the same as all other fragments).

To obtain these manipulated versions, the mean pitch was normalized to the mean of the eleven male speakers. Relatively small shifts in pitch (in semitones) were performed (using the PSOLA pitch manipulation implemented in the Praat software) on the male speakers, larger shifts were required for the female voices.[39] For the female speakers a gender transformation was carried out by decreasing the formant frequencies by 15%. Longer pauses were reduced to 500-ms length, and the remaining speech was linearly speeded up or slowed down such that the mean syllable duration (also called 'articulation rate', expressed in syllables per second) was the same for all speakers. We established, in a separate experiment (see § 3.6), that possible differences in sound quality (whether incurred in the original recordings or as a result of our signal manipulations) did not explain differences in judged intelligibility or judged similarity among the speech samples.

The 2 × 15 readings of the fable were recorded onto audio CD in four different random orders (A, B, C, D, where C and D were the reversed order of A and B). The 15 monotonized versions preceded the 15 versions with melody. At the beginning of the CD, as part of the instructions, we recorded the reading (with melody) of the fable in the dialect of the prospective listener group. This was done to make sure that the listeners would be perfectly familiar with the contents of the fable. In all, 60 different CDs were produced.

3.2.2 Listeners

In total 360 listeners participated in the experiment. For each of the 15 dialects a group of 24 native listeners was found (12 males and 12 females). These listeners satisfied the experimental requirements: They were born and had grown up in their respective native dialect-speaking areas. They had not traveled outside their hometown in their life. Ideally, their parents only speak their native dialect. Listeners were mono-dialectal so that they had little experience with any other Chinese dialects. Because of the popularity of Putonghua (Standard Chinese) and the exposure to TV programs, we preferably include listeners who are in the middle to older generation (ages between 40 and 60), although they may have had some familiarity with the Standard Mandarin language through primary education and later media exposure. Therefore, the younger generation is not eligible.

In order to check the qualifications of the listeners, I prepared a questionnaire for each of the candidate listeners. They were asked to fill in their background information

[39] PSOLA: Pitch Synchronous Overlap and Add, a technical method in speech synthesis processing, which allows the user to change the duration or fundamental frequency of speech waveforms without audibly affecting the spectral quality of the sounds (see e.g. Moulines & Verhelst 1995).

(those who are illiterate obtained the assistance from the experimenter we recruited). The questionnaire sheet is attached in Appendix 3.1

3.2.3 Procedure

For each locality of the 15 dialects, we enrolled an experimenter was contracted to be in charge of the experiment and play the CD to 24 listeners. Manuals with detailed instructions were both prepared for the experimenter and for the listeners, respectively. The experimenter was obliged to read and explain the instructions to the listeners very clearly before s/he began any run of the experiment. Besides the paper instruction, I also recorded spoken instructions for the experimenter and for the listeners at the very beginning of each CD. The first part of the CD was a demonstration of the example of the monotonized voice and the normal voice with the melody unchanged. On each CD, the native dialect of the listeners preceded the other 14 dialects. This arrangement is just for the listeners to get familiar with the story they would listen to on the CD. For each CD, the monotonized versions were followed by the normal pitch versions. As a result, each listener listened to 30 passages of the speech (15 monotonized and 15 normal pitch versions). Each CD was played through loudspeakers to six listeners (three females, three males), either individually or in small groups, in a quiet room with little reverberation.

We designed the answer sheet for the listeners to give their scores. On the answer sheet, each listener saw two tasks: one was an 11-point scale for the judgment of intelligibility, the other scale was for the judgment of similarity. Listeners were asked to express their judgment on scales from 0 to 10. '0' meant 'This dialect is completely different from my own dialect and 'I can understand none of the words', while '10' meant 'This dialect is exactly the same as my own' and 'I understand every word.' Consequently, listeners rated the materials twice: the first time they estimated how well they believed a monolingual listener of their own dialect, confronted with a speaker of the dialect in the recording for the first time in his/her life, would understand the other speaker. The second time the listener rated the similarity between his own dialect and the dialect of the speaker in the recording. Illiterate listeners orally communicated with the experimenter, who noted down their responses on the answer sheets.

In between fragments listeners were given 7 seconds to fill in their scores on both scales. The crucial point here is that the listeners were required to use their intuitive judgments without hesitation. There was no need for them to convert any words into their own dialect.

In all 21,600 judgments were collected and statistically analyzed through SPSS (Statistical Package for the Social Sciences). The results are illustrated in the next section (§ 3.2.4).

3.2.4 Results

I will now present the experimental results obtained from native listeners in our own data collection. In total, I obtained four types of results from the experiment: two intelligibility results (for monotonized and normal-pitch versions) and two similarity results (again, for monotonized and normal-pitch versions). As I mentioned above, for each dialect, we have 24 listeners, so that we collected 360 scores for each pair of dialects (a listener dialect and a speaker dialect). The scores were then averaged and the final mean scores for each pair of the dialects were used for further statistical computation. As explained in the introduction (Chapter One), the experimental results can be asymmetrical, because the perceived similarity and intelligibility between dialects A and B are not necessarily identical to that between dialects B and A. In the next sections, I will present the results in the order implied above.

3.2.4.1 Judged intelligibility

Table 3.1 presents the mean subjective intelligibility ratings (obtained from 24 listeners for per dialect) for each of the 225 pairs of dialects in our sample. The stimulus dialect (speaker dialect) is listed as the row variable against the dialect of the listener group in the columns of the matrix. The intelligibility judgments in Table 3.1 are based on monotonized stimuli only, so that differences among the various tone systems are obscured to a large extent (differences between tones may still be cued partially by temporal organization and intensity).

The data in Table 3.1 show that, generally, listeners who are exposed to their own dialect, rate the speaker highly intelligible, with mean scores between 8 and 10 on the 11-point rating scale. It was found earlier, by Gooskens and Heeringa (2004), that intelligibility judgments may be less than perfect for listeners responding to speakers of their own dialect. The speaker of the sample may deviate slightly from the specific language variety of the listeners, who typically form a very homogeneous group, hailing from one village or town. In our data, however, rather poor own-dialect ratings are found for Xiamen and Meixian, and especially for Xi'an. In fact, our Xi'an listeners indicate that they understand the speaker of the neighboring Taiyuan dialect better than their 'own' speaker. Possibly, therefore, our Xi'an listeners originated from the border area of Xi'an, so that they understood the Xi'an dialect less well. [40]

---

[40] Table 3.1 shows that Xi'an listeners indicated that they understood the speaker of the Jinan and Hankou dialects better than their 'own' speaker. In order to find possible reasons that can explain this result, I checked the questionnaire sheets for the Xi'an locality and found that seven listeners claimed that they were not exactly 'native' Xi'an but hailed from neighboring areas: six (out of 24) listeners came from Huxian, and one more from Lantian. Both Huxian and Lantian are towns near Xi'an city (which is the capital city of Shaanxi Province, China). I suspect that these seven listeners' judgments reduced the intelligibility ratings because they honestly judged that the speaker's dialect (Xi'an) they were listening to was not exactly the same as or as perfectly intelligible as their own dialects (Huxian or Lantian).

Table 3.1. Intelligibility ratings for monotonized stimuli spoken in 15 Chinese dialects as judged by groups of 24 listeners of the same 15 dialects. Double lines separate Mandarin from non-Mandarin dialects.

Speaker dialect (down)	Listener dialect(across)															
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an	Mean
Suzhou	**9.83**	2.13	.71	.46	1.13	.87	.63	1.04	.92	.46	.58	1.46	.38	1.21	.54	1.49
Wenzhou	1.63	**8.42**	.71	.33	1.29	.58	.38	1.42	1.67	.65	.38	1.33	.25	2.21	.88	1.48
Guangzhou	.00	.13	**10.00**	2.54	1.29	1.46	3.75	1.67	.96	.25	.04	.38	.13	.42	.67	1.58
Xiamen	.00	.29	.67	**7.58**	1.38	5.08	.38	.75	.54	.13	.25	.17	.13	.46	.50	1.22
Fuzhou	.08	.09	.46	1.08	**9.71**	1.17	.42	1.04	.50	.13	.29	.33	.17	.50	.50	1.10
Chaozhou	.00	.29	.63	1.96	1.79	**9.46**	.50	.42	.25	.42	.00	.17	.17	1.21	1.00	1.22
Meixian	.08	.33	.63	.50	1.08	1.25	**7.46**	1.63	1.08	1.13	.38	.71	.13	1.96	.88	1.28
Nanchang	5.13	1.92	.38	1.17	.96	1.33	.42	**9.63**	3.83	5.25	.42	3.29	1.29	3.63	.71	2.62
Changsha	5.88	2.54	.75	1.22	1.21	.71	.71	3.63	**9.63**	6.29	.88	4.58	3.25	4.92	1.50	3.18
Taiyuan	5.67	1.13	.63	.92	1.00	.58	.83	5.21	2.88	**9.50**	1.04	5.83	1.13	4.00	3.29	2.91
Beijing	8.33	4.50	.63	2.38	1.42	1.88	3.54	8.46	7.29	9.21	**8.92**	9.50	7.13	8.79	8.17	6.01
Jinan	7.00	4.46	.46	1.67	1.33	1.29	.79	6.63	6.29	8.25	3.29	**9.63**	3.33	7.58	7.04	4.60
Hankou	6.13	3.08	.33	.88	1.04	.67	.29	3.58	5.25	5.88	1.08	7.50	**9.75**	7.21	5.88	3.90
Chengdu	6.75	4.92	.58	1.83	1.25	1.63	.96	5.63	6.42	4.88	1.83	7.88	5.75	**10.00**	4.79	4.34
Xi'an	4.63	2.63	.79	1.21	1.17	.54	.33	2.96	3.50	4.71	1.21	7.88	2.75	5.88	**5.75**	3.06
Mean	4.08	2.46	1.22	1.72	1.80	1.90	1.43	3.58	3.40	3.81	1.37	4.04	2.38	4.00	2.81	

We used the average linkage method (which was used by Cheng 1997) to generate dendrogram (tree) in order to show the graphical affinity relationship between the pairs of the 15 target dialects. The mean scores were computed automatically by the SPSS software via the asymmetrical matrix in Table 3.1. A proximity matrix was obtained and shown in Appendix 3.2. The resulting tree (the agglomeration tree) based on these scores was illustrated in Figure 3.2.

Figure 3.2 shows that there is a clear primary split cutting the 15 dialects into two branches. The lower branch includes five non-Mandarin (Southern) dialects, the upper branch comprises all the Mandarin dialects plus four non-Mandarin (Southern) dialects: Changsha (Xiang), Nanchang (Gan), Suzhou and Wenzhou (Wu).

Before we compare this tree with the traditional Chinese dialect taxonomy, let us define some simple criteria in order to evaluate the degree of correspondence between the tree and the traditional dialect taxonomy. A strict criterion would require that the primary split into the upper and lower branches of the tree should perfectly correspond with the traditional division between Mandarin and non-Mandarin (Southern) dialects. Using this strict criterion, the tree contains four misclassifications: Nanchang, Changsha, as well as Suzhou and Wenzhou are incorrectly classified together with the six Mandarin dialects in the upper branch.

However, we may relax the criterion somewhat. It would not be unreasonable to separate Suzhou and Wenzhou off the upper branch comprising all Mandarin dialects and two non-Mandarin (Southern) dialects. Then we may add it to the lower branch forming a new cluster comprising seven non-Mandarin (Southern) dialects. In this case, a primary split between an upper branch with eight dialects that include all the Mandarin dialects plus two non-Mandarin (Southern) in the set of 15, and a lower branch that includes non-Mandarin (Southern) dialects only (marked in the braces). Therefore, the number of classification errors is two. These have been bolded in Figure 3.2. For the detailed internal cluster structure, we leave it to later discussion.

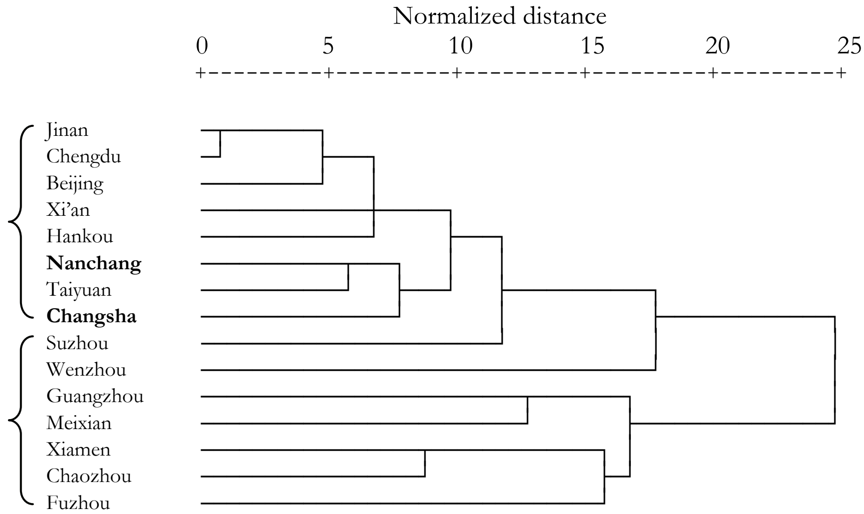

Figure 3.2 Dendrogram based on judged intelligibility scores with no pitch (monotonized stimuli), using Average Linkage (Between Groups) and the Euclidean distance measuring method. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

Let us next consider Table 3.2, which presents the same information as Table 3.1, but now for stimuli which were presented with full melodic information.

Table 3.2. Intelligibility ratings for stimuli with full melodic information spoken in 15 Chinese dialects as judged by groups of 24 listeners of the same 15 dialects. Double lines separate Mandarin from non-Mandarin dialects.

Speaker dialect (down)	Listener dialect (across)														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an	Mean
Suzhou	**10.00**	2.83	.38	1.25	1.21	.83	.46	1.25	.58	.50	1.63	1.92	.54	1.92	2.42	1.85
Wenzhou	1.25	**9.38**	.54	.88	1.38	.46	.71	1.54	2.33	.83	2.50	1.38	.54	1.46	1.00	1.75
Guangzhou	.04	.87	**10.00**	1.42	1.25	1.54	4.58	1.71	1.33	.21	.58	.71	.71	.79	1.54	1.82
Xiamen	.00	.71	.96	**9.08**	1.46	6.04	.54	1.13	.96	.50	1.17	.13	.17	.17	.58	1.57
Fuzhou	.13	.96	.67	.50	**9.83**	1.21	.33	.63	.67	.29	.58	.46	.21	.71	1.29	1.23
Chaozhou	.00	.83	.63	4.79	1.83	**9.71**	.42	.67	.29	.63	.17	.96	.08	.67	1.17	1.52
Meixian	.17	1.08	.38	1.50	2.00	1.88	**9.00**	2.08	.88	1.08	1.08	1.04	.25	.79	.71	1.59
Nanchang	5.08	2.71	.88	1.79	2.17	1.42	.88	**9.92**	4.04	5.04	1.96	4.58	1.46	3.13	2.71	3.18
Changsha	7.00	3.88	.50	2.96	1.71	1.79	.88	6.50	**10.00**	6.71	5.08	6.96	4.54	6.42	5.54	4.70
Taiyuan	6.21	2.75	.67	1.21	1.63	1.21	1.00	2.88	4.17	**10.00**	2.67	6.33	2.25	6.71	6.00	3.71
Beijing	8.58	8.91	.58	7.38	2.42	2.71	5.29	9.46	9.54	9.83	**9.08**	9.83	7.75	9.29	9.54	7.35
Jinan	7.58	5.38	.54	4.54	1.75	1.92	1.00	7.83	6.58	9.42	6.21	**10.00**	3.67	7.96	8.54	5.53
Hankou	7.00	4.00	.42	2.33	2.00	1.25	.92	6.88	6.96	7.75	4.96	8.33	**10.00**	7.67	7.79	5.22
Chengdu	6.54	5.25	.79	1.63	1.46	.83	1.13	6.33	7.04	7.38	4.88	8.00	4.58	**10.00**	6.33	5.07
Xi'an	6.21	3.63	.63	1.92	1.25	.96	.71	4.88	4.46	8.96	4.79	8.00	2.42	7.25	**9.58**	4.38
Mean	4.39	3.54	1.24	2.88	2.22	2.25	1.91	4.25	3.99	4.61	3.16	4.58	2.61	4.33	4.32	

In this condition, listeners who respond to a speaker of their own dialect rated the intelligibility of the speaker at 9 or more on the 11-point scale, and, in fact, 7 out of the 15 dialect speakers were judged to be perfectly intelligible by their own listener group. The Xi'an listeners now rate their own speaker's intelligibility at 9.58 – marginally better than their rating of the neighboring Taiyuan speaker; there are four dialects whose speaker is judged to be less intelligible than 9.5 by their own listeners. It would appear, therefore, that our Xi'an listeners preferred to judge their dialect as less intelligible and less similar to some other dialects (e.g. Jinan) without tone information. However, when they were listening to the normal dialect speech with pitch (tone information) untouched, they were aware that the speaker's tonal behavior made him/her very much an authentic speaker of their own dialect (Xi'an) although they also claimed that they could well understand their neighbor dialect Taiyuan with tone information (the mean score is 9.54) as well. I will refrain from further observations on the structure of the rating data. These can be analyzed in a more insightful manner after applying a cluster analysis.

Using the same method, we generated the dendrogram tree from the matrix of Table 3.2, which is illustrated in Figure 3.3. The proximity matrix can be seen in Appendix 3.3.

Again, if following the strict criterion, the tree primarily split the 15 dialects into two branches. The lower branch comprises five non-Mandarin (Southern) dialects, the upper branch includes all the six Mandarin dialects plus four non-Mandarin (Southern) dialects: Nanchang, Changsha, Suzhou and Wenzhou. In this case, the classification errors are four.

However, if we relax the criterion somewhat, it would not be unreasonable to separate off Suzhou and Wenzhou (the two Wu dialects in our sample) from the other eight dialects in the upper branch, and add it to the lower branch. In this case, an optimal split between an upper branch with eight dialects that include all the Mandarin plus two non-Mandarin (Southern) dialects in the set of 15, and a lower branch that includes seven non-Mandarin (Southern) dialects only (marked in the braces). Therefore, the number of classification errors is also two. These have been bolded in Figure 3.3.

This performance is the same as what was found above for the judgments based on monotonized information. There is a tendency for the tree in Figure 3.2 to reflect dialect subgroups better than its counterpart in Figure 3.3. Min dialects (Xiamen, Chaozhou, Fuzhou) form a coherent cluster in Figure 3.2 but not in 3.3.

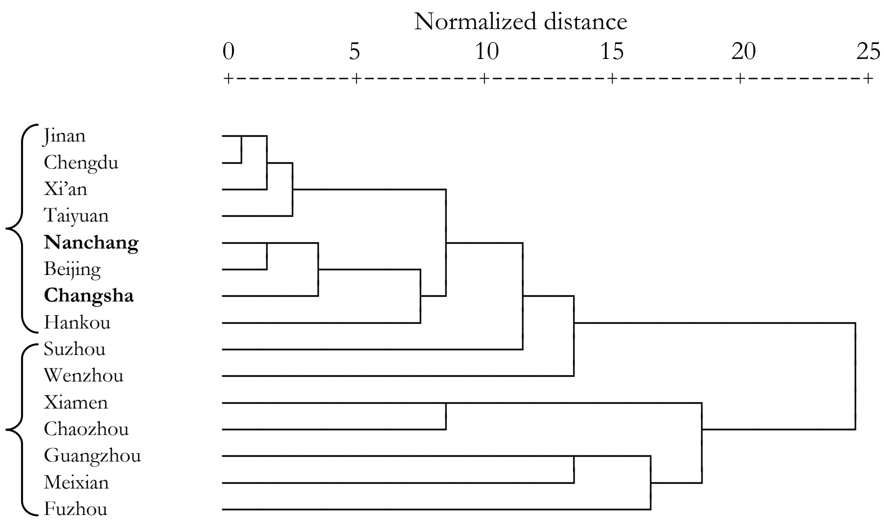

Figure 3.3 Dendrogram based on judged intelligibility scores with pitch untouched (melodic stimuli), using Average Linkage (Between Groups) and the Euclidean distance measuring method. Incorrectly classified dialects have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

3.2.4.2 Judged similarity

It is possible to formally quantify the degree of congruence between the subjective intelligibility and similarity ratings. This will be done as part of § 3.3, where I will compute the correlation coefficient for this pair of subjective measures of linguistics distance collected in this study. I will then be in a position to predict the subjective measures from each other. However, before attempting to correlate the two subjective measures with each other, we will first analyze the judged similarities among the 15 dialects in the form of tree structures.

I will now consider the subjective estimations of the structural similarity between the 15 dialects. These results are presented in Table 3.3 for monotonized stimuli and in Table 3.4 for the full-melodic versions.

Table 3.3. Similarity ratings for monotonized stimuli spoken in 15 Chinese dialects as judged by groups of 24 listeners of the same 15 dialects. Double lines separate Mandarin from non-Mandarin dialects.

Speaker dialect (down)	Listener dialect(across)															
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an	Mean
Suzhou	**9.50**	2.25	.25	.21	.00	.39	.33	.29	.63	.17	.50	.92	.04	.63	.25	1.09
Wenzhou	1.08	**8.70**	.25	.42	.04	.25	.21	.57	1.08	.70	.29	1.00	.08	1.17	.50	1.09
Guangzhou	.00	.13	**10.00**	1.63	.08	.96	4.38	.67	.39	.13	.00	.29	.08	.21	.17	1.27
Xiamen	.00	.30	.21	**6.83**	.25	4.70	.25	.25	.17	.13	.17	.21	.00	.08	.29	0.92
Fuzhou	.00	.43	.08	.71	**10.00**	1.17	.21	.17	.29	.08	.21	.21	.04	.17	.08	0.92
Chaozhou	.04	.33	.25	.33	.33	**9.13**	.38	.21	.29	.38	.00	.21	.08	.67	.42	0.87
Meixian	.04	.46	.29	.17	.08	.79	**8.17**	.79	.61	.67	.38	.46	.00	.75	.38	0.94
Nanchang	.71	.50	.21	.50	.00	.42	.21	**9.42**	2.38	2.00	.42	2.17	.67	1.92	.29	1.45
Changsha	.71	.42	.33	.35	.00	.54	.54	2.54	**8.67**	2.33	.83	2.83	1.42	3.00	.83	1.69
Taiyuan	.38	.17	.29	.29	.00	.13	.42	2.92	1.83	**9.58**	.63	3.92	.46	2.17	2.08	1.68
Beijing	.13	1.63	.21	.88	.00	.38	1.21	6.38	4.50	4.63	**8.71**	8.17	3.54	5.38	6.25	3.47
Jinan	.46	.42	.13	.38	.04	.26	.38	4.83	4.08	3.04	2.25	**9.21**	1.67	4.67	4.88	2.45
Hankou	.50	.88	.25	.13	.00	.57	.17	1.96	3.22	1.92	1.21	5.88	**9.88**	5.04	3.71	2.35
Chengdu	.50	.96	.21	.67	.00	.67	.54	3.39	4.13	2.13	2.25	6.13	2.75	**10.00**	3.29	2.51
Xi'an	.33	.75	.46	.08	.00	.13	.17	1.79	1.96	1.63	1.21	6.54	.92	3.67	**5.00**	1.64
Mean	0.96	1.22	0.89	0.91	0.72	1.37	1.17	2.41	2.28	1.97	1.27	3.21	1.44	2.64	1.89	

Inspection of Table 3.3 reveals a large measure of correspondence between the intelligibility ratings presented above. Again, most listener groups rated their 'own' speaker with scores of 8 or better. This time, there are two speakers who are judged to speak a dialect that is no more similar to the listeners' dialect than 6.8 (Xiamen) or even 5.0 (Xi'an). Comparison with the similarity ratings for the full-melodic versions (Table 3.4) shows that, again, the poor within group ratings for Xiamen and Xi'an disappear completely such that these speakers' dialects are considered to be highly similar to the listeners' dialects with scores close to 9 on the 11-point scale.

The resulting dendrogram tree generated from the matrix in Table 3.3 through the average linkage method is shown in Figure 3.4. The proximity matrix automatically computed by SPSS can be found in Appendix 3.4.

The tree in Figure 3.4 shows a primary split between a lower branch comprising seven non-Mandarin (Southern) dialects only and an upper branch including all Mandarin dialects plus two non-Mandarin (Southern) dialects: Changsha and Nanchang. The number of classification errors is two. The internal structures within the basic branches will be discussed later. Let us first move on to the judged similarity ratings based on the full tone information.

I will repeat the procedure for the responses based on speech samples with full pitch information. The resulting tree is in Figure 3.5.

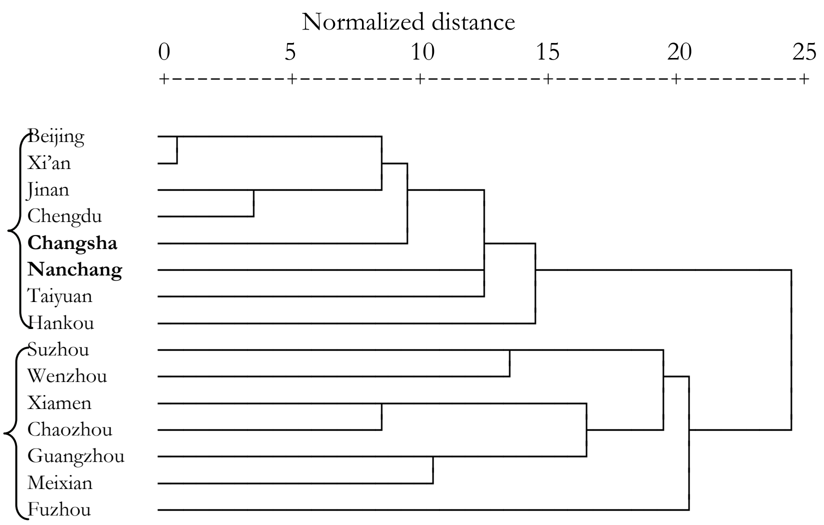

Figure 3.4. Dendrogram based on judged similarity scores with no pitch (monotonized stimuli), using Average Linkage (Between Groups) and the Euclidean distance measuring method. Incorrectly classified dialects have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The rating scores of judged similarity by 24 listeners for each of the 15 dialects based on the full melody version can be seen in Table 3.4. Table 3.4 reproduces the earlier effect that our listeners gave better scores when they were listening to the dialect speech with full melody information. The highest similarity ratings are found on the diagonal line in the matrix of Table 3.4, ranging from 8.6 to 10.00. This time, our Xi'an listeners judged the Xi'an speaker as 86% similar to their own dialect, i.e. nearly 10% more similarity than their judgment for the Beijing speaker.

Table 3.4. Similarity ratings for full-melody stimuli spoken in 15 Chinese dialects as judged by groups of 24 listeners of the same 15 dialects. Double lines separate Mandarin from Southern dialects. Double lines separate Mandarin from non-Mandarin dialects.

Speaker Dialect (down)	Listener dialect(across)														
	Suzhou	Wenzhou	Guangzho u	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an	Mean
Suzhou	**9.96**	2.54	.13	.63	.04	.58	.17	.38	.54	.42	1.50	.92	.04	1.13	1.42	1.36
Wenzhou	.92	**9.17**	.29	.38	.00	.04	.29	.63	1.08	.33	2.25	.67	.04	.92	.38	1.16
Guangzhou	.00	.41	**10.00**	.83	.17	.58	5.52	.67	1.04	.00	.63	.38	.08	.29	1.00	1.44
Xiamen	.00	.75	.50	**8.96**	.17	5.13	.17	.25	.50	.38	.92	.13	.04	.04	.13	1.20
Fuzhou	.04	1.00	.25	.30	**10.00**	.71	.25	.00	.35	.00	.50	.21	.00	.17	.79	0.97
Chaozhou	.00	.38	.33	3.96	.38	**10.00**	.17	.00	.13	.25	.25	.42	.00	.13	.46	1.12
Meixian	.08	.63	.08	1.29	.29	1.25	**9.13**	.75	.50	.38	.92	.58	.13	.33	.33	1.11
Nanchang	.79	2.83	.46	1.08	.00	.50	.58	**9.92**	2.82	.88	1.71	2.46	.13	2.00	1.46	1.84
Changsha	1.25	.38	.25	1.38	.00	.25	.50	4.67	**10.00**	1.38	4.48	3.92	2.25	4.13	3.33	2.54
Taiyuan	.21	1.29	.38	.65	.00	.08	.58	1.29	2.17	**10.00**	2.29	3.92	.83	4.42	3.83	2.13
Beijing	.21	2.00	.29	5.38	.08	.29	1.71	5.00	7.83	4.08	**8.96**	8.04	4.58	6.38	7.71	4.17
Jinan	.29	.21	.29	2.50	.00	.25	.54	5.42	4.58	2.50	5.57	**9.96**	1.67	**5.83**	6.38	3.07
Hankou	.38	.42	.13	1.33	.00	.58	.54	3.58	5.04	1.58	4.43	5.92	**10.00**	5.96	5.38	3.02
Chengdu	.38	1.33	.33	.88	.04	.00	.58	3.58	4.13	1.42	4.52	4.83	2.38	**10.00**	4.04	2.56
Xi'an	.38	.33	.29	.92	.00	.25	.50	3.00	2.79	2.54	3.70	5.87	1.21	5.46	**8.63**	2.39
Mean	0.99	1.58	0.93	2.03	0.74	1.37	1.42	2.61	2.90	1.74	2.84	3.22	1.56	3.15	3.02	

The tree is illustrated in Figure 3.5, the proximity matrix is in Appendix 3.5.

The results are roughly the same as those seen in Figure 3.4. The primary split in Figure 3.5 divides the 15 dialects into a lower branch comprising seven non-Mandarin (Southern) dialects and an upper branch including all the six Mandarin dialects plus two non-Mandarin (Southern) dialects: Changsha and Nanchang. Therefore, the number of classification errors is two. There is no difference in terms of dialect subgroups. Suzhou and Wenzhou are correctly seen as a coherent subcluster (Wu dialects) but this grouping is seen in both trees alike.

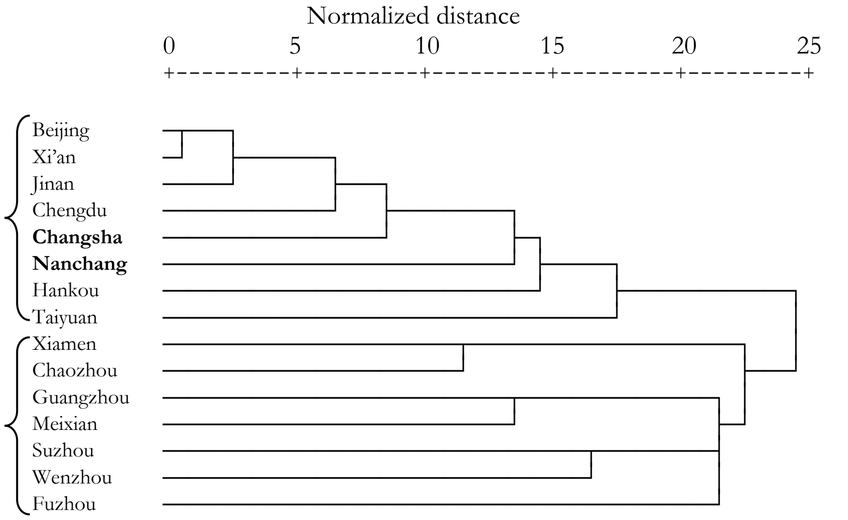

Figure 3.5 Dendrogram structure based on judged similarity scores with pitch untouched (melodic stimuli), using Average Linkage (Between Groups) and the Euclidean distance measuring. Incorrectly classified dialects have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

3.3 Correlation between judged intelligibility and judged similarity

From the Introduction part (Chapter One), we know that the distance/similarity between language pairs is often related to the (mutual) intelligibility between language pairs as well. The other way round, we argue that the more languages are similar/close to each other, the higher degree of (mutual) intelligibility between them can be expected. If two languages are distantly related, then they will be (mutually) unintelligible. In this sense, we expect mutual intelligibility to be strongly related to the distance/dissimilarity between language pairs. Gooskens & Heeringa (2004) found a fairly strong correlation (r = 0.67) for aggregate Levenshtein distances between fifteen Norwegian dialects and the distances as perceived by speakers of these dialects, which makes the assumption quite reasonable. Since we collected separate intelligibility and similarity judgments on our 15 Chinese dialects, we are in a position to check to what extent judged similarity and judged intelligibility do indeed coincide.

I computed the correlation coefficient between judged similarity and judged intelligibility for all combinations of our 15 Chinese dialects. Here we based the results on the responses collected for dialect speech samples with and without full melodic information. We computed the correlation coefficient three times for each melodic case.

The first time we used all 15 x 15 = 225 combinations of speaker and listener dialects, such that the intelligibility and similarity judgment for dialects A and B need not be the same as for dialects B and A (i.e. asymmetrical). The result is $r = .854$ ($p < .001$) for full melody information and $r = .883$ for the monotonous version respectively. These are high correlations, explaining 73% and 78% of the variance, respectively.

It may be objected, of course, that the resulting correlation coefficients may have been inflated by the fact that very high intelligibility and similarity ratings should be expected when listeners respond to a speaker who shares their own dialect. Therefore, I also computed the correlation coefficients after excluding cells along the diagonal of the matrix (210 combinations of speaker and listener dialects remain). Indeed, the correlation coefficients drop somewhat, to $r = .810$ for full melody information and $r = .841$ for the monotonous version – which still accounts for 66% and 71% of the variance, respectively.

Generally the judgment scores for intelligibility and similarity are highly correlated at $r = .854$ ($p < .001$) if all 225 combinations of speaker and listener dialects are included, and $r = .810$ ($p < .001$) when the 'own dialect' condition is excluded.

The third time I computed the coefficient for mutual intelligibility, i.e. on the mean intelligibility and mean similarity for every dialect pair AB and BA (105 combinations). Here, the correlation between intelligibility and similarity judgments is strongest of all, at $r = .888$ for the melody information and $r = .900$ for the monotonous information – which accounts for 79% and 81% of the variance, respectively.

It seems as if the correlation between intelligibility and similarity judgments is somewhat stronger when the judgments are based on monotonized speech samples than when the stimuli contain full tonal information. It might be the case, therefore, that removing a source of linguistic variability from the input stimuli (in this case tonal variation) allows the listeners to focus better on the remaining linguistic features.

3.4 Mutual intelligibility within and between Mandarin and non-Mandarin groups

In Chapters One and Two, we mentioned the persistent claim found in the literature that Mandarin dialects are mutually much better understood than are Southern (non-Mandarin) dialects. Also, it was claimed that Mandarin dialects are more intelligible to Southern (non-Mandarin) listeners than vice versa. These impressionistic claims can now be experimentally tested.

In Figure 3.6 I plotted the mean judged intelligibility (left-hand panel, A) and mean judged similarity (right-hand panel, B) for three groups of listeners.

The first group consists of listeners who listen to speakers of only their own dialect; this is the mean of the scores found along the main diagonal of the matrices in Tables 3.2 and 3.4. Both judged intelligibility and judged similarity are very high and close the maximum possible score of 10. There is no difference between the scores obtained by Mandarin and non-Mandarin (Southern) speakers.

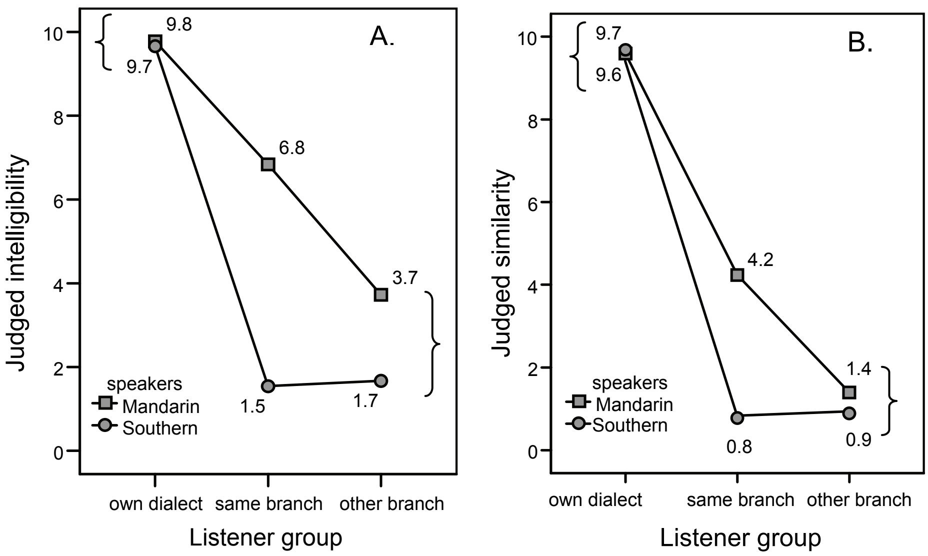

Figure 3.6. Judged intelligibility (on a scale from 0 to 10, panel A) and judged similarity (panel B) of Mandarin and non-Mandarin (Southern) speakers, as judged by listeners of the same dialect as the speakers ('own dialect'), by listeners of another dialect within the same branch ('same branch'), and by listeners of a dialect in the branch opposed to that of the speaker ('other branch'). Braces enclose means that do not differ from each other by a Scheffé test ($p < .05$).

The second group of listeners respond to dialect samples spoken in another (but not their own) dialect within the same main branch, i.e. Mandarin listeners respond to other Mandarin speakers, and non-Mandarin (Southern) listeners react to speakers of other non-Mandarin (Southern) dialects only. Here we see a very clear difference between Mandarin and non-Mandarin (Southern) speakers. A Southern (non-Mandarin) speaker is judged to be practically unintelligible by other non-Mandarin listeners (1.5 on the 10-point scale) whilst the Mandarin dialect speakers receive a mean intelligibility judgment close to 7 by other Mandarin listeners. The same effect is observed in the similarity ratings. The third group contains listeners who are exposed to speakers of dialects in the opposed main branch, i.e. Mandarin listeners responding to non-Mandarin (Southern) dialects, and vice versa. Now, we see that non-Mandarin (Southern) speakers are as unintelligible to Mandarin listeners as they were to other non-Mandarin (Southern) speakers. Also the non-Mandarin (Southern) speakers' dialects are judged to be as different from the listeners dialect by non-Mandarin (Southern) and Mandarin listeners alike. Mandarin speakers, however, are considered to be somewhat intelligible by non-Mandarin (Southern) listeners (3.7), although this score does not differ significantly from the 1.7 score in the opposed situation. In terms of judged similarity there is no difference: non-Mandarin (Southern) listeners consider the Mandarin speakers' dialects as different from their own (mean similarity rating of 1.4) as vice versa (Mandarin listeners responding to non-Mandarin/Southern speakers, with a mean similarity rating of 0.9).

A two-way Analysis of Variance with speaker dialect group (Mandarin, Southern) and listener group (own dialect, other dialect within same branch, dialect in other branch) as fixed factors reveals significance for both factors as well as for the interaction between the two, $F(1,219) = 37.3$ for the main effect of speaker group, $F(2,219) = 78.6$ for listener group, and $F(2,219) = 19.7$ for the interaction ($p < .001$ in all cases). Means in Figure 3.6 that are enclosed by the same brace, do not differ from each other by a post-hoc Scheffé test ($p < .05$). Approximately the same effects were found when judged similarity was the dependent variable, $F(1,219) = 20.4$ for the main effect of speaker group, $F(2,219) = 28.1$ for listener group, and $F(2,219) = 28.5$ for the interaction ($p < .001$ in all cases).

3.5 Conclusions

This chapter presented a first attempt at determining the degree of intelligibility between pairs of Chinese dialects from a set of 15. Six of the dialects have traditionally been classified as Mandarin dialects, the other nine are non-Mandarin (or: Southern) dialects. According to Chinese dialectologists (and lots of anecdotal evidence) Mandarin dialects are mutually intelligible to a much higher degree than are non-Mandarin (Southern)dialects. Also, Mandarin dialects are claimed to be more intelligible to non-Mandarin (Southern) dialect listeners than vice versa. The asymmetry may have language-internal and language-external causes. It may indeed be true that it is inherently easier for a non-Mandarin (Southern) dialect listener to recognize words and phrases in Mandarin than vice versa (language internal reasons) but it may also be that exposure to Standard Chinese, which is quite similar to Beijing dialect, through media and education, gave Mandarin varieties an advantage in our study. We will not try to disentangle these competing explanations in the present chapter (this matter is deferred to Chapter Six). Here, it is sufficient to note that the asymmetry has been claimed and then establish to what extent our results confirm the basic correctness of the impressionistic claims. From such comparisons we may also answer the question to what extent traditional taxonomies for Chinese dialects, as constructed by linguists, are reflected in our subjects' judgments of similarity and intelligibility among the dialects. The third question we wish to answer is whether listeners' judgments (on either dialect similarity or intelligibility) are sharper when tonal information is included or excluded from the auditory stimuli. Finally we consider the relationship between similarity and intelligibility judgments and ask ourselves to what extent these variables basically measure the same thing.

3.5.1 Asymmetry between Mandarin and Non-Mandarin dialects

Our results, presented in Figure 3.6A-B, first of all confirm the impressionistic claim that Mandarin dialects have greater mutual intelligibility than non-Mandarin (Southern) dialects. In later chapters (e.g. Chapter Six) we should expect to find that Mandarin dialects differ less from each other in terms of their vocabularies and sound systems than non-Mandarin (Southern) dialects do. Our results also show that, indeed, it is easier for listeners of non-Mandarin (Southern) dialects to understand Mandarin speakers than vice versa. The asymmetrical relationship is seen in the similarity and intelligibility judgments alike. As said before, the asymmetry may have language-internal as well as language-external causes, but before we try to establish the relative importance of these disparate causes (in Chapter Six) it is important that we establish that the asymmetry exists. The present chapter has done just that.

3.5.2 Convergence with linguistic taxonomy

From our experimental data we generated four agglomeration trees, each of which can be compared with traditional taxonomies postulated by Chinese dialectologists. The four trees (based on judgments of similarity and of intelligibility, with and without tonal information included in the stimulus speech samples, all lead to the same number of incorrect (or discrepant) classifications. In each tree a predominantly Mandarin group of dialects could be distinguished from a uniformly Southern group. In each case there were two incorrect (or at least discrepant) classifications: Nanchang and Changsha, which should be non-Mandarin (Southern) dialects, were incorrectly grouped with the (predominantly) Mandarin cluster. Geographically, these two dialects are located at a transition area, so that language contact may influence the mutual intelligibility and may explain the incorrect classifications.

The trees based on similarity judgments are somewhat more congruent with the linguistic taxonomy. Here the primary split between the upper and lower branch in the tree coincided with the optimal split between Mandarin and non-Mandarin (Southern) dialects. In the trees based on intelligibility judgments the (predominantly) Mandarin group was a subcluster ('relaxed grouping criterion'), leaving a non-integral set of non-Mandarin (Southern) dialects.

There are no systematic effects in the trees that can be related to dialect subgroups, with the exception of a tendency for Wu dialects to form a coherent two-member subcluster (Suzhou, Wenzhou). This behavior was observed in the trees based on similarity but not in the intelligibility trees. Although the differences are marginal, this might indicate that linguistic taxonomy is somewhat more closely related to laymen's intuition on similarity between the dialects than to their ideas on intelligibility.

3.5.3 Effect of tonal information

We have seen in the preceding sections that the effects found for stimuli with and without tonal information were largely the same. The high degree of convergence is also born out by the correlation coefficients that can be computed between intelligibility judgments based on samples with and without tonal information (r = 0.946, N = 225, $p < 0.001$) and for similarity judgments with and without tonal information (with exactly the same r = 0.946, N = 225, $p < 0.001$). Both in the case of intelligibility and of similarity judgments, the responses to versions with and without pitch information share 90% of their variance.

Figure 3.7 presents the mean scores for Intelligibility and Similarity judgments, for versions with and without tonal information. The data have been broken down according to the same organization as in Figure 3.6.

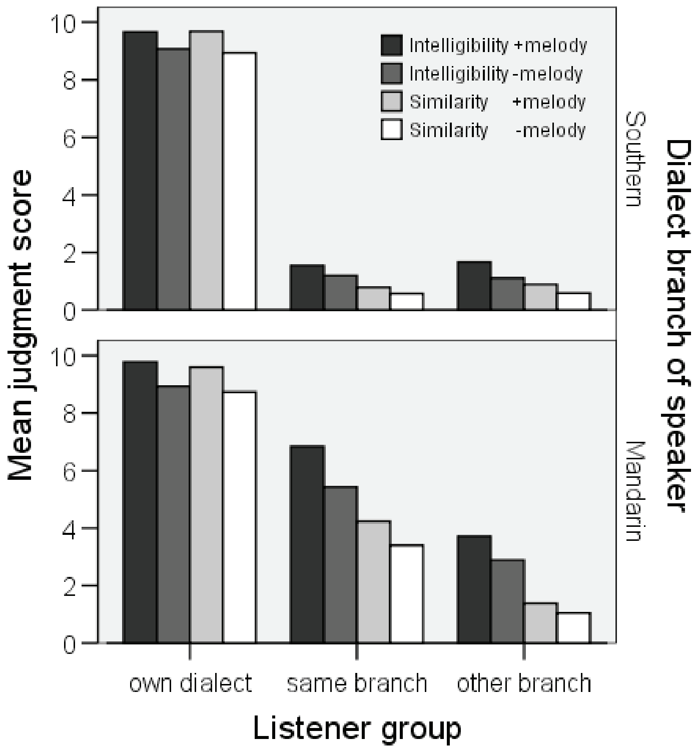

Figure 3.7. Mean judgments for intelligibility and similarity for speech samples with and without melody, broken down for Mandarin and Non-Mandarin (Southern) speakers and broken down further for three types of listeners: (i) listeners listening to speech samples of their very own dialect, (ii) listeners listening to samples of a dialects belonging to the same dialect group (Mandarin and Southern listeners listening to Mandarin and Southern dialects, respectively), and (iii) listeners listening to dialect samples belonging to the other dialect group (i.e. Mandarin listeners responding to Southern dialects and vice versa).

The point I want to make is that there is a small but systematic overall effect showing that judgments based on versions without pitch information tend to be lower than the corresponding judgments based on versions with full pitch information.[41] This

[41] Gooskens & Heeringa (2004) do not present the similarity judgments obtained for their mono-tonized Norwegian dialect samples. Yet, they do report that similarity judgments based on the non-manipulated speech samples can be predicted somewhat more successfully from objective

difference is, in fact, significant by paired t-tests, t(224) = 9.8 ($p < .001$) for intelligibility judgments and t(224) = 6.4 ($p < .001$) for similarity judgments. We interpret this effect as indicating that our listeners consider (artificially) monotonized speech as being more different from their own dialects, and less intelligible, than the same speech with natural melody, even if the natural melody deviates from expected tone patterns). In the final main section of this chapter we will test to what extent the degraded speech quality that resulted from the signal manipulation needed to monotonize speech samples contributed to this overall effect on the judgments.

3.5.4 Similarity versus intelligibility judgments

Our results bear out that judged similarity and judged intelligibility can be predicted from each other with considerable success (r^2 between 66 and 81%). Although this is a high degree of convergence, it does not mean that similarity and intelligibility are one and the same thing. There is still some 20 to 35% discrepancy between the two. It may well be the case, for example, that our listeners are well aware of substantial differences between their own variety and Beijing dialect but understand the Beijing (and related Mandarin dialects) rather well as it resembles the standard language. We will come back to this issue in Chapter Six, when we consider the question to what extent similarity and intelligibility judgments can be predicted from objective structural differences between pairs of dialects.

3.6 Testing possible artefacts of sound quality – a control experiment

3.6.1. Introduction

In the collection of our experimental data we presented speech samples to our listeners such that each of the Chinese language varieties was represented by a single speaker, either male or female. Individual speakers may differ in intelligibility even if they speak the same language variety. Moreover, some of our speakers were subjected to more extreme PSOLA manipulations than others, not to mention the fact that the four female speakers underwent a (digital) gender transformation. As a result of differences in individual voices and effects of subsequent manipulations, the sound quality of some dialect samples may have been better than that of others. In order to check whether differences in sound quality may have been of influence on the judgment of intelligibility, both by within and by across-dialect listeners, I ran a control experiment.

I collected perceived judgments of the sound quality of the 2 × 15 dialect samples (one monotonized, one with full melody) in the abstraction of intelligibility. This was done

(Levenshtein) distance measures, than the judgments based on monotonized samples. They also report that the range of judgment scores based on monotonized samples is somewhat compressed relative to that obtained for the original samples. In our data, the range of monotonized scores is also compressed but only at the high end of the scale (i.e. asymmetrical compression), indicating simply that monotonized samples are judged to be less intelligible and less similar to the listener's own dialect.

by playing the samples to listeners who did not understand a word of Chinese, in whatever dialect. Such listeners were easy to find at Leiden University among students and colleagues in LUCL (The Leiden University Centre for Linguistics), native speakers of various European languages with no working knowledge of Chinese. We reasoned that if no correlation could be established between the mean judgments of sound quality of dialect samples and their judged intelligibility – either within or across dialects – then our listeners must have been able to properly abstract away from actual recording and sound quality, and base their intelligibility judgments on some measure of linguistic distance.

3.6.2. Procedure

Twenty-five fellow linguists or students of linguistics at the Leiden University Centre for Linguistics individually listened to the 30 speech fragments, which were truncated after 10 seconds (in the next pause at an utterance boundary). Listeners were native speakers of Dutch, English, Greek, Italian, Polish, Hungarian or German; none had any working knowledge of any Chinese language. Ten listeners were female, the other 15 were male.

Stimuli were presented to the listeners individually in their own office or home through an internet application, in different random order for each listener. Listeners were instructed to judge the quality of the sound samples by imagining how intelligible the fragment would be if they were a native listener of the same language that was spoken in the fragment. The quality judgment was given on an 11-point rating scale, where 0 stood for 'extremely poor sound quality' and 10 represented 'perfect sound quality'.

3.6.3. Results

Figure 3.8 presents the mean quality judgments for each of the fifteen sample varieties, separately for versions with and without melodic information. The varieties are arranged from left to right in descending order of judged quality (full-melody version).

Figure 3.8 shows that the listeners clearly differentiated between samples of better and poorer sound quality. The intonated versions were judged between 8 and less than 5 on the rating scale, while the monotonized versions of the same samples were judged to have poorer sound quality by 3 to 4 points on the scale. The effect of melody is highly significant by a repeated measures Analysis of Variance, with melody and speaker dialect as within-subject factors, $F(1, 24) = 168.1$ (Huynh-Feldt corrected, $p < .001$). The effect of dialect is smaller, but still highly significant, $F(14, 258.8) = 10.8$ (Huynh-Feldt corrected, $p < .001$), as is the interaction between the factors, $F(14, 262.9) = 7.8$ (Huynh-Feldt corrected, $p < .001$). The interaction would seem to be due to the fact that the range of rating values is somewhat compressed for the monotonized stimuli.

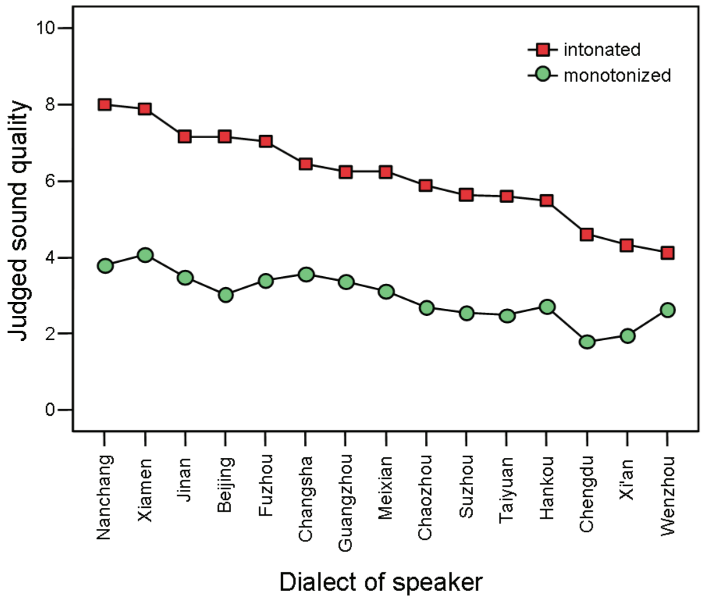

Figure 3.8. Judged sound quality of 15 samples of Chinese language varieties broken down by melodic version.

In spite of the interaction, it is quite clear from Figure 3.8 that there is a rather strong correlation between the intonated and monotonized versions. This is corroborated by a correlation coefficient of r = .884 (N = 15, p < .001) between the pairs of ratings obtained for the 15 dialect samples.

Crucially, we want to answer the question if the intelligibility judgments obtained in the main experiment can in any way be explained by observed differences in sound quality of the stimulus samples. To answer this question we first computed the correlation between judged sound quality and the intelligibility rating obtained from native listeners of the dialect samples only. Each group of 24 dialect listeners judged the intelligibility of two samples of the 'own' dialect, i.e. monotonized and with full melody. There were considerable differences in mean intelligibility ratings (Tables 3.3 and 3.4, respectively), which may have been caused by differences in sound quality of the samples. The correlation between sound quality and intelligibility, however is r = .205 (N = 30, ins.). The correlation is even poorer when computed for the two melodic versions separately, r = .189 (N = 15, ins.) for intonated samples and r = .098 (N = 15, ins.) for monotonized versions.

Correlations computed between judged sound quality and the overall intelligibility judgments across all listener groups are practically zero, r = .019 (N = 15, ins.) for monotonized versions, or even negative, r = −.195 (N = 15, ins.) for intonated samples and r = −.204 (N = 15, ins.) across both melodic versions.

As a final demonstration of the non-effect of sound quality, I present in Figure 3.9 the intelligibility ratings for all 30 samples given by Beijing listeners plotted against judged sound quality, and broken down for monotonized and intonated versions. This demonstration is the most convincing as the intelligibility ratings are not compromised by uncontrolled experience with the Beijing (i.e. Standard Mandarin) dialect. This selection of judges are native listeners of Beijing, and have no experience whatsoever with any of the other 14 dialects.

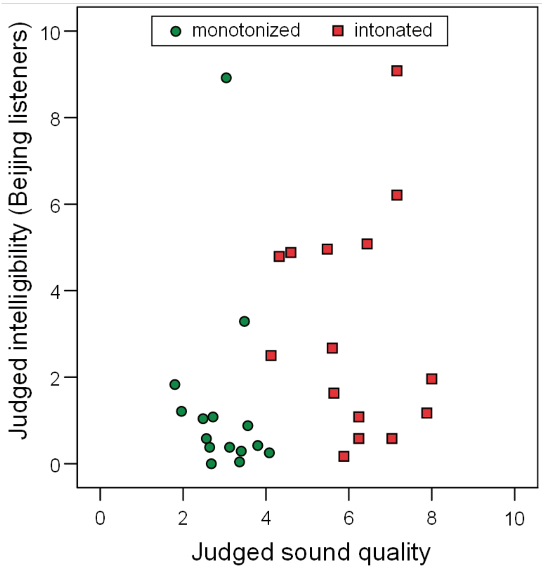

Figure 3.9. Intelligibility judgments by Beijing listeners only plotted against judged sound quality of the dialect samples, broken down by intonated and monotonized versions.

The figure shows quite clearly that there is not the slightest correlation between intelligibility and sound quality. We may therefore safely conclude that the Chinese listeners in our data collection were fully able to make their intelligibility and similarity judgments in the abstraction of the actual sound quality of the recordings they were exposed to.

Chapter Four

Mutual intelligibility of Chinese dialects: Functional tests

4.1 Introduction

In the preceding chapter (Chapter Three), we tested the mutual intelligibility between pairs of selected Chinese dialects through opinion judgment experiments.

We used opinion tests to obtain judged similarity and judged intelligibility of our 15 Chinese dialects by asking naïve raters for their intuitive judgments after listening to readings of the fable *The North Wind and the Sun* spoken in these dialects. The test results were based on 24×15 listeners' judgment scores for the passages read by 15 dialect speakers in both melodic and monotonized versions on the 11-point scale. The results (based on the lower triangle part of the matrix, i.e. the mean similarity of cross-diagonal cells, excluding diagonal cells) showed that judged similarity and judged intelligibility are highly correlated at $r = .888$ for the melody information and $r = .900$ for the monotonized information – which accounts for 79% and 81% of the variance, respectively (sign. at $p < .001$; Tang & Van Heuven, 2007; also § 3.2 in Chapter Three).

The dendrogram trees generated from the collected judgment scores were used to correspond with the traditional dialect taxonomy and no perfect reflection was found. Basically, by using relax criterion, some trees correctly reflected the primary split between the Mandarin and non-Mandarin (Southern) branches but not with the internal structures of the two branches. The sub-groups or internal clusters were not correctly classified. We cannot decide that the opinion tests can be used as a short-cut to establish the mutual intelligibility for Chinese dialects. The high correlation coefficient between the judgment similarity and judgment intelligibility indicated that it is hard to decide which test can be used as a substitute for mutual intelligibility testing (see Tang & Van Heuven 2007, Chapter Three). In this present chapter, I will test the true mutual intelligibility between pairs of Chinese dialects through functional tests.

Functional intelligibility scores were collected at the word and sentence level. I will compare these with each other and with opinion scores obtained earlier for the same set of 15 languages (reported in Tang & Van Heuven 2007; cf. Chapter Three). I will then decide to what extent opinion scores may serve as an acceptable substitute for functional intelligibility testing. In order to do so we will evaluate the functional and opinion scores against traditional dialect taxonomies proposed by Chinese linguists.

In Chapter One, I introduced two experimental approaches to test mutual intelligibility. When the selection of language pairs exceeds a certain number, (e.g. ten, yielding the 90 possible combinations for comparisons between dialects), the opinion testing method had better be used as shortcut for measuring. However, for this moment, it is undecided which subjective measure (opinion test or functional test) is more advantageous over the other one. In this chapter, I will aim to determine the degree of mutual intelligibility between pairs of Chinese dialects through functional experiments. I will correlate the two types of subjective results with each other. I will then decide to what extent opinion scores may serve as an acceptable substitute for functional intelligibility testing. In the next chapter (Chapter Five), we will consider various objective distance measures both from the published literature and computed by ourselves. Later in Chapter Six, I will correlate the two types of subjective results from the experiments with all the objective structural measures obtained in Chapter Five to see how well we can predict mutual intelligibility from various objective distance measures and which objective distance measure(s) contribute best to the prediction. Chapter Seven will give the conclusions.

4.2 Functional Experiments

In addition to the opinion tests, I also want to know to what extent *mutual* intelligibility of our 15 target Chinese dialects can be determined by functional tests. Furthermore, we want to find out what is the correlation coefficient between functional test scores and the scores obtained from the earlier opinion tests? How much do the functional and judgment results overlap with or deviate from traditional dialect taxonomy? Therefore, we aim at functionally testing how well a listener of language variety X actually understands a speaker of language variety Y (and vice versa). Specifically, we are interested in the percentage of words correctly translated from variety X to variety Y and vice versa. In order to obtain experimental data, we designed two tests: one is a so-called word-intelligibility test (counting the percentage of recognition/translation at the level of isolated words), the other is called sentence-intelligibility (testing the percentage of recognition/translation at the sentence level).

The word-intelligibility test was developed from scratch, and affords fast and economical testing the recognition of a large number of isolated words. Target words are not translated; instead recognition is tested through semantic multiple-choice categorization. Listeners are required to indicate to which of ten pre-given semantic categories a spoken word belongs. For instance, if the listener hears the word for 'apple', s/he should categorize it as a member of the category 'fruit'. Here, the assumption is that correct categorization can only be achieved if the listener correctly recognizes the target words. Since there are as many as ten semantic categories, the role of guessing should be negligible.

Word recognition in sentence context was tested by a Chinese version of the SPIN ('Speech Perception in Noise') test, which was originally developed for English by Kalikow, Stevens & Elliott (1977). In the SPIN test the listener has to write down only the last word in a number of short spoken sentences. In the materials I used, the identity of the final word was largely predictable from the earlier words in the sentence, so that this test addresses the efficient interaction of bottom-up (information from the speech signal) and top-down (expectations derived from earlier context) processes in continuous speech recognition. Earlier work has shown that this type of test is highly sensitive to differences in intelligibility due to different language backgrounds of speakers and listeners (Wang 2007).

One additional question that we hope to answer on the basis of the present chapter, is to what extent the recognition of isolated words (bottom-up information only) and of words in context (interaction of bottom-up and top-down information) are predictable from each other. If recognition of words in context is largely predictable from isolated-word recognition scores, the latter type of test will suffice for future work on functional *mutual* intelligibility testing in the Chinese language area.

4.2.1 Methods

4.2.1.1 The recordings

4.2.1.1.1 Recording materials: word and sentence selection

For the word part, we prepared a Swadesh-like list of 288 'core' standard Mandarin words.[42] These words are frequently used in daily life forming such categories as body part, family member, plant, fruit, house furnishing, article of clothing, word for orientation in time and space, animal, etc. The words all denote simple concepts commonly used in everyday life and thus they are assumed to be used in each of our 15 target Chinese dialects. I tried to avoid words with the same morphemes (Standard Mandarin-orientated only) in order to obviate repetition or priming identity effects. It is well known from the literature that prior recognition of a word or stem morpheme greatly facilitates subsequent recognition of the same word or morpheme (e.g. Morton 1969, Murrel & Morton 1974, Nooteboom 1981, Cutler & Donselaar 2001).

For the sentence part, I selected 70 sentences based on the high-predictability section in the SPIN ('Speech Perception in Noise') test sentence lists. In the SPIN test listeners have to write down the final word (target) of each sentence they hear.[43] Getting the final word is easier as the listener also correctly recognizes the earlier words in the sentence, as in *He wore his broken arm in a sling* (target underlined). The seventy sentences were selected on the basis of their applicability to the Chinese linguistic/cultural situation, and translated into Standard Mandarin. I tried to select the sentences

[42] A Swadesh list is a prescribed list of basic vocabulary developed by Swadesh in the 1940-50s, which is used in glottochronology (lexicostatistical dating). For details on the method see Swadesh (1972).

[43] There are two types of materials in the SPIN test. I only used the part that presents target words that are highly predictable from the earlier context (H sentences). I did not use the part with words that are not predictable from the context (L sentences), as in *We could have discussed the dust* (target underlined). Wang (2007) showed that the H part of the SPIN test was more sensitive to differences between speaker and listener groups with different degrees of listening comprehension in English.

maintained the structure of the SPIN sentences such that each Mandarin sentence ended with in final noun as it does in English.

4.2.1.1.2 Sound recordings

Thirty speakers were recorded, i.e. one male and one female native speaker of each of the 15 target dialects. All of these speakers were students at Chongqing Jiaotong University, China. They were born and bred in the dialect region they represented. They had moved to Chongqing as young adults. They returned to their dialect area on a regular basis, for at least two months in the summer and six weeks in the winter season. In Chongqing they were part of fairly large dialect communities, and in most cases the male and female speaker representing a particular dialect were a couple who had continued to speak the dialect in their own home when in Chongqing. Also, when the recordings were made, the male and the female speaker pair spent considerable time together, speaking the dialect, in order to prepare the translations.

Before the recording sessions, the speakers translated 288 isolated target words and the 70 sentences from Standard Mandarin into their own dialects. The translation was done by pairs of speakers (one male, one female) for each dialect independently. In case of the divergence between the two translators in some expressions of a particular dialect, the alternative that both speakers agreed was most typical of the local vernacular, was selected.

Using Adobe Audition running on a notebook computer, the words and sentences were then read from paper and recorded by the 30 speakers in individual sessions. Speakers were seated in a quiet office and wore a Shure SM10A head-mounted close-talking microphone. The air conditioner (it was high summer time in Chongqing, P. R. China at the time the recordings were made) was temporarily switched off during the recordings. Each speaker was required to read both the word part and sentence part in their own dialect (instead of Standard Mandarin) using the translations they had prepared themselves.

4.2.1.2 Listening test

4.2.1.2.1 Data segmentation and processing

I firstly segmented and labeled each individual word and sentence in the recordings and saved these as separate wave files.

For the word part, I finally extracted 150 words in ten lexical categories (eight main categories, two of which were subdivided):[44]

---

[44] The 150 words had been selected from a larger set of 288 core words. The original set was not compiled for the purpose of constructing a semantic categorization task. As a consequence it was not always simple to find clearly distinct semantic categories that could be filled with 15 clear instantiations of that category.

1. Body parts
2. Plants
    a. Sweet fruits/nuts
    b. Vegetables
3. Animals
    a. Four-legged
    b. Other (animals)
4. Textiles/fabrics/articles of clothing, apparel
5. Orientation in time/space
6. Natural phenomena
7. Perishables (food/drinks other than fruits and vegetables)
8. Verbs of action/things people do.

Appendix 4.1 presents the list of 150 target words (in Mandarin only), in characters and in Pinyin (Romanized Mandarin phonological spelling plus tones), glossed and subdivided into the ten semantic categories.

For the sentence part, I made a further selection of 60 sentences (from the original set of 70). These sentences basically satisfied the condition of having a noun in final position in each of the 15 dialects (with only very few exceptions, in which case the target word was a prefinal noun in some specific dialects). A full list of sentences in (Standard Mandarin), in Chinese characters and in Romanized Pinyin glosses (plus tone numbers), and English translations is given in Appendices 4.2 and 4.3.

4.2.1.2.2 Creating CDs

The intelligibility tests basically require word recognition. In word recognition tests it is imperative that a listener does not hear the same word (or morpheme) twice. A word (or morpheme) which is heard for the second (or third, fourth) time within an interval of up to a day, is recognized more successfully than the first time (e.g. Morton, 1969). In order to prevent such repetition or priming effects, the stimulus words and sentences have to be blocked over listeners, such that each listener hears each word only once, irrespective of the dialect of the speaker. Therefore, we worked out a completely balanced word and sentence stimulus order using a Latin Square design.[45] On the first CD (CD1) the 150 words were placed in a fixed random order (from nr. 1 to nr. 150). Every following word was spoken in a different dialect, so that every dialect was represented by 10 words. On the second CD (CD2) the words were presented in the same order with the exception that the presentation began with word nr. 150 which was then followed by words nr. 1 to nr. 149. As a result of this shift, every word on CD2 was spoken in a different dialect than on CD1. On the third CD (CD3) the first item was word nr. 149, the second was nr. 150, followed by words nr. 1 to nr. 148, and so on for CDs 3 to 15. Again, every word on CD3 was spoken in a different dialect than on the earlier CDs. CD15 started with word nr. 137, followed by words nr. 138 to

[45] For a general reference to the use of Latin Square designs, see e.g. Box et al. (1978).

nr. 150, and then followed by words nr. 1 to nr. 136. Through this rotation scheme we ensured that (i) each listener heard each of the words and sentences only once, (ii) each of the 15 listeners in one dialect group heard each version of a word in a different dialect, while (iii) at the same time every listener heard one-fifteenth of the materials in each of the 15 dialects (stimuli were blocked over listeners in a Latin square design).

Note, finally, that it was not possible to divide the materials evenly between male and female speakers in each dialect, since 15 is an odd number. In order to solve this small imbalance, half of the dialects were represented by 8 male and 7 female speakers, whilst the other half of the dialects were represented by 7 male versus 8 female speakers.

In all, 225 CDs (15 copies of 15 different CDs) were produced. On each CD, the word part preceded the sentence part. Ten words or ten sentences formed a track, with a pause between words or sentences of 7 seconds and with 11-s pauses between tracks. As a consequence, each CD contained 28 tracks including spoken instructions at the beginning, in the middle and at the end plus practice tracks containing 10 words and 10 sentences, respectively. Practice items were sampled from materials that were not selected as proper stimuli.

4.2.1.2.3 Answer sheets

For each CD, we prepared an answer sheet to match the corresponding stimulus tracks. There were 15 blocks of word stimuli and six blocks of sentence stimuli. For each block of words ten stimulus words were required to be categorized into one of the designated ten semantic categories. The categories were listed across the page. Listeners were asked to tick the appropriate box for each successive stimulus. The categories were repeated after every ten lines.

For each block of sentences, the final (or incidentally prefinal) words for each of ten stimulus sentences had to be written down in the listener's own dialect.

4.2.2 Procedure

For each dialect in the set of 15, a local contact person was contracted. In ten cases the local contact had also served as one of the two speakers of the dialect materials I used as stimuli. In the case of five other dialects neither the male nor the female speaker could make a trip to their dialect area, in which case we asked another contact person, one whom we had used in our earlier study.[46]

[46] In the case of these five contact persons, there may have been a difference between the exact town or village of the speaker of the dialect sample and that of the listeners recruited by the contact person. Due to this circumstance five listener groups possibly may have listened not to their very own dialect but to a neighboring dialect within the same dialect group. These five dialects are Nanchang (Gan family), Fuzhou (Min family), Xi'an (Northern Mandarin), Taiyuan (Northern Mandarin), and Chengdu (Southwestern Mandarin). Results show that, indeed, these five listener groups got poorer scores when responding to their 'own' dialect than the other ten

Each of the 15 contact persons, a native speaker of the dialect of the listener group targeted, was instructed to enlist 15 listeners who were monolingual rural dialect speakers in the age bracket between 40 and 65 and who had not traveled much and had never lived outside their own province. Ideally, the listeners should be selected from the larger groups of 24 listener subjects who participated in our first experiment (Tang & Van Heuven 2007). For the present experiment, however, subjects had to be literate – so that some substitutions had to be made. All local contact persons and the listener subjects were paid for their services. Most listeners belonged to the lower working class with fairly low level of education and professions of low status. Listeners filled in a questionnaire asking them about their language background, familiarity with other Chinese varieties, and some demographic details. A summary of the responses to the questionnaire is given in Table 4.1. There was a roughly equal split between male (N = 115) and female (N = 110) listeners. The mean age was well above 40 for most dialects; the Nanchang listener group, however, had a mean age of 36. With very few exceptions (seven listeners out of 225, and never more than two in one dialect group) all listeners declared to be monodialectal. Nevertheless, a majority of the listeners claimed to be able to speak Standard Mandarin (63%, including the 15 Beijing listeners), and most listeners claimed to be able to understand Standard Mandarin to a greater or lesser degree. This may have implications for the interpretation of the results of this study. We will return to this issue in later sections (also see Tang & Van Heuven 2009).

Listeners took part in the experiment in individual sessions. Each listener in a dialect group listened to a different CD, one of the set of 15 CDs. All listeners were required to both read the paper instructions and to follow the spoken instructions (in Mandarin) on the CD. Stimuli were presented through twin loudspeakers in a quiet room, often in the contact person's private home, using either a computer or a stereo set.

The isolated word recognition task was presented first. Here, the listener was required to tick one of ten boxes for each word representing the ten semantic categories/ subcategories (see § 4.3.1) every time a word was presented. For the subsequent sentence part, the listener had to write down the final or pre-final target word(s) in their own dialect after listening to each of the 60 sentences on the CD. Whether the target word was in final or in pre-final position was indicated explicitly for each sentence on the listener's answer sheet.

---

groups did. (see Table 4.2, shaded cells). The mean word scores were 63 versus 55% correct, whilst the sentence scores were 89 versus 75%. The former difference is not significant by a paired t-test, $t(13) = 0.9$ ($p = 0.173$, one-tailed) but the latter is, $t(13) = 1.8$ ($p < .050$, one-tailed).

Table 4.1. Summary of listener characteristics broken down by dialect group. Mean and Standard deviation of age in years. N males = number of male listeners (out of 15). Education (highest level attained): 0 = none at all, 1 = primary school, 2 = junior middle school, 3 = senior middle school, 4 = vocational college, 5 = university undergraduate, 6 = university graduate. Dialects = number of dialects spoken. Understanding of Standard Mandarin: 0 = not at all, 1 = poor, 2 = moderate, 3 = good. Speaking Mandarin: 0 = no, 1 = yes. Double lines separate Mandarin from non-Mandarin dialects..

	Age					Standard Mandarin	
Dialect	mean	SD	N males	Education	Dialects	Understanding	Speaking
Suzhou	44.20	3.59	7	2.27	1.07	2.67	0.87
Wenzhou	45.67	3.83	8	1.47	1.13	1.93	0.73
Guangzhou	46.67	3.77	8	2.20	1.13	2.67	0.93
Xiamen	45.47	13.81	10	1.20	1.00	0.73	0.40
Fuzhou	47.53	5.58	8	1.60	1.00	1.93	0.53
Chaozhou	49.33	6.95	8	0.73	1.00	0.87	0.13
Meixian	47.93	6.97	9	2.10	1.00	2.44	0.44
Nanchang	36.33	7.68	8	2.07	1.00	2.73	0.87
Changsha	48.33	4.94	7	1.73	1.00	2.27	0.20
Taiyuan	44.07	5.71	5	2.33	1.00	3.00	0.80
Beijing	42.20	4.36	9	2.87	1.00	3.00	1.00
Jinan	51.20	4.11	7	2.40	1.13	2.73	0.33
Hankou	46.80	4.96	8	0.67	1.00	2.27	0.33
Chengdu	42.67	14.88	6	3.80	1.00	2.80	1.00
Xi'an	48.53	4.10	7	2.93	1.00	3.00	0.87

After the last of the 60 sentences had been presented, the local contact person translated the 60 response words into Mandarin in the presence of the listener, asking the listener for clarification whenever necessary.

4.2.3 Results

In all, I collected 33,750 responses (15 × 150 × 15) for the word part and another 13,500 (15 × 60 × 15) for the sentence part. The dependent variable in the word-intelligibility test was the choice of semantic category. This choice was coded with a value from 1 to 10 and entered in a database, along with information on the dialect of the listener, dialect of the speaker and on the semantic category of the stimulus word. The correctness of the listener's choice was evaluated automatically by having the computer check whether the semantic category of the response matched that of the stimulus. From this information we computed a mean percentage of correctly classified (recognized) words for each of the 15 × 15 combinations of speaker and listener dialects, yielding 225 mean word recognition scores (see Table 4.2).

For the sentence intelligibility test, the procedure was less straightforward. As a native speaker of Chinese, I manually checked whether the sentence-final (or pre-final) target word was correctly translated back into Mandarin by the local contact person. If the translation was semantically equivalent to the target specified for the item, the response was considered correct. If the translation was incorrect or if no translation was given at all, the response was considered an error. From these data we computed 15 × 15 = 225 mean sentence-intelligibility scores, i.e. one mean score for each combination of speaker and listener dialect (see Table 4.3).

I will now first describe the analysis of the results for the word intelligibility test (§ 4.4.1), and defer the presentation of the results of the sentence intelligibility test to § 4.4.2.

I tested the mutual intelligibility of 15 Chinese dialects functionally at the level of isolated words (word intelligibility) and the level of sentences (sentence intelligibility). I collected data for each dialect by playing isolated words and sentences spoken in 15 Chinese dialects to 15 listeners. Word intelligibility was determined by having listeners perform a semantic categorization task whereby words had to be classified as one of ten different categories such as body part, plant, animal, etc. Sentence intelligibility was estimated by having the listener translate a target word in each sentence into their own dialect.

I obtained 47,250 data (15 × 150 × 15 for the word part and 15 × 60 × 15 for the sentence part). I firstly analyzed the isolated word intelligibility results based on scores given by 15 listeners for each of the 15 dialects. With the assistance of SPSS, I generated the dendrogram tree based on the matrix of the word-intelligibility scores as we did in Chapter Three. The tree was then compared with the traditional Chinese dialect taxonomy and the classification errors were counted. The same procedure was repeated to analyze the sentence-intelligibility scores. As I did in Chapter Three, I also test the impressionistic claims about the mutual intelligibility within and between Mandarin and non-Mandarin (Southern) dialects based on these collected data. In order to find out to what extent the word-intelligibility and sentence-intelligibility converge or deviate each other and whether the two experimental method(s) of mutual intelligibility tests (opinion judgment tests or functional tests) correlate significantly with each other, I firstly correlate these two functional results with each other, then I compute the correlation coefficients between all the subjective measures (opinion scores and functional scores). Results were validated against the traditional dialect classifications proposed by Chinese linguists.

4.2.3.1 The results from the isolated word intelligibility test

Table 4.2 presents the mean percentage of correctly classified (recognized) words for each combination of speaker dialect (listed in the rows) and listener dialect (listed column-wise).

Overviewing Table 4.2, we find that all the scores are always better than chance (i.e., 10% – given ten alternatives to choose from), therefore, we claimed that there is always some degree of intelligibility between Sinitic dialects/languages.

Table 4.2. Percent correctly classified words broken down by 15 speaker dialects and 15 listener dialects. Each mean is based on 150 responses (each of 150 words is heard once, with 10 different words for each of 15 listeners). Total number of responses is 225 × 150 = 33,750. Double lines separate Mandarin from Southern dialects. Double lines separate Mandarin from non-Mandarin dialects.

Speaker Dialect (down)	Listener dialect(across)															
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an	Mean
Suzhou	**65**	20	25	17	21	15	23	22	23	29	26	29	39	28	29	27
Wenzhou	23	**41**	17	19	17	17	18	21	15	24	25	25	28	18	19	22
Guangzhou	23	18	**55**	25	25	29	40	21	19	33	34	33	38	25	29	30
Xiamen	20	14	23	**39**	19	25	19	19	12	18	19	25	26	17	16	21
Fuzhou	17	18	17	18	**47**	14	17	16	15	22	20	23	24	20	16	20
Chaozhou	18	12	23	22	23	**68**	15	10	15	23	27	29	24	24	23	24
Meixian	31	24	35	24	23	25	**67**	31	27	43	43	43	41	37	31	35
Nanchang	27	26	30	25	29	22	41	**37**	29	47	51	48	57	41	42	37
Changsha	31	22	31	24	31	20	34	31	**48**	47	49	47	60	38	43	37
Taiyuan	33	30	30	29	31	21	36	36	30	**57**	59	64	55	50	48	41
Beijing	64	41	63	45	53	38	61	51	54	76	**83**	74	72	65	70	61
Jinan	40	22	31	22	36	19	39	39	31	59	61	**80**	58	51	55	43
Hankou	37	29	33	28	41	22	42	33	35	63	59	67	**81**	53	47	45
Chengdu	28	24	30	32	35	19	49	36	38	62	59	61	70	**72**	56	45
Xi'an	47	36	43	27	35	23	48	43	47	63	64	67	65	55	**59**	48
Mean	34	25	32	26	31	25	37	30	29	44	45	48	49	40	39	

We expect the highest scores in the cells along the main diagonal in Table 4.2 (bolded). These are the scores obtained by listeners who listen to speakers of their own dialect. Scores in off-diagonal cells should be poorer, as these cell means are based on listeners listening to speakers of a different dialect. Indeed, generally we do find the highest correct classification scores in the diagonal cells. The highest percentage correctly classified words is between Beijing speakers and listeners; Beijing listeners correctly recognized 83% of the words spoken in their own dialect and classified them into the right categories, the listeners of Jinan and Hankou dialects recognized the speakers of their own dialects as high as 80% and 81% respectively. Other listener groups were less successful. For instance, Xiamen and Nanchang listeners could not understand the speakers of their respective dialects very well, given the mean scores of 39% and 37%, respectively. On two occasions, in fact, the native dialect listener groups were outperformed by one of the other groups. This is the case for the native Nanchang group, which is outperformed by no fewer than seven non-native listener groups; for the Xi'an group, which is outperformed by four non-native groups and for the groups of Changsha and Taiyuan, which are respectively outperformed by two other groups.[47]

The data in Table 4.2 were then used to generate a dendrogram, using the average linking method that we also used in our earlier report (Tang & Van Heuven 2007, also Chapter Three). As a first step in the procedure, the matrix was made symmetrical by averaging corresponding cells above and below the diagonal, i.e., the cell contents of every pair of cells *i, j* and *j, i* were averaged. The tree structure that was generated is displayed in Figure 4.1.

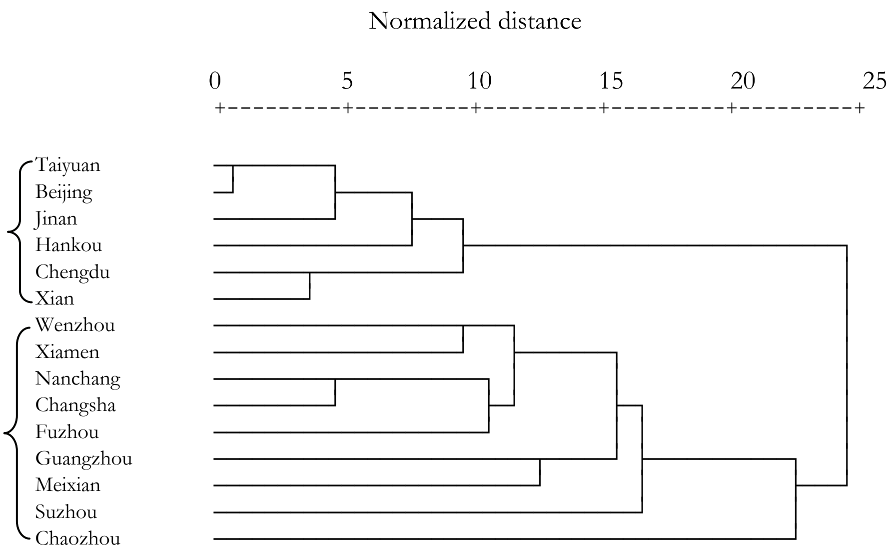

Figure 4.1. Dendrogram (using average linking between groups and Euclidean distance measures based on word intelligibility scores obtained for all 225 combinations of 15 speaker and 15 listener dialects.

The tree makes a primary split between the six Mandarin dialects, and a group of nine dialects which comprises all the non-Mandarin (Southern) dialects. This division concurs well with traditional taxonomies postulated by Chinese dialectologists. The dialects in the branches of Mandarin and non-Mandarin are braced. We will delay more detailed discussion of the internal cluster structure within the main branches until § 4.4.

[47] These four groups are among the set of five for which the speaker of the dialect materials did not hail from exactly the same town or village as the listeners (see Note 5).

4.2.3.2 Results from the sentence intelligibility test

Table 4.3 presents the results of the intelligibility test at the sentence level. Percent correctly translated target words are given for each combination of speaker and listener dialects.

Table 4.3. Percent correctly translated target words in sentences broken down by 15 speaker dialects and 15 listener dialects. Each mean is based on 60 responses (each of 60 sentence-final words is heard once, with 4 different words per dialect for each of 15 listeners). The total number of responses is 225 × 60 = 13,500. Double lines separate Mandarin from non-Mandarin dialects.

Speaker dialect (down)	Listener dialect(across)														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an	Mean
Suzhou	**77**	7	5	18	13	5	7	13	13	20	5	18	15	15	7	16
Wenzhou	5	**93**	5	12	3	2	7	10	2	7	2	10	8	7	2	10
Guangzhou	5	7	**92**	10	20	25	55	22	13	7	3	22	8	17	7	21
Xiamen	13	5	8	**97**	23	28	13	18	13	3	5	15	7	17	8	18
Fuzhou	3	3	2	17	**92**	7	3	8	5	0	0	7	2	0	3	10
Chaozhou	7	0	3	52	13	**98**	3	12	3	7	2	13	10	3	5	15
Meixian	13	2	12	28	17	20	**70**	25	18	10	3	25	15	25	8	19
Nanchang	28	13	20	25	27	17	33	**50**	32	35	18	53	43	37	23	30
Changsha	12	3	8	23	17	3	17	25	**93**	13	13	38	53	28	2	23
Taiyuan	63	35	45	63	57	25	55	68	68	**73**	77	92	92	85	73	65
Beijing	87	62	90	90	93	60	80	78	92	90	**98**	98	97	98	93	87
Jinan	52	27	32	48	48	15	40	60	70	75	77	**97**	83	82	67	58
Hankou	48	32	32	52	53	27	45	53	62	58	67	95	**100**	73	65	57
Chengdu	47	22	40	48	72	27	48	58	62	65	62	98	95	**95**	68	60
Xi'an	53	33	50	58	57	30	57	58	63	68	58	82	78	70	**67**	59
Mean	34	22	30	43	40	26	36	37	41	35	33	51	47	43	33	

Although the range of sentence scores is larger than that for semantic categorization (from 0 to 100%), the mean scores for own dialect are much better than that for semantic categorization (see the diagonals). It appears from the table that this sentence-level test was an easier task than the semantic categorization task with isolated words in the preceding section. The mean scores for the native dialect listener groups (listening to their own speakers) range between 50% and 98% correct (with a mean of 82%) for the nine non-Mandarin (Southern) dialects, and between 67% and 100% correct (with a mean of 88%) for the Mandarin dialect groups. The difference between the Mandarin and non-Mandarin (Southern) groups, however, fails to reach statistical significance, $t(13) = 0.745$ ($p = 0.469$).

On three occasions native listener groups are outperformed by non-native groups. This occurs in the Mandarin part of the table only, where native Taiyuan listeners happen to do as well as the Xi'an group and do more poorly than all other Mandarin groups. Chengdu native listeners do more poorly than two other groups, and Xi'an native listeners are second to four other groups.

Using the same procedure as in § 4.2.3.1, we generated a hierarchical cluster tree for the sentence-intelligibility results. The resulting tree structure is presented in Figure 4.2.

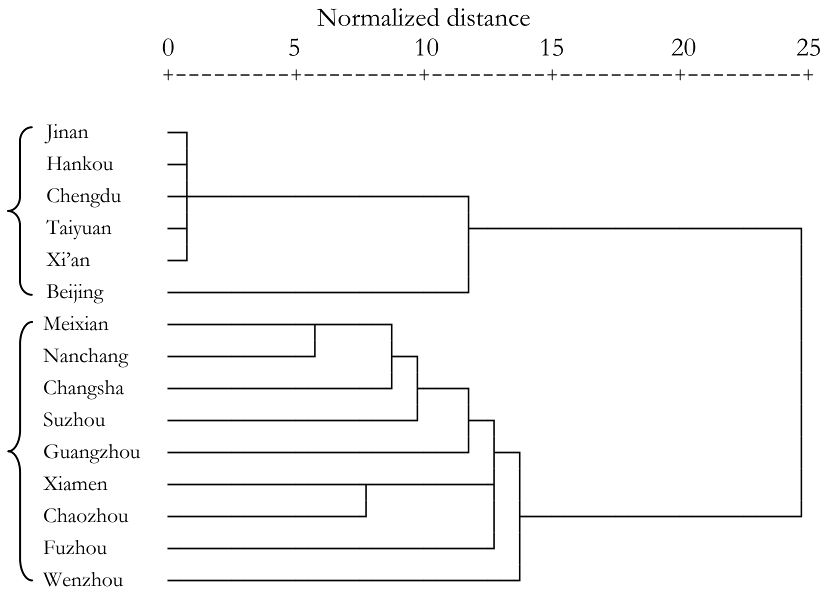

Figure 4.2. Dendrogram (using average linking between groups) and Euclidean distance measures based on sentence-level intelligibility scores obtained for all 225 combinations of 15 speaker and 15 listener dialects.

The sentence-level tree shows, again, a clean cut between the six Mandarin and the nine non-Mandarin (Southern) dialects.[48] As before, we will not deal with the internal structure of the dialects within the main branches. This matter will be taken up in § 4.4 where we will compare the clustering of the dialects in the trees above with the dialect taxonomy proposed by linguists. First, however, we will consider the question how well

[48] The agglomeration tree was generated by cases, i.e. using the speaker dialect as cases and listener dialects as variables. When the tree was generated from the variables, the classification of the dialects into Mandarin and non-Mandarin types was less successful. Different trees are generated from cases and variables when the similarity matrix is asymmetrical. For symmetrical matrices the difference does not apply.

the functionally determined word and sentence intelligibility scores can be predicted (in § 4.3) from our earlier judgment scores (on intelligibility and on linguistic distance).

4.2.3.3 Mutual intelligibility within and between (non-)Mandarin groups

In the Introduction part (Chapter One) we mentioned that Mandarin dialects are held to be more mutually intelligible amongst themselves than are non-Mandarin (Southern) dialects, and that Mandarin dialects are more intelligible to non-Mandarin (Southern) dialects than vice versa. Our results show that both impressions are borne out by the experimental data, both in terms of the word classification scores and of the sentence intelligibility test.

Tables 4.1 and 4.2 show that typically, listeners whose native dialect belongs to the Mandarin group were more successful in both the word classification and sentence translation tasks. The result is illustrated in Figure 4.4.

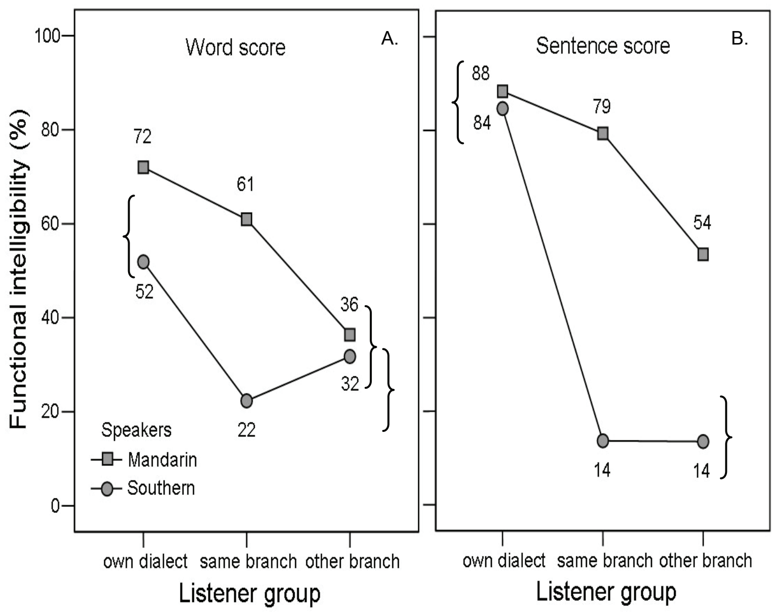

Figure 4.4. Intelligibility as a function of type of speaker dialect and of listener dialect at the word level (panel A) and at the sentence level (panel B). Braces enclose means that do not differ from each other by a Scheffé test ($p < .05$).

In the word classification task, the mean across the six Mandarin dialects is 72% correct, while the mean correct classification of the listeners with non-Mandarin (Southern) native dialects is 52%. The former is 20 percentage points better than the latter. These mean word recognition scores are summarized in Figure 4.4A. This figure also shows intelligibility of Mandarin and non-Mandarin (Southern) speakers for listeners within the same dialect group (i.e. Mandarin speakers and listeners not sharing the same dialect, and non-Mandarin/Southern speakers and listeners not sharing the same dialect) and for listeners in the other dialect group (Mandarin speakers and non-Mandarin/Southern listeners, or vice versa). A two-way ANOVA (analysis of variance) found the difference in mean percent correct between Mandarin and non-Mandarin (Southern) listener groups is highly significant, $t(13) = 3.1$ ($p = 0.008$, two-tailed).

The same result is found in word translation in sentence context. Figure 4.4B shows again, the *mutual* intelligibility is very good within the Mandarin dialects and very poor in the non-Mandarin (Southern) dialect branch. Non-Mandarin (Southern) dialects are as poorly intelligible to Mandarin listeners as they are to non-Mandarin (Southern) listeners. Mandarin speakers are fairly intelligible to non-Mandarin (Southern) listeners (54% intelligibility), and this effect largely remains even if we exclude Beijing speakers (48%). Main effects and the interaction in Figure 4.4B are significant, $F(1,219) = 165.1$, $F(2,219) = 94.8$ and $F(2,219) = 38.5$, respectively ($p < .001$).

Figure 4.4 shows that speakers and listeners within the Mandarin branch recognize many of the words in other Mandarin dialects (this is even the case when Beijing speakers are excluded, in which case the score is 59% instead of 61%, see discussion). Mandarin listeners get much poorer word recognition scores when listening to Southern dialects (36%). Southern listeners recognize a mere 22% of the words in other Southern dialects. They do, in fact, better on Mandarin dialects (32% correct; the same score is found even if we exclude Beijing speakers). A two-way Analysis of Variance with speaker dialect group (Mandarin, Southern) and listener group (own dialect, other dialect within same branch, dialect in other branch) as fixed factors reveals significance for both factors as well as for the interaction between the two, $F(1,219) = 120.1$, $F(2,219) = 61.8$ and $F(2,219) = 78.6$, respectively ($p < .001$ in all cases). Means in Figure 4.4 that are enclosed by the same brace, do not differ from each other by a post-hoc Scheffé test ($p < .05$).

One reason why Mandarin dialects are mutually more intelligible than are non-Mandarin (Southern) dialects could be that the former are intrinsically more similar than the latter. It is also the case however, that most Chinese listeners are familiar, through education and the media, with Standard Chinese, which is very similar to Beijing dialect. If we recompute the word and sentence intelligibility scores after eliminating the two Beijing dialect speakers, the results of our study are hardly affected, as we observed on several occasions in the previous section. The clearest way of testing the intrinsic greater similarity of Mandarin dialects is by including the scores obtained by Beijing listeners only (60 and 68% correct word and sentence scores) and comparing these with the scores obtained for the Southern listeners exposed to other non-Mandarin (Southern) dialects (22 and 14% correct recognition, cf. Figure 4.4 AB). Clearly, intelligibility of other Mandarin dialects for Beijing listeners is much better than *mutual* intelligibility within the non-Mandarin (Southern) dialects, $t(75) = 13.2$ and 11.0 for word and sentence scores, respectively ($p < .001$). These comparisons show that intrinsic linguistic similarity is a more important determinant here than the possible advantage of Mandarin dialects being close to the Standard language.

4.3 Correlations between subjective measures

So far we have obtained two kinds of subjective data experimentally, one is from the opinion tests, the other is from functional tests. Within the first type (see § 4.1, Chapter Three) we distinguish between judgments of (i) intelligibility and (ii) similarity between dialects. In the second type (the present chapter) we distinguish between functional intelligibility (iii) at the word level and (iv) at the level of the sentence. In the next sections we will consider the correlation structure in this set of variables. We will first examine, in § 4.3.1, the correlation between (iii) and (iv) on the basis of the data collected in the present study. In § 4.3.2 we will see to what extent the opinion scores are correlated with the functional scores.

4.3.1 Intelligibility at the word and sentence level

The results obtained from the word-intelligibility and the sentence-intelligibility tests presented above converge to a great extent. In order to quantify the degree of correspondence between the two methods of functional intelligibility testing, i.e. using isolated words versus words in sentence context, we established the correlation coefficient for all cells (including those on the main diagonal) in Table 4.2 and the corresponding cells in Table 4.3.

Figure 4.5 presents a scatterplot of the word (horizontal axis) and sentence-level (vertical axis) intelligibility scores. The correlation is high, viz. $r = .835$ (N = 225, $p < .001$) but clearly less than perfect: the coefficient of determination $r^2 = .697$ shows that 30% of the variance goes unaccounted for.

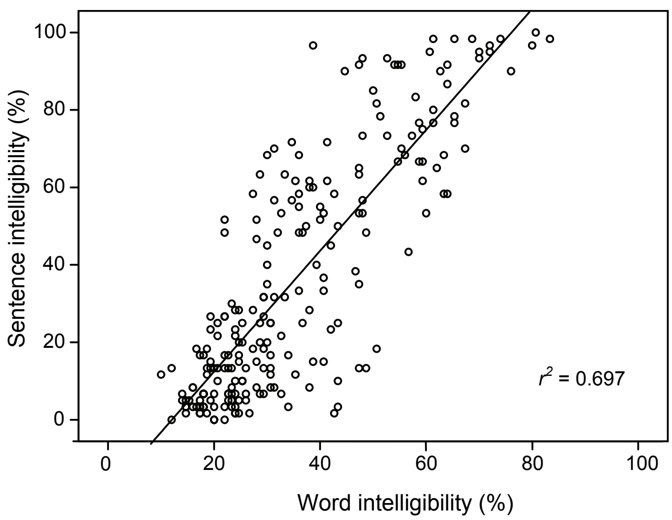

Figure 4.5. Scatterplot generated from mean scores based on the isolated word-level and sentence-level (full matrix).

In the introduction of Chapter One, we defined *mutual* intelligibility between two language varieties X and Y as the mean of the intelligibility of X to Y and of Y to X. Accordingly, we also computed the correlation coefficient for the word and sentence scores after averaging the contra-diagonal cells in the matrix (i.e., averaging the contents of every pair of cells {*i*, *j*} and {*j*, *i*}), which makes it a symmetrical matrix, of which only the non-redundant part ('lower triangle') is used in the computation of *r*. This procedure yields a higher correlation coefficient, $r = .928$ (N = 105, $p < .001$). The coefficient of determination is $r^2 = .86$, which means that still 14% of the variance is left unaccounted for.

It seems unclear, therefore, whether the word-intelligibility test (semantic categorization test) can be adequately used as a short-cut to functional intelligibility. For the moment we will assume that both the word-level and the sentence-level tests are needed. At some later stage, when we compare the test results with external data (objective measures of structural difference, traditional genealogies), we may be able to choose between the two and consider one a more valid or representative measure of intelligibility than the other.

4.3.2 Functional tests versus opinion tests

In our earlier study (Tang & Van Heuven 2007, Chapter Three), we collected opinion scores on intelligibility and on similarity between all pairs of our 15 dialects. The results revealed a very strong correlation between the two sets of opinion scores, especially when the correlation was computed on symmetrical matrices: $r = .949$ (N = 105, $p < .001$), which leaves only 10% of the variance unaccounted for (see also Table 4.6 in this paper).[49] From this, we drew the provisional conclusion that the two opinion scores are practically interchangeable. We will now determine to what extent the two sets of judgment data are correlated with intelligibility scores determined on the basis of functional test procedures. Obviously, opinion testing is much faster and easier to accomplish than functional testing. Therefore, if indeed the functional scores can be adequately predicted from the opinion scores, the latter type of testing will be preferred in the future – for reasons of economy.

Table 4.4 presents a correlation matrix for the four subjective measures at issue. From a range of opinion scores, I only selected the judgment scores obtained for the better sound quality (Tang & Van Heuven (2007). These were the opinion scores for

[49] When I used the symmetrical matrix of the input based on the melodic and monotonous versions, I obtained correlation values of $r = .888$ and .900 respectively (which leave 21% and 19% of the variance unaccounted for, respectively). If the complete, asymmetrical matrices are used to compute the correlation between judged intelligibility and judged similarity, I obtain $r = .854$ (N = 225, $p < .001$). This is still a high correlation but it leaves 27% of the variance unaccounted for (see Table 4.4).

intelligibility and similarity based on readings of the *North Wind and the Sun* fable with melodic information left unaffected.[50]

Table 4.4. Correlation coefficients of the four subjective measures, computed on the full matrices (N = 225). (the judgment data are based on the melodic versions only)

	Word intelligibility	Sentence intelligibility	Judged intelligibility
Sentence intelligibility	.835*		
Judged intelligibility	.722*	.772*	
Judged similarity	.652*	.692*	.854*

* Correlation is significant at the 0.001 level (2-tailed).

I computed the correlation coefficients for all six combinations of the four functional and opinion scores. I did this in three different ways. First (Table 4.4), we computed the coefficients on the complete (asymmetrical) data matrices (all 225 pairs of cells). It may be objected that scores in cells that lie on the main diagonal (where the listeners share the language of the speakers), will always be much better than scores on off-diagonal cells (where speakers and listeners have different language backgrounds). This might unduly boost correlation coefficients, so we produced a second set of correlation coefficients (Table 4.5) after excluding cells on the main diagonal. The third set of correlations (Table 4.6) was computed on the non-redundant parts (lower triangles) of the matrices after they had been made symmetrical by averaging the contents of all contra-diagonal cells $\{i, j\}$ and $\{j, i\}$.

Table 4.4 reveals three effects. First, the highest correlation coefficients are found between variables of the same type. That is to say, correlations between two opinion scores or between two functional test scores are better than correlations for cross-type test scores (from functional to opinion score or vice versa). Second, Table 4.4 shows that intelligibility judgments are a better predictor of the functional test scores than similarity judgments are. Third, functional intelligibility at the sentence level can be somewhat better predicted from opinion scores than at the word level.

Let us now see, in Table 4.5, if the same effects re-appear when we eliminate combinations of speaker and listener groups who share the same dialect. Table 4.5 shows that, indeed, all correlations are somewhat lower than in Table 4.4, but the relative differences are unaffected.

[50] Judgment scores were generally lower, and less clearly structured, when monotonized versions of the fable were presented (using PSOLA analysis and resynthesis). The data on monotonized versions are omitted from the present chapter.

Table 4.5. Correlation coefficients of the four subjective measures, using asymmetrical data matrices (N = 210) excluding combinations of speaker and hearers sharing the same language variety. (the judgment data are based on the melodic versions only)

	Word intelligibility	Sentence intelligibility	Judged intelligibility
Sentence intelligibility	.820*		
Judged intelligibility	.681*	.723*	
Judged similarity	.608*	.639*	.810*

* Correlation is significant at the 0.001 level (2-tailed).

The last step in the analysis is to consider, in Table 4.6, the correlation coefficients obtained after averaging the intelligibility scores for contra-diagonal cells (representing *mutual* intelligibility rather than just intelligibility).

Table 4.6. Correlation coefficients of the four subjective measures, using the symmetrical means (lower triangle, N = 105) of the matrix. (the judgment data are based on the melodic versions only)

	Word intelligibility	Sentence intelligibility	Judged intelligibility
Sentence intelligibility	.928*		
Judged intelligibility	.772*	.818*	
Judged similarity	.738*	.779*	.888*

* Correlation is significant at the 0.001 level (2-tailed).

A general effect seen in Table 4.6 is that all correlation coefficients are better than in the earlier two tables. Probably, the averaging over contra-diagonal cells eliminates some noise from the data, so that the correlation coefficients are improved. The results in Table 4.6 should be compared with those of Table 4.5, since in both these tables diagonal cells were excluded. Crucially, in spite of the overall boost of the correlation coefficients, the same three effects that were observed in Tables 4.4 and 4.5, are found in Table 4.6 as well.

4.4 Discussion

Observing the tree structures generated from the mean scores obtained from the judgment (opinion) and functional tests of *mutual* intelligibility, we found no perfect reflection of traditional taxonomy for Chinese dialects proposed by dialectologists.

In the trees based on opinion scores (see Tang & Van Heuven 2007, Chapter Three) we see that Changsha and Nanchang dialects are consistently wrongly parsed as Mandarin members – whereas actually they are traditionally classified into the non-Mandarin (Southern) branch. However, in the two tree structures based on the functional tests, Changsha and Nanchang are correctly classified as non-Mandarin (Southern) dialects and consistently go together, i.e., they make up an identifiable sub-cluster in both trees. A survey of the traditional literature on Chinese dialectology indicates that the two dialects belong to different dialect groups, viz. Xiang, and Gan, respectively. The other sub-groups in the Mandarin branch and in the non-Mandarin (Southern) branch are not perfectly reflected either, especially in the word-intelligibility tree.

These results may seem a little disappointing at first sight. However, we have to recognize the fact that the traditional dialect classifications are based on the characteristics of phonological (including tonal) differences between Chinese dialects. The degree of mutual intelligibility between pairs of these dialects (groups) was never tested. The results of our *mutual* intelligibility experiment tests not only established the degree of mutual intelligibility between dialects but also reflected their intrinsically phonological differences within the dialect subgroups to some extent as well. In the next chapter, we will compute various linguistic distance measures (lexically and phonologically) based on different sources of databank and use them to predict the mutual intelligibility of Chinese dialects concerning the traditional dialect taxonomy.

First, we find one early cluster for Nanchang and Changsha in both functional trees, although these two represent different groups (Gan and Xiang, respectively). This means the two dialects are more mutually intelligible than other dialects, which can be reasonably explained. These two dialects have more contacts because of their geographic proximity.

But why were these two dialects judged as dialects in the Mandarin branch in the opinion trees? We know that Nanchang and Changsha dialects are at a transitional area between the Mandarin and the non-Mandarin (Southern) branches. They are spoken in an area surrounded by Mandarin and non-Mandarin (Southern) dialects (see figure 3.1 in Chapter Three). As a consequence, they might share some features with Mandarin and others with non-Mandarin (Southern) dialects, so that their classification is unstable. Norman (1999) used vocalism as a basis for dialect classification and found that New Xiang (Changsha) shows some Mandarin characteristics while Old Xiang (Shuangfeng) which preserved voiced obstruent initials to some degree are inconsistently either like Mandarin or Gan dialects. Norman claimed that the status of Xiang as a separate dialect group should be reevaluated. In our earlier study Mandarin (Beijing) listeners claimed they could understand at least 50% Changsha dialect but almost nothing about Nanchang dialect. (whilst Nanchang and Changsha listeners indicated they could understand Beijing dialect very well, 94.6% and 95.4% respectively). The present functional intelligibility tests, especially the word-intelligibility test, show that Nanchang and Changsha are, in fact, relatively easier than other non-Mandarin (Southern) dialects to understand for Mandarin (Beijing) listeners (51% for Nanchang and 48% for Changsha respectively).

Second, within the Min group, the internal difference and uniformity are reflected by the mutual intelligibility to some extent. For instance, because of their internal differences, the Min group is subdivided into several subgroups (cf. Figure 2.2 in

Chapter Two). But there is uniformity between these subgroups to make them form a Min group (Figure 2.1 in Chapter Two). According to *the Language Atlas of China*, the South Min, Puxian and the East Min share some common features and then form the Eastern cluster whilst the North Min, the Central Min share some other common features and they form the Western cluster (Wurm et al. 1987). To some extent, our subjective trees (viz. the judged intelligibility and similarity trees as well as the functional sentence-intelligibility tree) reflected the internal difference and uniformity of Min group: (i) Fuzhou (East Min) did not form a cluster with Xiamen and Chaozhou (the South Min) at the same level, which shows their different degree of *mutual* intelligibility, viz. Xiamen and Chaozhou are more mutually intelligible than to Fuzhou but (ii) Fuzhou is added to the cluster of Xiamen and Chaozhou at a higher level, which shows that, as Min members, they are more mutually intelligible to each other than to other dialects. [51]

Third, from the word-intelligibility and sentence-intelligibility trees, we can see that Meixian (Hakka/Kejia group) is rather close to both Guangzhou (Yue) and to Nanchang (Gan). The same relationship was seen in the trees based on Cheng's LSI and PCI (cf. Tang & Van Heuven 2007, Chapter Five). [52] Cheng's trees show that Meixian (Kejia) is lexically more similar to Guangzhou (Yue) than to other dialects but shares more phonological correspondence with Nanchang (Gan). These findings also can be explained reasonably. We already know that the Kejia dialect literarily means 'the guest language'. This dialect was formed and affected by Gan during the first immigration period, so it shares many common features with Gan (see also Chapter Two), and then it was influenced by Yue during the second immigration period. Actually, Kejia is an interlanguage between Gan and Yue so that Kejia listeners can understand both Guangzhou (Yue) and Nanchang (Gan) to some extent. That is why some dialectologists proposed the Gan-Ke(jia) group or Yue-Gan-Ke supergroup (Li 1937, Lau 2002).

Fourth, we do find a clear cluster of Suzhou and Wenzhou (in the Wu group) in the trees of judged-similarity on both melodic and monotonized versions but not in the intelligibility trees (judgment intelligibility melodic and monotonized trees, functional word & sentence-intelligibility trees). This can be explained as Suzhou and Wenzhou share more similar phonological features as Wu members, but they are not really quite mutual intelligible to each other. Traditionally, the Wu group comprises dialects of Northern Wu (e.g. Suzhou) and Southern Wu (Wenzhou). According to *the Language Atlas of China*, dialects in the Wu group are geographically between the Jianghuai Mandarin (to their north) and the Min groups (to their south). The northern Wu dialects are heavily influenced by the neighboring Mandarin dialects whilst the southern Wu dialects share some features with the Min dialects. Thus, in some cases, it might not

---

[51] Our opinion test scores were based on readings of the fable *The North Wind and the Sun* in different dialects, i.e. on connected speech rather than isolated words. The internal structure of the Min group is also correctly reflected by C. C. Cheng's LSI and PCI trees (for more detail cf. Tang & van Heuven 2007, and Chapter Five).

[52] I remind the reader of the fact that LSI and PCI are terms I coined myself. Cheng (1997) used different names, viz. 'Lexical Affinity' for LSI and – quite confusingly – 'Mutual intelligibility' for PCI.

be easy to determine their classifications. As for the cluster of Wenzhou and Xiamen in functional word-intelligibility tree, it might be because South Wu (Wenzhou) shares some common features with Min dialects.

Fifth, in spite of the high correlation between the two subjective tests, the two functional intelligibility tests yield higher correlations with each other than the two opinion judgment tests do. Moreover, the functional measures reflect the traditional dialect taxonomy better than opinion tests. The sentence-intelligibility test reflects the traditional dialect taxonomy best (although word-intelligibility is highly correlated with sentence-intelligibility). We explain this result from the fact that words in context supply more communicative information for mutual intelligibility than isolated words, in the Chinese dialects situation. There may be at least two reasons why this is so. First, the isolated word test is not just a word-recognition test; it also involves the additional task of semantic classification, which may introduce a source of statistical noise (error) into the data. Listeners may be quite able to recognize a word in another dialect and yet fail to come up with the correct classification for the word. This problem does not arise in the sentence-intelligibility test. Second, in real life isolated words are the exception rather than the rule. Listeners are used to hearing words in connected speech, and to using earlier context to narrow down the range of recognition candidates. It can be argued, therefore, that the results of the sentence-intelligibility reflect natural speech intelligibility better than the rather contrived semantic categorization task.

I would like to end this discussion section by relating our findings to research done on European languages using the same or a similar methodology that we adopted for the present study on Chinese dialects. Gooskens (2007) determined mutual intelligibility among three West-Germanic languages (Frisian, Dutch, Afrikaans) and separately among three Scandinavian languages (Norwegian, Swedish, Danish). In the latter study intelligibility was measured functionally through five comprehension questions on a short news item. We are not familiar with any studies that allow a systematic comparison of mutual intelligibility among a fairly large number of language varieties (e.g. 15) using both opinion testing and functional tests. Opinion testing is generally proposed as a feasible short-cut when running full-fledged functional intelligibility tests are unpractical. From the literature on intelligibility testing in speech technology we know that native listeners have very accurate intuitions (opinions) on the intelligibility of talking computers (see Van Bezooijen & Van Heuven 1997), so that the use of opinion testing as a shortcut to functional testing seems warranted in that area of application. It is an open question, of course, if the same conclusion would apply to the field of dialectology. Our study would be the first that allows a direct comparison of the value of opinion testing and functional testing of intelligibility in the context of dialectology. It would appear from our results that there is a large measure of correspondence between opinion tests and functional tests. We feel, however, that the correlation between the two types of tests is not good enough to recommend the indiscriminate use of opinion tests as a substitute for functional test procedures. When the resources are available mutual intelligibility should be tested functionally. The results of our functional tests agree clearly better with the general picture that emerges from linguistic taxonomies of the Sinitic dialects in our study. Such a clear correspondence could not be established in our earlier study in which we related the dialectological taxonomy to intelligibility measures derived from opinion tests.

4.5 Conclusions

We end this chapter with provisional and specific conclusions on the following three aspects: (i) the intelligibility within and between Mandarin versus non-Mandarin (Southern) dialects, (ii) the intelligibility correlations between the word and sentence level, and (iii) correlations between the opinion tests and the functional tests. More general conclusions will be presented after predicting mutual intelligibility from various structural distance measures and relating the test results to the traditional dialect taxonomy, in Chapters Six and Seven.

The following results were obtained.

(i) The mutual intelligibility within Mandarin dialects is intrinsically higher than that within non-Mandarin (Southern) dialects, both at word and sentence level. Non-Mandarin (Southern) listeners understand the Mandarin dialects consistently better than non-Mandarin (Southern) dialects.

(ii) All subjective measures significantly correlate with one another. The two types of subjective measures significantly correlate with each other, either in the same type data or cross-type data (e.g. judged similarity versus judged mutual intelligibility and functional word versus sentence intelligibility). Word intelligibility and sentence intelligibility are correlated with each other with r^2-values between .70 and .86, depending on the size of the data matrices used as input (asymmetrical matrix and non-redundant part of a symmetrical matrix, respectively).

(iii) All the results correspond with traditional dialect taxonomy to some extent. Functional intelligibility measures reflect the traditional dialect taxonomy better than opinion scores. Functional sentence-intelligibility scores conform more closely to the traditional dialect classification.

Our provisional overall conclusion is that the degree of mutual intelligibility can be subjectively determined by both opinion and functional tests. Functional intelligibility measures better reflect Chinese dialect classifications than opinion scores. Functional sentence-intelligibility test results conform best to traditional Chinese taxonomy. It may be argued that functional sentence intelligibility reflects the real language situation in China, but only to some extent. The test results show that dialects within a family group are not necessarily more mutually intelligible than certain dialects that belong in different branches. As can be seen on the geographic map, dialects in adjacent provinces (e.g. Nanchang and Changsha) or in transitional areas of the same province (as for Meixian and Guangzhou), may have rather high degrees of mutual intelligibility, even though they belong to different groups and even though their linguistic structures may differ considerably. Language contact may have facilitated the communication between these groups of language users (and made the linguistic distances smaller).

Chapter Five

Collecting objective measures of structural distance

5.1 Introduction

In the preceding two chapters I established the affinity among a set of 15 Chinese dialects by subjective methods. In Chapter Three, affinity was estimated by asking native listeners of each of the 15 Chinese dialects to judge (i.e. give their subjective opinion on) the intelligibility and linguistic distance between the same 15 Chinese dialects and their own. The judgments were based on listening to a reading of the fable of *the North Wind and the Sun* read in each of the 15 Chinese dialects. The results showed that listeners had well-developed intuitions on how intelligible they thought each of the other dialects was and how much it deviated from their own dialect.

In Chapter Four, I applied a different methodology. Here I did not ask listeners to judge intelligibility and linguistic distance but I submitted them to functional intelligibility tests. Listeners were presented a large number of speech samples, both isolated words and short sentences, and were asked to classify or translate these in order to show that they had actually recognized the words and understood the sentences.

In both approaches, i.e. opinion testing (Chapter Three) and functional testing (Chapter Four), the degree of (mutual) intelligibility and affinity could be expressed between the members of each of the 225 possible combinations of speaker and hearer dialects. When speakers and hearers shared the same dialect, mutual intelligibility and judged affinity was high. When the native dialect of the listeners deviated from that of the speaker, intelligibility and affinity scores dropped. Crucially, the scores allowed us to generate tree structures (dendrograms) that express affinity relationships among the 15 Chinese dialects.

Comparison of the various tree structures derived in Chapters Three and Four with linguistic taxonomies proposed by Chinese linguists (in Chapter Two) indicated that, overall, functional intelligibility scores obtained from a sentence understanding task, agreed best with such traditional taxonomies. Moreover, some issues in the classification of Chinese dialects, e.g. whether Taiyuan is a Mandarin or a non-Mandarin (Southern) – Jin dialect, could be settled in a rather straightforward manner if mutual intelligibility is accepted as a valid measure of dialect affinity.

In the present chapter I will collect additional data that may shed light on the degree of affinity between pairs of dialects within our set of 15. This time, however, the data will be statistical measures collected or computed on the sound (and tone) inventories and/or the vocabularies of the dialects. The measures do not involve human speakers or listeners and they are not the results of experiments using human subjects.

Typically, the structural measures of dialect affinity fall into one of two categories. The first type is what may be called *lexical affinity*. This is a measure that expresses the extent to which two dialects are lexically the same, i.e. share the same words (sound shapes) for the same concepts. Lexical affinity is roughly equivalent to the proportion of cognates (words having the common etymological origin) shared between the vocabularies of two dialects (or languages). Of course, setting up criteria to decide whether two forms in two dialects are cognate, is not a trivial task. Intuitively, we all feel that the following words for the concept 'moon' are cognate (i.e. derived from one and the same older form in a parent language): /muːn/, /maːn/, /mwonə/ and /moːnt/ in English, Dutch, Frisian and German, respectively. Here each word has the same consonants /m/ and /n/ before and after the stressed vowel, which is always a back vowel. Such forms can reasonably have developed over time from the same ancestor form, e.g. */moːna/. It is much harder to see how these four forms could be cognate with their counterparts in Romance languages such as French (/lyn/), or Spanish/Italian (/luna/). In Chinese dialectology, cognates are defined as words derived from the same root word and thus having the characteristics of similar sounds with the same conceptual meaning, similar conceptual meanings with the same sound or being identical to both the sound and the conceptual meaning (homophones and synonyms at the same time). Concerning Chinese dialects, there is the additional fortuitous circumstance that it still uses the ancient writing system that uses one character for a concept-sound correspondence, regardless of how the sound shape has developed since Middle Chinese. So, whenever concepts are written with the same character in two Chinese dialects, the sound shapes denoted by the character are lexically treated as cognate.

The second type of structural measure is often called *phonological affinity*. This measure is defined on the lexical subset of cognates shared between two dialects. It expresses, in one way or another, how much the sound shapes of the cognates resemble each other. In the case of Chinese dialects there is the added complication that sound shapes do not only differ in their segmental make up, i.e. in the sequence of vowels and consonants, but also differ in terms of their tonal make up.

It is possible to set up affinity measures at other, higher, linguistic levels. For instance, for many language groups it would make eminent sense to study morphological and syntactic affinity. In Chinese dialectology, morphemes and words are basically the same thing. Chinese has often been called a language with no morphology: Every (simplex) word contains one morpheme and every morpheme is a word. Functions that are carried by inflections in Western languages, such as tense, gender and number markers on verbs, are expressed in separate words in Chinese, which is therefore called an isolating language. Also, surface syntactic structures, at least of simple, basic sentences, do not differ very much between Chinese dialects. In other words, because Chinese dialects evolved from the common parental language, there always existed the phonological correspondences and they share cognates to some extent between pairs of dialects. For these reasons we will concentrate on measures of lexical and phonological affinity and largely ignore higher-level structural differences.

I collect these structural measures of affinity to serve as predictors in the next Chapter (Six), where we will try to predict (mutual) intelligibility – as determined in opinion and functional tests – in an attempt to establish the relative importance of lexical, segmental and tonal differences for intelligibility and thereby for language affinity.

In the present chapter I will collect measures of lexical and phonological affinity. Some measures will be copied from the literature on Chinese dialects, other measures I will compute myself from available language resources. Each time, the measure will be used to generate an affinity tree for the 15 target dialects, which will be compared with traditional dialect taxonomies. Only if the basic split between Mandarin versus non-Mandarin (Southern) dialects is correctly reflected in the tree, will I use the measure as a viable predictor of mutual intelligibility in Chapter Six.

5.2 Measures of lexical affinity

Measures of lexical affinity are based on the assumption that when language varieties are less close the number of cognate words decreases, which will strongly affect *mutual* intelligibility.

As briefly explained above, by lexical affinity we mean the degree to which two languages or dialects share the same vocabulary. Lexical affinity between two languages is high if the two languages use the same (or nearly the same) sound shapes to denote the same concepts. Lexical affinity equals zero if there is not a single concept in the two languages that is expressed by (nearly) the same sound shape. Lexical affinity is obviously related to mutual intelligibility. If two languages share a large proportion of their vocabulary, it will be relatively easy for a listener with language A to understand a speaker of language B, and vice versa.

There are at least two complications. The first is that the sound shapes denoting the same concept in two languages typically differ somewhat. Although we may have clear intuitions when two sound shapes are sufficiently similar to consider them still to be basically the same word, it is a very difficult task to lay down iron-clad decision rules. The notion ‘cognate’ plays a crucial role in the definition of lexical affinity. Words are cognates in two languages if they descend from the same word (sound shape) in a common parent language. The decision whether or not a sound shape in language A and its counterpart in language B have descended from a common origin, is made by etymologists. For Western languages we would not decide on cognateship ourselves but consult etymological dictionaries that list the ancestry of the words in the language. Such resources are also available for Chinese dialects, so that the problem of establishing cognateship can be circumvented. Moreover, as explained above, the Chinese writing system provides a heuristic to determine cognateship: when the same meaning-sound correspondence in two dialects is written with the same character(s), the words are cognate.

The second complication is that two sound shapes may be related but the meaning in language A may have grown different from that in language B. Take, for example, the word *knight* in English and its cognate *knecht* in Dutch. In Dutch the word has kept its ancient (Old Germanic) meaning of servant. In English, the word was used only to refer to persons who served immediately under the king, so that it came to denote a nobleman. Words that sound alike but have different meanings (or words diverging in meaning when language varieties developed separately) are called false friends in foreign language teaching. False friends are believed to be detrimental to mutual intelligibility. I will assume that the false-friend situation is very rare in Chinese dialects, and therefore choose to ignore the problem.

I will describe an attempt at establishing lexical affinity for my 15 target dialects. This attempt was not made by me but relies on existing literature. Chinese dialectologist and computational linguist Chin-Chuan Cheng (C. C. Cheng or Cheng) devoted his career to establishing affinity measures for Chinese dialects. He began his work in the seventies, and continued to publish on the topic well into the nineties. The 18 dialects he targeted on affinity measures are a superset of my 15 except Taiyuan and Hankou, so that Cheng's publications provide numerical indexes of affinity in a number of domains (e. g. lexical, phonological) that we can readily copy.[53] In the next section I will describe Cheng's measure of lexical affinity, and see how well it relates to traditional Chinese dialect taxonomy.

5.2.1 Cheng's Lexical Affinity Index

Since the 1970s, C.C. Cheng has collected and computed structural measures on Chinese dialects based on a database called *Hanyu Fangyan Cihui* (Chinese Dialect Word List, henceforth *Cihui*, Beijing University 1964). From the 1980s onward, he attempted to express the degree of lexical affinity between pairs of dialects in his set of 18 through quantitative measures (see note 1).

More specifically, the *Cihui* is a lexical database which contains 905 common words in Standard Mandarin and the equivalents for the same concept (very often but not always expressed by cognate words) in 18 dialect localities. The presence (assigned the value 1) and the absence (assigned the value 0) of cognates to express the same concept in each pair of dialects was taken as a measure of lexical affinity. Cheng listed the occurrences of all expressions (words) for the same 905 concepts in the 18 dialects side by side. This produces a table with 905 (concepts) × 18 dialects = 16,290 cells. In actual fact, only 6,454 variants occurred. Lexical affinity between two dialects is defined as the number of concepts expressed in both languages by cognate words as a proportion of the union of the vocabulary samples of the two dialects. Cheng (1997) derives his lexical affinity measure as follows. Logically, four possibilities exist for the expression of the same concept C in two dialects (see Table 5.1a). In situation (a), the concept is expressed by

[53] The 18 dialects C. C. Cheng targeted are: Beijing, Jinan, Shenyang, Xi'an, Chengdu, Kunming, Hefei, Yangzhou, Suzhou, Wenzhou, Changsha, Nanchang, Meixian, Guangzhou, Yangjiang, Xiamen, Chaozhou, Fuzhou. Thirteen of these dialects overlap with our selection of 15 (Taiyuan and Hankou were not among Cheng's set).

cognate forms in the two dialects. In situation (b), a concept is expressed by a word x in dialect A but does not occur (not even as a non-cognate form) in dialect B. Situation (c) arises in the reverse case, when a concept is expressed in dialect B by a word y but does not occur (not even as a non-cognate) in dialect A. In situation (d), neither dialect A nor dialect B have a word for concept C.

Table 5.1a. Possible occurrence of concepts in two dialects.

		Is concept expressed by cognate in dialect B ?	
		yes	no
Is concept expressed in Dialect A?	yes	a	b
	no	c	d

The affinity measure defined by Cheng is the proportion of cases *a* in the two lists of 905 possible expressions of concepts relative to the total set of words for concepts in dialects A and B taken together (i.e. the union of the two vocabularies, not counting any of the 905 concepts that have no expression in the two dialects together). In this dissertation, I call this lexical affinity measure as Lexical Affinity Index (LAI, henceforth for abbreviation) and it can be expressed in a formula as follows: [54]

LAI = a / (a + b + c)

So, if dialect A shares 500 cognate words with dialect B, and if 305 concepts occur in dialect A but not B (and vice versa), another 100 concepts have no word in either A or B, the LAI measure is computed as 500 / (500 + 305 + 305) = 0.451. This example is shown in Table 5.1b. Note that *b* and *c* must be equal, and that *d* is excluded from the LAI formula.

Table 5.1b. Hypothetical example of occurrence of concepts in two dialects.

		Is concept expressed by cognate in dialect B ?	
		yes	no
Is concept expressed in Dialect A?	yes	500	305
	no	305	100

I copied the submatrix of percent shared cognates for all pairs of dialects that were also included in our set of 15 (in fact, 13 dialects of Cheng's 18 dialects are shared by our study, see note 1). The result was first published in 1982 and reiterated in 1986 and 1991 (Cheng 1982, 1986, 1991). The most recent version was published as Cheng (1997), which version was also made available through the internet. Appendix 5.1a is the matrix containing the LAI values for all pairs of 13 dialects (copied from Cheng

[54] In earlier publications (e.g. Tang & van Heuven 2007, 2008) I called this measure LSI, short for Lexical Similarity Index.

1991: 96). Note that the table is a symmetrical matrix since the number of cognates shared between dialect A and B is identical to that shared between dialect B and A. I omitted the redundant upper triangle of the matrix in Appendix 5.1a.

Appendix 5.1a presents the lexical affinity index computed by the above formula for all pairs of the 13 languages in our dialect sample that were also included in the set of 18 studied by Cheng (unfortunately, the two dialects Taiyuan and Hankou were not available from the literature). To facilitate later comparison with other affinity measures I have included rows and columns for the missing dialects, i.e. Taiyuan and Hankou, but left the cells empty.

Finally, Cheng processed the LAI index with a cluster analysis to graphically represent the distance/closeness relationship of these dialects by using the (unweighed) average linking method.[55] Cheng used the LAI values to generate a hierarchical tree structure (dendrogram or agglomeration schedule) which illustrates the (sub)grouping of these dialects both numerically and visually. Following Cheng's method, we generated a lexical affinity tree using the same (unweighed) average linking method (see Figure 5.1). The proximity matrix underlying the tree can be found in Appendix 5.1b.

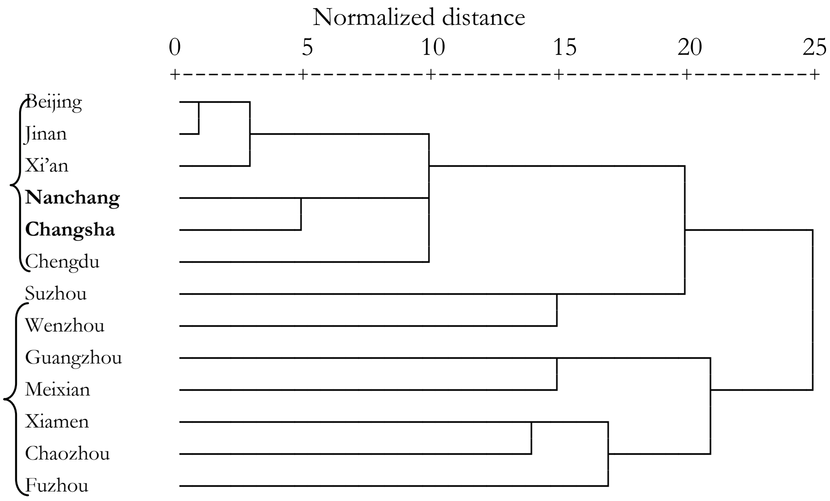

Figure 5.1. Dendrogram based on lexical affinity using Average Linkage (Between Groups) and the Euclidean distance measure used because of the interval variables. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

---

55 Cluster analysis first establishes a group by finding the pair of dialects having the minimum distance. Then the next minimally distant pair is found, the average distance between the two pairs is calculated and linked with another minimally distant pair and so on.

5.2.2 Lexical affinity tree versus traditional dialect taxonomy

Figure 5.1 shows a primary split between an upper branch with eight dialects that include all the Mandarin dialects in the set of 13, and a lower branch that includes five non-Mandarin (Southern) dialects only. Within the upper branch a cluster of another two non-Mandarin (Southern) dialects is seen (Suzhou, Wenzhou). The remaining cluster of six dialects comprises all the Mandarin dialects (Beijing, Jinan, Xi'an, Chengdu) in our set of 13 but the cluster is polluted with yet another two-member subcluster of non-Mandarin (Southern) dialects (Nanchang, Changsha).

Let us define some simple criteria in order to evaluate the degree of correspondence between the tree and the traditional dialect taxonomy. A strict criterion would require that the primary split into the upper and lower branches of the tree should perfectly correspond with the traditional division between Mandarin and non-Mandarin (Southern) dialects. Using this strict criterion, the tree contains four misclassifications: Nanchang, Changsha, as well as Suzhou and Wenzhou are incorrectly classified together with the six Mandarin dialects in the upper branch.

However, we may relax the criterion somewhat. It would not be unreasonable to separate off the cluster comprising Suzhou and Wenzhou (the two Wu dialects in our sample) from the other six dialects in the upper branch, and add it to the lower branch. In this case, a primary split between an upper branch with six dialects that include all the Mandarin dialects in the set of 13, and a lower branch that includes seven non-Mandarin (Southern) dialects only (marked in the braces). Within the upper branch a cluster of another two non-Mandarin (Southern) dialects is seen (Nanchang and Changsha). Therefore, only two classification errors remain: the integral cluster containing the six Mandarin dialects still contains the non-Mandarin (Southern) dialects Nanchang and Changsha. These have been bolded in Figure 5.1.

5.3 Measures of phonological affinity

5.3.1. Introduction

In the preceding section I explained that mutual intelligibility between two languages is expected to increase as the languages share a larger proportion of their vocabularies, i.e. as a larger number of the words in the two languages are cognate. At the same time, however, we conceded that cognates are not necessarily identical sound shapes. In fact, it is hardly ever the case that cognates in two languages have identical sound shapes. Generally, the more sounds are different between a cognate pair, the more difficult it will be for listener A to recognize the word when spoken in language B. Obviously, it will be easier for an English listener to recognize the Dutch cognate /kɑt/ as English *cat* /kæt/ (where only one sound differs) than to recognize /kas/ as *cheese* /tʃiz/ (where all three sounds are different).

In the example given here, a pair of sounds in the two languages is either the same or different, in a categorical way. In the pair /kɑt/ ~ /kæt/ only the vowel differed whilst the onset and coda consonants were held to be identical. Whether two sounds are same or different was decided on the basis of a broad phonemic transcription. In the actual communication between a Dutch speaker and an English listener, other, more subtle phonetic differences will also play a role. For instance, the onset /k/ in Dutch is not aspirated so that it sounds rather different than the aspirated plosive [kʰ] the English listener would expect to hear. Phonological difference between cognate words in two languages can be determined from a broad transcription and from a (more or less) narrow transcription.

A second complication is that not all differences between two sounds are necessarily equally large. The difference between /ɑ/ and /æ/ in the English-Dutch cognate pair for *cat* is smaller than the difference between the vowels in the cognates for *cheese,* /i/ and /a/. The difference between the members of a cognate pair may be computed such that sounds that are more different contribute more to the overall distance between the cognates. Such differential measures may also pertain to phonetic differences between corresponding sounds such as presence versus absence of aspiration.

To complicate matters further, it is often the case that cognates differ in the number of sounds. In the cognate pair for *knee*, the Dutch word has three phonemes /kni/ but the English counterpart /niː/ has only two: the /k/, which is still reflected in the English spelling, is not pronounced. It has been suggested (Heeringa 2004) that the absence/presence of a sound might compromise the recognition of a word less than the substitution of one sound for another. As a result, some researchers have chosen to weigh sound substitutions between cognate pairs more heavily than insertions/deletions.

In the following sections I will attempt a number of ways to quantify the difference in sound shapes of cognates in my 15 Chinese dialects. In doing so I will explore several avenues. My first attempt will be a simple comparison of the sound inventories of the languages, reasoning that mutual intelligibility will be better as two languages share a larger number of phonemes in their inventories. This can be done for the complete inventory but I may also make the comparison separately for vowels, consonants and tones. In its crudest form, the inventories of segments and tones can be compared across dialects on a binary basis, i.e. we just check whether or not a phoneme or tone is shared between two dialects. However, we may also select a representative sample of words (or morphemes) from the lexicon and see how often a particular sound or tone occurs in the word list. This would yield the same information as the binary count of co-occurring segments and tones, but now they are weighed by their lexical frequency. I was fortunate to have at my disposal a computer-readable database that contains fairly narrow phonetic transcriptions (including lexical tones) of 764 cognates in 40 Chinese dialects, which included all of my 15 dialects. I used this database to compute phonological distance measures.

A more complex comparison will involve the computation of the (average) difference in sound shapes in the list of 764 cognates in my 15 Chinese dialects. I did this separately for segments (i.e. the string of vowel and consonant symbols) and for the tones of the words in the database. Moreover, these computations were done twice in each domain: once any difference between two sounds was considered to be equally important, the second time the difference between two corresponding sounds was weighed by their perceptual distance. Details of the procedures followed will be explained below.

The difference between cognate pairs will be established by applying the so-called Levenshtein algorithm. This procedure yields a string distance measure that is based on the number of string operations (insertion, deletion, substitution) needed to convert the phonetic transcription of a word in language A to its counterpart in language B (or vice versa). Insertions/deletions and substitutions may be weighed differently; also, certain substitutions may contribute more to the word distance than others, depending on the perceptual distance between the sounds involved in the substitution.

It is important to point out here that the Levenshtein distance between two words, and between the means computed across an entire vocabulary, are symmetrical. That is, the distance between language A and B is the same as that between B and A. This is not a proper reflection of reality. Very often we find that it is easier for language A to be understood by listeners of language B, than the other way around. For instance, Portuguese listeners understand Spanish quite well but Spanish listeners have a hard time when listening to Portuguese. C. C. Cheng defined a computational measure for phonological distance that does reflect asymmetries between the sound systems of two languages. In the section (§ 5.2.5) on phonological distance, I will explain Cheng's (1997) procedure and use his phonological distance measure as a supplement to my own.

5.3.2 Distance between dialects based on sound inventories

In this section I will examine the sound and tone inventories of our 15 dialects, and see to what extent these differ from each other. I will then check to what extent the traditional dialect taxonomies reflect differences and similarities in the inventories of sounds (initials, nuclei, finals, codas) and tones.

The inventories of the 15 dialects were copied from the surveys provided by Yan (2006) and checked against the website maintained by Campbell (Campbell 2009, see http://www. glossika.com/en/dict/faq.php#1) The lists of segmental sound symbols and tones are included in appendices 5.2-5.7. Taking the cue from work by Cheng (1997), who computed the sameness and difference of the inventory elements in cognate words, I then drew up lists containing all the different initials, nuclei, finals, codas, and tones across the set of 15 dialects. In each list I specified for each entry (in the rows) for each of the 15 dialects (in the columns) whether the particular sound or tone was or was not part of the inventory. When the sound was in the inventory, this was indicated by a '1', when it was absent from the inventory, a '0' was entered. On such data matrices affinity trees can be generated, using Euclidean distance and binary cell contents. As before I will evaluate how well the tree agrees with traditional dialect taxonomy by determining the number of errors in the classification of dialects in terms of the primary split between Mandarin and non-Mandarin (Southern) dialects.

5.3.2.1 Initials

Across the 15 dialects I found a total of 37 different onset consonants. Each of the 15 dialects also allows the initial to be absent (or 'empty'). The zero initial was not included in the list (and if it had been, it would not have contributed to any distinction among the dialects). The number of onset consonants based on the inventory of the 15 dialects varied between 17 (Guangzhou, Hankou, Meixian, Nanchang, Xiamen) and 29 (Wenzhou). The dendrogram derived from the initials table (Appendix 5.2a) is presented in Figure 5.2. The proximity matrix computed from the binary data is included as Appendix 5.2b.

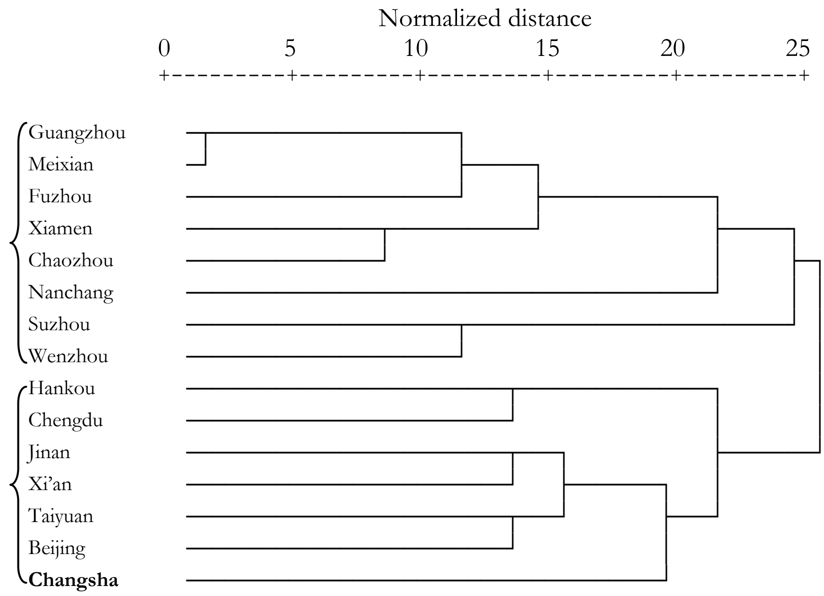

Figure 5.2. Dendrogram based on inventory of initials for 15 dialects, using Average Linkage (Between Groups) and Euclidean distance between binary variables. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

Following the same criterion as the LAI tree, I will compare this tree structure with the traditional taxonomy concerning only the primary split between the Mandarin and the non-Mandarin (Southern) branches instead of more elaborate reflections of the internal structure of each branch.

Observing the tree structure generated from the initials inventory of the 15 dialects, we see that it reflects the primary split between the Mandarin and non-Mandarin (Southern) branches very well. The upper part of the tree parses all the non-Mandarin (Southern) dialects together. The lower part of the tree comprises all the Mandarin dialects except

Changsha which is traditionally classified as a member of the Xiang group and should therefore be considered a non-Mandarin (Southern) dialect. Therefore, this tree yields just one classification error (Changsha is bolded in the graph as I did in the LAI tree).

I will do the same procedure about the vocalic nuclei, the coda, the tone, the finals (or rhymes, including the nuclei and the coda) separately in the following steps, again, all these affinity trees will be compared with the traditional dialect taxonomy in terms of the primary split. Each time the error will be counted and finally the errors will be summed and the closing remarks will be given.

5.3.2.2 Vocalic nuclei

The total number of different vocalic nuclei (including medials or glides) across the 15 dialects was 78. The maximum number of different nuclei that made up the inventory of a single dialect was 25 (Fuzhou) and the minimum was 14 (Meixian). I refer to Appendix 5.3a for details. Figure 5.3 shows the dendrogram generated from the nucleus inventories. The proximity matrix underlying the tree can be found in Appendix 5.3b.

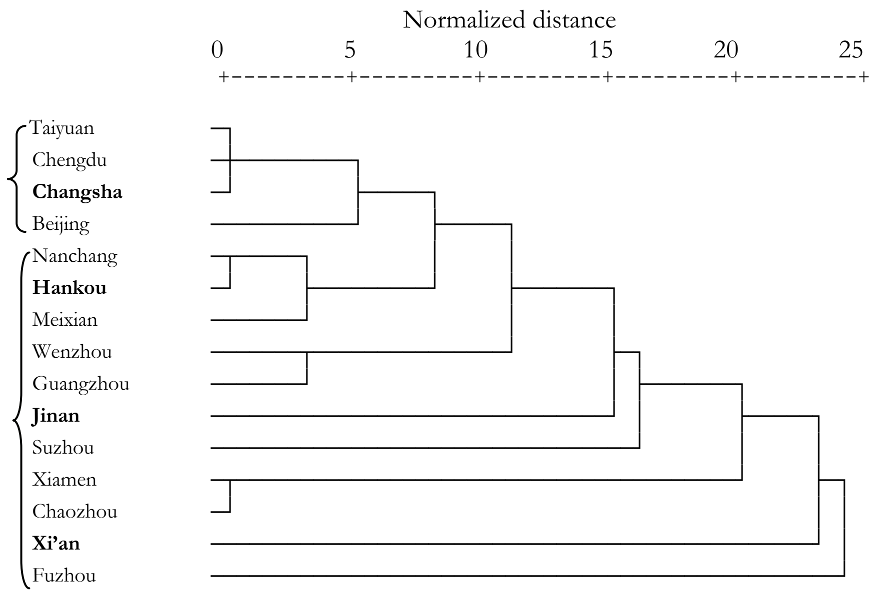

Figure 5.3. Dendrogram based on inventory of nuclei for 15 dialects, using Average Linkage (Between Groups) and Euclidean distance between the binary variables. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The tree structure generated from the vocalic nuclei in effect reflects the taxonomy very poorly. Strictly speaking, it does not reflect the primary split at all. There is no clean division between the Mandarin and the non-Mandarin (Southern) branches. In fact, it is just an add-on structure. If we follow the lenient criterion, we may reluctantly argue that there is a group at the upper part which comprises most of the Mandarin dialects but excludes Hankou, Jinan and Xi'an. Instead, a non-Mandarin (Southern) dialect (Changsha) is wrongly embedded in the candidate Mandarin branch. However, the other group (the lower part of the tree) can be the candidate non-Mandarin (Southern) dialect group but also contains three Mandarin dialects (Hankou, Jinan and Xi'an). In this case, the number of classification errors is four. I will refrain from further comments and move on to the affinity tree based on the codas.

5.3.2.3 Codas

The total number of codas occurring in our sample of 15 dialects amounts to 15. The smallest number of different codas is two, i.e. the two nasals /n, ŋ/ (for Beijing, Chengdu, Fuzhou, Jinan) whilst the largest number of coda consonants in any dialect was eleven (Xiamen). The table containing the distribution of coda consonants in the 15 dialects is given in Appendix 5.4a. Figure 5.4 shows the dendrogram derived from the occurrence of codas in the inventories. The proximity matrix underlying the tree can be found in Appendix 5.4b.

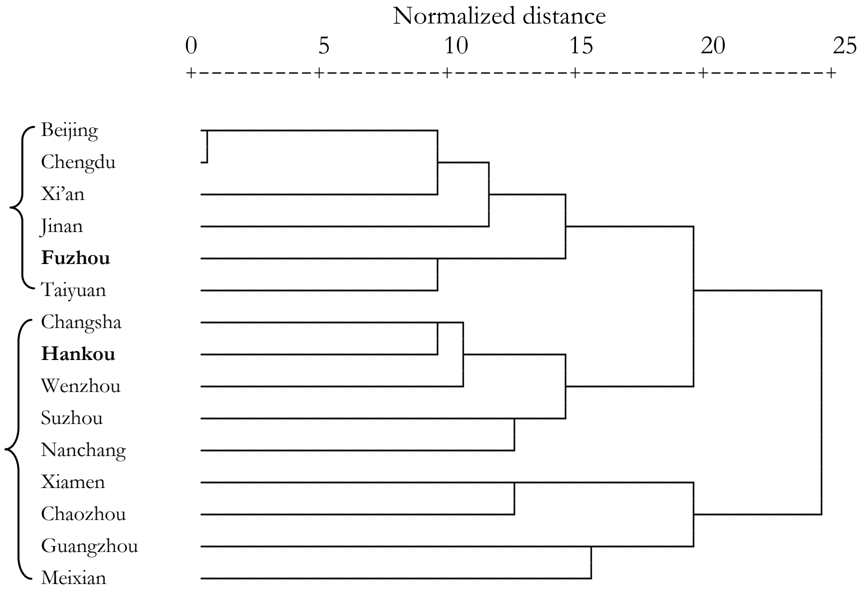

Figure 5.4. Dendrogram based on inventory of codas for 15 dialects, using Average Linkage (Between Groups) and Euclidean distance measure between binary variables. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The optimal binary split in this tree is between an upper cluster of four containing mainly Mandarin dialects, and the remainder of the tree, which includes all non-Mandarin (Southern) dialects except Wenzhou, and which is polluted with three Mandarin dialects (Chengdu, Xi'an and Taiyuan). This brings the number of classification errors to a total of four.

It is rather amazing that the tree should not reflect the primary split any better. Even if we split the 15 dialects using just the number tones in the inventory, this would get us an almost perfect division between Mandarin (maximally five tones) and non-Mandarin (Southern) dialects (more than five tones).

The affinity tree based on the codas is relatively better than the nucleus tree but clearly not so good as the one based on initials. We find a group at the upper part of the structure which comprises most Mandarin dialects plus a single non-Mandarin (Southern) one (Fuzhou). A second group at the lower part of the structure comprises most non-Mandarin (Southern) dialects but contains also one Mandarin dialect (Hankou). Accordingly, the number of classification errors is two.

5.3.2.4 Tones

Tones were transcribed as sequences of maximally three digits, each of which could assume a value between 1 and 5. Here '1' refers to the lowest tone in the speaker's range and '5' to the highest tone (Chao 1928). The assumption is that any word tone in a Sinitic language can be transcribed within this notation system. The tones may consist of single, double or triple digits. The number of digits in the transcription roughly corresponds to the duration of the tone. A three-digit tone is always a contour tone, i.e., a tone that does not remain flat throughout its duration; an example would be the 'dipping' Tone 3 in Mandarin, which is transcribed as 214. In the inventory I distinguish between short and longer level tones, so that, for example, '5' and '55' are considered to be different tones. '0' refers to the neutral tone. The table listing the tone inventories for the 15 dialects can be found in Appendix 5.5a.

The total number of tones occurring in our sample is 28 (including the toneless or neutral tone). Within single dialects the number of tones varies between four (all Mandarin dialects except Taiyuan) and as many as nine (Guangzhou). Figure 5.5 presents the tree structure derived from the tone inventories. The proximity matrix underlying the tree can be found in Appendix 5.5b.

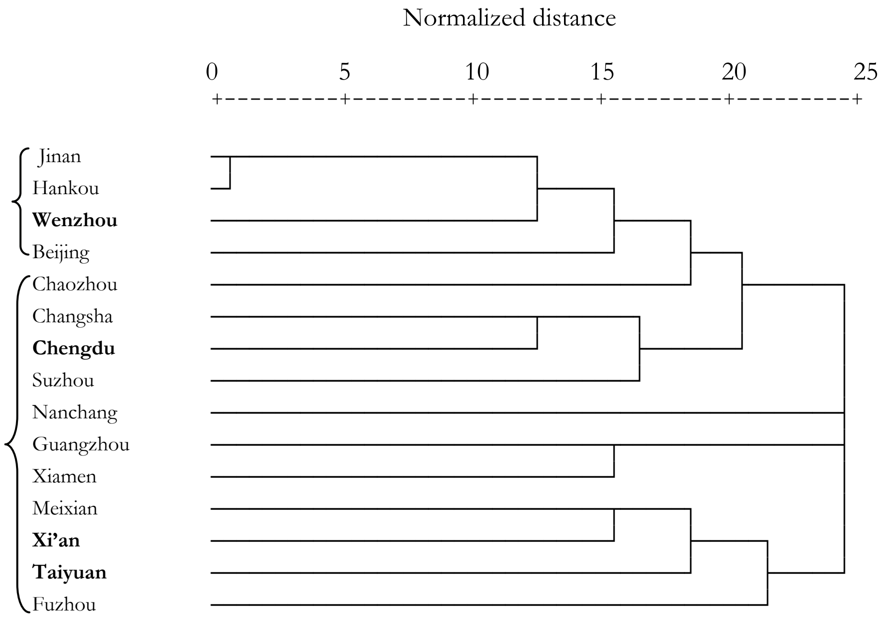

Figure 5.5. Dendrogram based on tone inventory for 15 dialects, using Average Linkage (Between Groups) Euclidean distance measure between binary variables. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

Let us now move on to considering the inventories of more complex (and therefore more diversified) sound structures, i.e. Finals (rhymes, i.e. nuclei + codas).

5.3.2.5 Finals

I tabulated all the finals that were listed for each of the 15 dialects, as these could be found in the literature.[56] Finals are combinations of nuclei (including medials) and codas, i.e. the string of segments that is left over from a syllable when the initial is stripped from it – and disregarding any tonal differences.

The total list contains 390. different finals. The smallest and largest number of finals occurring in any one dialect is 35 (Wenzhou) and 95 (Guangzhou), respectively. For details see Appendix 5.6a. The tree structure generated from the finals inventories is seen in Figure 5.6. The proximity matrix underlying the tree can be found in Appendix 5.6b.

[56] We have no resources that list for each of the 15 dialects which different combinations of segments and tones occur as syllables. The *Zihui* comes closest with 2,270 entries but this does not nearly list all the possibilities that exist in the set of 15 dialects.

This is a typical add-on tree. There is no clear hierarchical split into branches at all. This structure very poorly fits the traditional taxonomy. Accordingly, there are many classification errors. However, if we apply the relaxed criterion, we can still find a group which comprises mainly Mandarin dialects (but also includes two non-Mandarin/ Southern dialects, i.e. Changsha and Wenzhou), and another group which reflects all the non-Mandarin (Southern) dialects plus two Mandarin dialects (Jinan and Xi'an). Therefore, the number of classification errors is four.

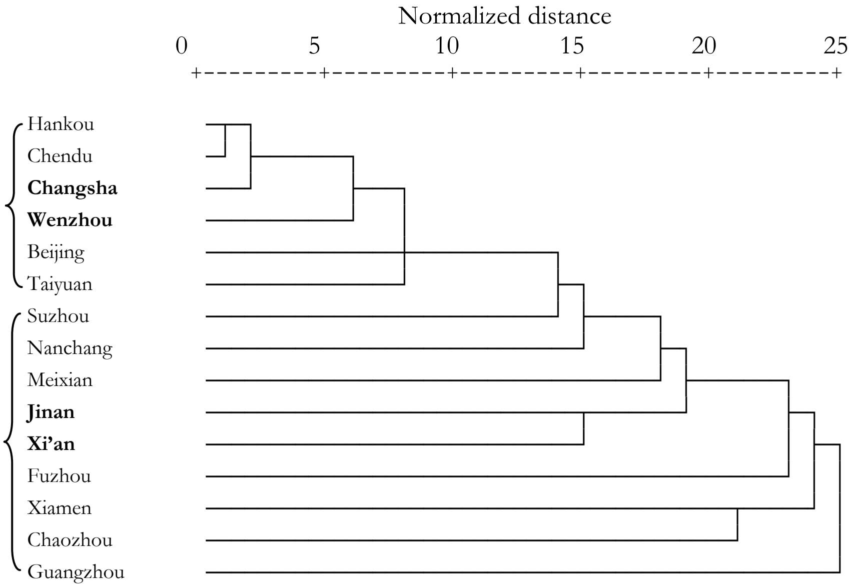

Figure 5.6. Dendrogram based on rhyme inventory for 15 dialects, using Average Linkage (Between Groups) and Euclidean distance measure between binary variables. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

5.3.2.6 Combining initials and codas

So far, we have seen that two of the phonological distance measures based on inventories were reasonably successful as indicators of dialect taxonomy, at least in terms of the primary division into Mandarin and non-Mandarin (Southern) dialects. These successful measures were the patterning of the inventory of (i) initials (onset consonants), and (ii) coda consonants. These measures yielded dendrograms with one and two misclassifications, respectively. All other measures led to at least four misclassifications.

The numbers of different onsets and codas across the fifteen dialects are very limited, i.e. 37 and 15, respectively. In this subsection I would like to explore the possibility to get a better classification if the inventory of initials and codas are combined, so that the total size of the inventory is 37 + 15 = 52 (see Appendix 5.7a). The dendrogram that results from this operation is presented in Figure 5.7. The proximity matrix underlying the tree can be found in Appendix 5.7b.

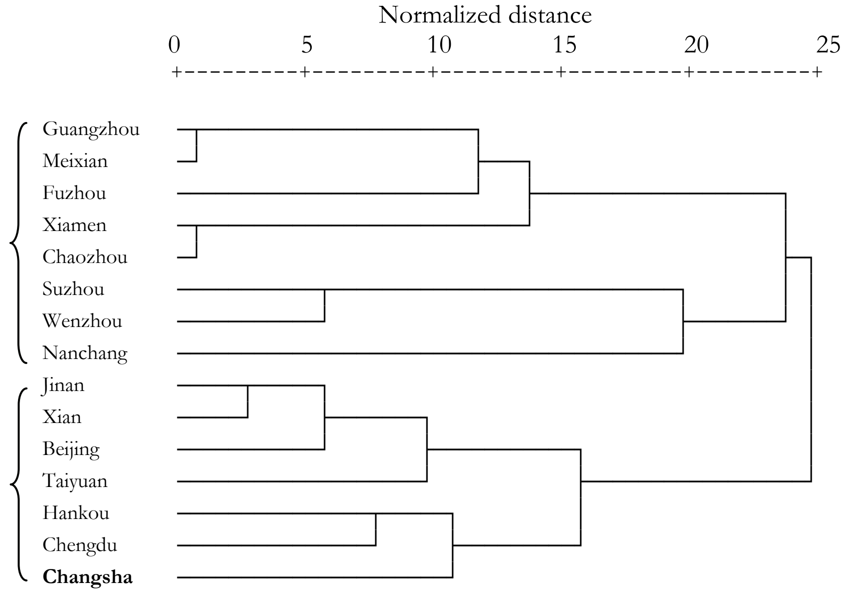

Figure 5.7. Dendrogram based on initial+coda inventory for 15 dialects, using Average Linkage (Between Groups) and and Euclidean distance measure between binary variables. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The primary split in this tree is basically the same as the one we found for the initials only. There are two clear integral groups with just one classification error, i.e. Changsha, a non-Mandarin (Southern) dialect that is incorrectly parsed with the six Mandarin dialects.

A further comparison of the initials and the initials+codas trees, however, reveals that the latter tree is somewhat better when it comes to the representation of the internal clustering within the two main branches. Jinan, Xi'an and Beijing plus Taiyuan form an identifiable cluster (Northern Mandarin, cf. the linguistic map on the internet) and so do Chengdu and Hankou (South-Western Mandarin). In the non-Mandarin (Southern) branch we see the two Wu dialects (Wenzhou, Suzhou) correctly form a cluster at a very low level. We also find a South Min cluster (Xiamen and Chaozhou) at the first level (for details on the dialect taxonomy, cf. Chapter Three).

5.3.2.7 Concluding remarks

Generally, generating dendrograms from inventories of segments and tones does not afford a viable way of classifying the dialects. Comparison of the resulting trees with the traditional dialect taxonomies reveals very little agreement between the two types of information. One notable exception, however, is in the initials. Here a very convincing split could be observed between Mandarin and non-Mandarin (Southern) patterning of elements in the inventories, with just one erroneous classification: Changsha was classified with the Mandarin group. Slightly better results were obtained by combining the information provided by the two best-performing criteria, i.e. initials and coda consonants.

By way of summary, Table 5.2 lists the number of classification errors according to the various tree structures. The structures are listed in ascending order of the number of types in each structure. There is no clear tendency for the less diversified types of structure to yield better fitting trees except the case in terms of tone (r = 0.465 (p = 0.353, ins.), ϱ = 0.309 (p = 0.552, ins.).

Table 5.2. Amount of information on inventory taken into account and number of classification errors of dialects into Mandarin versus non-Mandarin (Southern) groups.

Type of information	N of types	Misclassifications
Codas	15	2
Tones	28	4
Initials	37	1
Initials + codas	52	1
Vocalic nuclei	78	4
Finals	391	4

I will not decide here which measure of phonological distance corresponds best with traditional taxonomies of Chinese dialects. Let us first see to what extent more adequate measures of phonological affinity might be obtained by applying more sophisticated procedures than just comparing inventories of sounds and subsyllabic sound structures.

5.3.3 Weighing sound structures by their lexical frequency

In this part, I will make use of another existing resource, which is the word list contained in the dialect sound database of Modern Chinese.[57] This database includes

[57] In Pinyin it is called 'Xiandai Hanyu Fangyan Yinku' (Dialect Sound Database of Modern Chinese). This database was compiled after the publication of the *Language Atlas of China*. It was first compiled and published as series of 40 volumes, i.e. one book volume plus one cassette tape per dialect, (cf. Hou Jingyi 1994). More recently, it was made integrally available by the Chinese Academy of Social Sciences (CASS) on CD-ROM (cf. Hou Jingyi 2003). The dialect classification used was adopted from the *Language Atlas of China*.

forty Sinitic dialects. For each dialect, there are five main parts contained: (i) the inventory of sounds, (ii) segmental and tonal transcription of the common vocabulary, (iii) the principal regularities of word formation, (iv) syntactic examples of grammar rules with phonetic transcription and (v) a phonetic transcription of a reading of the fable 'The North Wind and the Sun.' Furthermore, each dialect comprises three to four appendices describing the general introduction of the (sub)group dialects, the survey of their representative dialect including its sound, vocabulary, grammatical characteristics and the homophone list with narrow phonetic transcriptions.[58] The five main parts for each dialect were sound-recorded by either male or female native speaker(s). The work was done by Chinese linguists in the Institute of Linguistics of CASS (Chinese Academy of Social Sciences) (cf. Hou Jingyi 1994, 2003). Henceforth, I will call this the CASS database. The list I use is contained in the database on a CD-Rom (Hou 2003). It contains 764 morphemes in Modern Chinese. For each morpheme, the dialectal variant (or variants) — sound shape(s) — in each of the forty dialects is/are listed. That makes 764 × 40 = 30,560 items. For each of variant, a segmental and tonal transcription has been entered. Segmental transcriptions are fairly narrow; tones are specified in terms of the 3-digit scheme proposed by Chao (1928).

The 40 dialects contained on the CD are a superset of my set of 15. I extracted the phonetic transcriptions of the 764 lexical items in each of my 15 dialects, and converted these into a format that could be processed by conventional tools such as Excel and SPSS. As it happens, the forms used for the 764 words are cognate with the (reference) form in Beijing in all the 15 dialects, with just very few exceptions. When a dialect does not use a cognate, the entry is left empty in the CASS database. Non-cognate forms occur in five dialects only, viz. Nanchang (29), Meixian (6), Fuzhou (2), Changsha, and Xiamen (1).[59] The non-cognates were simply disregarded.

I then split up the transcriptions into separate segmental and tonal representations, and made a further split in the segmental transcriptions in terms of onsets (initials), finals (rhymes). The latter were further subdivided into vocalic nuclei (including glides) and codas.

The frequencies of the various segmental parts and of the tones were then computed by SPSS. The frequencies are between 0 and 764. The basic data look very much like the inventories examined in the preceding sections, with one important difference: whereas the inventories merely specify the presence ('1') or absence ('0') of an item in a

[58] The forty dialects are: Shanghai, Suzhou, Hangzhou, Wenzhou (Wu group); Guangzhou, Nan'ning, Hongkong (Yue group); Xiamen, Fuzhou, Jian'ou, Shantou, Haikou, Taibei (Min group); Meixian, Xinzhu (Hakka group); Nanchang (Gan group); Changsha, Xiangtan (Xiang); Shexian, Tunxi (Hui group); Taiyuan, Pingyao, Huhehot (Jin group); Beijing, Tianjin, Ji'nan, Qingdao, Nanjing, Hefei, Zhengzhou, Wuhan, Chengdu, Guiyang, Kunming, Harbin, Xi'an, Yinchuan, Lanzhou, Xi'ning, Urumqi (Mandarin Group). My 15 dialect selection is a proper subset of these 40 dialects.

[59] Although, in principle, the occurrence of non-cognate forms could be used to compute a measure of lexical affinity among the 15 dialects (as was done in § 5.1), I decided against this on the grounds that the differences in number of cognates are too small in the present dataset (in fact 10 dialects have no non-cognates at all, so that these would be grouped together by any agglomeration schedule.

dialect, the data will now specify the frequency of an item in the list of 764 items. The frequency results will be used to generate tree structures, which can be compared with the traditional dialect taxonomy as I did in § 5.2.2. Again, dialect classification errors will be counted and the least erroneous trees will be further compared with Cheng's affinity trees, inventory trees (in this chapter), and with my own experimental trees (in Chapters Three and Four). The comparisons will be in next chapter (Chapter Six).

5.3.3.1 Lexical frequency of initials in the CASS database

In order to get the frequencies for all the initials (onsets) of dialectal variants of the 764 words in each of the 15 dialects, I separated the 11,460 (764 × 15) initials/onsets from the segmental syllabic structure (initial+finals). When a word began with a vowel, its initial/onset was specified as empty (or zero). In total, there are 38 onsets across my 15 dialects. Then I counted the frequencies for each initial in each of the 15 dialects.[60] These frequency measures were then used to generate a hierarchical tree structure by using between group linkage method via SPSS as Cheng and I did before. Since the frequencies are counts rather than interval numbers, I selected the chi-square option as the appropriate distance measure. The tokens of the initials/onsets for 15 dialect and the frequency will be listed in Appendix 5.8a. The resulting tree structure is illustrated in Figure 5.8. The proximity matrix underlying the tree can be found in Appendix 5.8b.

This tree reflects the primary split of the Mandarin and non-Mandarin (Southern) groups fairly well. Six Mandarin dialects are comprised by a predominantly Mandarin upper branch which also incorrectly includes two non-Mandarin (Southern) members (Nanchang, Changsha). The lower non-Mandarin branch comprises all the other non-Mandarin dialects. Within each of the basic branches, perfect sub-groups are also found. For example, Hankou and Chengdu form the South-western cluster in Mandarin, Suzhou and Wenzhou make up a correct Wu cluster, and the three Min dialects (Fuzhou, Chaozhou and Xiamen) are also correctly grouped together. Following the earlier criterion, the number of classification errors yielded by this tree is two.

[60] As explained in note 58, the case of empty cognate was filtered out before executing the frequency computing through SPSS.

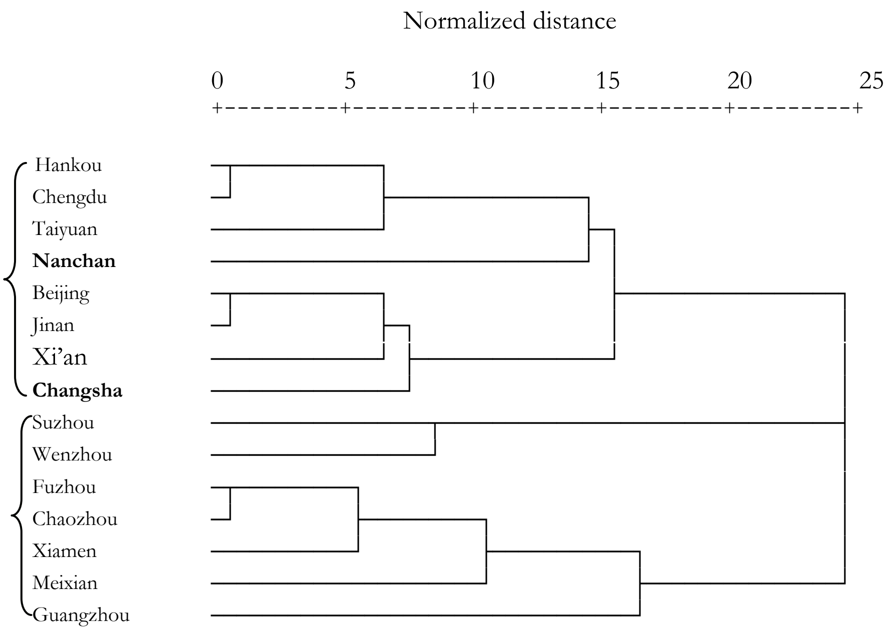

Figure 5.8. Dendrogram based on onset frequency for 15 selected dialects from CASS database, using Average Linkage (Between Groups) and chi-square as the distance measure because of the count variables. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

5.3.3.2 Lexical frequency of finals in the CASS database

Next, I extracted all the finals from the database and determined their lexical frequency in the 15 lists of 764 items. In all, 262 different finals were found in the database (see Appendices 5.9a and 5.9b). The highest lexical frequency was found for the final /a/, which occurred 98 times in Wenzhou. The agglomeration tree derived from the frequencies of finals is shown in Figure 5.9. The proximity matrix underlying the tree can be found in Appendix 5.9b.

Although the tree is not particularly well structured, it can be divided (using the more relaxed criteria) into an upper part that comprises all the Mandarin dialects plus Changsha and Wenzhou and a lower part that exclusively contains non-Mandarin (Southern) dialects. The number of classification errors amounts to two.

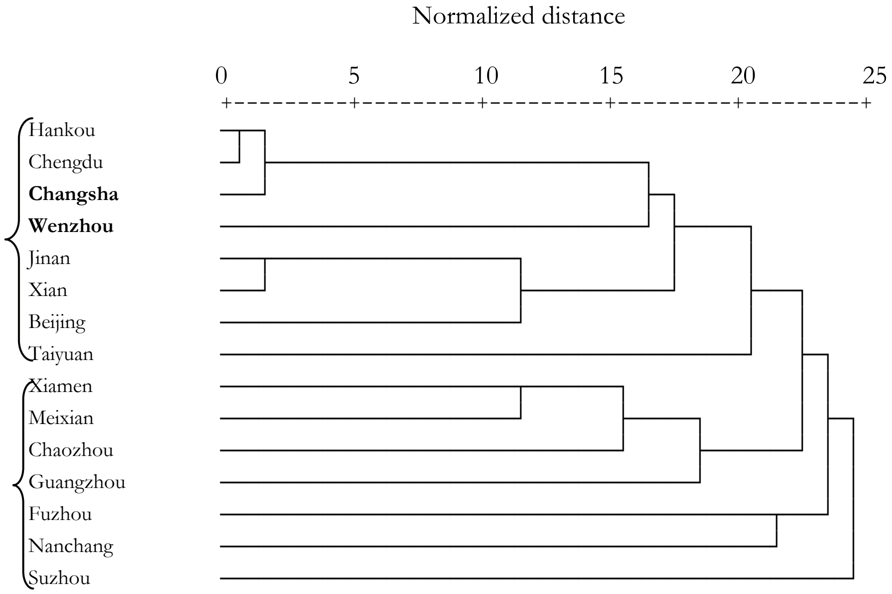

Figure 5.9. Dendrogram based on lexical frequency of finals for 15 selected dialects from CASS database, using Average Linkage (Between Groups) and chi-square as the distance measure. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

5.3.3.3 Lexical frequency of codas in the CASS database

The last part I did was to separate the coda from the other segmental parts. This procedure is followed as previous. However, as we know, in Sinitic languages, there are narrow restrictions on possible codas. As a result, only a very limited set of phonemes are qualified for the coda position. In my selected 15 dialects, finally there are 11 codas in total (including the zero or empty coda). In most cases, of course, the coda position is empty. The most frequent non-empty coda is the velar nasal, which occurs 222 times in the list for Chaozhou (for details see Appendices 5.10a and 5.10b). The coda frequency tree is in Figure 5.10.

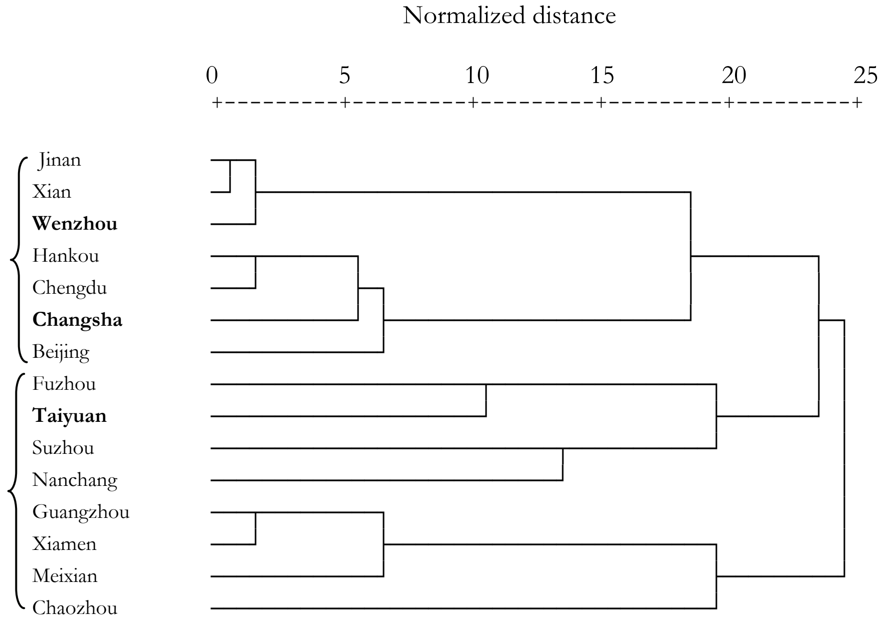

Figure 5.10. Dendrogram based on coda frequencies for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and chi-square as the distance measure. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

Apparently, the coda tree yields poorer results than the initials and finals. The primary split classified eleven dialects into the upper branch, whereas only four dialects fall into the lower branch. By using the more relaxed criterion, the optimal lower branch still only comprises seven non-Mandarin (Southern) dialects; the other two non-Mandarin (Southern) dialects (Wenzhou, Changsha) go to the upper branch which includes all the Mandarin dialects, except Taiyuan. The number of classification errors is therefore three.

5.3.3.4 Lexical frequency of tones in the CASS database

As before, the total number of different tones across the set of 15 dialects is 28 (see Appendices 5.11a and 5.11b). Some tones are much more frequent than certain others. The most frequent tone in any dialect is the 213 tone in Hankou. The well-known four tones of Beijing Mandarin occur in rather uneven frequencies in our list of 764 items: Tone 1 ('55'): 206, Tone 2 ('35'): 207, Tone 3 ('214'): 95 and Tone 4 ('51'): 256. On the basis of the lexical tone frequencies a tree structure was generated that is shown in Figure 5.11.

The tree does not afford a meaningful split into subgroups that correspond to dialect groups. Several solutions are possible to the problem of how to cut up the tree into dialectologically meaningful parts.

The primary split in the tree is between the bottom four dialects (two Mandarin and two non-Mandarin (Southern) dialects) and the others, yielding six classification errors in all. Alternatively we split off an upper (though embedded) branch comprising the upper four dialects, which again includes two Mandarin and two non-Mandarin (Southern) dialects, yielding six errors. Using a more relaxed grouping criterion, we may add the embedded cluster containing Jinan, Hankou and Nanchang to the lower branch. This would yield one group of seven with four Mandarin dialects, and another group of eight containing mainly non-Mandarin (Southern) dialects. The number of classification errors would still be as high as five. Clearly, then, lexical frequencies of tones do not reflect any linguistic taxonomy.

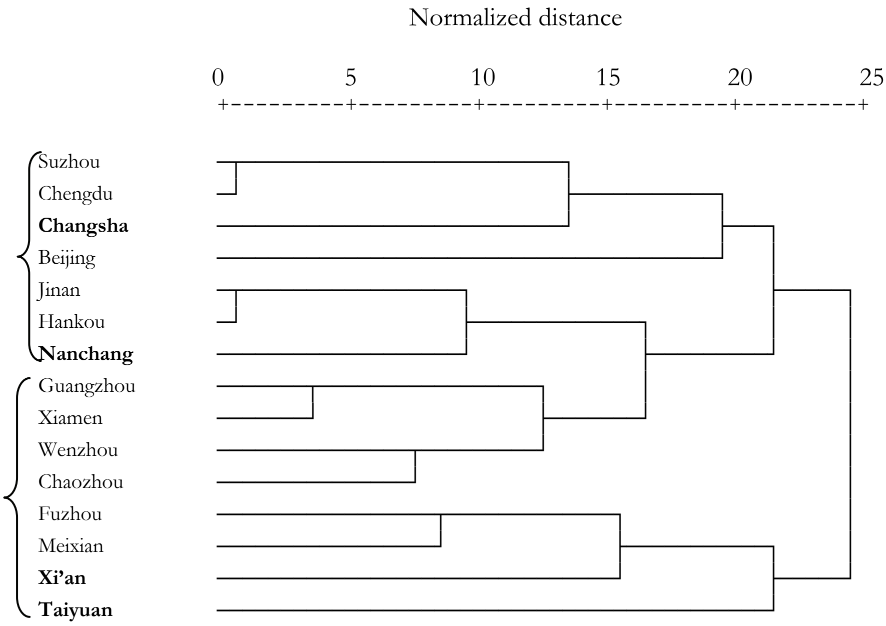

Figure 5.11 Dendrogram based on tone frequencies for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and chi-square as the distance measure. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

5.3.3.5 Lexical frequency of vocalic nuclei in the CASS database

I computed the frequencies of the vocalic nuclei in the CASS database, by counting the number of different sound shapes that remained after removing the coda from the finals. In total, there are 121 different nuclei (see Appendices 5.12a and 5.12b). The lexical frequency of nucleus types ranges from 0 to 177, with the highest frequency found for /i/ in Xiamen. The resulting tree is presented in Figure 5.12.

The nucleus tree reflects the primary split between the Mandarin and non-Mandarin (Southern) dialects well compared with the traditional dialect taxonomy although less better than the onset (initial) frequency tree concerning the internal subclusters. There is a clear cut between an upper group which contains all the Mandarin dialects plus one non-Mandarin (Southern) (Changsha) and a lower group which exclusively comprises all the non-Mandarin (Southern) dialect except Changsha. The number of classification errors is one.

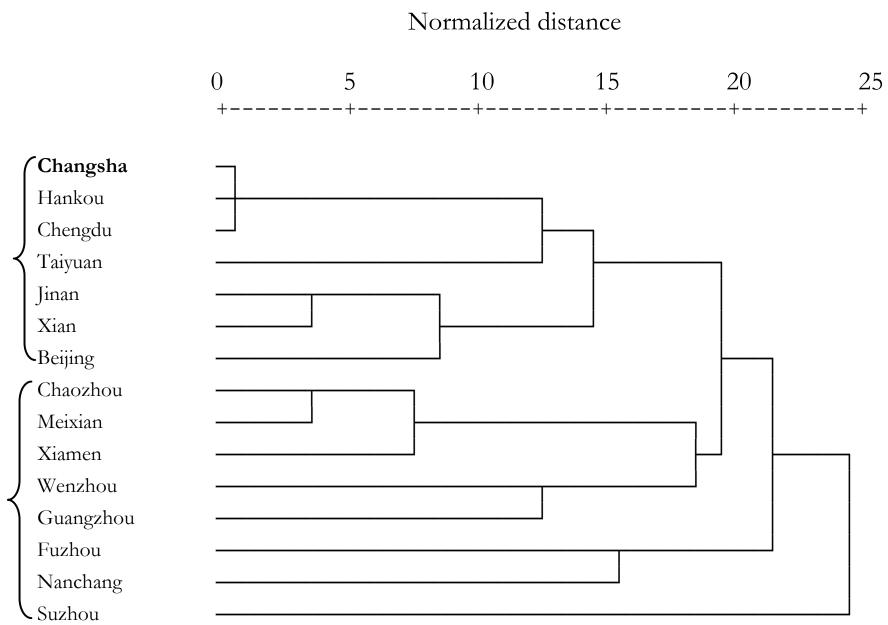

Figure 5.12. Dendrogram based on nucleus frequencies for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and chi-square as the distance measure. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The next procedure will deal with the combinations: the onset + finals and onset + finals + tones, respectively.

5.3.3.6 Lexical frequency of onset-final combinations in the CASS database

The first combination contains segmental parts only, i.e. onsets plus finals. I added the two sets of frequency data together yielding a total of 300 lexical frequencies (38 onsets plus zero onset and 262 finals in narrow transcription) (see Appendices 5.13a and 5.13b). Following the same steps as before, a hierarchical tree structure was then generated; it is shown in Figure 5.13.

The primary split reflects the traditional dialect taxonomy reasonably well. It suggests an upper branch comprising all Mandarin dialects plus one non-Mandarin (Southern) dialect (Changsha). All other non-Mandarin (Southern) dialects are parsed with the lower branch. The tree yields one classification error.

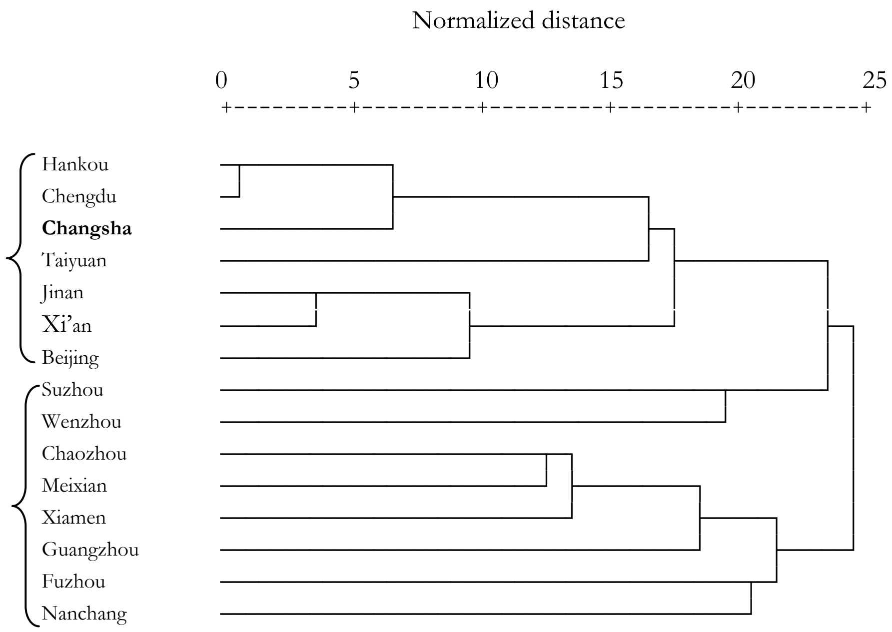

Figure 5.13. Dendrogram based on onsets + finals combination frequencies for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and chi-square as the distance measure. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

5.3.3.7 Frequency of onset-final-tones combinations in the CASS database

The last frequency count will be the union of the segmental and tonal parts: that is, the onsets along with the finals and the tones together. There are 328 lexical frequencies when the combination of onsets + finals (300) are added to 28 tone frequency tokens (see Appendices 5.14a and 5.14b). The tree structure generated can be seen in Figure 5.14.

The agglomeration tree contains an (embedded) upper branch that contains mainly Mandarin dialects (plus Changsha). The remainder of the tree contains all other Southern dialects plus one Mandarin dialect (Taiyuan). The total number of classification errors is two.

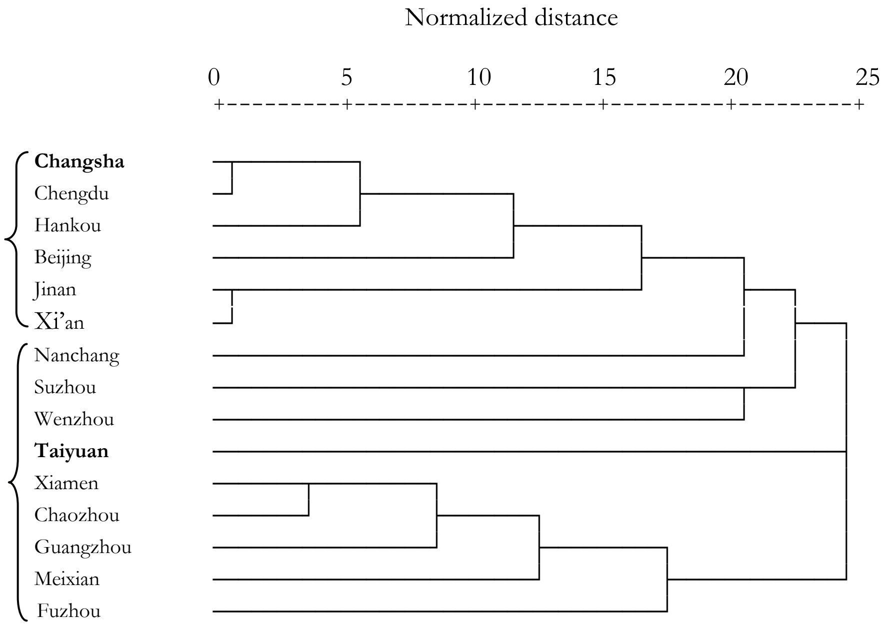

Figure 5.14. Dendrogram based on onsets + finals + tones combination frequencies for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and chi-square as the distance measure. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

5.3.3.8 Concluding remarks on the trees based on lexical frequencies

Viewing all the frequency results based on CASS data, as summarized in Table 5.3, we find that the best reflection of the traditional dialect taxonomy lies in the nucleus frequency tree (1 error). Reasonable trees are obtained from the lexical frequencies of onsets, of finals and of their combination. As before, when tones are taken into consideration, the error numbers increase.

I correlated the number of types at each linguistic level (and combinations of levels) with number of classification errors. Again, the result shows that there is no significant correlation, $r = -.559$ ($p = 0.192$, ins.), $\varrho = -.636$ ($p = 0.125$, ins.).

Table 5.3. Amount of information on CASS frequency taken into account and number of classification errors of dialects into Mandarin versus non-Mandarin (Southern) groups.

Type of information	N of types	Misclassifications
codas	11	3
tones	28	5
onsets	38	2
Vocalic nuclei	121	1
finals	263	2
Onsets + finals	300	1
Onsets + finals + tones	328	2

5.3.4 Levenshtein distance measures

As explained above, the Levenshtein distance is a string distance measure that is based on the number of string operations (insertion, deletion, substitution) needed to convert the phonetic transcription of a word in language A to its counterpart in language B (or vice versa).

The Levenshtein distance measure has proven to be successful for measuring phonetic distances between Dutch dialects (Heeringa 2004), and successfully validated against perceived distances between pairs of Norwegian dialects obtained experimentally (Gooskens & Heeringa 2004). May Levenshtein distance-based dialect distances between Chinese dialects be considered as a good approximation of the perceptual distances judged by dialect speakers?

We computed Levenshtein distances using the LO4 software package developed at Groningen University.[61] We computed Levenshtein distances between all pairs in our set of 15 dialects, once with and once without applying some perceptual weighing of sound differences. In the unweighed distance measure, any difference between two sounds is considered of equal weight. When perceptual weighing was applied, we used the number of distinctive feature levels that differed between two sounds as the weighing criterion. Here insertions and deletions were weighed at 50% of the maximum distance between either two consonants or between two vowels. Details of the weighing procedure can be found in Appendices 5.15a and 5.15b.

[61] The software can be downloaded from http://www.let.rug.nl/kleiweg/LO4/

Levenshtein distances were computed for the CASS database of 764 common morphemes in each of our 15 dialects. A problem in the case of Chinese dialects is that we have no way of knowing how tonal differences should be weighed against segmental differences. For this reason we decided to compute Levenshtein distances separately for the segmental and tonal properties of the morphemes. We will then later compare to what each of these domains contributes to intelligibility scores (judgment or functional test scores) and allows us to reconstruct linguistic taxonomies.

5.3.4.1 Segmental Levenshtein distance, unweighed

The IPA transcriptions provided in the CASS database were stripped of all diacritic marks, leaving only sequences of base symbols. Then tonal information (3-digit tone sequences) were deleted from the transcriptions. The remaining broad phonemic transcriptions were then submitted to the Levenshtein algorithm included in the LO4 package using default settings, i.e. all segmental substitutions counted as 1 unit of distance and insertions and deletions as half a unit. The distance matrix resulting from this procedure can be found in Appendix 5.15c. The tree structure derived from the matrix is shown in Figure 5.15. The proximity matrix underlying the tree can be found in Appendix 5.15d.

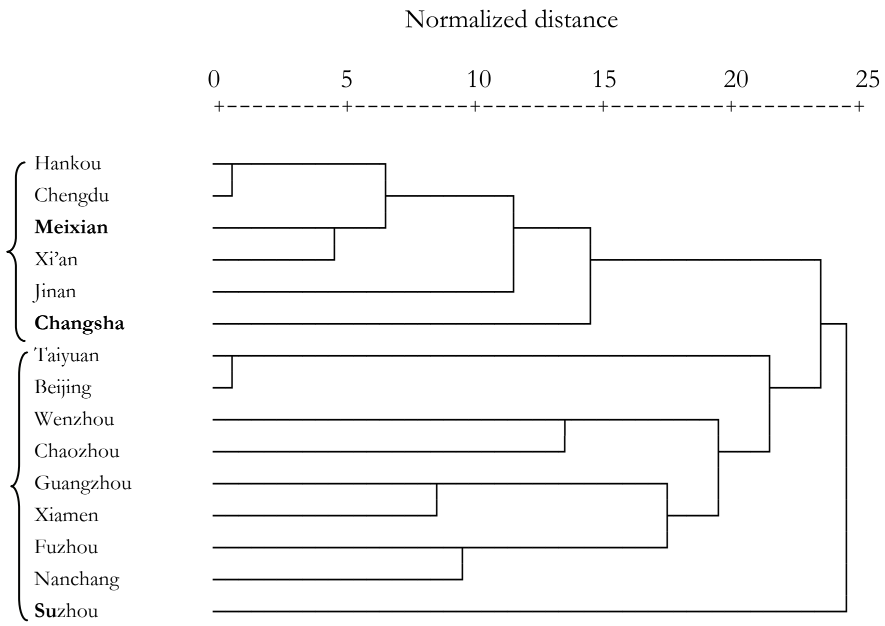

Figure 5.15. Dendrogram based on segmental Levenshtein distance for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and Euclidean distance measure because of the interval variables. No feature weighing was applied. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The tree can best be split in a lower branch comprising six non-Mandarin (Southern) dialects, and an upper part containing all six Mandarin dialects plus two incorrectly classified non-Mandarin (Southern) dialects. The affiliation of Suzhou is undecided between the two parts, so that the total number of classification errors amounts to 2.5.

5.3.4.2 Segmental Levenshtein distance, perceptually weighed

We ran the Levenshtein algorithm a second time, this time instructing the program to weigh all segment substitutions by the number of distinctive features that had opposed values. For details on the feature weighing procedure I refer to Appendix 5.16. As a result of the weighing, substituting /p/ for /ŋ/, for instance, would yield a greater distance than /p/ for /b/. The distance matrix generated by the LO4 software can be found in Appendix 5.16a. The tree structure derived from the matrix is presented in Figure 5.16. The proximity matrix underlying the tree can be found in Appendix 5.16b.

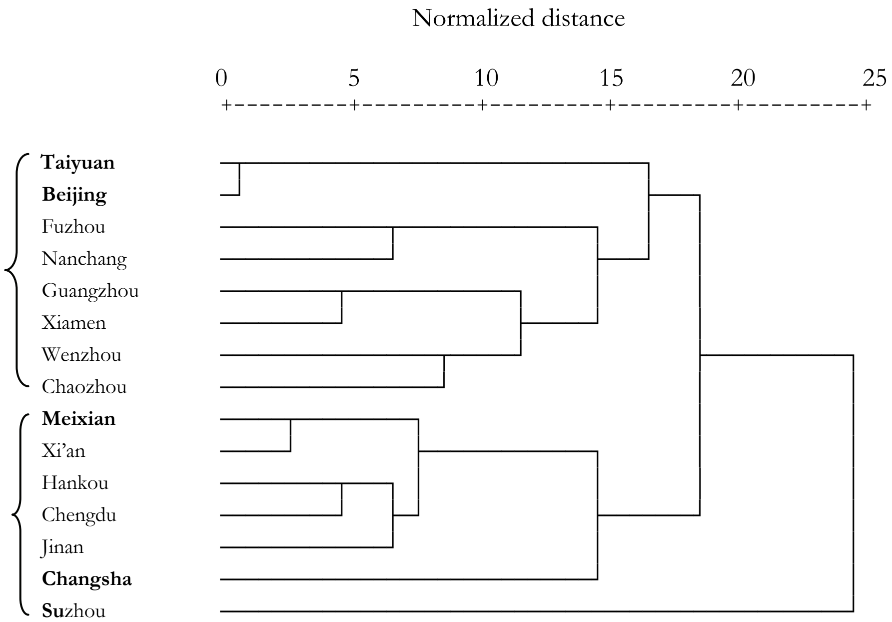

Figure 5.16. Dendrogram based segmental Levenshtein distance for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and Euclidean distance measure because of the interval variables. Sound differences were weighed by the number of features that differed between any pair of sounds. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The best split is between the upper eight dialects (mainly non-Mandarin/Southern but erroneously containing also Mandarin dialects Taiyuan and Beijing) and the lower seven comprising mainly Mandarin dialects (but also incorrectly including two non-Mandarin dialects – Meixian and Changsha). The affiliation of Suzhou cannot be decided, as it attaches as an isolate at the top level, so that the total number of classification errors yielded by this tree is 4.5. The feature-weighing operation has not led to any visible advantage here.

5.3.4.3 Tonal distance, unweighed

The tones in the CASS database are transcribed using the three-digit five-level system developed by Chao (1928). In this system the four tones of Mandarin are transcribed as 55 (high level tone), 35 (mid-rising tone), 214 (low dipping tone) and 51 (high falling tone). The tone digit sequences can be treated as strings with a maximum length of three, on which Levenshtein distances can be computed. When one member of a pair of tone strings is a single digit, this digit will be matched with the leftmost digit of a two-digit tone, and with the second digit of a three-digit tone. When a three-digit tone is compared with a shorter tone sequence, the second digit of the triplet (three-digit tone) will be matched with the first digit of the shorter string.

The Levenshtein distances between all six pairs of the four tones in Beijing Mandarin, for example, would be as follows (Table 5.4):

Table 5.4. Example: Levenshtein distances computed for all pairs of Mandarin tones.

Tone pair	Members	String operations	Relative Levenshtein distance	
(1 – 2)	55	1 substitution /	1 / 2	0.50
	35	2 alignments		
(1 – 3)	55	1 indel, 2 substitutions /	2.5 / 3	0.83
	214	3 alignments		
(1 – 4)	55	1 substitution /	1 / 2	0.50
	51	2 alignments		
(2 – 3)	35	1 indel, 2 substitutions /	2.5 / 3	0.83
	214	3 alignments		
(2 – 4)	35	2 substitutions /	2 / 2	1.00
	51	2 alignments		
(3 – 4)	214	1 indel, 2 substitutions /	2.5 / 3	0.83
	51	3 alignments		

In our analysis, of course, we did not determine the mean distance between the tones in the inventory of a single dialect (the differences between the tones in an inventory should always be large), but we computed the Levenshtein distance between the tone strings listed in the CASS lists for each of the 764 morphemes in any pair of dialects.

The distance matrix resulting from this analysis is included in Appendix 5.17a and the proximity matrix input in Appendix 5.17b. The hierarchical tree computed from the matrix is presented in Figure 5.17

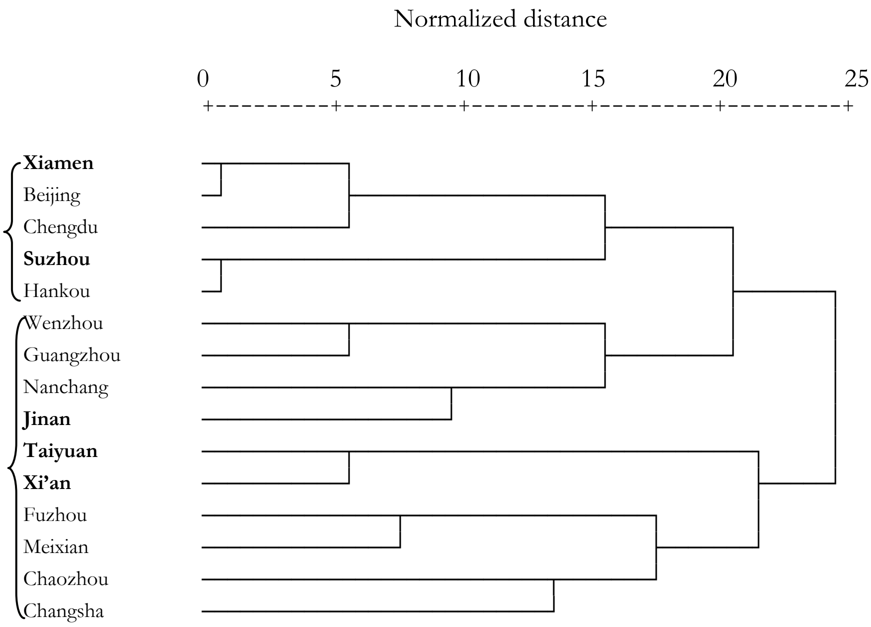

Figure 5.17. Dendrogram based on tonal Levenshtein distance (string matching for 3-digit tone transcriptions) for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and Euclidean distance measure. No weighing was applied. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The results seen in Figure 5.17 are highly confusing. There appears to be a more or less random spread of Mandarin and non-Mandarin (Southern) dialects over the branches in the tree, more often than not with Mandarin and a non-Mandarin (Southern) dialect as leaves on a binary subbranch. The best way I see to split the tree is to divide it into an upper group of five (mainly Mandarin) and a lower group of ten non-Mandarin (Southern) dialects (polluted with three Mandarin dialects). The total number of classification errors is 5.

Clearly, a simple string-edit distance on the tone transcriptions does not afford any insightful classification. We will now undertake a second attempt at computing a string distance measure for tone transcriptions. We will still use the number of string-edit operations as the distance measure but now the symbols in the strings are chosen so as to reflect some of the auditory characteristics of tone-language listeners. Following Yang & Castro (2009), who showed that results obtained with this method correlated best with mutual intelligibility of Tibeto-Burman and Tai-Kadai languages, we transformed the three-digit tone strings to sequences of two symbols. The first symbol (letter) represents the onset of the tone, the second the contour shape. We assume that Sinitic languages can be adequately described with three onset tone levels, viz. high (H), mid (M) and low (L). We further distinguished five contour types, viz. level (L), fall (F), rise (R), rise-fall (henceforth called peaked = P) and fall-rise (henceforth dipping = D). We computed Levenshtein distances on these onset-contour strings.

The resulting distance matrix is included in Appendix 5.18a; the proximity matrix is in Appendix 5.18b; the agglomeration tree is shown in Figure 5.18.

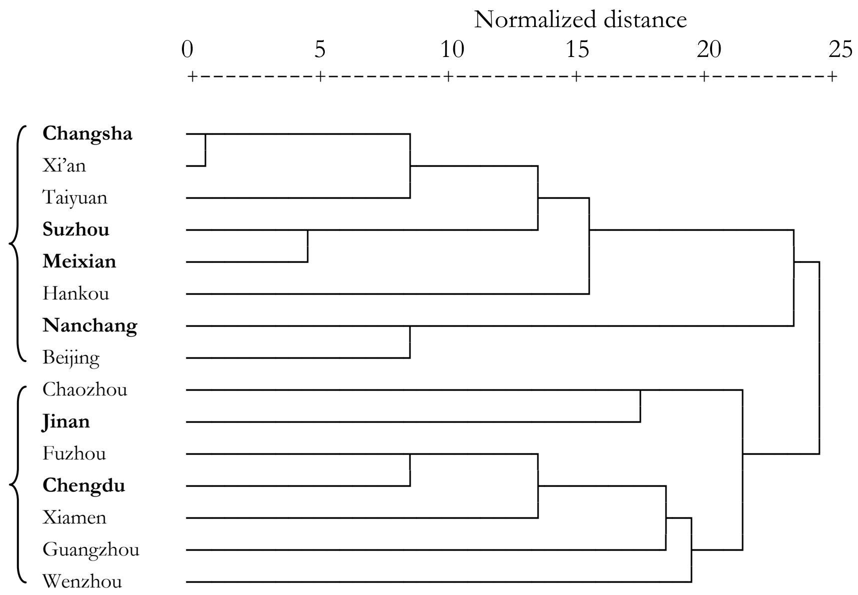

Figure 5.18. Dendrogram based on tonal Levenshtein distance (string matching of onset tone + tone change transcription, see text) for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and Euclidean distance measure. No feature weighing was applied. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

Again, this yields a very confusing tree, with a more or less random scatter of Mandarin and non-Mandarin (Southern) dialects. There are several equally poor solutions, which all result in six classification errors. We illustrate just one solution, which adopts the primary split between the upper group of eight and a lower group of seven dialects.

It is obvious that even a more sophisticated string-edit measure does not provide a handle on dialect affinity. We therefore made a third attempt which is explained in the next section.

5.3.4.4 Tonal distance, perceptually weighed

Both procedures used so far to compute the string distance between the tone digits are very crude and unrealistic, since they attach equal importance to any substitution of two tone levels or tone changes, whether the difference between the two pitches (or changes) is large or small. In order to come up with a more realistic distance metric for tone comparisons, we also computed a distance metric after perceptually weighting the various dimensions underlying the tonal space. Taking our cue from Gandour & Harshman (1978) we used five dimensions (tone features) as follows:

(a) *(Average) height*. We computed an average height (pitch) h for a tone as the mean of the (maximally) three tone digits. If $h > 3.5$ height was set to 5, if $h < 2.5$ it was set to 1; all values between (and including) 2.5 and 3.5 were set to 3.
(b) *Direction* was specified with three levels. Direction was defined on the last two digits in the tone string. Direction was set to 0 if the string contained just one digit or if there was no change in pitch level on the last two digits. Any falling pitch (on the last two digits) was given the value 1, and any rising pitch 2.
(c) *Duration* was also specified with three levels. Depending on the number of tone digits present in the string, duration was 1, 2 or 3 timing units (morae). Three-morae tones are always of the complex contour type (peaked or dipping), so that this feature covers more than just duration.
(d) *Slope* can be either steep or not. Steep slopes are found on tone strings with a difference of 3 or more tone levels (either up or down) on the last two digits. Steep slopes were specified as ‘1’, all non-steep slopes as ‘0’.
(e) The last feature specified was *extreme endpoint*. It was specified as ‘1’ if the final digit was either 1 or 5, and as ‘0’ for any other final digit.

This choice of values reflects different weight for three groups of features. Pitch is specified between 1 and 5 (spanning a range of 4), direction and duration are scaled between 0 and 2 (spanning a range of 2) whilst slope and extremity are either 0 or 1 (range of 1). As a result pitch : {duration, direction} : {slope, extremity} = 4:2:1.

The largest possible difference between any two tones specified with up to three digits in our perceptual weighing system is 10, i.e. the sum of the maximal differences for every feature, as exemplified in the following Table 5.5:

Therefore, we defined the perceptual distance between any two tone strings as the sum of the (implicitly weighted) feature differences divided by 10. As a result the perceptual distance between any tones is a fraction between 0 (no difference) and 1 (maximally different).

Table 5.5. Example of computation of perceptual distance between two tones.

Tone string	Height	Direction	Duration	Slope	Extremity	total
5	5	0	1	0	1	
214	1	2	3	1	0	
$\|\Delta\|$	4	2	2	1	1	10

Applying this computational procedure to the 764 tone strings for all pairs of our 15 dialects yielded a distance matrix (perceptually weighed tone distance) which is included in Appendix 5.19a and Appendix 5.19b is the proximity matrix. From this a tree structure was derived (using between-groups linkage and Euclidean distances) as shown in Figure 5.19.

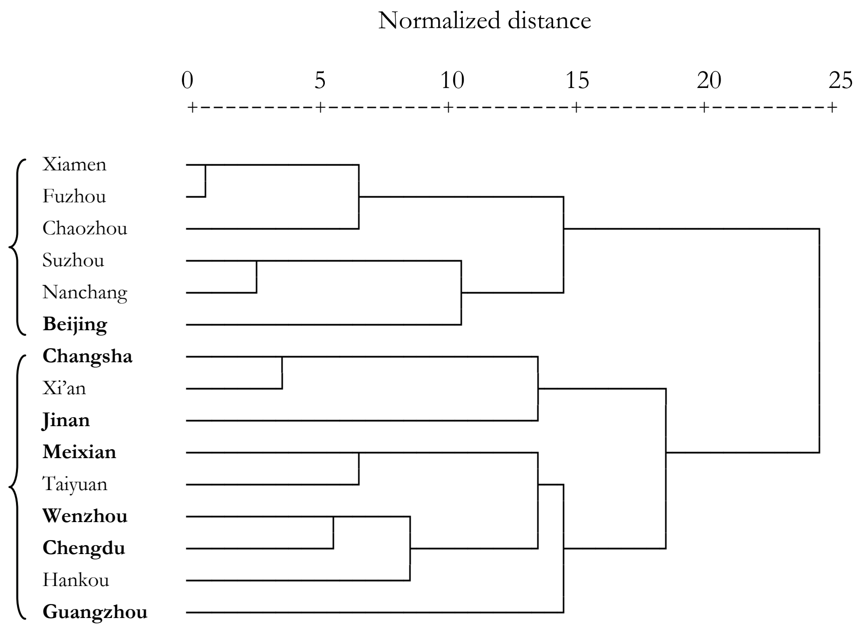

Figure 5.19. Dendrogram based on tonal distance (perceptually weighed, see text) for 15 selected dialects from the CASS database, using Average Linkage (Between Groups) and Euclidean measure as we did consistently in previous computation method. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The primary split in the above tree is also the optimal classification. The upper group comprises non-Mandarin (Southern) dialects only plus Beijing, whilst the lower branch contains the five remaining Mandarin dialects but also – erroneously – four non-

Mandarin (Southern) dialects. The number of classification errors is five in total. So, even our most sophisticated tone distance measure leaves us with highly unsatisfactory results.

5.3.4.5 Conclusions with respect to Levenshtein distance

In the above sections we attempted several measures of linguistic distance between pairs of our 15 target dialects, based on comparisons of phonetic transcriptions on a basic set of 764 highly frequent and productive morphemes shared by these dialects. We computed Levenshtein string edit-distance as a measure of affinity between dialects. In Table 5.6 I summarize the number of classification errors obtained with each of the five measures I computed.

Table 5.6. Number of classification errors yielded by various string-edit distance computations performed on the 764-item morpheme set of CASS.

Measure	Misclassifications
Segments, unweighed	2.5
Segments, feature-weighed	4.5
Tones, plain string-edit	5.0
Tones, string edit on onset+change	6.0
Tones, feature-weighed	5.0

Although such distance measures have often been reported to yield reasonable correspondences with traditional dialectological taxonomies (Heeringa 2004), the results of our attempts are disappointing. The crudest measure, i.e. unweighed segmental differences, leaves us with 2.5 incorrectly classified dialects, but any more sophisticated measure yields high numbers of classification errors. Other, conceptually and computationally easier distance measures (see Table 5.2), in fact, afforded better division of the Chinese dialects into Mandarin and non-Mandarin (Southern) than any of the string-edit distance measures presented here. As a case in point, just one single classification error was found when the criterion was simply the inventory of onset consonants shared between two dialects.

Correspondence in the tonal domain seems to be an especially poor criterion for determining affinity between Chinese dialects. Even our most sophisticated tone distance measure leaves us with highly unsatisfactory results. This should not come as a surprise to any expert on Chinese dialectology. The eight tones in Middle Chinese have developed along very different lines in the various dialects. The same level tone in MC (Middle Chinese) may have changed into a higher or lower level tone in one dialect, but into a rise, or fall in another dialect, or even into an even more complex shape. When we look at the tone inventories of the Chinese dialects today, there is no way to relate the present-day tone inventories to their counterparts in the ancestral language. The relationships between Modern Chinese tones (in any dialect) and their historical counterparts seems random. One way to come to grips with this problem is to determine the complexity of the (arbitrary) rule set that would be needed to convert the tones of one dialect into their counterparts in an other dialect (or to those in Middle Chinese). Such an attempt has been undertaken by Cheng (1997). We will review his results in § 5.3.5.

5.3.5 Measures published in the literature

Since 1980s, Cheng (1986, 1988, 1991, 1997) began to measure phonological affinity among 17 Chinese dialects based on the *Hanyu Fangyan Zihui [Word list of Chinese dialects]* (Beijing University 1962, 1989).[62] The *Zihui* provided transcriptions of over 2,700 words across 17 Chinese dialects, including the Middle Chinese phonological categories plus tone information. Cheng listed the initials (onset consonants) in all the dialects and tabulated their frequency of occurrence in each dialect. His earlier work on frequency of occurrence ignored the different reflections from historical sources (Middle Chinese). Later he separately tabulated the phonemes from different Middle Chinese origins in order to maintain the historical relations (Cheng 1991). In the case of multiple transcriptions for the same lexical entry (reflecting, for example, the literary-colloquial contrast), the colloquial pronunciation (or the first alternative listed in the case of multiple alternatives) was chosen so that consistency in pronunciation was guaranteed.

On this dataset of digital transcriptions, Cheng computed five measures of phonological affinity between all pairs of his 17 dialects. The first measure is based on the correlation of the lexical frequencies of the initials only (470 different types).[63] The second measure uses the lexical frequencies of the finals (rhyme portions of the syllables, 2770 different types). The third measure only considers the lexical frequencies of the tone transcriptions (133 different tone transcriptions). The fourth measure is based on the segmental transcription of the initials and finals combined (470 initials + 2770 finals = 3240 different transcriptions). The fifth and last measure is the combination of the previous one plus the 133 tone transcriptions (3373 different transcriptions). We will now present the results for each of these five measures of phonological affinity, and see how well each measure reflects the linguistic taxonomy proposed for the dialects. As said before, we will only check to what extent the primary division between Mandarin and non-Mandarin (Southern) dialects is reflected in the trees generated from the affinity measures.

[62] These 17 dialects are Beijing, Jinan, Xi'an, Taiyuan, Hankou, Chengdu, Yangzhou, Suzhou, Wenzhou, Changsha, Shuangfeng, Nanchang, Meixian, Guangzhou, Xiamen, Chaozhou, and Fuzhou. This selection of the 17 dialects is the superset of our 15 dialects (only Yangzhou and Shuangfeng are excluded)

[63] 'Lexical frequency' refers to the number of times a particular phenomenon occurs in the lexicon (here the Zihui list of 2,770 entries), where each occurrence in the list counts as 1. This is in contrast with the notion of 'token frequency' where the occurrences are counted in a corpus of texts and where the same lexical item may occur multiple times.

5.3.5.1 Phonological affinity based on initials

In the digital *Zihui* there were 470 different initial types, of which the occurrences could be counted in the word list. Correlation coefficients were then computed for each pair of dialects on the basis of the lexical frequencies of the 470 initial types. Using the same average linking method, an affinity tree was generated from the matrix scores of the correlation coefficients (as indicators of the degree of closeness between pairs of dialect). Appendix 5.20a present the correlation coefficients *r* based on the lexical frequency of initials, the proximity matrix is Appendix 5.20b. I only copied the scores of the fifteen dialects that overlapped with my 15 target dialects (Yangzhou and Shuangfeng were excluded). Figure 5.20 is the tree structure generated from Appendix 5.20.

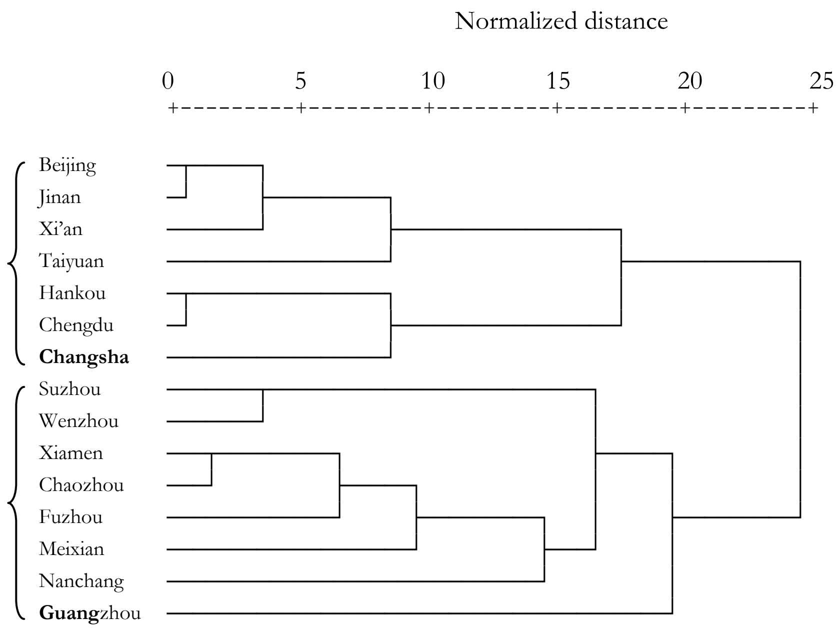

Figure 5.20. Dendrogram based on correlation of lexical frequencies of syllable-initials, using Between Groups linkage and chi-square measure because of the counts variables and chi-square as the distance measure. Incorrectly classified dialects have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces. The classification errors are bolded.

This tree reveals that Guangzhou is very different from all other 14 dialects in the sample. For the evaluation of the suitability of the initials as a measure of dialect affinity, I propose that we ignore the isolate attachment of Guangzhou, and consider this part of the lower branch. We then see binary split between seven dialects in the upper branch and eight in the lower branch. The upper branch contains all the Mandarin dialects plus Changsha, whilst the lower branch contains non-Mandarin (Southern) dialects only. This leaves us with just one plainly incorrect classification (Changsha, bolded in Figure 5.20). The classification of Guangzhou is ambiguous in this tree. Since it is an isolate, it can be classified with either group. I suggest that the total number of incorrect classifications is therefore 1.5 (the half error is half bolded, i. e. by the first syllable of Guangzhou).

5.3.5.2 Phonological affinity based on finals

Cheng treated the finals along the same lines as explained for initials (§ 3.3.1) and reflected their historical correspondences as well. The rhyme (*yun*), the lip rounding (*kai/he*, literally openness/closeness), and division (*deng*) in Middle Chinese were reflected by the entire final unit (medial, main vowel, and ending) in modern Chinese and these was taken into consideration in Cheng's alignment of occurrence distribution. Altogether, there were 2,770 cases of occurrence patterns tabulated. The correlation coefficients can be found in Appendix 5.21 and the affinity tree is presented in Figure 5.21

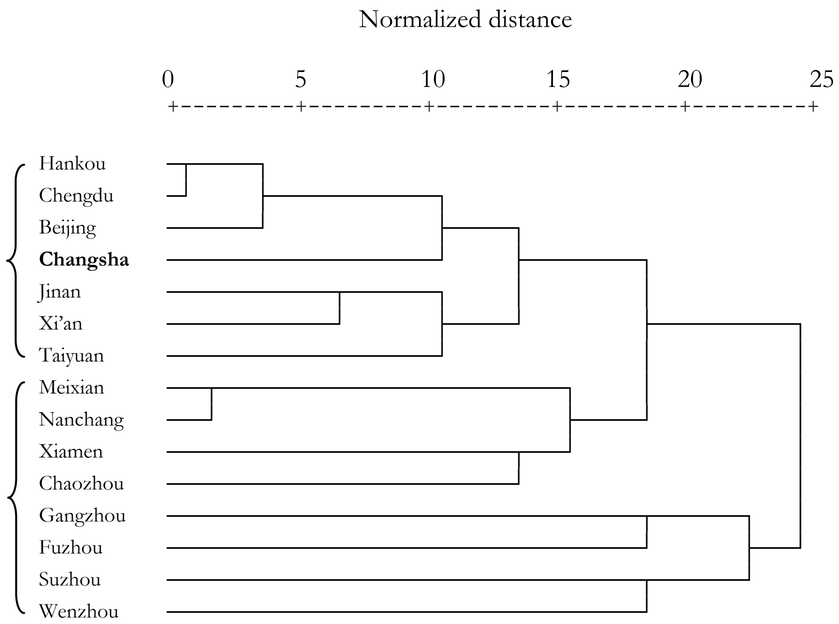

Figure 5.21. Dendrogram based on correlation of lexical frequencies of syllable-finals, using Between Groups linkage and chi-square measure because of the interval variables and Euclidean distance as the measure. Incorrect classifications have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces. The classification errors are bolded.

This tree breaks down into an upper branch of eleven dialects and a bottom branch of four. The 11-dialect cluster then breaks up into one lower cluster that comprises non-Mandarin (Southern) dialects only, and an upper cluster containing all Mandarin dialects plus one non-Mandarin (Southern) dialect, which – again – happens to be Changsha. Using the relaxed criterion, I suggest that the lexical frequency tree based on syllable-finals yields one classification error, i.e. Changsha, which was also found in the tree based on syllable-initials.

5.3.5.3 Phonological affinity based on tone transcription

The cases on tone affinity were more complicated compared to the initials and finals. Cheng applied the similar treatment taking the four Middle Chinese tones *ping*, *shang*, *qu*, *ru* and the three Middle Chinese initial classes, namely, voiceless consonants, voiced obstruents and sibilants into consideration. In total, he distinguished 133 different tone transcriptions. Their frequencies were counted in the 17 dialects. The correlation coefficients and the linked tree structure are given in Appendix 5.22 and Figure 5.22.

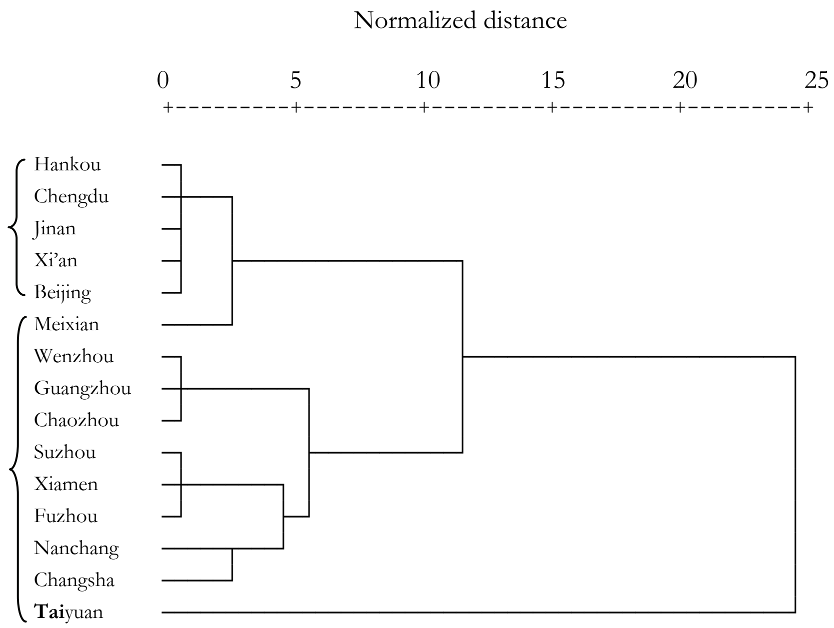

Figure 5.22. Dendrogram based on correlation of lexical frequencies of tones, using Between Groups linkage and chi-square measure because of the interval variables and Euclidean distance as the measure. Incorrect classification are bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

Let us ignore the isolate primary branch for Taiyuan for the moment. The tree then splits into an upper cluster of five Mandarin dialects plus Meixian. The lower branch comprises all other non-Mandarin (Southern) dialects. Using the more relaxed criterion, however, we may consider the upper five dialects as one group (Mandarin) and moving Meixian to the non-Mandarin group. The classification of Taiyuan is undecided, as it is parsed neither with the non-Mandarin (Southern) cluster nor with the (almost exclusively) Mandarin cluster. As a practical solution, I suggest, as before, that this yields a total number of 0.5 incorrect classifications. The undecided status of Taiyuan is marked in Figure 5.22 by highlighting only its first syllable.

5.3.5.4 Phonological affinity based on initials+ finals combined

The composite results of initial and finals was also analyzed (see Appendix 5.23). The corresponding affinity tree is given in Figure 5.23.

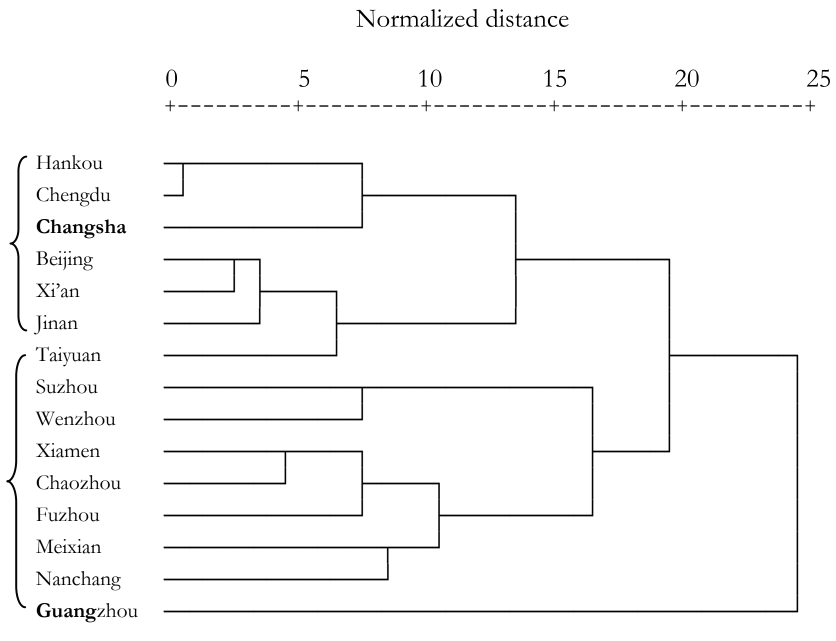

Figure 5.23. Dendrogram based on correlation of frequencies of combinations of syllable-initials and syllable-finals, using Between Groups linkage and chi-square measure because of the interval variables and Euclidean distance as the measure. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces. The classification errors are bolded.

This tree reflects the division into Mandarin and non-Mandarin (Southern) dialects quite well. The upper cluster comprises all six of our Mandarin dialects plus a non-Mandarin (Southern) dialect (Changsha), whilst all non-Mandarin (Southern) varieties are found in the lower part of the tree except Changsha. There is, however, one problem, which is that Guangzhou is parsed as an isolate, so that, as was also the case when the grouping was done on the basis of syllable-initials alone, its classification remains undecided. As before we mark this as 1.5 classification error.

5.3.5.5 Phonological affinity based on segmental + tonal transcriptions

The last part of Cheng's work on affinity measures based on the phonology of the syllable is the correlation between the lexical frequencies obtained for all segmental types (i.e. all initials and all finals) augmented with the 133 tone types. The total composite results combined with the initials, finals and tones are presented in Appendix 5.24 and the corresponding affinity tree in Figure 5.24.

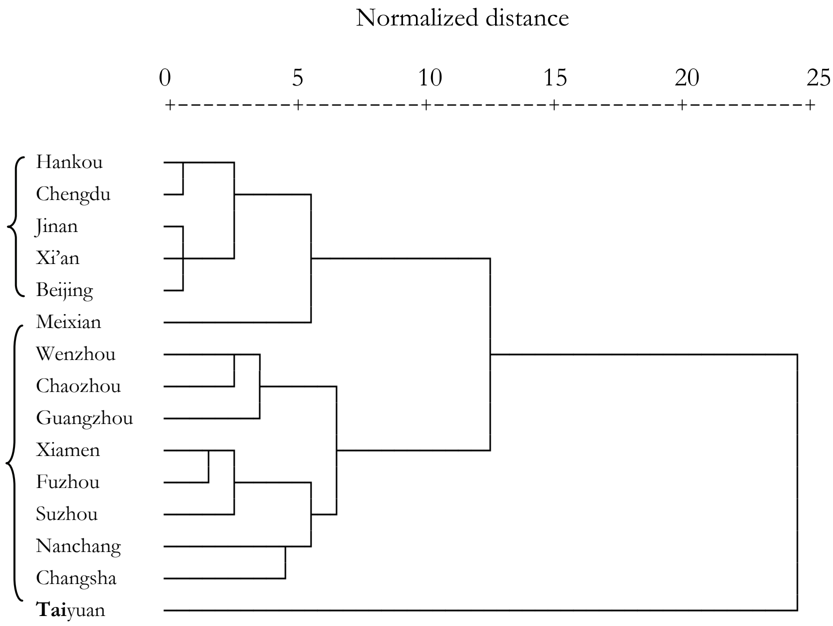

Figure 5.24. Dendrogram based on correlation of frequencies of combinations of syllable-initials, syllable-finals and syllable-tones together, using Between Groups linkage and chi-square measure because of the interval variables and Euclidean distance as the measure. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces. Classification errors are bolded.

Although this tree is based on more, and more diverse, information concerning syllable structure, it does not afford a clean split of the 15 dialects into the Mandarin and non-Mandarin (Southern) groups. First, the status of Taiyuan remains undecided since it is parsed as an isolate before any clusters are apparent. Within the remaining 14 dialects there is an upper cluster with five Mandarin dialects plus non-Mandarin (Southern) Meixian. The lower cluster contains non-Mandarin (Southern) dialects only. Using a lenient criterion, I suggest that this tree contains 0.5 misclassification, i.e. Taiyuan – which should be Mandarin according to the traditional taxonomy but is undecided. The other five Mandarin dialects constitute an integral cluster, so that any dialect not in this cluster would automatically be classified as non-Mandarin (Southern).

Using the very crude counting of wrong and undecided classifications yielded by the affinity trees, it seems that there is a clear relationship between the amount of information taken into account when building the tree and the quality of the classification. This relationship can be seen in Table 5.7 which lists the amount of information, and number of classification errors.

Table 5.7. Amount of information on syllable structure taken into account and number of classification errors of dialects into Mandarin versus non-Mandarin (Southern) groups.

Type of information	N of types	Misclassifications	Undecided	Total errors
Tones	133	0	1	0.5
Initials	470	1	1	1.5
Finals	2770	1	0	1
Initials + Finals	3240	1	1	1.5
Initials + Finals + Tones	3373	0	1	0.5

Table 5.7 shows that there is no tendency for the classification performance of Cheng's phonological affinity measures to get better (fewer errors) as the number of types counted increases. The correlation coefficient between the two variables is $r = 0.033$ ($p = 0.959$, ins.) or $\rho = 0.000$ ($p = 1.000$, ins.).

5.3.5.6 Cheng's phonological affinity based on correspondence rules

In the phonological affinity measures presented and evaluated above, the distance between dialect A and B is always the same as that between B and A. The distance matrices are therefore symmetrical around the diagonal (for this reason, the redundant upper halves of the matrices were not printed). Although such symmetrical measures capture at least part of the affinity between pairs of dialects, it is obvious that they do not account for the full complexity of *mutual* intelligibility. Very often it is the case that listeners of dialect (or language) A understand speakers of dialect (language) B better than vice versa. Danes understand Swedes better than the other way around, and the same relationship has been found between Portuguese and Spanish. One important cause of such asymmetrical relationships is that the corresponding sounds in two languages may be different. For instance, Dutch and English share many cognates. When the final obstruent in the English cognate is voiced it will always be voiceless in the Dutch counterpart. This relationship can be captured with a single rule. So it is relatively easy for a Dutch listener to abstract this regularity from just a few examples. On the other hand, it is much more difficult to establish the correspondence for final obstruents when an English listener listens to a Dutch speaker. For some words he has to replace the final voiceless obstruent by its voiced counterpart in English, for others he should not. In this (tiny) part of the phonology English is easier for Dutch listeners than vice versa.

In his more recent work, Cheng (1997) recognized this problem, and proposed a complex algorithm that computes the degree of regularity in the sound correspondences that hold between the sets of cognates shared by pairs of Chinese dialects.

In order to measure the asymmetrical phonological distance between pairs of dialects, Cheng's first step was to establish the correspondence patterns based on syllable-words. With the assistance of the DOC, Cheng (and colleagues) segmented the syllable into the traditional categories of initial, medial, nucleus, ending and tone for a more precise tabulation of correspondence (for details on the DOC, I refer to Cheng 1997).[64] Firstly, each of these elements was given an equal weight of 0.2, so that the five elements add up to a maximum of 1. For each cognate, the corresponding items for each of the five elements in a cognate were extracted. As a result, each cognate had five corresponding paired items. The items of all the cognate words were then tabulated, giving a lexical frequency count of the correspondences. Then the two categories of each tabulated item were compared to determine whether the correspondence constitutes signal or noise. If the correspondence patterns are entirely the same or mostly identical, then they are treated as recognition enhancement and thus as 'signal' in communication, otherwise, they are recognition interference and thus constitute 'noise' in communication. Cheng calculated the mean of each pair of corresponding elements and determined the weight value according to their attributes of signal or interference compared to the mean (for details, I refer to Cheng 1993). He assigned weight values ranging from +/– 0.05 to +/–0.20 according to their corresponding distribution. These weight values should be multiplied by the frequency of occurrence. The result for each corresponding pattern pair is the cumulative sum. At the end of the calculation of all syllables, the sum is divided by the total number of correspondence item to yield a fraction between 0 and 1. This sum obtained is interpreted as unidirectional distance/similarity between dialect A and B. Then, another unidirectional distance/similarity between dialect B and A was calculated following the same procedure. The correspondence patterns for all combinations of non-identical speaker and listener dialects among the set of 17 were established, making 272 pairs (17 × 17 = 289 minus 17 cases where speaker and listener dialect is the same). The distance/similarity measure (the phonological correspondence index, i. e. PCI, cf. Tang and Van Heuven 2007, 2008)

[64] DOC stands for Dictionary On Computer. It is an encoded computer file based on the data in the *Hanyu Fangyin Zihui* (Chinese Dialect Character Pronunciation List, henceforth *Zihui*), which was compiled by Futang Wang, in the Linguistic section of the Chinese Linguistics & Literature department at Beijing University in 1962 and revised in 1989. In the initial *Zihui* the sound shapes were collected of over 2,700 words in 17 dialects (see note 2). Later the materials were revised and extended to 20 dialects in 1989. For each word, the Middle Chinese phonological categories are also provided. Accordingly, the *Zihui* is reputed for its historical depth, geographical breadth, and coverage of tones (1962, 1989). Thus, it is worthy of being explored by means of a computer database (DOC) for various linguistic researches. (Cheng 1991, 1992, 1993a, 1993b, 1995, 1996, 1997 etc). So the current DOC file is based on the first edition of the *Zihui* with various modifications over years.

was yielded by taking the mean of the two unidirectional values for each pairs of dialects.

Appendix 5.25 copies the indexes from Cheng (1992). His set of 17 dialects is a proper superset of our 15 target dialects, so we just omitted the two dialects (Yangzhou and Shuangfeng) that did not occur in our own sample.

Appendix 5.25 shows that the intelligibility between dialects A and B is never exactly the same as between B and A. There is, in fact, a small but statistically significant effect, that the phonological correspondence between the dialects listed in the rows in Appendix 5.25 with the dialects listed in the columns is better (by 0.009 point) than their contradiagonal counterparts. More importantly, however, there is an overall tendency for the PCI scores in contradiagonal cells to be strongly correlated, $r = 0.874$, N = 105 (i.e. excluding the diagonal) and $p < 0.001$, so that mean of the AB and BA scores is never very different from either AB or BA. This, in turn, allows us to compute an affinity tree using the method of average linking – which presupposes a symmetrical matrix.

Figure 5.25 is the tree structure generated from Appendix 5.25 by cluster analysis. Before I generated the tree, I averaged the contradiagonal cells *i, j* and *j, i*.

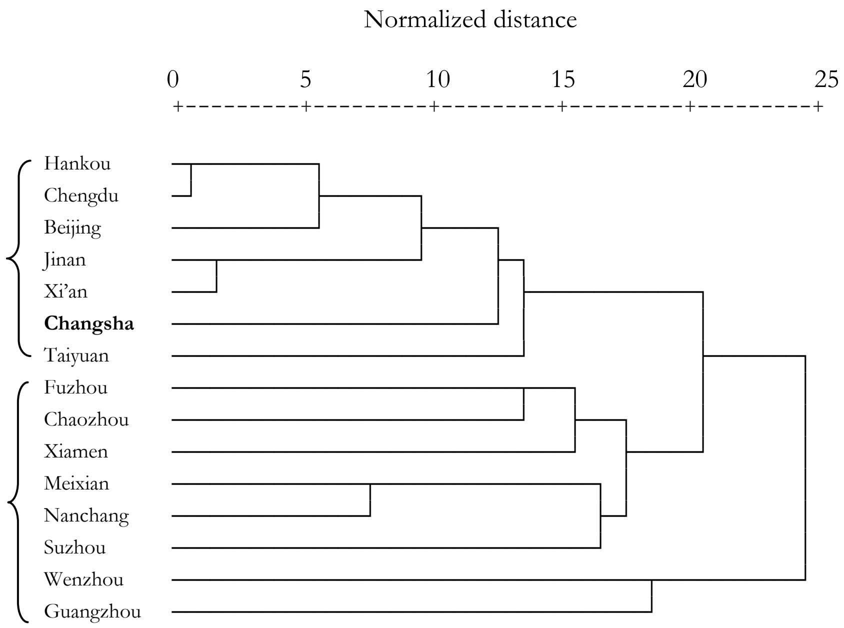

Figure 5.25. Dendrogram based on phonological correspondences indexes using Average Linkage (Between Groups), using Between Groups linkage and Euclidean distance as the measure. Incorrectly classified dialects (as determined by relaxed criteria, see text) have been bolded. The optimal split between Mandarin and non-Mandarin (Southern) dialects is indicated by braces.

The primary split in this dendrogram is between a lower cluster containing eight non-Mandarin (Southern) languages and an upper cluster with all six Mandarin dialects plus Changsha. Changsha is attached to the cluster in such a way that it cannot be separated from it without breaking up the cluster, so it should count as a full classification error.

In conclusion we may briefly summarize this review of Cheng's work on Chinese dialect affinity by saying that none of his measures perfectly reflects the primary division of Chinese dialects into Mandarin versus non-Mandarin (Southern) branches. Later, in chapter seven, we will consider the question how well these same measures may serve to predict *mutual* intelligibility among the 15 dialects.

5.4 Conclusions

In this chapter we have reviewed a fairly large number of measures computed to capture some objective distance between pairs of Chinese dialects. One distance measure is based on lexical similarity between dialects, all other measures relate to aspects of the sound structures of the dialects. The lexical affinity measure was copied from the literature. We had no resources at our disposal to determine lexical affinity between pairs of our 15 dialects ourselves. Phonological distance measures were either copied from existing literature or computed by ourselves, either on phonological inventories published in the literature or on a digital database that we obtained from the Academy of Social Sciences, Beijing. No information was available on syntactic distance between the dialects.

We applied a very crude criterion in order to determine the viability of the various distance measures. We assume that a criterion can be valid only if it is at least reasonably successful in reproducing the primary split between Mandarin and non-Mandarin (Southern) dialects. Using this criterion we will now select the more successful distance measures and list these in Table 5.8. The largest number of classification errors we obtained for any criterion, was six. We will consider a measure viable if it yields a classification into Mandarin and non-Mandarin (Southern) with fewer than three cases, i.e. we will keep only the distance measures that yield less than half the maximum number of errors found.

The best classification of the 15 dialects in our sample is afforded by a compound measure proposed by Cheng (1997), i.e. the affinity based on the lexical frequencies of onsets, finals and tones combined as counted in a 2,770 item list of common words in Chinese. This measure yields a perfect split of the dialects into Mandarin and non-Mandarin (Southern), with the exception of one dialect, Taiyuan, whose status remained ambiguous. I should point out, however, that in spite of the apparent success of this measure, it fails to reflect any of the internal taxonomy within the two main branches of Sinitic dialects.

Table 5.8. Summary of most successful objective distance measures (see text).

Source	Measure	Classification		
		Error	Undecided	Total
Inventories	Initials	1		1
	Codas	2		2
	Onsets + codas	1		1
CASS Lexical frequencies	Onsets	2		2
	Vocalic nuclei	1		1
	Finals	2		2
	Onsets + finals	1		1
	Onsets+finals+tones	2		2
CASS Levenshtein distance	Segments (unweighed)	2	1	2.5
Cheng (1997) Zihui word list, Phonological affinity	Onsets	1	1	1.5
	Finals	1		1
	Onsets + finals	1	1	1.5
	Tones	0	1	0.5
	Onsets + finals + tones	0	1	0.5
	Phonological correspondence	1		1
Cheng (1997) lexical affinity	Percent shared cognates	2		2

In the next chapter I will use these objective distance measures as predictors of mutual intelligibility. I will then correlate the affinity measures between each pair of dialects with the mutual intelligibility scores obtained in chapters three and four. In doing so, I will not only use the classification of the dialects into the two main categories but implicitly the subclassification will also play a role in the analysis.

Chapter Six

Predicting mutual intelligibility

6.1 Introduction

In earlier chapters I collected data within several distinct domains, all of which may be related with intelligibility among Chinese dialects. I started, in Chapter Three, by collecting from native listeners of 15 Chinese dialects judgments of linguistic similarity and intelligibility of these dialects. This enterprise yielded 225 combinations of speaker and listener dialects for which I now have a score for judged linguistic similarity and for judged intelligibility. We established, in Chapter Three, that judged intelligibility can be predicted rather well from judged linguistic similarity (and vice versa) with r = 0.888 with symmetrical input of the matrix based on the melodic version, N = 105 (cf. Chapter Three). We would now like to know how well these judgment scores (intelligibility and similarity) can be predicted from objective distance measures such as those that we collected in Chapter Five.

Next, in Chapter four, I collected functional intelligibility scores for the same set of 225 combinations of speaker and listener dialects, using separate tests to target intelligibility at the isolated-word and at the sentence level. We then established, first of all, that these two functional intelligibility measures converged with r = 0.928; such convergence was expected since word intelligibility is a prerequisite to sentence intelligibility. Second, we wanted to know the extent to which functional intelligibility of Chapter Four (the 'real thing') could be predicted from the 'quick and dirty' judgment tests of Chapter Three. If near-perfect prediction is possible, we will not have to apply cumbersome functional tests in the future, but may rely on the more convenient judgment tests. The results reported in Chapter Four reveal that the correlation between the functional word and sentence intelligibility scores and the intelligibility judgment scores is good (r = 0.772 and 0.818, respectively) but not good enough to advocate the unqualified use of judgment testing as a more efficient substitute for functional testing. In the present chapter we wish to determine how well the functional intelligibility scores, both at the word and at the sentence level can be predicted from objective distance measures such as those collected in Chapter Five. Moreover, I also want to know whether or not functional test scores can be better predicted from objective distance measures than opinion scores.

Finally, in Chapter Five, I collected a large number of so-called objective measures, all of which contain some information on similarity between (pairs of) Chinese dialects. I computed structural similarity measures based on a simple comparison of the sound and tone inventories of the 15 dialects, with and without weighing the sound units for their lexical frequency. I also determined to what extent words in all pairs of dialects are pronounced the same, separately for segmental and tonal aspects. This work was based on lists of phonetic transcriptions of 764 words (basic morphemes) in each of the 15 dialects made available by the Chinese Academy of Social Sciences. I also copied from the literature published measures of structural similarity between all pairs of my 15 dialects (Cheng 1997), determined on a much larger list of 2,770 words (or rather concepts) occurring in the dialects. Among the various measures published by Cheng there is one that deserves special attention: this is the only measure I have for lexical similarity among the dialects (percent cognates shared); all other measures relate to differences in sound structure (vowels, consonants, tones). I would now like to know to what extent all these structural similarity measures impart the same information, and, even more importantly, if these objective measures of similarity between dialects allow us to predict the experimentally-based intelligibility and similarity scores (the latter was only obtained from judgments).

The present chapter is an attempt to answer the various questions identified here. Before we go on, however, we need to discuss certain properties of the data. The great majority of the objective similarity measures collected in Chapter Five are symmetrical. That is to say, the distance between (the speaker's) dialect A and (the listener's) dialect B is the same as the distance between B and A. This may be compared with the travel time to go from village A to village B in a flat country: it will take as much time to travel from A to B as is does to travel back from B to A. However, some distance measures in our data are asymmetrical. This applies especially to all the measures we collected in chapters three and four, which involved the use of human subjects. It may well be the case that a speaker dialect A is more intelligible to a listener with dialect B than vice versa. Such asymmetries are well documented for certain European languages. For instance, we know that Portuguese listeners understand spoken Spanish quite well but Spanish listeners have great problems understanding spoken Portuguese (Jensen 1989). This is comparable to the travel distance between one village in a valley and another village high up in the mountains: it takes longer to walk uphill than downhill. When we relate the various measures discussed above to each other, we will make a simplification in the sense that we will treat all distances between speaker and listener dialects as symmetrical. We do this since our main interest is in mutual intelligibility between pairs of Chinese dialects. Mutual intelligibility was defined by Cheng (1997) as the mean of the intelligibility of speaker A for listener B and of speaker B for listener A. Obviously, if the intelligibility of A and B is not the same as that between B and A, averaging the AB and BA intelligibility scores eliminates the asymmetry. The averaging operation is performed on all pairs of contradiagonal cells i, j and j, i in the 15 (speaker dialects) by 15 (listener dialects) = 225 cells in the score matrices we collected. We then delete the redundant part of the matrices, keeping only the non-redundant lower triangle (without the main diagonal), and use the remaining 105 scores in the comparisons.

Comparisons between variables (intelligibility scores, linguistic similarity scores – whether subjective or objective) will proceed on the basis of correlation analysis. That is, I will compute correlation coefficients for all pairs of variables and present these in correlation matrices in order to identify groups of variables that provide the same or nearly the same information. In this way we will be able to address the questions raised above. At the end of each section we will consider possibilities to predict experimental intelligibility scores not from just one structural measure at the time but from multiple information sources together, using multiple regression analysis.

6.2 Predicting subjective ratings from objective measures

In Chapter Five we collected a total of 25 objective variables, all of which cover some aspect of linguistic similarity between pairs of our 15 Chinese dialects. One variable (LAI, Lexical Affinity Index) specifically targets the similarity in the lexicons of the dialects, i.e. the percentage of cognate word forms shared between a pair of languages (from a list of 2,770 concepts commonly used in Chinese languages, Cheng 1997), All other variables are concerned with aspects of the sound structure. Some of the variables list very simple properties, e.g. which onsets consonants are shared between the phoneme inventories of two dialects, whilst others (i. e. Phonological Correspondence) are of a more complex nature, such as the complexity of the rule system needed to convert phonological forms in one dialect to their cognates in the other dialect. Also, quite a few variables are of a compound nature. For instance the variable initials + codas is the aggregate of two variables that are also considered separately, viz. initials (37) and codas (15). Obviously, these two separate variables will correlate strongly with the aggregated variable, the initials more strongly so than the codas because the former take up a larger proportion of the aggregated variable.

We computed the correlation coefficients between each of these 25 objective distance measures and the two opinion scores (i.e. judged intelligibility and judged similarity) we collected in Chapter Three. In fact, each opinion score comes in two different guises: once intelligibility and similarity were judged on the basis of fully intoned speech samples, a second set of judgments was collected on the basis of monotonized versions of the samples. We also computed the correlation coefficients between any pair of objective distance measures (so-called intercorrelations). The full correlation matrix is included in Appendix 6.1. Here we will single out the more successful predictors of judged intelligibility and judged similarity from among the set of objective distance measures. We will also try to prune the set of correlations and zoom in on interesting predictors that correlate with the judgments in a non-trivial way.

6.2.1 Single predictors of judgment scores

The first group of predictors was obtained from Cheng (1997)'s frequency counts of initials, finals and tones in the 15 dialects plus two compound measures. The tonal frequency counts contain virtually no information that helps to predict the judgment scores. Also, the compound measures are no better predictors than at least one of the simplex measures that underlie them. This leaves just two of Cheng's frequency measures as potentially useful predictors of the judgments: the lexical frequencies of initials and of finals, with correlation coefficients around $r = 0.500$ for the initials and somewhat above $r = 0.600$ for the finals. Similarity judgments correlate better with these predictors than intelligibility judgments, and finals afford better predictions than onset frequency. There is no systematic difference between fully intonated versus monotonized speech. The two remaining lexical frequency measures (initials and finals) are moderately intercorrelated (r = 0.523), so that together they may afford substantial better prediction of the judgment scores than each predictor by itself.

The next two potentially useful predictors are Cheng's Lexical Affinity Index (LAI) and the Phonological Correspondence Index (PCI).[65] These two measures correlate very well with the judgment scores, with r-values around 0.850 for LAI and between 0.710 and 0.757 for the PCI. These two predictors are intercorrelated at r = 0.761, so that together they may yield a fairly good prediction of the judgment scores. The lexical frequency measures intercorrelate at low values with LAI and PCI, so that we may expect improved prediction of the judgment scores when we attempt to enter all these variables in a multiple regression. We will do this at the end of this section.

The third set of objective measures of linguistic distance between pairs of our 15 dialects is based on shared elements in the phonological inventories of the dialects. Generally correlation coefficients between these inventory-based measures and the judgment scores are fairly low for onsets and nuclei (r < 0.400) and only moderate for the codas (r-values around 0.550) and tone inventories (r-values around 0.450). These measures are only weakly intercorrelated with the earlier distance measures we discussed, so that some further improvement may be expected if we include the inventory based measures in a multiple regression analysis.

We now come to the objective distance measures that I computed myself on the 764-item common morpheme lists provided by the Chinese Academy of Social Sciences. Distance measures based on tones (also in compound measures) are poorly correlated with the judgment scores. I will not consider these tone-based distance measures any further. Distance measures based on onsets, nuclei and codas, as well as larger phonological units such as finals (nucleus+coda) correlate with the judgment scores at r-values between 0.400 and 0.600). The best prediction of judgment scores in this group is afforded by the compound measure initials+finals (r-value around 0.675 for judged similarity and around 0.600 for judged intelligibility). It is rather likely, however, that equally good prediction of the judgments can be obtained when we enter initials and finals as separate predictors in a multiple regression analysis. For this reason I will not consider the compound measure any longer.

The Levenshtein distances, finally, correlate disappointingly poorly with the judgment scores. The only Levenshtein distance that reaches significant correlation values is the segment-based rather than tone-based, without feature weighting. The r-values do not exceed 0.350. Also, there is hardly any correlation between the Levenshtein distances and any of the other objective distance measures.

---

[65] I remind the reader of the fact that PCI (in later text) are terms I coined myself. Cheng (1997), quite confusingly, used a different name for PCI, viz. *Mutual Intelligibility.*

6.2.2 Multiple prediction of judgment scores

In this section I will present the results of multiple regression analyses, in an attempt to determine how well the judgment scores collected in Chapter Three can be predicted from the objective distance measures we collected in Chapter Five. I performed linear regression analyses on each of the four judgment scores separately. Rather than predicting these criterion variables from single predictors, we now include all potentially relevant objective distance measures. The algorithm will determine which distance measures make a significant contribution to a better prediction of the judgment scores. Only those distance measures that make such a contribution are kept in the analysis. This excludes all parameters that are compounds of simplex distance measures. The beta weights attached to the remaining predictors express the relative importance of the predictor in the overall prediction.

In fact, I ran two MR (Multiple Regression) analyses for each independent variable. In the first analysis I entered all the predictors simultaneously. This yields very high multiple R^2 values but these are based on a large number of predictors, many of which do not make a significant contribution to the prediction. Therefore I also ran an incremental stepwise MR analysis, limiting the number of predictors to only those that make a significant contribution. On the basis of the raw correlation matrix (Appendix 6.1) the best predictor is selected. Then, on the basis of a partial correlation with the residual, the next-best predictor is selected, and so on, until no predictor can be found that still makes a significant improvement to the prediction of the criterion. In the results summarized below the number of predictors that contributed significantly was never in excess of four.

Table 6.1 presents the results for the MR analyses. The four criterion variables are specified in the leftmost column. For both the simultaneous entry and the stepwise solution the predictors with the associated R^2 and β-coefficients are presented in the next six columns. Only predictors have been specified that make a significant contribution ($p < .05$) to the overall R^2. No intermediate R^2 values have been specified for the simultaneous entry solution; instead I present the total R^2 when all predictors were included.

Table 6.1 shows that, generally, judged intelligibility scores can be somewhat better predicted from objective linguistic distance measures than similarity judgments. Including all (non-compound) predictors trivially leads to even better prediction of the criterion. The stepwise and simultaneous entry solutions converge in the case of intelligibility judgments: here the same predictors are successful and significant irrespective of the solution chosen. The prediction of the Similarity judgments comes up with different predictors. In all cases the strongest and most successful predictor is the Lexical Affinity Index (percentage of cognates shared between the dialects, as published by Cheng 1997). The additional contribution of other distance parameters (always based on phonological properties of the languages) is very modest in comparison. Nevertheless, a useful and consistent contribution is made by a fairly simple and straightforward distance measure such as the inventory of tones. Strikingly, this parameter also contributes significantly to the prediction of judged similarity and intelligibility if monotonized versions of the speech samples were presented.

Table 6.1. Results of Multiple Regression Analyses, predicting judgment scores from non-compound objective measures of linguistic distance. CC: Data from Cheng (1997), Inv: our own data on sound inventories of Chinese dialects, CA: lexical frequencies based on the Academy of Social Sciences database. Further see Chapter Five.

	Simultaneous entry				Stepwise entry		
	Predictors	R^2	β		Predictors	R^2	β
Similarity monotonous	CC_LAI		.668		CC_LAI	.727	.599
	Inv_Initials		.196		+CA_Finals	.768	–.124
					+CC_PCI	.783	.208
	All	.855			+Inv_Tones	.798	–.134
Similarity intonated	CC_LAI		.725		CC_LAI	.744	.773
	Inv_Tones		–.267		+CA_Nucleus	.784	–.161
	Inv_Initials		–.242		+Inv_Tones	.798	–.133
	All	.848					
Intelligibility monotonous	CC_LAI		.626		CC_LAI	.728	.683
	Inv_Tones		–.247		+Inv_Tones	.780	–.218
	Inv_Finals		–.445		+Inv_Finals	.814	–.213
	CA_Finals		.495				
	CA_Codas		–.316				
	All	.880					
Intelligibility intonated	CC_LAI		.723		CC_LAI	.764	.707
	Inv_Tones		–.297		+Inv_Tones	.809	–.201
	Inv_Finals		–.307		+Inv_Finals	.844	–.218
	CA_Codas		–.209				
	All	.889					

6.2.3 Single predictors of functional scores

In Chapter Four, we collected the subjective intelligibility data based on the functional word-intelligibility and sentence-intelligibility tests. In this section, we will correlate these results with the objective measures we analyzed in Chapter Five.

The raw correlation coefficient between each of the 27 objective linguistic similarity measures and the functional intelligibility scores at the word and at the sentence level are included in Appendix 6.1. Following the same principle as in § 6.2.1, we will now determine the best, and most promising, single linguistic distance measures in each of four types of data: (i) sound inventories, (ii) overall measures of lexical and phonological similarity published by Cheng (1997), (iii) lexical frequencies of phonological units published by Cheng (1997), (iv) lexical frequencies of similar sound units derived from the CASS transcriptions and (v) string distance measures (Levenshtein) determined on the same collection of transcriptions.

Within the similarity measures based on the sound inventories, finals, and especially coda elements (rather than vocalic nuclei) shared between dialects provide the best predictors of functional intelligibility (*r*-values around .500). Tones shared in the inventories are intermediate (around .400), and least successful predictors are shared initials (onsets) with *r*-values on the order of *r* = .250 (marginally significant).

Much better predictions are obtained from the published measures in Cheng (1997). Both the lexical and the phonological affinity correlate with word and sentence intelligibility with *r*-values between .740 and .772. We also note that the intercorrelation between lexical and phonological similarity is still low enough (*r* = .761) to make multiple prediction a worthwhile undertaking (§ 6.2.4).

We now come to the simpler types of measures published by Cheng (1997). Among this group of objective distance measures the shared finals stands out with *r*-values around *r* = .720. Correlation coefficient for other phonological units are poorer, and no correlation at all is obtained for shared tones.

The distance measures we derived ourselves from our lists of sound inventories in the 15 dialects reflect the same tendencies that were apparent in Cheng (1997). Again, the best correlations are found for shared finals (codas rather than nuclei), whilst shared initials (onsets) and tones are poorer predictors.

Also, when we consider the distance measures computed on the lexical frequencies of the sound units in the CASS transcriptions of 764 basic morphemes, we find the best (but not good) correlation for shared finals (*r*-values around .425), slightly poorer correlations for onsets, nuclei and codas (*r*-values between .360 and .400) and the poorest correlation for tones (around *r* = .220). Distance measures based on string-edit procedures correlate least with functional intelligibility scores (insignificant or marginally significant *r*-values between .038 and .326).

6.2.4 Multiple prediction of functional scores

We will now attempt multiple regression analyses for the functional intelligibility scores along the lines sketched above in § 6.2.2. for the judgment scores. The results of these analyses are presented in Table 6.2. Again, results are shown separately for predictors that were entered simultaneously, and for stepwise solutions.

Table 6.2. Results of Multiple Regression Analyses, predicting functional intelligibility scores (word level, sentence level) from non-compound objective measures of linguistic distance. CC: Data from Cheng (1997), Inv: our own data on sound inventories of Chinese dialects, CA: lexical frequencies based on the Academy of Social Sciences database. Further see Chapter Five.

	Simultaneous entry				Stepwise entry		
	Predictors	R^2	β		Predictors	R^2	β
Intelligibility Word level	CC_LAI		.615		CC_LAI	.593	.708
	CC_Finals		.247		+CC_Finals	.650	.265
	Inv_Tones		–.410		+CC_Onsets	.749	.722
	Inv_Initials		–.306		+Inv_Initials	.790	–.277
	CA_Onsets		.581		+CC_PCI	.810	.332
	CA_Tones		.353				
	Leven-TC		–.112				
	All	.883					
Intelligibility Sentence level	CC_LAI		.571		CC_LAI	.548	.621
	CC_Finals		.278		CC_Finals	.612	.405
	Inv_Tones		–.612		CA_Onsets	.680	.663
	Inv_Initials		–.410		Inv_Finals	.725	–.498
	CA_Onsets		.696		Inv_Tones	.759	–.646
	CA_Tones		.481		CA_Tones	.816	.475
					CA_Finals	.846	.621
					Leven_weight	.855	.101
	All	.877					

With simultaneous entry of all predictors we obtain high R^2 values of .883 and .877 for word and sentence intelligibility, respectively, at least when all (non-compound) predictors are included. However, only seven objective distance measures make a significant contribution in the prediction of word intelligibility, and a proper subset of six out of these seven, with very similar beta-weights (also listed in Table 6.2), recur in the prediction of intelligibility at the sentence level. This is more or less to be expected since word and sentence level intelligibility were found to be highly correlated (r = .928).

When we attempt stepwise entry of predictors the highest R^2 value obtained for word intelligibility is .810 with five predictors and .855 for sentence intelligibility with eight predictors. The first two predictors (CC_LAI and CC_Finals) are the same as in the simultaneous entry solution, with roughly the same beta weights but from the third predictor onwards the results diverge. By and large, these results indicate that a fairly good prediction of word and sentence intelligibility can be obtained (R^2 of .650 and .612, respectively) from just two predictors, one that covers lexical distance (percent cognates shared) and one that covers phonological distance, i.e. lexical frequency of finals (syllable rhymes) in Cheng's (1997) count based on a 2,270 item word list.

6.3 Conclusions

The basic question underlying the present chapter is how well can we predict mutual intelligibility between pairs of Chinese dialects from objective measures of linguistic distance. We will now try to answer this question on the basis of the results presented in this chapter. These results are summarized in Table 6.3.

Table 6.3. Summary of findings.

Criterion variable (down)	Type of MR analysis			
	simultaneous		stepwise	
	# of predictors	R^2	# of predictors	R^2
Judged similarity (monotonized)	2	.750	4	.798
Judged similarity (intonated)	3	.788	3	.798
Judged intelligibility (monotonized)	5	.837	3	.814
Judged intelligibility (intonated)	4	.847	3	.844
Functional intelligibility words	7	.846	5	.810
Functional intelligibility sentences	6	.816	8	.855

From Table 6.3 we may draw a number of provisional conclusions. First, it would appear that, generally, similarity judgments are somewhat more difficult to predict from linguistic distance measures than intelligibility scores (whether judgments or functionally determined). The difference, however, is marginal in the case of the stepwise analyses, but larger in the simultaneous entry solution (R^2 values below .8 for similarity judgments, and above .8 for all other criterion variables.

Second, functional intelligibility scores, whether determined for single words or for (short) sentences cannot be better predicted than judged intelligibility. There does seem to be a tendency, however, that prediction of judged intelligibility requires fewer predictor variables.

On the basis of these observations we cannot decide whether judged or functionally determined intelligibility is more amenable to prediction from objective measures. This does not seem to provide a basis to choose between judgment tests and functional tests as the preferred method of measuring mutual intelligibility.

At the end of Chapter Four (see also Appendix 6.1) we noted that judgment scores and functional intelligibility scores correlate substantially but not perfectly between r = .73 and r = .82 depending on the specific pair of scores. Even in the best case, the r-value of .82 leaves 33% of the variance unaccounted for. Therefore the two types of intelligibility measures cannot be used interchangeably. We must either include both measures in future research or find a principled way of choosing between the two types of measurement. One way to settle the problem would be to see which of the two measurements can be predicted better from objective measures of linguistic distance. However, as we have now seen, this approach does not provide an answer. In the last chapter of this dissertation we will attempt to make a principled choice using yet another criterion, i.e. by validating the two types of intelligibility measures against traditional claims (intuitions) made by Chinese dialectologist. The mutual intelligibility measure that best reflects the traditional dialect taxonomy will then be the preferred type.

Chapter Seven

Conclusion

7.1 Summary

In the introduction of the present dissertation (Chapter One), we state the fact that dialects are speech varieties from the same parent language family and address the questions: can the distance between pairs of dialects be used to distinguish the dialect from language? Can the distance between pairs of dialects be measured subjectively and objectively based on various criteria? The objective measures are actually structural differences and thus are multi-dimensional. The subjective measures are intelligibility tests. In the real speech situation, the intelligibility between two speech varieties is asymmetrical, that is, the intelligibility between variety X and Y is not necessarily identical to that between variety Y and X. We use the mutual intelligibility (the mean intelligibility of X to Y and Y to X) as the measure to compute the distance between pairs of speech varieties. Specifically, *mutual* intelligibility is practically defined as the mean intelligibility of speaker A and listener B and *vice versa*. It is not an absolute (all-or-nothing) measure but a scalar value (e.g. on a scale from 0 to 100%). Thus it can be used to measure distance between speech varieties. If the two varieties are close enough then the mutual intelligibility is supposed to be high and vice versa. The other way around, if the mutual intelligibility between two varieties is sufficiently high, then the two varieties are dialects of the same parent language. The mutual intelligibility tests can be performed through experiments. The resulting scores from the structural measures and the experimental tests are used to correlate with each other in order to find out the best prediction parameters. Dendrograms (trees) can be generated from the results of structural measures and the experimental tests. These trees illustrate the affinity and mutual intelligibility between pairs of speech varieties and thus are used to express the genetic relationship of dialects. The better, reasonable measures can be found out when we correspond these cladistic trees with traditional language taxonomy proposed by linguists.

Structural measures can be computed multi-dimensionally because the language varieties differ in many linguistic aspects. The computation can be based on lexical affinity (e.g., percentage of shared of cognates) and various phonological distance measures (for example, Levenshtein distance) involving both weighed or unweighed parameters. Mutual intelligibility can be established through experimental tests: judgment (or: opinion) tests and functional tests. Opinion tests determine how well the hearer thinks s/he understands the other language/dialect. The measure is based on the collected opinion scores given by listeners on the scalar range when they are listening to fragments of dialect speech. Functional tests determine how well the hearer actually understands the other language/dialect. The measure is counting the percentage of correctly recognized or translated words. Mutual intelligibility testing is one-dimensional. Thus methodologically, mutual intelligibility is argued to be the ideal criterion to differentiate dialects from language(s) and to validate the genealogical relationship between speech varieties.

Judgment testing is simpler and more straightforward compared to the functional tests. It is often used as a shortcut of mutual intelligibility tests. If the results from the two kinds of testing methods sufficiently correlate with each other, that means they actually measure the same property and judgment testing is ideally suited as a substitute of functional testing for the sake of simplicity and economy.

In the present dissertation, I focussed on determining the mutual intelligibility between pairs of Chinese dialects. I employed various structural measures (both obtained from the literature and developed by myself) as predictors to validate the mutual intelligibility tests. The objective distance measures were focused on the lexicon affinity, phonological affinity and sound shape (e.g., shared frequency of segmental elements and tonal patterns) from several database sources. We copied the lexicon affinity index and phonological correspondence index from the literature (mainly from C. C. Cheng based on the DOC database). We also computed other objective structural measures (such as the frequency of inventory, the frequency of segmental phonemes and tonal pattern based on the shared cognates, Levenshtein distance of the sound shape, etc.) based on several sources of database (the phonological inventory of each of our 15 selected dialects, the CASS data base) by ourselves.

We target at establishing the degree of mutual intelligibility between pairs of Chinese dialects via experiment tests. We tested how much our selected 15 Chinese dialects are mutually intelligible and to what extent the mutual intelligibility between pairs of these dialects can be predicted from the linguistic distance measures, which testing method is best predicted. We also tested impressionistic claims and anecdotes consistently found in the literature that Mandarin dialects have greater mutual intelligibility than non-Mandarin (Southern) dialects and non-Mandarin (Southern) speakers understand Mandarin dialects better than vice versa. All the computational measures were correlated with each other and translated into affinity trees. These trees can be compared with traditional Chinese dialect taxonomies.

7.2 Answers to research questions

In this section, I will recapitulate the research questions in Chapter One: Introduction and offer the integrated answers to them as follows.

7.2.1 The correlation between judged (mutual) intelligibility and similarity

Viewing the previous chapters (viz., Chapters Three and Four), we summarize that between the pairs of our selected target 15 Chinese dialects, the judged intelligibility is significantly correlated with the judged similarity based on melodic and monotonized data versions, with full matrix, off-diagonal and lower part of the triangle matrix, the *r*-value is .854, .883; .810, .841; .888, and .900, respectively. There is systematic consistence that the monotonized data produce higher correlation values. However, as we read in § 3.5.3, the results found for stimuli with and without pitch information were largely the same; the degree of convergence is exactly the same, both between the judged intelligibility with and without tonal information and judged similarity with and without tonal information ($r = .946$, N =225, $p < 0.001$). They account for 90 percent of the variance. Therefore, we conclude that in the Chinese dialectal situation, tonal information does not play a significant role in mutual intelligibility. This conclusion is also supported by Qin (2007) who tested the intelligibility of Standard Mandarin (i.e. Beijing dialect) in which the original tone contours had been replaced by Dutch rise-fall pitch accents on every content word in the same set of SPIN sentences that were used in my own experiments to Dutch listeners. Qin found that substituting Dutch pitch accents for Mandarin tones reduced the intelligibility for Mandarin listeners by a mere 4 percent.

7.2.2 Mutual intelligibility within and between (non-)Mandarin dialects

Generally speaking, in Chinese linguistic situation, there exists mutual intelligibility between each pair of dialects. The degree of *mutual* intelligibility can be established through experiments (both by judgment opinion tests and functional word-intelligibility and sentence-intelligibility tests).

The results from both the opinion judgment tests and the functional tests support the claims of asymmetry in the mutual intelligibility between Sinitic dialects. There exists some degree of mutual intelligibility between pairs of Sinitic dialects. In general, Mandarin dialects are more intelligible to non-Mandarin (Southern) dialects listeners than vice versa. The dialects within Mandarin branch are more mutually intelligible than the dialects within non-Mandarin (Southern) branch are. Functional tests confirm the impressionistic claims more convincingly than judgment opinion tests. Sentence-intelligibility test is the best supporter.

Two reasons might explain this phenomenon. Firstly, it might be because Mandarin dialects are intrinsically more similar to each other than non-Mandarin (Southern) dialects are. Secondly, it might also be the case, however, that most Chinese listeners (including Southern dialect speakers) are familiar, through education and the media, with Standard Chinese, which represents intrinsic properties of Mandarin dialect to large extent.

7.2.3 Mutual intelligibility predicted from objective distance measures

7.2.3.1 Correlation between subjective tests

All of the experimental results collected from the judgment opinion tests and functional tests correlate with each other to large extent. Firstly, judged intelligibility can be predicted fairly well from judged linguistic similarity (and vice versa) with r = 0.888 with symmetrical input of the matrix, N = 105. Secondly, two functional intelligibility measures (the intelligibility based on isolated word and word in sentence context) converge very well with r = 0.928. Thirdly, the correlation between the functional word and sentence intelligibility scores and the intelligibility judgment scores is good (r = 0.772 and 0.818, respectively) but not good enough to advocate the unqualified use of judgment testing as a more efficient substitute for functional testing (cf. Chapter Four, Table 4.6)

7.2.3.2 Predicting subjective results from objective measures

We collected 25 objective variables which cover some aspect of linguistic distance or similarity between pairs of our 15 Chinese dialects. These variables are either copied from the literature (cf. C.C. Cheng) or computed by ourselves. One variable (LAI, Lexical Affinity Index) is about the similarity in the lexicons of the dialects, i.e., the percentage of cognate word forms shared between a pair of languages (from a list of 2,770 concepts commonly used in Chinese languages, Cheng 1997). All other variables are concerned with aspects of the sound structure, either about the simple properties, e.g., which onsets consonants are shared between the phoneme inventories of two dialects, or of a more complex nature (i.e. Phonological Correspondence), such as the complexity of the rule system needed to convert phonological forms in one dialect to their cognates in the other dialect. Also some compound nature of a few variables is concerned. (see Chapter Five)

The correlation between various objective structural distance is complex and it is not the concern of all possible variables in this study. What we are interested in is to what extent the mutual intelligibility between dialects can be predicted from these objective distance measures. We only select those useful variables to predict the mutual intelligibility between pairs of Chinese dialects. We consider the prediction from both the single predictors and multiple predictors.

7.2.3.2.1 Single predictors of judgment and functional scores

The *mutual* intelligibility of Chinese dialects can be predicted from the structural measures to some extent. All of these results are used to correlate with each other through multiple regression analysis.

Prediction of judgment scores. Firstly, the lexical frequencies of initials and of finals from C.C. Cheng's computation based on the *Cihui* database is useful as single predictor for judgment scores, with correlation coefficients around r = 0.500 for the initials and somewhat above r = 0.600 for the finals. Similarity judgments correlate better with these predictors than intelligibility judgments, and finals afford better predictions than initials frequency. The two lexical frequency measures (initials and finals) are moderately intercorrelated (r = 0.523).

Secondly, Cheng's Lexical Affinity Index (LAI) and the Phonological Correspondence Index (PCI) based on the DOC database are found to be two good predictors. These two measures correlate very well with the judgment scores, with *r*-values around 0.850 for LAI and between 0.710 and 0.757 for PCI. These two predictors are intercorrelated at $r = 0.761$.

Thirdly, distance measures based on onsets, nuclei and codas, as well as larger phonological units such as finals (nucleus+coda) from the CASS database correlate with the judgment scores at *r*-values between 0.400 and 0.600. The best prediction of judgment scores is afforded by the compound measure initials+finals (*r*-value around 0.675 for judged similarity and around 0.600 for judged intelligibility).

Any correlation based on tonal measures or some Levenshtein distance is poor. We will not use these measures any more in the later multiple regression analysis.

Prediction of functional scores. Firstly, finals, especially coda elements (rather than vocalic nuclei) shared between dialects based on the sound inventories provide the best predictors of functional intelligibility (*r*-values around .500). Tones shared in the inventories are intermediate (around .400), and least successful predictors are shared initials (onsets) with *r*-values on the order of $r = .250$ (marginally significant).

Secondly, the lexical and the phonological affinity correlate well with word and sentence intelligibility with *r*-values between .740 and .772. The intercorrelation between lexical and phonological similarity is .761.

Thirdly, as for the simpler types of measures based on Cheng's computation on the 2,770 word list (Cheng 1997), the shared finals stands out with *r*-values around $r = .720$. Correlation coefficient for other phonological units are poorer, and no correlation at all is obtained for shared tones. The same tendencies are found from distance measures we derived ourselves from our lists of sound inventories in the 15 dialects. Again, the best correlations are found for shared finals (codas rather than nuclei), whilst shared initials (onsets) and tones are poorer predictors.

Fourthly, we find the comparatively best (but not good) correlation for shared finals (*r*-values around .425) in the distance measures computed on the lexical frequencies of the sound units in the CASS transcriptions of 764 basic morphemes, slightly poorer correlations for onsets, nuclei and codas (*r*-values between .360 and .400) and the poorest correlation for tones (around $r = .220$). Also, string-edit procedures correlate least with functional intelligibility scores (insignificant or marginally significant *r*-values between .038 and .326).

7.2.3.2.2 Multiple predictions of judgment and functional scores

Multiple predictions of judgment scores. I performed linear regression analyses on each of the judgment scores separately. Rather than predicting these criterion variables from single predictors, we now include all potentially relevant objective distance measures in order to find out which distance measures make a significant contribution to a better prediction of the judgment scores.

An incremental stepwise MR analysis found out generally, judged intelligibility scores can be somewhat better predicted from objective linguistic distance measures than similarity judgments. In all cases the strongest and most successful predictor is the Lexical Affinity Index (percentage of cognates shared between the dialects, as published by Cheng 1997). The additional contribution of other distance parameters (always based on phonological properties of the languages) is very modest in comparison. Nevertheless, a useful and consistent contribution is made by a fairly simple and straightforward distance measure such as the inventory of tones. Strikingly, this parameter also contributes significantly to the prediction of judged similarity and intelligibility if monotonized versions of the speech samples were presented.

Multiple prediction of functional scores. Following the same procedure as the multiple prediction of judgment scores, we found the results that a fairly good prediction of word and sentence intelligibility can be obtained (R^2 of .650 and .612, respectively) from just two predictors, one that covers lexical distance (percent cognates shared) and one that covers phonological distance, i.e. lexical frequency of finals (syllable rhymes) in Cheng's (1997) count based on a 2,270 item word list.

Conclusion. A number of conclusions about the prediction of the mutual intelligibility are drawn. First, in general, similarity judgments are somewhat more difficult to predict from linguistic distance measures than intelligibility scores (whether judgments or functionally determined). Second, functional intelligibility scores, whether determined for single words or for (short) sentences cannot be predicted better than judged intelligibility. There does seem to be a tendency, however, that prediction of judged intelligibility requires fewer predictor variables. Third, concerning the prediction power observed, we cannot decide whether judged or functionally determined intelligibility is more amenable to prediction from objective measures. This does not seem to provide a basis to choose between judgment tests and functional tests as the preferred method of measuring *mutual* intelligibility. Fourth, we can make a principled choice using another criterion, i.e. by validating the two types of intelligibility measures against traditional claims (intuitions) made by Chinese dialectologist. The *mutual* intelligibility measure that best reflects the traditional dialect taxonomy will then be the preferred type. The final conclusion is that sentence-intelligibility can be better predicted from the objective distance measures.

7.3 The status of Taiyuan

One of the controversial problems for Chinese dialect classification is whether Taiyuan is a Mandarin or a non-Mandarin dialect. Traditionally, Taiyuan is considered as a Mandarin dialect. The most possible reason is that Taiyuan is geographically situated in the north part of China where most Mandarin dialects are distributed (Mandarin is also called 'Northern dialects'). However, most recently, some dialectologists proposed to separate a Jin group represented by Taiyuan from the Mandarin branch. The claim is that dialects in this Jin group kept *Ru* tones which are normally found in most of the non-Mandarin groups.

From the dendrogram trees obtained from the various distance measures, Taiyuan was classified into the possible branch of Mandarin based on the lexical and phonological characteristics. The only exception that Taiyuan was in the non-Mandarin branch occurred when the tone affinity was computed. This testifies the statistical results of shared similarity based on the overall lexical and phonological data in Campbell's webpage, which claims that the most intelligible language variety to Mandarin is Jin (61 % similarity). (Campbell, 2009: http://www.glossika.com/en/dict/faq.php#1)

All of our experimental trees also give us the results that Taiyuan is a Mandarin dialect. Another supportive reason for us to insist Taiyuan as a Mandarin dialect is that we also found *Ru* tones in other Mandarin dialects such as some dialects in the Eastern and South-Western groups of the Mandarin branch (for example, Hefei and Yangzhou, which are not included in our 15 sample dialects). Therefore, we argue that there is no straightforward reason to branch off a Jin group from the Mandarin branch. We would like to follow the traditional Chinese dialect taxonomy that Taiyuan is one of the Mandarin dialects.

7.4 Relating mutual intelligibility to traditional Chinese dialect taxonomy

Generally speaking, in Chinese dialect situation, we claim that there exists *mutual* intelligibility between pairs of Chinese dialects to some extent, and the degree of *mutual* intelligibility can be established through experiments (both by judgment opinion tests and functional word-intelligibility and sentence-intelligibility tests). All the experiments results comparably highly correlate each other and can be predicted from the various linguistics distance measures to some extent respectively. LAI (Lexicon Affinity Index) and PCI (Phonological Correspondence Index) are the best predictor parameters for mutual intelligibility between pairs of Chinese dialects. The affinity trees generated from the computed affinity scores obtained objectively and subjectively were validated with the traditional Chinese dialect taxonomy based on the primary Mandarin versus non-Mandarin split and their internal cluster structures. To a large extent, the dendrograms obtained from the collected mutual intelligibility testing scores correspond with the primary split of the Mandarin and non-Mandarin (Southern) dialect branches suggested by Chinese linguists, some even converge well with the internal sub-groups or clusters. To some extent, the claim that mutual intelligibility can be used to argue for the dialect classification and to validate the taxonomy is confirmed.

The final conclusion is that sentence-intelligibility can be better predicted from the objective distance measures. Better results are expected from the future research based on larger selected dialect samples and general conclusions about the mutual intelligibility will be obtained when we overview all the dialects in the future.

7.5 Remaining questions

Our experimental results confirm that mutual intelligibility can be argued as a convincing criterion to illustrate the genetic relationship between speech varieties to some extent. We may also explain the reason why Indo-European language varieties are called languages but Sinitic varieties are dialects as the definition based on different methodologies. The western definition of languages is based on the distance measures between varieties (both objectively and subjectively), while Chinese dialects were defined on the historical change of Middle Chinese (MC) and are traditionally classified based on the phonological characteristics. These impressive classifications can be tested through computing the structural distance measures and through establishing the degree of mutual intelligibility following the research methodology on western language speeches. The language phylum can be reconstructed and illustrated by dendrograms (cladistic trees) generated from the matrix scores through hierarchical cluster analysis.

Our research results also show that the contribution of tone information to mutual intelligibility is not so important, at least to the construction of dendrogram trees. Further research is expected to support this result. We also expect further research mutual intelligibility testing involving the morphological and syntactic levels in the future.

References

Anderson, H. (2005). Intelligibility testing (RTT) between Mendankwe and Nkwen. Summer Institute of Linguistics, Dallas, Texas. *Electronic Survey Reports* 2005-002: 30.

Bao, Zhiming (1999). *The Structure of Tone*. New York: Oxford University Press.

Beijering, K., C. Gooskens, W. Heeringa (2008). Modelling intelligibility and perceived linguistic distances by means of the Levenshtein algorithm, in M. van Koppen, B. Botma (eds.) *Linguistics in the Netherlands 2008*. Amsterdam: John Benjamins, 13-24.

Beijing University (1962). *Hanyu fangyin zihui [Chinese Dialect Character Pronunciation List]*. Beijing: Wenzi Gaige Chubanshe (Philology Reform Publish house).

Beijing University (1964). *Hanyu fangyin cihui [Chinese Dialect word List]*. Beijing: Wenzi Gaige Chubanshe (Philology Reform Publish house).

Beijing University (1989). *Hanyu fangyin zihui di'er ban [Chinese Dialect Character Pronunciation List, second edition]*. Beijing: Wenzi Gaige Chubanshe (Philology Reform Publish house).

Beijing University (1995). *Hanyu fangyin cihui di'er ban [Chinese Dialect word List, second edition]*. Beijing: Wenzi Gaige Chubanshe (Philology Reform Publish house).

Bezooijen, R. van, R. van den Berg (1999). Word intelligibility of language varieties in the Netherlands and Flanders under minimal conditions, in R. van Bezooijen, R. Kager (eds.) *Linguistics in the Netherlands 1999*. Amsterdam: John Benjamins, 1-12.

Bezooijen, R. van, V. J. van Heuven (1997). Assessment of speech synthesis, in D. Gibbon, R. Moore, R. Winksi (eds.) *Handbook of standards and resources for spoken language systems*. Berlin/New York: Mouton de Gruyter, 481-653.

Biggs, B. (1957). Testing intelligibility among Yuman languages. *International Journal of American Linguistics* 23, 57-62.

Blalock, H. M., Jr., A. B. Blalock, eds. (1968). *Methodology in Social Research*. New York: McGraw-Hill.

Boersma, P., D. Weenink (1996). *Praat, a system for doing phonetics by computer, version 3.4*. Report 132, Institute of Phonetic Sciences, University of Amsterdam.

Bolognesi, R., W. Heeringa (2002). De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten [The influence of dominant languages on the lexicon and phonology of Sardic dialects]. *Gramma/TTT: Tijdschrift voor Taalwetenschap* 9, 45-84.

Box, G. E. P., W. G. Hunter, S. J. Hunter (1978). *Statistics for Experimenters*. New York: John Wiley & Sons.

Bradley, C. H. (1968). Dialect intelligibility testing: the Mixtec case, in H. M. Blalock, A. B. Blalock (eds.) *Methodology in Social Research*. New York: McGraw-Hill.

Brye, E., E. Brye (2002). *Rapid appraisal and intelligibility testing surveys of the Eastern Beboid group of languages (Northwest Province)*. SIL Electronic Survey Reports 2002-019. http://www.sil.org/silesr/2002/019/

Casad, E. H. (1969). The Choapan Zapotec dialect survey, in H. M. Blalock, A. B. Blalock (eds.) *Methodology in Social Research*. New York: McGraw-Hill.

Casad, E. H. (1974). *Dialect intelligibility testing.* Summer Institute of Linguistics Publications in Linguistics and Related Fields, 38. Norman: Summer Institute of Linguistics of the University of Oklahoma.

Cavalli-Sforza, L., W. S.-Y. Wang (1986). Spatial distance and lexical replacement. *Language* 62, 38-55

Chambers, J. K., P. Trudgill (1980). *Dialectology*. Cambridge: Cambridge University Press.

Chao, Yuan-Ren (1928). *Studies in the Modern Wu Dialects*. Tsinghua College Research Institute Monograph 4, Beijing.

Chao, Yuan-Ren (1934). The non-uniqueness of phonemic solutions of phonetic systems. In *Zhongyang yanjiuyuan lishi yuyan yanjiu jikan*, (1992). vol. 4, No. 4. Also in *Zhongguo xiandai yuyan xue de kaituo he fazhan — Zhao Yanren yuyanxue wenxuan*. Beijing: Qinghua University Press. 中央研究院历史语言研究所集刊 4 本 4 分 载于《中国现代语言学的开拓和发展 — 赵元任语言学文选》北京清华大学出版社.

Chao, Yuan-Ren (2006). Chinese Language, in Wu Zong-ji and Zhao Xin-na (eds.) *Linguistic Essays by Yuan-Ren Chao*. Beijing: The Commercial Press.

Chen, M. (1973). Cross-dialectal comparison: a case study and some theoretical considerations. *Journal of Chinese Linguistics* 1, 1-13.

Cheng, Chin-Chuan (1973). *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, Chin-Chuan (1982). A quantification of Chinese dialect affinity. *Studies in the Linguistic Sciences* 12, 29-47.

Cheng, Chin-Chuan (1986). Contradictions in Chinese language reform. *International Journal of the Sociology of Language* 59, 87-96.

Cheng, Chin-Chuan (1987). A quantification of Chinese dialect affinity, in Chinese Language Society of Hong Kong (ed.) *Wan Li Memorial Volumes*. *English Volume*s 73-100. Hong Kong: Joint Publishing Co.

Cheng, Chin-Chuan (1988). 漢語方言親疏關係的計量研究 [Computational research on the affinity between Chinese dialects]. *中國語文[Chinese Languages]* 203, 87-102.

Cheng, Chin-Chuan (1990). Syllable-based dialect classification and mutual intelligibility. Paper presented at the First International symposium on Chinese Languages and Linguistics, Taipei, Taiwan.

Cheng, Chin-Chuan (1991). Quantifying affinity among Chinese dialects, in W. S.-Y. Wang (ed.) *Languages and dialects of China*. Journal of Chinese Linguistics Monograph Series 3. Berkeley CA: University of California, 78-112.

Cheng, Chin-Chuan (1992). Syllable-based dialect classification and mutual intelligibility. *Chinese Languages and Linguistics (I): Chinese Dialects*. Symposium Series of the Institute of History and Philology, Academia Sinica 2, 145-177.

Cheng, Chin-Chuan (1993a) (With Zhiji Lu) Chinese dialect affinity based on syllable initials. *Studies in the Linguistic Sciences* 15, 127-148.

Cheng, Chin-Chuan (1993b). *Quantifying dialect mutual intelligibility*. Technical Report No. LLL-T-21-93. Language Learning Laboratory, College of Liberal Arts and Sciences, University of Illinois at Urbana-Champaign.

Cheng, Chin-Chuan (1995). Calculation of Chinese dialect mutual intelligibility. *Selected papers from the Fourth International Symposium on Teaching of Chinese*. Beijing: Beijing Language Institute Publishing House, 387-397

Cheng, Chin-Chuan (1996). Quantifying dialect mutual intelligibility, in J. Huang, C. Li (eds.) *New horizons in Chinese linguistics*. Boston: Kluwer Academic Publishers, 269-292.

Cheng, Chin-Chuan (1997). Measuring relationship among dialects: DOC and related resources**.** *Computational Linguistics & Chinese Language Processing* 2, 41-72

Cheng, Chin-Chuan (1998). Extra-linguistic data for understanding dialect mutual intelligibility. *Proceedings of the Annual Meeting of Pacific Neighborhood Consortium*. Taipei, Taiwan: Computer Center, Academia Sinica, 419-422

Cheng, Chin-Chuan (2002). Language intelligibility as a constraint on phonological change. *Dialect Variations in Chinese*. Papers from the Third International Conference on Sinology, Linguistics section, 81-95

Crawford, J. C. (1967). Theoretical and methodological beginnings of dialect survey programs. Lecture presented to the Dialect Survey Field Workers' Conference. Cuernavaca, Morelos, Mexico.

Cutler, A., W. van Donselaar (2001). 'Voornaam' is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44, 171-195.

Delsing, L.-O., K. Lundin-Åkesson (2005). *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska*. [Does language keep the Nordic countries together? A research report on how well young people understand Danish, Swedish and Norwegian]. Köpenhamn: Nordiska ministerrådet.

Diekhoff, G. (1992). *Statistics for the Social and Behavioral Sciences*. Dubuque, IA: Brown.

Duanmu, San (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Forrest, R. A. D. (1948). *The Chinese Language*. London: Faber and Faber.

Gandour, J., R. Harshman (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Journal of the Acoustical Society of America* 62, 693-707.

Gordon, R. G., Jr., ed. (2005). *Ethnologue: Languages of the world*, Fifteenth edition. Dallas, TX: Summer Institute of Linguistics, Dallas, Texas International.

Gooskens, C., R. van Bezooijen (2002). The role of prosodic and verbal aspects of speech in the perceived divergence of Dutch and English language varieties. In J. Berns, J. van Marle (eds.) *Present-day Dialectology. Problems and Findings*. Berlin: Mouton de Gruyter, 173-192.

Gooskens, C., W. Heeringa (2004). Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change* 16, 189-207.

Gooskens, C. (2007). The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development* 28, 445-467.

Gooskens, C., R. van Bezooijen (2008). Linguistic distance between Flemish and Norwegian around 1900: subjective views and objective measurements, in P. Broomans, G. Jensma, H. Vandevoorde, Maarten van Ginderachter (eds.) *The beloved mothertongue. Ethnolinguistic nationalism in small nations: inventories and reflections*. Groningen Studies in Cultural Change 33. Leuven: Peeters, 121-139.

Gooskens, C., V. J. van Heuven, R. van Bezooijen, J. J. A. Pacilly (2009). Is Danish an intrinsically more difficult language to understand than Swedish? in B. Heselwood, C. Upton (eds.) *Proceedings of New Methods in Dialectology XII.* Bamberger Beiträge zur englischen Sprachwissenschaft Series. Bern: Peter Lang (submitted)

Gooskens, C. (2009). Non-linguists' judgments of linguistic distances between dialects. *Nordic Journal of Linguistics* (submitted)

Heeringa, W. (2004). *Measuring dialect pronunciation differences using Levenshtein distance.* Doctoral dissertation, University of Groningen.

He, Jiuying (1995). *A History of Chinese Modern Linguistic.* Guangdong Education Press.

Hickerson, H., G. D. Turner, N. P. Hickerson (1952). Testing procedures for estimation transfer of information among Iroquois dialects and languages. *International Journal of American Linguistics* 18, 1-8.

Hou, Jinyi (1994). *Xiandai hanyu fangyan yiku [The sound databank of Modern Chinese].* Shanghai: Shanghai Education Press.

Hou, Jinyi (2003) *Xiandai hanyu fangyan yiku [The sound databank of Modern Chinese]* (CD-ROM version, in Chinese). Shanghai: Shanghai Education Press.

Hsieh, Chiao-min (1973). *Atlas of China.* Los Angeles: McGraw-Hill.

Hsieh, Hsin-I. (1973). A new method of dialect subgrouping. *Journal of Chinese Linguistics.* 1, 64-92.

Hsieh, Hsin-I. (1990). In search of a grammatical foundation for dialect subgrouping. *Proceedings of the First International Symposium on Chinese Language and Linguistics.* Taipei: Academic Press, 146-167.

Huang, Jinhu (1987). *Hanyu Fangyanxue [Chinese Dialectology].* Xiamen University Press.

Hout, R. Van, H. Münstermann (1981), Linguistische afstand, dialekt en attitude [Linguistic distance, dialect and attitude]. *Gramma* 5, 101-123.

Jensen, J. B. (1989). On the mutual intelligibility of Spanish and Portuguese. *Hispania* 72, 848-852.

Nerbonne, J., P. Kleiweg, W. Heeringa, F. Manni (2008). Projecting dialect differences to geography: Bootstrap clustering vs. noisy clustering, in: C. Preisach, L. Schmidt-Thieme, H. Burkhardt, R. Decker (eds.) *Data Analysis, Machine Learning, and Applications. Proceedings of the 31st Annual Meeting of the German Classification Society.* Berlin: Springer, 647-654.

Kalikow, D. N., K. N. Stevens, L. L. Elliott (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America* 61, 1337-1351.

Karlgen, B. (1962). *Sound and Symbol in Chinese* (revised edition). Hongkong University Press, OUP, 16-31.

Kessler, B. (1995). Computational Dialectology in Irish Gaelic.*Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics.* Dublin: ACL, 60-67.

Kirk P. L. (1970). Dialect intelligibility testing: the Mazatec Study. *International Journal of American Linguistic*s 36, 205-211

Krishnamurti B., L. Moses, D. Danforth (1983). Unchanged cognates as a criterion in linguistic subgrouping. *Language* 59, 541-68.

Kroeber, A. L., C. D. Chretien (1937). Quantitative classification of Indo-European Languages. *Language* 13, 83-103.

Kroeber, A. L. , C. D. Chretien (1939). The statistical technique and Hittite. *Language* 15, 69-71.

Kruskal, J. B. (1999). An overview of sequence comparison, in D. Sankoff, J. Kruskal (eds.) *Time warps, string edits, and macro molecules; The theory and practice of sequence comparison*. Stanford: ASLI, 1-44.

Language Atlas of China (1987). Jointly compiled by the Chinese Academy of Social Sciences and the Australian Institute of Humanities. Hong Kong: Longman (English version).

Lau, Chun-fat (刘镇发) (2002). Hanyu fangyan de fenlei biaozhun yu 'Kejiahua' zai Hanyu fangyan fenlei shang de wenti 汉语方言的分类标准与'客家话'在汉语方言分类上的问题 [Criteria for Chinese dialect classification and the problem of the position of the 'Hakka dialect' in Chinese dialect grouping]. *Journal of Chinese Linguistics* 30, 82-96.

Lee, Mei W (1987, A-1, backside). Languages in China. *The Language Atlas of China*. (English version)

Li, Fang-kuei (李芳桂) (1937). *Languages and Dialects of China*. The Chinese Yearbook. Shanghai: Commercial Press (Reprinted in *Journal of Chinese Linguistics* 1.1 1–13, 1973).

Li, Rong (1987, A-1), Languages in China. *The Language Atlas of China* (English version).

Li, Rong, (1987, A-2), Chinese Dialects in China. *The Language Atlas of China* (English version).

Li, C., S. A. Thompson (1981) *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Lu, Zhiji, Chin-Chuan Cheng (1985). Chinese dialect affinity based on syllable initials. *Studies in the Linguistic Sciences* 15, 127-148.

Lü, Shuxiang (1980). *Yuwen Changtan 《语文常谈》 [Informal Notes on Language]*. Beokomg: Sanlian Shudian.

Merriam, C. H. (1926). The classification and distribution of the Pit River Indian tribes of California. Smithsonian Misc. Coll., 78, No. 3.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review* 76, 165-178.

Moulines, E., W. Verhelst (1995). Time-domain and frequency-domain techniques for prosodic modification of speech, in W. B. Kleijn, K. K. Paliwal (eds.) *Speech coding and synthesis*. Amsterdam: Elsevier Science, 519-555.

Murrell, G. A., J. Morton (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology* 102, 963-968.

Nerbonne, J., W. Heeringa, E. van den Hout, P. van der Kooi, S. Otten, W. van de Vis (1996). Phonetic Distance between Dutch Dialects, in G. Durieux, W. Daelemans, S. Gillis (eds.) *Papers from the sixth CLIN meeting*. University of Antwerp, Center for Dutch Language and Speech, 185-202.

Nooteboom, S. (1981). Lexical retrieval from fragments of spoken words: Beginnings vs endings. *Journal of Phonetics* 9, 407-424.

Norman, J. R. (1988). *Chinese*. Cambridge University Press.

Norman, J. R. (1999). Vocalism in Chinese dialect classification, in R. VanNess Simmons (ed.) *Issues in Chinese Description and Classification*, *Journal of Chinese Linguistics*. Monograph 15, 193-203.

Ogura, M. (1990). *Dynamic dialectology: A study of language in time and space*. Tokyo: Kenkyusha.

Olmsted, D. L. (1954). Achumawi-Atsugewi non-reciprocal intelligibility. *International Journal of American Linguistic*s 20, 181-84.

Oxford English Dictionary (2009). online links: http://dictionary.oed.com/ and http://dictionary.reference.com/

Petyt, K. M. 1980. *The Study of Dialect: An introduction to dialectology. The Language Library.* Edited by David Crystal. London: Deutsch.

Pierce, J. E. (1952). Dialect Distance Testing in Algonquian. *International Journal of American Linguistic*s 18, 208-218.

Pierce, J. E. (1954). Crow versus Hidatsa in dialect distance and in glottochronology. *International Journal of American Linguistic*s 20, 134-36.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., Slowiacek, L. M. (1985a). Speech perception, word recognition and the structure of the lexicon. *Speech Communication* 4, 75 – 95.

Pisoni, D., B. Greene, H. Nusbaum (1985b). Perception of synthetic speech generated by rule. *Proceedings of the IEEE* 73, 1665-1676.

Pisoni, D., B. Greene, H. Nusbaum (1985c). Some human factors issues in the perception of synthetic speech. *Proceedings Speech Tech '85*, New York, 57-61

Preston, D. R. (1989). *Perceptual dialectology. Nonlinguists' views of areal linguistics.* Dordrecht: Foris.

Qin, Qing (2007). Understanding foreign-accented language: Experiments on the relative contribution of segments and prosody in Dutch-accented Mandarin. M. Phil thesis, Leiden University Centre for Linguistics.

Ramsey, S. R. (1987). *The Languages of China*. Princeton: Princeton University Press.

Reed, D. W., J. L. Spicem (1952). Correlation Methods of comparing idiolects in a transition area. *Language* 28, 348-60.

Stoltzfus, R. D. (1967). Tzotzil Survey Report. Ms.

Swadesh, M. (1950). Salish internal relationships. *International Journal of American Linguistic*s 16, 157-167.

Swadesh, M. (1972). What is glottochronology? in M. Swadesh, *The origin and diversification of languages*. London: Routledge & Kegan Paul, 271-284.

Tang, Chaoju, V. J. van Heuven (2007). Mutual intelligibility and similarity of Chinese dialects, in B. Los, M. van Koppen (eds.) *Linguistics in the Netherlands 2007*. Amsterdam: John Benjamins, 223-234.

Tang, Chaoju, V. J. van Heuven (2008). Mutual intelligibility of Chinese dialects tested functionally, in M. van Koppen, B. Botma (eds.) *Linguistics in the Netherlands 2008*. Amsterdam: John Benjamins, 145-156.

Tang, Chaoju, V. J. van Heuven (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119, 709-732.

Ting, Pang-Hsin (1982). Hanyu Fangyan qufen de diaojian. 汉语方言区分的条件 [Phonological features for classification of the Chinese dialects]. *Tsing Hua Journal of Chinese Studies* 14, 257-273.

Voegelin, C. F., E. W. Voegelin (1946). Linguistic considerations of Northeastern North America, in F. Johnson (ed.) *Papers of the Robert S. Peabody Foundation for Archaeology* 3.

Voegelin, C. F., Z. S. Harris (1951). Methods for determining intelligibility among dialects of natural languages. *Proceeding of the American Philosophical Society* 95, 322-329.

Voiers, W. D. (1977). Diagnostic evaluation of speech intelligibility, in M. E. Hawley (ed.) *Speech intelligibility and speaker recognition.* Stroudsburg, PA: Dowden, Hutchinson & Ross, 374-387.

Wang, Hongyan (2007). *English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English.* LOT Dissertation Series 147. Utrecht: LOT.

Wang, Hongyan, V. J. van Heuven (2007). Quantifying the interlanguage speech intelligibility benefit. *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken*, 1729-1732.

Wang, Li (1963) *Hanyu Yinyun*. 汉语音韵 *[Chinese Phonology]*. Beijing: Zhonghua Shuju.

Wang, Li (1981). *Zhongguo Yuyanxue shi. 中国语言学史 [History of Chinese Linguistics].* Taiyuan: Shanxi Renmin Chubanshe.

Wang, Li (1985). *Wangli Wenji. 王力文集 [Collections of Wang Li's Articles].* Ji'nan: Shangdong Jiaoyu Chubanshe.

Wang, W. S-Y. (1987). Representing language relationships, in H. M. Hoenigswald, L. F. Wiener (eds.) *Biological metaphor and cladistic classification. An interdisciplinary perspective.* Philadelphia, PA: University of Pennsylvania Press, 243-256.

Wang, W. S-Y. (1996). Linguistic diversity and language relationships, in J. Huang, C. Li (eds.) *New horizons in Chinese linguistics*. Boston: Kluwer Academic Publishers, 235-267.

Wang, Yude (1960). Chugokugo dai hogen no bunretsu nendai no gengo nendai gakuteki shitan [Lexicostatistic estimation of time depths of five major Chinese dialects]. *Gengo Kenkye* 38, 33-105.

Wolff, H. E. (1964). Synopsis of the Ogoni languages. *Journal of African Linguistics* 3, 38-51.

Wurm, S. A., B. T'sou, D. Bradley, Li Rong, Xiong Zhenghui, Zhang Zhenxing, Fu Maoji, Wang Jun, Dob (1987). *Language Atlas of China.* Jointly compiled by the Chinese Academy of Social Sciences and the Australian Institute of Humanities. Hong Kong: Longman.

Yan, M. Mian (2006). *Introduction to Chinese Dialectology*. LINCOM Studies in Asian Linguistics. München: LINCOM,

Yang, C., A. Castro (2008). Representing tone in Levenshtein distance. *International Journal of Humanities and Arts Computing* 2 (in press).

Yuan, Jiahua, (1960). *Hanyu Fangyan Gaiyao [An outline of the Chinese dialects]*. Beijing: Wenzi Gaige Chubanshe.

Yuan, Jiahua (1989). *Hanyu Fangyan Gaiyao. [The survey of Chinese dialects]*. Wenzi Gaige Chubanshe.

Zhang, Zhenduo (1900). *Qinxian Fangyanzhi 沁县方言志 [The Qinxian dialect]*. Taiyuan: Shanxi Gaoxiao Lianhe Chubanshe.

Zhang, Zhenxing (1997). Reread Chinese Linguistic Atlas. *Dialect* 4, 241-248. (in Chinese)

Zhou, Zhenhe (1991). Migrations in Chinese history and their legacy on Chinese dialects, in W. S.-Y. Wang (ed.) *Languages and dialects of China.* Journal of Chinese Linguistics Monograph Series 3. Berkeley CA: University of California, 29-51.

Samenvatting

Dit proefschrift brengt de mate van onderlinge verstaanbaarheid in kaart van 15 Sinitische (Chinese) taalvariëteiten aan de hand van experimenten waarin zowel functionele als opinietests zijn afgenomen. De correlatie tussen de resultaten van de twee typen testmethoden werd berekend om te zien in hoeverre de tests dezelfde eigenschappen meten. De resultaten van de twee experimentele tests werden ook gecorreleerd met objectief bepaalde taalkundige afstandsmaten (hetzij gepubliceerd in de literature hetzij door mijzelf berekend) om te zien in hoeverre de onderlinge verstaanbaarheid voorspeld kan worden uit deze afstandsmaten. De scores uit de experimenten en de afstandsmaten zijn gebruikt om hiërarchische structuren te genereren (boomstucturen) die de verwantschapsrelaties tussen de betreffende taalvariëteiten tot uitdrukking brengen. Het voorstel is dat die methode om onderlinge verstaanbaarheid te bepalen de voorkeur geniet waarvan de boomstructuren beter overeenkomen met traditionele taxonomieën van talen en dialecten zoals die zijn bedacht door taalkundigen.

Hoofdstuk een bevat de Inleiding. Hierin leg ik uit waarom we de afstand tussen paren taalvariëteiten (d.w.z. dialecten of verwante talen) willen meten en hoe we dat zouden kunnen doen. De afstand tussen taalvariëteien is een potentieel belangrijke parameter om talen van dialecten te kunnen onderscheiden. Als de afstand klein is, d.w.z. als de variëteiten veel op elkaar lijken, dan zullen taalvariëteiten gekwalificeerd worden als dialecten van dezelfde taal maar als de afstand groter is, dan worden zij beschouwd als verschillende talen. Taalkundige afstand kan subjectief en objectief worden gemeten. De objectieve meetmethode kwantificeert de structurele afstand tussen taalvariëteiten langs meerdere dimensies (bv. klankstructuur, woordbouw, zinsbouw, c.q. allerlei deelaspecten daarvan). De subjectieve aanpak is om de mate van (onderlinge) verstaanbaarheid tussen paren taalvariëteiten te bepalen. De subjectieve benadering is conceptueel eenvoudiger omdat de algemene mate van (onderlinge) verstaanbaarheid kan worden uitgedrukt langs één dimensie.

Onderlinge verstaanbaarheid kan bepaald worden aan de hand van functionele methoden en van opinietests. De functionele aanpak test in hoe verre luisteraars *daadwerkelijk* taaleenheden (woorden) herkennen in gesproken stimuli en hoe veel zij *aantoonbaar* begrijpen van de boodschap die door deze eenheden in hun specifieke volgorde wordt uitgedrukt. Opinietests daarentegen gaan na hoe goed luisteraars (moedertaalsprekers van taal A) *denken* dat zij een spreker van een (verschillende) taal B zouden kunnen verstaan of begrijpen. Als we de onderlinge verstaanbaarheid van talen of dialecten willen vaststellen, moeten we beseffen dat het aantal vergelijkingen dat we moeten maken, exponentieel toeneemt met het aantal taalvariëteiten.Vergelijken we 10 variëteiten dan zijn $10^2 = 100$ verschillende combinaties van spreker- en luisteraargroepen nodig. Als het aantal te onderzoeken taalvariëteiten 15 bedraagt, dan stijgt het aantal combinaties tot $15^2 = 225$. Daar komt nog bij dat als een luisteraar moet laten zien hoe goed deze 10 of zelfs 15 verschillende taalvariëteiten verstaat, deze nooit een tweede keer mag luisteren naar dezelfde tekst, ten einde leer- en herhalingseffecten te vermijden. Zulke effecten kunnen we uitsluiten door gebruik te maken van zeer grote groepen luisteraars (zodat die opgesplitst kunnen worden in veel verschillende maar overigens volkomen gelijkwaardige luisteraargroepen die ieder worden blootgesteld aan een verschillende taalvariëteit) of door veel verschillende maar volmaakt gelijkwaardige verzamelingen testmaterialen te ontwerpen. Om dergelijke buitensporig arbeidsintensieve werkwijzen te vermijden kunnen we functionele verstaanbaarheidstests wellicht ook vervangen door opinietests (ook wel: beoordelingstests), zeker als het aantal te vergelijken taalvariëteiten groot is. Het onderhavig onderzoek heeft onder andere als doel om na te gaan in hoeverre opinietoetsen een valide alternatief vormen voor de bewerkelijker functionele verstaanbaarheidstests.

Twee verschillende criteria zijn gebruikt om de validiteit te bepalen van de functionele en opiniemethode van verstaanbaarheidsmeting. Het eerste criterium gaat na hoe goed de resultaten van de onderscheiden verstaanbaarheidsbepalingen voorspeld kunnen worden uit objectief bepaalde maten van structureel verschil tussen de betreffende taal-variëteiten. Het tweede validiteitscriterium is de mate van overeenkomst tussen (i) de verwantschapsbomen die voor de variëteiten gegenereerd kunnen worden op basis van de functionele of opiniemetingen en (ii) de zgn. taalstambomen of taxonomieën die voor de betreffende variëteiten zijn opgesteld door taalkundigen (dialectologen).

Dit proefschrift wil een bijdrage leveren aan de bepaling van onderlinge verstaanbaarheid van Sinitische dialecten en bevestigt diverse ideeën via experimentele functionele en opiniemeetmethoden: (1) De Mandarijndialecten zijn *intern* beter onderling te verstaan dan de Zuidelijke dialecten (2) De Zuidelijke dialecten zijn slechter te verstaan voor luisteraars van Mandarijndialecten dan omgekeerd (3) De betwiste Jin-dialecten (vertegenwoordigd door Taiyuan) kunnen beter worden ingedeeld bij de Mandarijn-dialecten dan bij de Zuidelijke dialecten (4) Een groot aantal structurele afstandsmaten zijn berekend voor de Sinitische dialecten op basis van bestaande databases, en met behulp van multipele regressieanalyse zijn daaruit de beste predictoren van onderlinge verstaanbaarheid opgespoord (5) Blijkens mijn testresultaten is een functionele verstaanbaarheidsbepaling op zinsniveau het meest valide omdat de verwantschapsboom die deze methode oplevert, het best overeenkomt met traditionele Chinese dialect-indelingen.

Hoofdstuk twee verschaft achtergrondinformatie over de taalsituatie in China. Het Chinese taallandschap is uiterst divers. Een veelheid van ‘dialecten’ (naar westerse maatstaven eigenlijk verschillende talen) wordt in China gesproken en hun indeling verschilt nogal, afhankelijk van de specifieke criteria die worden gehanteerd. Dit proefschrift richt zich op 15 dialecten binnen de Sinitische taalfamilie. Twee belangrijke taalindelingen zijn als referentiepunten gebruikt. Tussen deze indelingen bestaat volledige overeenstemming, op één punt na: (1) Beide indelingen baseren zich op historische veranderingen in fonologische (klankvormelijke) eigenschappen, d.w.z. veranderingen in de klinkers en medeklinkers en in de woordtonen, zoals de vereenvoudiging van de beginmedeklinkers, het wegvallen van de glottisslag, de opkomst van het stemhebbend/stemlooscontrast aan het begin van de lettergreep, de splitsing van de oorspronkelijke vier woordtonen in een hoge en een lage reeks (Yin- versus Yangregister) en de daaropvolgende versmelting of juist verdere doorsplitsing van tonen (2) Er is overeenstemming dat er een primaire opdeling is in Mandarijn- en Zuidelijke dialecten. Mandarijndialecten hebben hun plofklanken verloren aan het einde van de lettergreep en zij hebben minder woordtooncategorieën bewaard. Er is echter geen overeenstemming over verdere indeling van de Mandarijn- en Zuidelijke dialecten in termen van subgroepen (3) Van de Mandarijndialecten wordt beweerd dat zij *intern* beter onderling te verstaan zijn dan de Zuidelijke dialecten, omdat de eerstgenoemde in de loop van de geschiedenis minder uit elkaar gegroeid zijn – los nog van de invloed die het Standaardmandarijn, (dat is gebaseerd op het dialect van Beijing) heeft (4) Mandarijndialecten zijn naar wordt beweerd verstaanbaarder voor luisteraars van Zuidelijke dialecten dan omgekeerd (5) De mate van onderlinge verstaanbaarheid binnen en tussen Mandarijn- en Zuidelijke dialecten is nog niet vastgesteld; er is zelfs geen test waarmee de onderlinge (on)verstaanbaarheid binnen en tussen de Mandarijn- en Zuidelijke dialecten kan worden bepaald (6) De indeling van de Jin-dialecten is omstreden maar kan mogelijk vastgesteld op basis van de experimentele gegevens die voortkomen uit onderlinge verstaanbaarheidstests.

De volgende Chinese dialecten zijn opgenomen in de onderhavige studie: zes Mandarijndialecten, onderverdeeld in een Noordelijke groep (Beijing, Jinan, Xi'an, Taiyuan) en een Zuid-Westelijke groep (Chengdu, Hankou), alsmede negen niet-Mandarijn- (Zuidelijke) dialecten, onderverdeeld in zes groepen, die elk vertegenwoordigd zijn door een tot drie dialecten: Suzhou, Wenzhou (Wu-dialecten), Nanchang (Gan-dialect), Meixian (Hakka-dialect), Xiamen, Fuzhou, Chaozhou (Min-dialecten), Changsha (Xiang-dialect), en Guangzhou (Yue-dialect). Waar deze dialecten worden gesproken, is te zien op de kaart op het kaft van dit boek. De indeling van Taiyuan bij de Mandarijntalen is slechts voorlopig.

Hoofdstuk drie beschrijft de experimenten waarin ik de onderlinge verstaanbaarheid heb bepaald aan de hand van de opiniemeetmethode. Ik heb gebruik gemaakt van bestaande opnamen van de fabel *De Noordenwind en de Zon*, die werd voorgelezen door sprekers van de 15 geselecteerde dialecten. Elke opname werd (met de computer) gemanipuleerd zodat alle sprekers een mannenstem hadden met dezelfde spreeksnelheid (en dezelfde pauzeduren), en met dezelfde (gemiddelde) toonhoogte. Met behulp van de PSOLA toonhoogtemanipulatiefunctie in de Praat-software werden van elke opname twee versies gemaakt, een met en een zonder toonhoogte-informatie (melodie). De geïntoneerde (met melodie) en de monotone (zonder melodie) versies dienden om een schatting te kunnen maken van het belang van informatie over woordtoon. Voor elk van de 15 dialecten werden 24 (mono-dialectale) luisteraars geworven. Zij kregen instructie om de sprekers te beoordelen op 11-puntschalen (steeds op een schaal van 0 tot 10) op twee parameters: (i) verstaaanbaarheid en (ii) overeenkomst met het eigen dialect van de luisteraar. Op de beoordelingsschalen betekende '0' 'volkomen onverstaanbaar' en 'geen enkele overeenkomst' terwijl '10' de interpretatie had van 'volmaakte verstaanbaarheid' en 'perfecte overeenkomst'. In totaal werden 15 luisteraargroepen (één groep per dialect) × 15 spekerdialecten = 225 combinaties van spreker- en luisteraardialecten getest. Elke luisteraargroep omvatte 24 luisteraars, en elk sprekerdialect werd vertegenwoordigd door één enkele spreker steeds in twee versies (gemonotoniseerd, geïntoneerd). Op basis van de resulterende 21,600 scores zijn agglomeratiebomen geconstrueerd.

De resultaten geven aan dat er geen perfecte groepering van dialecten ontstaat op basis van de vier boomstructuren die door de experimenten worden opgeleverd, als we de traditionele dialectindeling als maatstaf hanteren. Allereerst wordt de belangrijkste tweedeling tussen de Mandarijn- en Zuidelijke dialecten niet helemaal correct weerspiegeld, ook de interne groepering binnen de hoofdtakken is onstabiel. Twee dialecten die traditioneel worden geclassificeerd als Zuidelijk (Nanchang en Changsha), werden samengenomen met de Mandarijndialecten. Op de tweede plaats wordt de noch opdeling binnen het Mandarijndialectcluster in Noordelijke en Zuid-Westelijke dialecten, noch die binnen het Zuidelijke cluster, bv. in Wu- en Min-dialecten, weerspiegeld. Grofweg vond ik de volgende resultaten: (1) De primaire splitsing tussen Mandarijn- en Zuidelijke dialecten wordt alleen teruggevonden in de agglomeratieboomstructuren als de criteria losser gehanteerd worden. (2) Taiyuan (de vertegenwoordiger van de Jin-dialecten, waarvan de indeling nog steeds betwist is, wordt consistent gegroepeerd bij de Mandarijndialecten. (3) De correlatie tussen beoordeelde overeenkomst en beoordeelde verstaanbaarheid is in alle gevallen (volledige matrix, matrix zonder hoofddiagonaal, niet-redudante halve matrix na middeling van contra-diagonale cellen) behoorlijk hoog (coëfficiënten tussen $r = .810$ en $r = .888$ voor de geïntoneerde spraakfragenten, en van $r = .841$ tot $r = .900$ voor de gemonotoniseerde versies). De gemonotoniseerde spraakfragmenten leverden daarbij iets hogere correlaties op. (4) Meer in detail laten de grafieken met gemiddelden van beoordeelde overeenkomst en verstaanbaarheid opgesplitst naar drie typen luisteraargroepen het volgende zien:

a. Zowel beoordeelde verstaanbaarheid als beoordeelde overeenkomst zijn heel hoog en liggen dicht tegen de maximumscore van 10 als luisteraars reageren op de spreker van hun eigen dialect ('own dialect' in de grafieken).
b. Een heel duidelijk verschil is zichtbaar tussen sprekers van de Mandarijn- en de niet-Mandarijn- (Zuidelijke) dialecten wanneer luisteraars andere dan hun eigen dialecten moesten beoordelen binnen dezelfde hoofdgroep als die van hun eigen dialect ('same branch'). Sprekers van Zuidelijke (niet-Mandarijn-)dialecten worden als praktisch onverstaanbaar beoordeeld door luisteraars van andere Zuidelijke dialecten (1.5 op de schaal van 1 tot 10) terwijl de sprekers van de Mandarijndialecten van luisteraars van andere Mandarijndialecten een verstaanbaarheidsbeoordeling krijgen van bijna 7 op de schaal. Hetzelfde effect is te zien in de overeenkomstbeoordelingen.
c. Als luisteraars de dialecten beoordeelden van sprekers van variëteiten uit de tegenovergestelde hoofdgroep ('other branch'), dan blijken de Zuidelijke dialectsprekers even onverstaanbaar voor Mandarijndialectduisteraars als voor luisteraars met een ander Zuidelijk dialect. Op dezelfde manier worden de dialecten van Zuidelijke sprekers als even afwijkend van het eigen dialect aangemerkt door luisteraars van andere Zuidelijke dialecten als door die van Mandarijndialecten. Toch worden de Mandarijndialectsprekers als enigszins verstaanbaar beoordeeld door Zuidelijke luisteraars (3.7); het omgekeerde is echter niet het geval (score van slechts 1.7). In termen van beoordeelde overeenkomst is er in het geheel geen verschil: Zuidelijke dialectluisteraars vinden de Mandarijndialecten even afwijkend van hun eigen dialect (gemiddelde overeenkomstbeoordeling van 1.4) als omgekeerd: Mandarijndialectluisteraars geven de Zuidelijke dialecten een overeenkomstcijfer van 0.9.

Om de mogelijke invloed na te gaan van individuele verschillen in geluidskwaliteit en van de kwaliteitsvermindering van het spraaksignaal als gevolg van de digitale manipulaties (toonhoogteveranderingen, en bij drie sprekers een vrouw-manverandering) heb ik een controle-experiment uitgevoerd. Ik heb beoordelingen opgevraagd van de geluidskwaliteit, d.w.z. de eigenschap die uitspraakkwaliteit van de spreker en de opnamekwaliteit overkoepelt, van de 2 × 15 dialectfragmenten (zowel van de gemonotoniseerde als van de geïntoneerde). Ik heb daarbij geabstraheerd van verstaanbaarheid door de fragmenten aan te bieden aan luisteraars aan de Universiteit Leiden, d.w.z. studenten en collega's aan het Leiden University Centre for Linguistics. Dat waren uitsluitend sprekers van Europese talen; zij hadden geen gebruikskennis van het Chinees.

De resultaten wijzen uit dat deze luisteraars de geïntoneerde versies consequent beter vonden klinken (5-8 op de schaal) dan de gemonotoniseerde versies (3-4 op de schaal). De correlatie tussen de geïntoneerde en gemonotoniseerde versies is hoog met $r = .884$ (N = 15, $p < .001$). De correlatie tussen geluidskwaliteit en verstaanbaarheid (deze laatste zoals vastgesteld in het hoofdexperiment) was echter laag en insignificant, $r = .205$ (N = 30, ins.) als de verstaanbaarheid van de geïntoneerde en gemontoniseerde fragmenten was beoordeeld door luisteraars van het 'eigen dialect'. De correlatie was was nog slechter als de zij apart werd berekend voor de geïntoneerde ($r = .189$, N = 15, ins.) en de gemonotoniseerde ($r = .098$, N = 15, ins.) versies. De correlatie tussen beoordeelde geluidskwaliteit en verstaanbaarheid (zoals beoordeeld door alle luisteraargroepen in het hoofdexperiment) is vrijwel nul, $r = .019$ (N = 15, ins.) voor gemonotoniseerde versies of zelfs negatief, $r = -.195$ (N = 15, ins.) voor geïntoneerde spraakfragmenten en $r = -.204$ (N = 15, ins.) gemiddeld over beide versies per fragment. Als we alleen de verstaanbaarheidsbeoordelingen van luisteraars van het Beijingdialect (dat het meest lijkt op het Standaardmandarijn) in de beschouwing betrekken, blijkt evenmin een een correlatie tussen geluiddkwaliteit en verstaanbaarheidsbeoordeling.

De conclusies de ik trek in dit hoofdstuk zijn dan:

(1) Er is altijd enige mate van (beoordeelde) onderlinge verstaanbaarheid tussen paren Chinese dialecten.

(2) De veronderstelde asymmetrie in onderlinge verstaanbaarheid tussen Mandarijnen niet-Mandarijndialecten is experimenteel aangetoond en statistisch hard te maken.

(3) De primaire splitsing tussen Mandarijnen- en Zuidelijke dialecten wordt op hoofdlijnen weerspiegeld in de resultaten van het experiment. De subclassificatie van dialecten binnen de hoofdgroepen correspondeerde echter niet met de traditionele dialectindeling voor de Chinese talen.

(4) Er is geen significant verschil tussen de beoordelingen van fragmenten met en zonder tooninformatie. Kennelijk heeft woordtoon maar een geringe invloed op onderlinge verstaanbaarheid. De iets systematischer resultaten bij de gemonotoniseerde fragmenten kunnen we begrijpen als we aannemen dat luisteraars hun aandacht beter kunnen richten op de overige spraakeigenschappen als de toonhoogteinformatie uit het signaal is verwijderd.

(5) De hoge correlatie tussen beoordeelde overeenkomst en beoordeelde verstaanbaarheid duidt erop dat deze twee schalen in de praktijk dezelfde eigenschap meten.

(6) Het ontbreken van enige correlatie tussen beoordeelde evrstaanbaarheid en beoordeelde geluidskwaliteit van de fragmenten garandeert date er in het hoofdexperiment heen artefact is opgetreden van verschillen in geluidskwaliteit. De beoordeling van verstaanbaarheid en overeenkomst door mijn Chinese luisteraars is dus niet gebaseerd op de de geluidskwaliteit van de opnames.

In **hoofdstuk vier** heb ik de onderlinge verstaanbaarheid bepaald tussen Chinese dialecten met bulp van functionele tests, waarbij ik verstaanbaarheidscores heb verzameld op woord- en op zinsniveau. Deze twee typen resultaat zijn vergeleken met elkaar en ook met de scores n.a.v. het eerdere beoordelingsexperiment. De resulterende boomstructuren zijn daarnaast gevalideerd tegen de traditionele dialectindelingen.

Voor de woordverstaanbaarheidstest heb ik 150 woorden geslecteerd uit het basisvocalulaire en onderverdeeld in tien betekeniscategorieën (die in het dagelijks leven gebruikt worden, zoals lichaamsdeel, familierelatie, etenswaar, enz.), met 15 woorden per categorie. Voor de zinsverstaanbaarheidstest heb ik 60 SPIN-zinnen (Speech Perception in Noise test) gebruikt, die ik vanuit het Engels in het Mandarijn heb vertaald. Daarbij heb ik alleen zinnen geselecteerd met een contextueel hoogvoorspelbaar doelwoord aan het zinseinde (ook na vertaling) en die betrekking hadden op situaties die ook in de Chinese samenleving alledaags zijn. Twee moedertaalsprekers (een man, een vrouw) werden aangezocht voor elk van mijn 15 dialecten. Ieder paar vertaalde de 150 woorden en 60 zinnen in hun eigen dialect (consensusvertaling) en maakte vervolgens een geluidsopname waarin zij de woorden en de zinnen in hun eigen dialect voorlazen. Voor ieder van de 15 dialecten zijn 15 CD's aangemaakt met verschillende stimulusvolgordes (en met de woorden en zinnen in aparte tests) geblockt volgens een Latin-sqare design. De sets van 15 verschillende CD's werden afgespeeld voor 15 luisteraars (één CD per luisteraar) voor elk van de 15 dialecten. In de woordverstaanbaarheidtest werd de luisteraar gevraagd ieder woord in te delen in de bestpassende betekeniscategorie (met gedwongen keuze uit de tien voorgegeven categorieën). In de zinsverstaanbaarheidstest was de taak om het laatste woord van elke zin te vertalen in het eigen dialect.

In totaal werden 33.750 responsies (15 × 150 × 15) verzameld n.a.v. de woordstimuli en nog eens 13.500 (15 × 60 × 15) voor de zinsstimuli. Ik behandel eerst de wordresultaten. Statistische analyse wijst het volgende uit: (1) er bestaat altijd enige mate van onderlinge verstaanbaarheid tussen paren Chinese dialecten, (2) luisteraars herkenden het hoogste percentage woorden in hun eigen dialect, (3) Mandarijndialectsprekers werden altijd beter verstaan dan de Zuidelijke dialectsprekers door zowel niet-Mandarijnluisteraars als door Mandarijnluisteraars (de gemiddelden zijn atijd hoger) (4) de primaire splitsing in Mandarijn- en niet-Mandarijndialecten wordt correct weerspiegeld door de boomstructuur die uit de woordverstaanbaarheidsmatrix wordt gegenereerd. Taiyuan wordt andermaal ingedeeld bij de Mandarijngroep. De zinsverstaanbaarheid vertoont dezelfde structuur met dien verstande dat de gemiddelde scores over de hele linie hoger liggen dan bij de woordverstaanbaarheid.

De asymmetrie in onderlinge verstaanbaarheid tussen Mandarijn- en niet-Mandarijn wordt dus opnieuw bevestigd, nu door functionele verstaanbaarheidstests. De gemiddelde woordverstaanbaarheid over de zes Mandarijndialecten ('own dialect') is 72%, maar zakt bij de Zuidelijke (niet-Mandarijn)luisteraars tot 52%. Mandarijnluisteraars verstaan sprekers van andere Mandarijndialecten vrij goed (61%) maar Zuidelijke luisteraars verstaan andere Zuidelijke dialectsprekers slecht (22%). De onderlinge verstaanbaarheid tussen Mandarijn- en Zuidelijke dialectsprekers/luisteraars is over en weer (symmetrisch) slecht (32 to 36%).

Hetzelfde resultaat vinden we bij de zinsverstaanheid. Onderlinge verstaanbaarheid uit heel goed binnen de Mandarijndialecten maar erg slecht in de Zuidelijke (niet-Mandarijn) tak. Zuidelijke dialectsprekers zijn even onverstaanbaar voor Mandarijnluisteraars als voor andere Zuidelijke dialectluisteraars. Daarentegen zijn sprekers van Mandarijndialectem tamelijk goed verstaanbaar door Zuidelijke luisteraars (54% verstaanbaarheid); dit effect blijft grotendeels overeind ook als we de sprekers van het Beijing-dialect uitsluiten (48%).

Woord- en zinsverstaanbaarheid zijn sterk gecorreleerd, namelijk $r = .835$ (N = 225, $p < .001$). Bij berekening op alleen de niet-redundante halve matrix (symmetrisch gemaakt via middeling over contra-diagonale cellen) stijgt de correlatie zelfs tot $r = .928$ (N = 105, $p < .001$).

Om na te gaan of opinietests (die sneller en meer economisch zijn te organiseren) kunnen dienen als vervangingsmiddel voor functionele verstaanbaarheidstests, ben ik nagegaan hoe goed de functionele scores voorspeld kunen worden uit die van de opinietests (beoordelingstests). De analsye laat zien dat (1) beoordeelde verstaanbaarheid een betere voorspeller is van de functionele verstaanbaarheid dan beoordeelde dialectovereenkomst, (2) functionele verstaanbaarheid op zinsniveau beter te voorspellen is dan op woordniveau. De correlatiecoëfficiënten waren het hoogst voor de symmetrisch gemaakte have matrices, die naar ons idee de beste benadering zijn van onderlinge verstaanbaarheid: beoordeelde verstaanbaarheid (en niet zo zeer beoordeelde dialectovereenkomst) correleert met functionele zinsverstaanbaarheid (meer dan met woordverstaanheid) met $r = .818$ (N = 105, $p < .001$)

De volgende conclusies zijn uit dit hoofdstuk te trekken:

(1) Onderlinge verstaanbaarheid tussen Chinese dialecten kan adequaat worden vastgesteld met behulp van functionele testmethoden.

(2) De onderlinge verstaanbaarheid binnen de Mandarijndialecten is intrinsiek beter dan die tussen de niet-Mandarijn (Zuidelijke) dialecten, zowel op woord- als op zinsniveau. Zuidelijke niet-Mandarijnluisteraars verstaan Mandarijndialecten consistent beter dan niet-Mandarijndialecten.

(3) De primaire tweedeling tussen Mandarijn- and niet-Mandarindialecten wordt correct weerspiegeld maar dit geldt niet voor de verdere indeling van dialecten binnen de dialectgroepen.

(4) Alle subjectieve maten zijn significant met elkaar gecorreleerd. De r-waarde is consistent hoger tussen subjectieve maten van dezelfde soort (beoordelingsmaten onderling, funcionele maten onderling) dan over de soorten heen.

(5) Alle resultaten kloppen tot op zekere hoogte met de traditionele Chinese dialectindeling. Functionele verstaanbaarheidsmaten weerspiegelen de taalkundige dialectindelingen beter dan de beoordelingsmaten. De beste benadering van de traditionele Chinese dialectindeling wordt verkregen aan de hand van functionele verstaanbaarheid op zinsniveau.

In hoofdstuk vijf heb ik een groot aantal objectieve afstandsmaten berekend op mijn Sinitische dialecten. Deze afstandsmaten waren ofwel beschikbaar in gepubliceerde literatuur of ze zijn door mij zelf berekend. Eén afstandsmaat is gebaseerd op lexicale overeenkomst tussen dialecten, alle andere maten betreffen aspecten van de klankstructuur van de dialecten. Alle maten werden gebruikt om verwantschapsbomen mee te genereren, die vervolgens alleen op hoofdlijnen zijn vergeleken met traditionele dialect-indelingen. Het doel was om alleen die maten te selecteren die in ieder geval de primaire tweedeling in Mandarijn- en niet-Mandarijndialecten reproduren, te selecteren als kansrijke predictoren om te gebruiken in hoofdstuk zes, waar ik wil beslissen welke maat (of combinatie van maten) de beste voorspeller is van onderlinge verstaanbaarheid tussen Sinitische dialecten.

De objectieve maten van structurele overeenkomst vallen uiteen in twee categorieën: lexicale verwantschap em fonologische verwantschap. Lexicale verwantschap kwaltificeert in welke mate twee dialecten dezelfde woordengebruiken voor dezelfde concepten (zgn. cognaten, woorden met een gemeenschappelijke etymologie), d.w.z. de proportie cognaten dat gedeeld wordt door de vocabularies van twee dialecten (of talen). Fonologische verwantschap wordt gedefinieerd op de alleen de verzameling cognaten die twee dialecten gemeenschappelijk hebben. Dit type maten vertelt ons iets over hoeveel de klankvormen (de klinkers en medeklinkers, en in het Chinees ook de tonen) van de cognaten op elkaar lijken.

Ik heb de lexicale verwantschapsmaat voor 13 dialects (een strikte deelverwamenling van mijn 15 dialectem) uit de literatuur kunnen overnemen (op basis van werk van Cheng 1982, 1986, 1991, 1997). Ik heb vervolgens een lexicale verwantschapsboom gegenereerd. Deze boom levert twee indelingsfouten op: het cluster dat de zes Mandarijndialecten bevat, wordt verontreinigd door twee niet-Mandarijn (= Zuidelijke) dialecten, nl. Nanchang and Changsha.

Fonologische afstandsmaten zijn op een aantal manieren afgeleid, hetzij overgenomen uit de literatuur hetzij door mij zelf berekend op gepubliceerde databases pf op een digitale database die voor mij toegankelijk is gemaakt (zie onder).

(1) Eén verzameling van overeenkomstmaten is berekend op de fonologische invaentarissen van de 15 dialecten, apart voor initiëlen ('welke beginmedeklinkers komen voor in beide talen?'), nuclei ('welke klinkers komen voor in beide talen?') , coda's ('welke medeklinkers komen in beide talen voor aan het einde van de lettergreep?'), finalen (welke rijmdelen van letergrepen komen voor in beide talen?'), woordtonen ('welke tonen komen voor in beide dialecten?') alsook een aantal samengestelde maten zoals het gezamenlijk voorkomen van bepaalde medeklinkers aan zowel begin als einde van de lettergreep. Van al deze maten weerspiegelde alleen de boomstructuur die kon worden opgetrokken n.a.v. de medeklinkers die in beide dialecten aan het begin van lettergrepen kunnen voorkomen, de primaire tweedeling tussen Mandarijn- en niet-Mandarijndialecten op overtuigende wijze. Er was slechts één classificatiefout: Changsha werd ten onrechte ingedeeld bij de Mandarijndialecten.
(2) Een tweede set maten was gebaseerd op de leciale frequentie van de verschijnselen genoemd onder (1). Deze frequenties zijn geteld op een lexicale database van 764 woorden (in al mijn dialecten die is damengesteld door onderzoekers aan het Department of Linguistics van de Chinese Academy of Social Sciences in Beijing (CASS). Hier weerspiegelde de verwantschapsboom op basis van het aantal

gemeenschappelijke vocalische nuclei (gewogen naar lexicale frequentie) de traditionele dialectindeling het best.

(3) Ik heb voor elk paar van mijn 15 dialecten Levenshtein string-edit afstanden berekend op de CASS-database van 764 alledaagse woorden met behulp van het LO4 softwarepakket dat is ontwikkeld aan de Rijksuniversiteit Groningen, eenmaal met en eenmaal zonder perceptieve weging van klankverschillen. De berekeningen heb ik apart uitgevoerd voor segmentele en tonale eigenschappen van de woorden. De resultaten wijzen uit dat de Levenshteinafstanden teleurststellen. Als we andermaal de traditionele dialectindeling als maatstaf nemen, dan is het geringste aantal classificatiefouten 2.5 (voor ongewogen klankverschillen). Vooral overeenkomst in het tonale domein lijkt een slecht criterium om verwantschap tussen Chinese dialecten mee vast te stellen.

(4) Het laatste type afstandsmaten heb ik gekopieerd uit publicaties van Cheng, Zijn berekeningen van fonologische verschillen zijn uitgevoerd op de *Zihui* database *[Woordenlijst van Chinese dialecten],* die transcripties bevat van ruim 2.700 woords parallel voor 17 Chinese dialecten, waaronder al mijn 15 dialecten. Vijf fonologische overeenkomstmaten zijn berekend: (op basis van) de frequentie van de initialen, de frequentie van de finalen (syllaberijmdelen), de frequentie van de woordtonen, de frequentie van intialen en finalen gecombineerd, en de frequenties van intialen, finalen en woordtonen gecombineerd. Een zesde, veel complexere maat beoogde de graad van fonologische overkomst te vatten voor de cognate delen van de woordenschat in elk paar dialecten. Deze maat, die ik de Phonological Similarity Index (PCI) heb genoemd, drukt in beginsel uit hoeveel formele regels nodig zijn om de fonetische transcripties van alle cognaten in het ene dialect om te zetten in correcte transcripties in het andere dialect. Deze maat is asymmetrisch omdat de regelset die nodig is om symboolreeksen van dialect A om te zetten in die van B meer (of minder) complex kan zijn dan de set die nodig is om de reeksen van dialect B om te zetten in die van A.

De beste classificatie van de 15 dialecten in mijn steekproef wordt opgeleverd door een samengestelde maat die is voorgesteld door Cheng (1997), nl. de overeenkomst in de lexicale frequenties van de initialen (beginmedeklinkers) de finalen (rijmdelen) en de woordtonen gecombineerd, zoals geteld in de lijst van 2.770 alledaagse woorden in Chinees. Deze maat levert een perfecte tweedeling op in Mandarijn en niet-Mandarijndialecten, met uitzondering van één, Taiyuan, waarvan de indeling onbestemd bleef. Ook hier echter was de maat niet in staat om ook maar enigszins de verdere onderverdeling van de dialecten binnen de hoofdtakken te achterhalen.

In **hoofdstuk zes** heb ik de objectieve afstandsmaten die ik in hoofdstuk vijf heb verzameld als voorspellers van onderlinge verstaanbaarheid. I heb de overeenkomstmaten tussen elk paar dialecten gecorreleerd met de onderlinge verstaanbaarheidsscores die ik heb gevonden in hoofdstukken drie en vier. De volgende observaties kunne worden gedaan aan de hand van de regressieanalyse: (1) Overeenkomstbeoordelingen zijn wat moeilijker te voorspellen uit taalkundige afstandsmaten dan de verstaanbaarheidsscores (ongeacht of die gebaseerd waren op opinietests fof functionele tests). R^2-waarden lagen onder de .8 voor overeenkomstbeoordelingen maar boven de .8 voor alle andere criteriumvariabelen. (2) Functionele verstaanbaarheidsscores (zowel op het woord- als op het zinsniveau bepaald) zijn minder goed te voorspellen dan verstaanbaarheidsopinies. Daarom kunnen we niet beslissen of beoordeelde of functioneel vastgestelde verstaanbaarheid zich het best leent voor voorspelling uit objectieve maten. Toch concludeer ik dat functioneel bepaalde verstaanbaarheid de voorkeur verdient als testmethode maar dat is omdat de boomstructuren die deze methode oplevert het best overeenkomen met de traditionele Chinese dialectindeling.

In **hoofdstuk zeven** zet ik mijn belangrijkste bevindingen nog eens op een rij en probeer ik antwoord te geven op de onderzoeksvragen uit de inleiding.

(1) Mijn experimenten wijzen uit dat beoordeelde verstaanbaarheid significant gecorreleerd is met beoordeelde taalkundige overeenkomst tussen dialecten. In de Chinese dialectsituatie speelt tonale informatie slechts een ondergeschikte rol bij onderlinge verstaanbaarheid, wat blijkt uit de vrijwel identieke resultaten die ik vond bij spraakfragmenten met en zonder oonhoogteinformatie.

(2) Er is altijd wel enige mate van onderlinge verstaanbaarheid tussen Chinese dialecten. De mate van onderlinge verstaanbaarheid kan worden vastgesteld aan de hand van experimenten (zowel beoordelings-/opinietests als functionele woord- en zinsverstaanbaarheidstests). Mandarijndialecten zijn beter te verstaan voor niet-Mandarijnluisteraars dan omgekeerd. Dialecten binnen de Mandarijnhoofdgroep zijn onderling beter te verstaan dan de dialecten binnen de Zuidelijke (niet-Mandarijn)-hoofdgroep. Functionele verstaanbaarheidstests bevestigen de impresionistische claims uit de literatuur duidelijker dan beoordelings-/opninietests. De veronderstelde asymmetrie is het duidelijkst zichtbaar in de resultaten vann de functionele zinsverstaanheidstest.

(3) Alle experimentele resultaten (van de beoordelings/opinietests en van de functionele tests) zijn onderling sterk gecorreleerd, beoordeelde verstaanbaarheid met overeenkomstbeoordeling (r = .888, N = 105), woordverstaanbaarheid met zinsverstaanbaarheid (r = .928.), functionele woordverstaanbaarheid met beoordeelde verstaanbaarheid (r = .772) en functionele zinsverstaanbaarheid met beoordeelde verstaanbaarheid (r = .818).

(4) De onderlinge verstaanbaarheid van Chinese dialecten (zowel opiniescores als functionele testscores) kunnen tot op zekere hoogte voorspeld worden uit taalstructurele afstandsmaten. Overeenkomstbeoordelingen zijn wat moeilijker te voorspellen uit taalkundige afstandsmaten dan de verstaanbaarheidsscores (al dan niet functioneel bepaald). Functionele verstaanbaarheidsscores, bepaald aan de hand van losse woorden of van (korte) zinnen, zijn niet beter te voorspellen dan verstaanbaarheidsoordelen.

(5) De betwiste status van de Jin-dialecten (hier vertegenwoordigd door Taiyuan) kan worden beslecht op basis van mijn experimenten. Alle boomstructuren n.a.v. de experimenten geven aan dat Taiyuan een Mandarijndialect is. Onze stelling is daarom dat er geen duidelijke reden is om de Jin-groep buiten de tak van de Mandarijndialecten te houden.

(6) De dendrogrammen die gegenereerd zijn op basis van de subjectief en objectief verkregen verwantschapsscores, zijn gevalideerd tegen de traditionele Chinese dialectindeling, in eerste instantie alleen op de primaire tweedeling tussen Mandarijn- en niet-Mandarijndialecten, en daarna ook nog wel enigszins aan de hand van secundaire onderverdeling binnen de hoofdtakken. De dendrogrammen die zijn verkregen uit de onderlinge verstaanheidsscores geven in het algemeen een correcte weergave van de primaire tweedeling in Mandarijn- en niet-Mandarijndialec-

ten zoals voorgesteld door Chinese taalkundigen; in sommige gevallen wordt ook een deel van de interne structuur binnen de hoofdgroepen teruggevonden. De claim dat onderlinge verstaanbaarheid gebruikt kan worden bij dialectclassificatie en kan dienen als criterium ter validatie van dialectindelingen is – in elk geval gedeeltelijk – waargemaakt. Mijn eindconclusie is dat functionele verstaanbaarheidsmetingen (vooral indien bepaald op zinsniveau) de traditionele Chinese dialectindeling zoals voorgesteld door taalkundigen, het best weerspiegelen.

(7) Verder onderzoek is nodig dat is gebaseerd op een ruimere keuze van dialecten en met tests die ook de hogere taalkundige , bv. syntactische, niveaus aanspreken.

Summary

This dissertation establishes the degree of mutual intelligibility between 15 target Sinitic language varieties through experiments administering both functional and opinion tests. The correlation between the results from the two test methods was calculated to see how much these two methods test the same property. The results from the two experimental tests were also correlated with objectively collected linguistic distance measures (either published in the literature or computed by myself) to see how much the mutual intelligibility can be predicted from the distance measures. The scores based on the experiments and distance measures are used to generate hierarchical structures (tree structures) expressing the affinity relationships between these language varieties. It is proposed that the mutual intelligibility testing method is to be more preferred as the tree structures generated from its results correspond better with traditional language and dialect taxonomies set up by linguists.

Chapter One is the Introduction. In this part, I explain why we need to measure the distance between pairs of language varieties and how we could measure the distance between pairs of language varieties. Distance between language varieties is a potentially important parameter to distinguish dialect(s) from language(s). If the distance is small, i.e., if two language varieties resemble each other a lot, these varieties are likely to be classified as dialects of the same language; otherwise, they are two different languages. Linguistic distance can be measured objectively and subjectively. The objective method is to quantify the structural difference between pairs of language varieties, along multiple dimensions. A subjective approach is to determine the degree of (mutual) intelligibility between pairs of language varieties. The subjective approach is conceptually simpler, because the overall degree of (mutual) intelligibility can be expressed along a single dimension.

Mutual intelligibility can be tested by functional methods and opinion methods. The functional approach tests to what extent listeners actually recognize linguistic units in the spoken stimuli and understand the message expressed by them. Opinion methods test how well listeners (natives of language A) *think* they would understand a speaker of a (different) language B. When establishing the mutual intelligibility between languages or dialects (language varieties is often used as the superterm), the number of comparisons to be made grows exponentially with the number of varieties. If 10 varieties are compared, $10^2 = 100$ different combinations of speaker and listener groups are required. When the number of varieties is 15, the number of combinations is as high as $15^2 = 225$. Moreover, when one listener has to show how well s/he understands 10 or even 15 different language varieties, s/he should never listen to the same text twice, in order to avoid learning or repetition effects. Such effects can be overcome by enlisting very large numbers of listeners (so that they can be split into many different but perfectly equivalent subgroups each of which is exposed to a different language variety) or by devising many alternative sets of perfectly equivalent test materials. To avoid this extremely laborious process, opinion (or judgment) testing can (and has been)

used as a shortcut for functional testing, especially in situations in which where the number of language pairs is large. The present research aims to find out to what extent the short-cut through opinion testing is valid.

Broadly speaking, two different criteria were adopted in order to establish the validity of the functional and opinion testing methods of intelligibility testing. The first criterion asks how well the results of either intelligibility testing method can be predicted from objectively determined measures of structural difference between the language varieties concerned. The second validity criterion is the degree of congruence between (i) affinity trees generated for the varieties from either the functional or opinion test scores and (ii) the so-called cladistic trees or taxonomies drawn up for the varieties by linguists (dialectologists).

Generally speaking, this dissertation contributes to establish the degree of mutual intelligibility between Sinitic dialects and confirms several questions via experimental functional and opinion approaches: (1) The Mandarin dialects are more *internally* mutual intelligible than Southern dialects. (2) The Southern dialects are less intelligible to Mandarin dialects than vice versa. (3) The debated Jin dialects (represented by Taiyuan) are more reasonably classified into the Mandarin dialects than into Southern dialects. (4) More structural distance measures on Sinitic varieties were computed based on the databank and the relatively better predictors were decided by multiple regression techniques. (5) Functional approach at the sentence level is tested to be more valid to correspond with the traditional Sinitic dialect taxonomy.

Chapter Two provides background information on the language situation in China. The language environment in China is diverse. A wealth of dialects (often distinct languages) are spoken in China and their classification varies depending on the specific criteria used. This dissertation aims at 15 dialects belonging to the Sinitic stock. Two major taxonomies were used as reference points. There is major consensus and one (minor) discrepancy between these two taxonomies: (1) The taxonomies are based on the historical changes of phonological features, i.e., phonetic changes and tone evolution, for example, the simplification of initial consonants, the loss of glottal stops, the appearance of the voice-voiceless initial consonants, the Yin-Yang split (high versus low tone register) of four original tone categories, the subsequent merging or further splitting of tones. (2) It is agreed that there is primary split between Mandarin and Southern dialects. Mandarin dialects have lost the final stop consonants and retained fewer lexical tone categories. Southern dialects kept the final stops and retained richer tone inventories. The sub-groupings for both Mandarin and Southern branches are not consensually agreed. (3) The Mandarin dialects are claimed to be *internally* more mutually intelligible than Southern dialects because of the greater intrinsic uniformity among Mandarin dialects, independently of the influence of Standard Mandarin (based on Beijing dialect). (4) Mandarin dialects are claimed to be more intelligible to Southern dialects than vice versa. (5) The degree of mutual intelligibility within and across the Mandarin and Southern branches is not yet established and there is no test to validate the mutual (un)intelligibility within and across the Mandarin and Southern dialects. (6) The classification of Jin dialects is controversial but can possibly be settled on the basis of experimental data taken from mutual intelligibility testing.

The following Chinese dialects were included in the present study: six Mandarin dialects, subdivided into a Northern group (Beijing, Jinan, Xi'an, Taiyuan) and a South-Western group (Chengdu, Hankou), and nine non-Mandarin (Southern) dialects, divided into six groups, each represented by one to three dialects: Suzhou, Wenzhou (Wu dialects), Nanchang (Gan dialect), Meixian (Hakka dialect), Xiamen, Fuzhou, Chaozhou (Min dialects), Changsha (Xiang dialect), and Guangzhou (Yue dialect). The location of the dialects can be seen on the cover of this book. The classification of Taiyuan with the Mandarin group is provisional only.

Chapter Three describes the mutual intelligibility testing experiments via the opinion approach. I used existing speech passages of the fable *The North Wind and the Sun*, read by native speakers of the selected 15 dialects. Each of the passages was manipulated so that the voices suggested the same (male) gender and had the same speaking rate (with the same pause duration), and the same mean pitch. Two versions of each passage were made, one with and one without pitch information, using the PSOLA pitch manipulation function in the Praat software. The melodic and monotonized versions of the dialect samples were used to estimate the influence of the tone information. Twenty-four (mono-dialectal) listeners for each of the 15 dialects were recruited. They were instructed to rate speakers on 11-point scales (ranging from 0 to 10) on two parameters: (i) intelligibility and (ii) similarity to the listener's own dialect. On the rating scales '0' stood for 'complete unintelligibility' and 'no similarity at all', '10' represented 'perfect intelligibility' and 'complete similarity'. In total 15 listener groups (one group per dialect) × 15 speaker dialects = 225 combinations of speaker-listener dialects were tested. Each listener group comprised 24 listeners, each speaker dialect was represented by one speaker in each of two versions (monotonized, intonated). The 21,600 scores collected were used to generate agglomeration trees.

The results show that no perfect grouping of target dialects is formed according to the four trees produced by the experiments, compared to the traditional taxonomies. Firstly, the primary split between Mandarin and Southern dialects is not ideally reflected, the internal grouping structures are unstable. Two dialects traditionally classified as Southern (Nanchang and Changsha) were parsed with the Mandarin dialects. Secondly, neither of the sub-clusters in Mandarin branch, i.e., South-Western Mandarin nor the sub-clusters in the families of Wu, Min is clearly reflected. Broadly, I obtained the following results: (1) The basic Mandarin-Southern split was correctly reflected in the agglomeration trees, after some adjustment of criteria. (2) Taiyuan (the representative of Jin dialects, whose classification is still undecided) is consistently grouped with the Mandarin dialects. (3) The correlation coefficients between judged similarity and judged intelligibility in all cases (off-diagonal, the full, lower part matrix, with and without pitch information, respectively) are reasonably high (correlation coefficients between $r = .810$ and $r = .888$ for intonated versions, and from $r = .841$ to $r = .900$ for the monotonized versions). The monotonized speech samples consistently produced better correlations. (4) In more detail, the plotted figures based on mean judged intelligibility and mean judged similarity, broken down by three groups of listeners, show that:

a. Both judged intelligibility and judged similarity are very high and close the maximum possible score of 10 when listeners have the same dialect as the speakers: 'own dialect'.

b. A very clear difference between Mandarin and non-Mandarin (Southern) speakers can be seen when listeners judged speakers of another dialect within the same branch: 'same branch'. A Southern (non-Mandarin) speaker is judged to be practically unintelligible by other non-Mandarin listeners (1.5 on the 10-point scale) whilst the Mandarin dialect speakers receive a mean intelligibility judgment close to 7 by other Mandarin listeners. The same effect is observed in the similarity ratings.
c. When listeners judged the dialects spoken by the speaker of varieties opposed to their own branch ('other branch'), non-Mandarin (Southern) speakers are as unintelligible to Mandarin listeners as they are to other non-Mandarin (Southern) listeners. Also, the non-Mandarin (Southern) speakers' dialects are judged to be as different from the listeners dialect by non-Mandarin (Southern) and Mandarin listeners alike. Mandarin speakers, however, are considered to be somewhat intelligible by non-Mandarin (Southern) listeners (3.7), the opposed situation is not the case (only 1.7). In terms of judged similarity there is no difference: non-Mandarin (Southern) listeners consider the Mandarin speakers' dialects to be as different from their own (mean similarity rating of 1.4) as *vice versa* (Mandarin listeners responding to non-Mandarin/Southern speakers, with a mean similarity rating of 0.9).

In order to test the possible influence of individual difference of sound quality and the possible deteriorations caused by the gender transformation and other sound manipulations, I ran a control experiment. I collected perceived judgments of the sound quality, i.e. the overall property comprising the articulatory quality of the speaker and the recording quality) of the 2 × 15 dialect samples (both monotonized and melodic) in the abstraction of intelligibility by playing the samples to listeners at Leiden University among students and colleagues in the Leiden University Centre for Linguistics who were native speakers of various European languages with no working knowledge of Chinese.

The results show that the listeners reliably judged the melodic versions (5-8 scale range) to have better sound quality than monotonized versions (3-4 scale range). The correlation between the intonated and monotonized versions is strong with $r = .884$ (N = 15, $p < .001$). However, the correlation between sound quality and intelligibility (as established in the main experiment) was poor and insignificant, $r = .205$ (N = 30, ins.) when the intelligibility of melodic and monotonized samples was judged by 'own-dialect' listeners. The correlation was even poorer when computed for the two melodic versions separately, $r = .189$ (N = 15, ins.) for intonated samples and $r = .098$ (N = 15, ins.) for monotonized versions. Correlations computed between judged sound quality and the overall intelligibility judgments across all listener groups are practically zero, $r = .019$ (N = 15, ins.) for monotonized versions, or even negative, $r = -.195$ (N = 15, ins.) for intonated samples and $r = -.204$ (N = 15, ins.) across both melodic versions. The result with Beijing-only listeners shows no correlation between the sound quality and judged intelligibility blocked by melodic and monotonized versions.

Conclusions in this chapter are as follows:
(1) There always exists some degree of judged mutual intelligibility between pairs of Sinitic dialects.

(2) The asymmetry of mutual intelligibility between Mandarin and non-Mandarin dialects is experimentally found and statistically confirmed.
(3) The primary Mandarin-Southern split is basically reflected by the experimental results. However, the sub-clusters did not accurately correspond with the traditional dialect taxonomy.
(4) There is no significant difference of the judged results with and without tone information. Apparently, tone only has a minor influence on mutual intelligibility. The slightly more systematic results with monotonized versions can be explained when we assume that listeners focus better on the remaining linguistic features when the pitch information was removed in the mutual intelligibility testing.
(5) The high correlation between the judged similarity and judged intelligibility indicates that these two scales actually measure the same property.
(6) The absence of any correlation between judged intelligibility and judged sound quality insures that there is no artifact of sound quality in our experiment. The Chinese listener judgments of intelligibility and similarity were not based on the actual sound quality of the recordings they were exposed to.

Chapter Four measures the mutual intelligibility between Sinitic dialects using functional tests. I collected functional intelligibility scores at the word and sentence level. I compared the two results with each other, as well as with the earlier judgments results in the opinion experiments. The hierarchical structures were also validated against the traditional dialect taxonomies.

For the word-intelligibility test I selected 150 basic words subdivided into ten semantic categories (used regularly in daily life such as body parts, family member, food, etc.), with 15 words in each category. For the sentence intelligibility test I used 60 SPIN (Speech Perception in Noise test) sentences, which I had translated from English into Mandarin. I selected only sentences which had the contextually predictable target word in final position after translation, and which dealt with situations that are also applicable to Chinese society. Two native speakers (one male, one female) for each of the 15 sample dialects were recruited. They translated the 150 words and 60 sentences into their own dialect (consensus translation) and then recorded their readings of the words and sentences. 15 CDs for each of 15 dialects were created with different stimuli orders (and words and sentences in separate tests) blocked by Latin Square design. Fifteen copies for each of 15 CDs were played to 15 listeners for each of 15 sample dialects. In the word-intelligibility test, listeners were asked to classify each word into the best fitting semantic category (with forced choice from ten categories). In the sentence-intelligibility test, the task was to translate the final word in each sentence into their own dialect.

In all, I collected 33,750 responses (15 × 150 × 15) for the word stimuli and another 13,500 (15 × 60 × 15) for the sentence stimuli. I will first deal with the word results. Statistical analysis shows that: (1) again there is always some degree of mutual intelligibility between pairs of Sinitic dialects, (2) listeners recognized the highest percentage of words in their own dialects, (3) Mandarin speakers were always better understood by both non-Mandarin and other Mandarin listeners (the means are always higher), (4) the primary split is correctly reflected in the tree structure generated from the word intelligibility matrix. Taiyuan is again in the Mandarin branch. The sentence intelligibility results show the same structure except that the mean scores are generally higher than in the word intelligibility.

Again, the asymmetry of mutual intelligibility between Mandarin and non-Mandarin is confirmed by the functional tests. Mean word intelligibility across the six Mandarin dialects is 72%, while the mean correct classification of the listeners with non-Mandarin (Southern) native dialects is 52%. Mandarin listeners understand speakers of other Mandarin dialects rather well (61%) whilst Southern listeners understand speakers of other Southern dialect very poorly (22%). The reciprocal intelligibility between Mandarin and Southern dialect speakers and listeners is symmetrically poor (32 to 36%).

The same result is found for sentence intelligibility. Mutual intelligibility is very good within the Mandarin dialects and very poor in the non-Mandarin (Southern) dialect branch. Non-Mandarin (Southern) dialects are as poorly intelligible to Mandarin listeners as they are to non-Mandarin (Southern) listeners. Mandarin speakers are fairly intelligible to non-Mandarin (Southern) listeners (54% intelligibility), and this effect largely remains even if we exclude Beijing speakers (48%).

Word and sentence intelligibility were strongly correlated, viz. $r = .835$ (N = 225, $p < .001$). When I used the non-redundant part (or 'lower triangle') of the score matrix, which makes the intelligibility scores symmetrical after averaging the contradiagonal mutual intelligibility'), the correlation increases to $r = .928$ (N = 105, $p < .001$).

In order to find out whether opinion testing (faster and more economical) can be used as a substitute of functional testing, I did a series of computations to see how well the functional results can be predicted from the opinion results. The analysis reveals that (1) Judged intelligibility is a better predictor of the functional test scores than judged similarity is, (2) functional intelligibility at the sentence level can be somewhat better predicted from opinion scores than word intelligibility. The best correlation coefficients were obtained in the case of non-redundant symmetrical ('lower triangle') matrices, capturing mutual intelligibility: judged intelligibility (rather than judged similarity) is correlated with functional sentence-intelligibility (rather than with word-intelligibility) at $r = .818$ (N = 105, $p < .001$).

The following conclusions were drawn in this chapter:

(1) Mutual intelligibility between Sinitic dialects can adequately be established through functional testing.
(2) Mutual intelligibility within Mandarin dialects is intrinsically higher than that within non-Mandarin (Southern) dialects, both at word and sentence level. Non-Mandarin (Southern) listeners understand Mandarin dialects consistently better than non-Mandarin (Southern) dialects.
(3) The primary split between Mandarin and non-Mandarin is correctly reflected, but no perfect reflection is found for the sub-clusters of the dialectal families.
(4) All subjective measures significantly correlate with one another. The r-value is consistently higher between the subjective measures of the same type (judged intelligibility versus judged similarity, functional word intelligibility versus functional sentence intelligibility) than across types (using opinion measures to predict functional measures and *vice versa*).

(5) All the results correspond with traditional dialect taxonomy to some extent. Functional intelligibility measures reflect Chinese dialect classifications better than opinion scores. Functional sentence-intelligibility test results conform best with traditional Chinese dialect taxonomy.

Chapter Five collects a number of objective distance measures computed on Sinitic dialects. These distance measures are available from the published literature or were computed by myself. One distance measure is based on lexical similarity between dialects, all other measures relate to aspects of the sound structures of the dialects. All the measures were used to generate affinity trees, which were then (crudely) compared with traditional dialect taxonomies. The aim was to select those measures which produced at least a correct primary split between Mandarin and non-Mandarin (Southern) dialects as viable predictors of mutual intelligibility to be used in Chapter Six, where I will decide which measure (or ensemble of measures) is the best predictor of mutual intelligibility between Sinitic dialects.

Overall, the objective structural similarity measures fall into two categories: lexical affinity and phonological affinity. Lexical affinity captures how much two dialects share the same words for the same concepts (cognates, words having the common etymological origin), i.e. the proportion of cognates shared between the vocabularies of two dialects (or languages). Phonological affinity is defined on the lexical subset of cognates shared between two dialects. It expresses how much the sound shapes (segmental make-up and tonal make-up) of the cognates resemble each other.

I copied the lexical affinity measure for 13 dialects (a proper subset of my 15 target dialects) from the literature (work by Cheng 1982, 1986, 1991, 1997). I generated a lexical affinity tree using the same average linking method as used in the literature. The tree produces two classification errors: the integral cluster containing the six Mandarin dialects incorrectly also contains two non-Mandarin (Southern) dialects, i.e. Nanchang and Changsha.

Phonological distance measures were derived in a number of ways, either copied from the literature or computed by myself, either on phonological inventories published in the literature or on a digital database that was made available to me (see below).

(1) One set of similarity measures was computed on the phonological inventories of the 15 dialects, separately for initials (which onset consonants are shared by two dialects?), nuclei (which vowels are shared by two dialects?), codas (which syllable-final consonants have the same occurrence in two dialects?), finals (which syllable rhymes are shared by two dialects), tones (which tones are shared by two dialects?), as well as a number of composite measures, such as the shared occurrence of both onsets and coda consonants. From all these measures, only the tree based on shared initial consonants (onsets) reflects a convincing split between Mandarin and non-Mandarin, with just one classification error: Changsha was incorrectly grouped with the Mandarin dialects.
(2) A second set of measures was based on the lexical frequencies of the same phenomena as in (1). The frequencies were computed on a lexical database of 764 words compiled by researchers at the Department of Linguistics of the Chinese

Academy of Social Sciences in Beijing (CASS). Here the affinity tree based on the number of shared vocalic nuclei (weighed for lexical frequency) reflects the traditional dialect taxonomy best.

(3) Levenshtein string-edit distances based on the CASS database of 764 common morphemes in each of our 15 dialects were computed by using the LO4 software package developed at Groningen University, once with and once without applying some perceptual weighing of sound differences. As before, I did the computations separately for the segmental and tonal properties of the morphemes. The results show that the Levenshtein distance measures yield disappointing results. Again using traditional dialect taxonomy as a validation criterion, the least number of classification errors is 2.5 (unweighed segmental differences). Correspondence in the tonal domain seems to be an especially poor criterion for determining affinity between Chinese dialects.

(4) The last type of distance measures was copied from Cheng's publications. His computations of phonological differences were done based the *Zihui* database *[Word list of Chinese dialects],* which provides transcriptions of over 2,700 words across 17 Chinese dialects, which include all of my 15 dialects. Five measures of phonological affinity between all pairs of dialects were computed: the initials frequency, the finals frequency, tones frequency), frequency of onsets and rhymes combined, and the frequencies of onsets, rhymes and tones combined. A sixth, and much more complex, measure aimed to capture the degree of phonological similarity between the cognate parts of the lexicons of the target dialects. This measure, which I call the Phonological Correspondence Index (PCI), basically expresses how many formal rules are needed to convert phonetic transcriptions of a cognate in one dialect to its counterpart in another dialect. The measure is asymmetrical as the rule set needed to convert strings from dialect A to B may be more (or less) complex than the set needed to convert the same strings from dialect B to those of A.

The best classification of the 15 dialects in my sample is afforded by a compound measure proposed by Cheng (1997), i.e. the affinity based on the lexical frequencies of onsets, finals and tones combined as counted in the 2,770-item list of common words in Chinese. This measure yields a perfect split of the dialects into Mandarin and non-Mandarin (Southern), with the exception of one dialect, Taiyuan, whose status remained ambiguous, yet it fails to reflect any of the internal taxonomy within the two main branches of Sinitic dialects.

In **Chapter Six**, I used the objective distance measures collected in Chapter Five as predictors of mutual intelligibility. I correlated the affinity measures between each pair of dialects with the mutual intelligibility scores obtained in chapters three and four. The following observations could be made from the results of the regression analysis: (1) Similarity judgments are somewhat more difficult to predict from linguistic distance measures than intelligibility scores (whether judgments or functionally determined). R^2 values were below .8 for similarity judgments but above .8 for all other criterion variables. (2) Functional intelligibility scores (both at word level and at sentence level) are less well predicted than judged intelligibility. Therefore, we cannot decide whether judged or functionally determined intelligibility is more amenable to prediction from objective measures. However, we conclude that functionally determined intelligibility is the preferred method of intelligibility testing as functional measures allow tree structures to be generated that reflect the traditional taxonomy of Chinese dialects.

Chapter Seven reviews the main findings and tries to answer the research questions asked.

(1) My experiments show that judged intelligibility is significantly correlated with judged linguistic similarity between dialects. In the Chinese dialect situation, tonal information plays a minor role in mutual intelligibility, as is indicated by the essentially similar results obtained for speech samples with and without pitch information.

(2) There always exists some mutual intelligibility between Chinese dialects. The degree of mutual intelligibility can be established through experiments (both by judgment/opinion tests and functional word-intelligibility and sentence-intelligibility tests). Mandarin dialects are more intelligible to non-Mandarin (Southern) dialects listeners than *vice versa*. Dialects within the Mandarin branch are more mutually intelligible than are dialects within the non-Mandarin (Southern) branch. Functional tests confirm impressionistic claims in the literature more clearly than judgment/opinion tests. The claimed asymmetry is most clearly seen in the results of the functional sentence-intelligibility test.

(3) All of the experimental results (from the judgment/opinion tests and functional tests) correlate with each other to a large extent, judged intelligibility versus similarity (r = .888, N = 105), word intelligibility versus sentence intelligibility (r = .928.), functional word versus judgment intelligibility (r = .772), functional sentence versus judgment intelligibility (r = .818).

(4) The mutual intelligibility of Chinese dialects (both opinion scores and functional test scores) can be predicted from the structural measures to some extent. Similarity judgments are somewhat more difficult to predict from linguistic distance measures than are intelligibility scores (whether judgments or functionally determined). Functional intelligibility scores, whether determined for single words or for (short) sentences, cannot be predicted better than judged intelligibility.

(5) The debated status of Jin dialects (represented by Taiyuan) can be settled on the basis of my experiments. All of the experimental trees indicate that Taiyuan is a Mandarin dialect. Therefore, we argue that there is no straightforward reason to branch off a Jin group from the Mandarin branch.

(6) The dendrograms generated from the computed affinity scores obtained objectively and subjectively were validated against the traditional Chinese dialect taxonomy based on the primary Mandarin versus non-Mandarin split and their internal cluster structures to some extent. However, to a large extent, the dendrograms obtained from the collected mutual intelligibility test scores correspond with the primary split of the Mandarin and non-Mandarin (Southern) dialect branches suggested by Chinese linguists; some even reflect part of structure in terms of internal sub-groups or clusters. The claim that mutual intelligibility testing can be used in dialect classification and to validate dialect taxonomies is – at least partly – confirmed. The final conclusion is functional intelligibility results (especially the sentence-intelligibility result) reflects best the traditional Chinese dialect taxonomy proposed by linguists.

(7) Further research based on a larger selection of dialect samples and experimental testing also at higher, e.g. syntactic, linguistic levels, is needed in future work.

摘要

本论文通过功能测试法和意见调查法两种实验方法来测定 15 种汉语方言之间的互懂度，并通过计算两种实验结果的相关系数来检测两种实验方法在多大程度上测定的是相同的属性。两种实验结果分别和收集到的（文献上已经发表的或者是我自己计算的）客观语言距离测量值再进行相关性研究，来检测语言之间的互懂度在多大程度上能够通过语言之间的客观距离预测出来。实验得出的数据和计算出的客观语言距离用来生成层级（树形）结构图，表明语言之间的亲疏关系。比较而言，在认定语言的亲疏关系时，互懂度测试法被认为更值得提倡因为其树形结构图与传统语言学家认定的语言系谱图更吻合。

第一章是导论部分。这一部分解释了为什么要测量语言配对之间的距离和怎样测量语言配对之间的距离，语言之间的距离是一个区分方言的重要的潜在参数。如果距离小，也即两种语言非常相似，那么它们就极有可能是一种语言的方言，否则，它们就是两种不同的语言。语言之间的距离可以通过客观法和主观法测定。客观法就是计算语言配对之间的结构差异，这种测量是多维测量。主观法就是确立语言配对之间的互懂度，主观法从概念上说更简单，因为综合的互懂度可以在单维层面上表示。

互懂度可以通过功能实验法和意见调查法来测定。功能法测量在多大程度上听音人能实际辨认出说话人语言的语言单位，并理解这些语言单位所表达的信息。意见调查法测试听音人（语言 A 的本族语者）认为他们能听懂多少说话人的语言（语言 B）。在确立语言或方言之间的互懂度时，语言配对之间的比较数量是随着语言数量呈指数增长的，如果要比较 10 对语言，那么就有 100 个说话人和听音人之间的语言比较对；如果有 15 种语言，那么就有 225 个语言比较对。另外，要测出听音人对 10 种或 15 种不同的语言听懂了多少，他/她就不能对同一段材料听两遍，即避免重复效应或学习效应。重复效应或学习效应也可以通过征募大量的听音人来克服，这样就可以把听音人完全对等地分成许多小组，每一个小组能保证听不相同的材料，或者通过设计大量完全对称的测试材料。为了避免这样极大的劳动量，人们就用意见调查法作为功能实验法的捷径，尤其是在语言配对的数量相当大的情况下。本论文将找出在多大程度上意见调查法能作为功能实验法的捷径。

进一步说，我们还用两种不同的标准来检验功能实验法和意见调查法这两种互懂度测试法的有效性。第一种标准检验互懂度测试的结果能在何种程度上由语言配对之间的客观结构差异预测出来，第二种有效性标准是检验由功能实验法或意见调查法结果生成的语言亲属关系图和由语言学家或方言学家画出的语言亲缘关系图或系谱图的叠合度。

总体说来，本论文促成确定汉语方言之间的互懂度，并通过功能实验法和意见调查法两种方法对以下问题予以确认：（1）官话方言内部之间比南方方言内部之间更互懂，（2）南方方言对官话方言的互懂度要小于官话方言对南方方言的互懂度，（3）争议中的晋方言（以太原话为代表）划为官话方言比划为南方方言更合理。（4）利用方言数据库计算出更多的汉语方言之间的客观语言结构距离，并通过多元回归技巧，找出相对更好的预测变量。（5）功能实验法在句子层面上测试的互懂度结果与传统的汉语方言系谱图能更有效地吻合。

第二章提供了中国语言的背景知识，中国的语言环境具有多样性。大量的方言（往往也是区别性很大的语言）在中国使用着，它们的分类也因为使用的具体标准不同而各异。本论文针对属于汉语语系的 15 种方言，使用两种语言系谱图作为参照点。这两种系谱图之间绝大部分一致，只有一个小的分歧：（1）两种系谱图都是建立在语言音系特征的历史变化之上的，即语音的变化和音调的演变，比如首辅音的简化，喉塞音的消失，首辅音的清浊对比，四种原始调类的阴（高音域）阳（低音域）分化，以及随后音调的进一步合并和分离。（2）两种系谱图一致同意汉语方言存在官话方言和南方方言两大基本分类。官话方言失去了词尾促音，只保留了较少的音调数目；南方方言保留了促音，并保留了丰富的音调库目。但二者在官话方言和南方方言的各自的内部分支上未能达成一致。（3）官话方言内部之间被认为比南方方言内部之间更互懂，因为排除标准官话（以北京方言为基础）的影响，官话方言之间有更大的内在一致性。（4）相对于南方方言对官话方言而言，官话方言对南方方言更可懂。（5）官话方言或南方方言各内部之间的互懂度和官话方言与南方方言之间的互懂度还从未确立，也没有进行过有效性验证。（6）晋方言的划分问题一直处于争论之中，但有可能通过互懂度测试的实验数据得以解决。

本研究包括以下的汉语方言：六种官话方言，其中北京，济南，西安，太原属于北方官话，成都和汉口属于西南官话；九种南方方言，其中苏州和温州属于吴方言，南昌属于赣方言，梅县属于客家方言，厦门、福州、潮州属于闽方言，长沙属于湘方言，广州属于粤方言。这些方言的地理位置可以从本书的封面上看到，官话方言中太原话的划分是临时的。

第三章描述了用意见调查法测试互懂度的实验。我用的是现成的、由 15 种方言的本族发音人阅读的寓言故事《北风和太阳》的有声材料。每种方言的材料都用 Praat（语音处理）软件中的 PSOLA（语音合成法）音调处理功能进行处理，使得所有的声音都是男性声音，说话的语速相同（有相同的停顿），有相同的平均音高，每种方言制成两个版本，一种有音调信息（韵律版），另一种没有音调信息（无韵律版），用来评估音调信息对互懂度的影响。每种方言征募 15 个单方言（只会一种方言）的听音人，他们需要在互懂度和相似性两个参数上对说话人从 0 到 10 的 11 个点值范围内进行估值判断。在估值尺度上，“0”表示“完全不互懂”和“没有任何相似性”，“10”表示“绝对互懂”和“完全相似”。统计起来，共有 15 组听音人方言（每个方言一组）和 15 组

说话人方言，即 225 种说话人-听音人方言对组合需要测试。每个听音组有 24 个听音人，每种方言说话人有两个版本（韵律版和无韵律版），这样就产生 21，600 个判断分值，用来生成聚合树形图。

结果显示，与传统的语言系谱图比较，由实验结果产生的四个树形图，没有一个树形图完美地生成了所研究方言的关系结构。首先，官话方言和南方方言的基本分界没有理想地反映出来，两大方言分支的内部结构不稳定，两种传统的南方方言被诠释成官话方言；其次，没有一个方言分支的次方言分支（方言群或方言丛），如官话分支的西南官话，南方方言的吴、闽方言次分支等清楚地被反映出来，具体结果如下：（1）官话-南方方言的基本分界在调整标准后可以从树形图上正确地反映出来，（2）太原话（晋方言代表，其分类一直悬而未决）总是被组合到官话方言中，（3）判断相似性和判断互懂度的相关系数在任何情况下（分别从韵律版和无韵律版两种版本中得出的全矩阵，三角矩阵，排除对角线矩阵中）都合理地高度相关（相关系数在韵律版时分别从 r=.810 到 r=.888, 在无韵律版时从 r=.841 到 r=.900），无韵律版总是产生更好的相关性，（4）更具体地说，如果以三组不同的听音人为对象，用平均判断互懂度和平均判断相似性互为坐标，而得出的散点图显示：a.当听音人的方言和说话人的方言一样时，即“听自己的方言”，判断互懂度和判断相似性都非常高，接近可能的最大值 10。b.当说话人方言和听音人的方言同属一个分支的时候（相同语支），就能看出官话方言和南方方言之间有明显差异。南方方言说话人被其他南方方言听音人判断为不可懂，（在 10 分的值上，平均互懂值为 1.5），而官话方言的说话人却从其他官话方言的听音人那里得到接近 7 分的平均互懂度值，同样的效果也能在相似性的判断值上看到。c. 当听音人听跨语支的方言的时候（其他语支），官话方言听音人听不懂南方方言说话人，就象南方方言听音人听其他南方方言说话人一样听不懂。同样地，南方方言被其他南方方言听音人和官话方言听音人判断为和自己的方言不一样，而官话方言说话人却被南方方言听音人判断为在一定程度上可懂（平均可懂值为 3.7），但相反的情形却不存在（即南方方言说话人不被北方方言听音人判断为可懂，其可懂值只有 1.7）。就判断相似性而言，没有上述基于听音人的差别：南方方言的听音人认为官话方言说话人的方言和他们自己的方言不一样（平均相似值为 1.4），就如同官话方言听音人判断南方方言说话人的方言不同于自己的方言（其平均相似值为 0.9）。

为了检测声音质量的个体差异可能导致的影响，以及因为转变说话人的性别和其他方面的声音处理而可能导致的质量磨损，我进行了一个控制实验，收集了对声音质量的感知判断值，即包括说话人的发音质量和录音质量的综合特质判断值。我一共收集了 30 个材料的判断值，这些材料分别从互懂度测试材料的韵律版和无韵律版的听音材料中截取，并播放给莱顿大学语言中心的学生和同事听（这些听音人的母语为各自的欧洲语言，没有任何的实际汉语知识）。

结果显示，听音人的判断是可信的，他们判断韵律版的材料（其判断分值从 5 到 8）比无韵律材料（其判断分值为 3 到 4）有更好的声音质量。韵律版和无

韵律版高度相关，r=.884(N=15, p < .001)。但是，当韵律版和无韵律版的互懂度由“自己方言”听音人判断时，声音质量和互懂度值（由主实验确立的）之间的相关性却很差，并且无显著性，r = .205 (N = 30, ins.)。分开计算韵律版和无韵律版，相关性更差，韵律版为 r = .189 (N = 15, ins.)，无韵律版为 r = .098 (N = 15, ins.)。声音质量判断值和基于所有听音组的综合互懂度判断值之间的相关系数几乎为零（r = .019 ，N = 15, ins. 无韵律版）， 甚至为负数（r = −.195，N = 15, ins. 韵律版），r = −.204 ，N = 15, ins.， 基于两个版本）。我们只用北京方言听音人的实验结果来计算声音质量和判断互懂度之间的相关性，结果显示声音质量和判断互懂度（韵律版和无韵律版）之间没有相关性。

本章的结论如下：
（1）汉语方言配对之间总是存在一定程度的判断互懂度值。
（2）官话方言和南方方言之间互懂度的非对称性通过实验和统计得以证实。
（3）官话方言和南方方言之间的基本分界在实验结果中得以显示，而各自的内部次分支却没能准确地和传统的方言系谱图相符合。
（4）实验结果表明，有语音信息和无语音信息的判断结果之间没有显著的差异性。很显然，音调只对互懂度产生很小的影响。稍微偏高的无韵律版结果可以解释为在互懂度测试中，当音调信息被去掉后，听音人更集中在对剩下的语言特征的判断上。
（5）判断相似性和判断互懂度之间的高相关性显明这两种尺度实际测量的是同一种属性。
（6）判断互懂度和判断声音质量之间无相似性确保了在我们的实验中，声音质量没有受到人为因素的影响。方言听音人对互懂度和相似性的判断不是建立在判断录音质量上的。

第四章通过功能实验法测量汉语方言之间的互懂度。我在词汇和句子两个层面上收集了功能实验的互懂度值。我将两种结果互相比较，也和前面的意见调查法的判断结果进行比较，并同样将层级结构图和传统的方言系谱图进行有效性验证。

我选择了 150 个基本词汇作为词汇互懂度的测试材料，这些词汇按照语义分为 10 类（比如日常生活中常用的身体部位，家庭成员，食物等），每一类 15 个单词。我选择 60 个用于噪音背景下的语音感知实验句子（SPIN for Speech Perception in Noise Test）作为句子互懂度 实验材料，我将这些英语句子翻译成汉语普通话，我只选择翻译之后能从上下文预测出句子最后一个单词，并且也符合现代汉语的表达习惯的句子。每一种方言征募了两名（一男一女）发音人，他们把 150 个单词和 60 个句子翻译成他们自己的方言（翻译要一致），然后录制他们对本方言词汇和句子的发音。每一种方言制成 15 张 CD 碟，每一张碟上分别录有 15 种方言的单词和句子，且通过拉丁方阵设计，确保各张碟上的方言顺序不一样，每个方言选用 15 个听音人（每个听音人听一张不同的 CD）。在词汇测试中，听音人要求将所听到的单词划入最适合的语义类别

中（必须在 10 个分类选择项中选择一个）；在句子测试中，听音人的任务是把每句的最后一个单词翻译成他们自己的方言。

我总共收集了 33，750 个词汇结果（15 × 150 × 15）和 13，500 个句子结果（15 × 60 × 15）。我首先计算了词汇结果，统计分析结果如下：（1）再一次表明汉语方言配对之间总是存在一定程度的互懂度。（2）听音人对自己方言辨认的百分比最高。（3）官话方言说话人总能被南方方言和其它官话方言的听音人更好地听懂（平均值总是更高）。（4）两大方言的基本分界能正确地从词汇互懂度矩阵生成的树形图上反映出来。太原话仍被划分为官话方言。句子互懂度的结果生成同样的结构图，只是平均值普遍高于词汇互懂度。

官话方言和南方方言之间互懂度的非对称性再一次被功能实验结果证实。在听音人听相同方言分支的时候（即官话方言听音人听其他官话方言，男方方言听音人听其它南方方言），六种官话方言之间的平均词汇互懂度为 72%，而南方方言之间，听音人对词汇的平均正确划分率为 52%，官话方言听音人能很好地听懂其他官话方言说话人（可懂度为 61%），而南方方言听音人却不能很好地听懂其他南方方言说话人（可懂度只有 22%）。跨方言的互懂度，即官话-南方方言和南方-官话方言之间的互懂度都不高（分别为 32%和 36%）。

同样的结果也在句子互懂度测试中发现。官话方言之间的句子互懂度很好，而南方方言之间的句子互懂度很差，男方方言对官话听音人来说，就如同南方听音人听南方方言一样不可懂。官话方言对于南方方言听音人存在较好的可懂度（54%），即使我们去掉北京方言说话人，仍然有 48%的可懂度。

单词和句子互懂度之间高度相关，相关系数为 $r = .835$ ($N = 225, p < .001$)。当我去掉矩阵中的冗余信息，用双边可懂度的对称平均值（互懂度值），即“下三角”矩阵来计算时，相关性更高，$r = .928$ ($N = 105, p < .001$)。

为了找出意见调查法（更快也更经济）是否能用作功能实验法的替代方法，我进行了一系列运算，以便找出在什么程度上功能法的结果能从意见调查法结果上预测出来。分析表明：（1）比较相似判断值而言，互懂度判断值是功能测试法更好的预测变量。（2）比较词汇互懂度，功能测试的句子互懂度能更好地从意见调查法中得到预测。最好的相关系数是涉及互懂度的，在无冗余的对称矩阵中得到的：判断互懂度（而不是判断相似值）与功能测试的句子互懂度（而不是词汇互懂度）很好地相关，$r = .818$ ($N = 105, p < .001$)。

本章结论如下：
（1）汉语方言之间的互懂度可以通过功能测试法充分建立起来。
（2）官话方言之间比南方方言之间天然存在更高的词汇和句子互懂度，南方方言听音人总是能够更好地听懂官话方言而不是南方方言。
（3）官话方言和南方方言之间的基本分界在功能实验结果中仍然得以显示，而各自的内部次分支却没能得到准确反映。

（4）各种主观测试的结果都显著地互为相关，在同类主观测试中（判断可懂度和判断相似性之间，功能词汇互懂度和功能句子互懂度之间），两两参数的相关 r 值比跨类测试（用意见调查法结果预测功能实验法结果，或者反过来）更高。
（5）所有的测试结果都在一定程度上和传统的方言系谱图相符合。功能实验法的结果比意见调查法的结果更好地反映了中国方言的分类。功能句子互懂度测试结果最好地符合传统的方言系谱图。

第五章收集了很多汉语方言之间的客观距离值。这些客观距离值或者是从出版文献中获得，或者是通过自己计算获得，其中一种是方言之间的词汇相似值，其余的测量值都是关于方言之间的语音结构方面的。所有的测量值都用来生成亲缘关系树形图，并大致和传统的方言系谱图相比较。目标是筛选出至少能产生官话方言和南方方言分界线的测量结果，用于第六章对互懂度预测的可变量，在第六章，我将决定哪种测量或者整体测量是汉语方言之间互懂度的最好预测变量。

总体来说，对客观结构相似性的测量分为两个范畴：词汇相似性和音系相似性。词汇相似性是指两种方言共享多少同源词，即两种方言的词汇表中，其同源词的比例是多少。音系相似性是指两种方言所共享的同源词的相似程度，即同源词的语音组成和音调组成的相似性。

我从文献中复制了 13 种方言（全部包含于我的 15 种方言之中）的词汇相似值（郑 1982，1986，1991，1997）。我用和文献中一样的平均列联法，生成了词汇亲缘树形图。树形图显示两个分类错误：聚合六种官话方言的整体方言群也错误地包含了两种南方方言，南昌和长沙。

语音距离测量值是通过好几种方法获得的，或者从文献中复制，或者我自己计算，或者利用出版文献的语音表，或者从我找到的电子音库中获得（详情见下）。
（1）一组相似测量值是根据 15 种方言的语音表计算出来的。以音节为单位，从首辅音（哪些首辅音是两种方言共享的？），腹元音（两种方言共有的元音有哪些？），尾辅音（哪些音节末辅音共同出现在两种方言中？），韵音（哪些韵音是两种方言都有的？），调音（两种方言共有的调音有哪些？）等五部分分别计算，同时计算出某些复合值，比如，两种方言中共同出现的首辅音和尾辅音的组合。在所有这些计算结果中，唯有音节首辅音的树形图反映出令人信服的官话方言和南方方言的分界线，只有一个分类错误：长沙方言被错误地划给官话方言。
（2）第二组测量值是关于词汇频率的计算，和（1）一样，也是从五个方面进行计算。词汇频率计算基于中国社科院语言研究所编撰的方言数据库里的 764 个单词。由此得出的计算结果里面，韵腹树形图（乘以词汇频率的权重系数）最好地吻合了传统的方言系谱图。

（3）中科院方言数据库中关于 15 种方言的 764 个常用语素的 Levenshtein 算法距离是用荷兰格罗宁根大学开发的 LO4 软件包计算出来的。每个距离运算两次，一次运用语音感知差异的权重系数，一次不运用。同上面一样，我将每个语素的音段单位和音调单位分开计算，结果显示，Levenshtein 算法距离令人失望，用传统的方言系谱图作为有效性的标准进行检验，最少的分类错误是 2.5（未加权的音段单位距离），调域方面的对应在确立汉语方言的亲疏关系时，是一个特别不佳的指标。
（4）最后一类距离测量是复制于郑锦全先生的文献中。他对于语音差异的计算基于汉语方言字汇调查表，此表提供了 17 种方言的 2，700 个单词的音标，我研究的 15 种方言全部包含于此。在五个方面对所有方言对之间的音系关系作了计算：词首频率，音韵频率，音调频率，词首+加音韵的组合频率，以及词首+音韵+音调的组合频率。第六个测量计算更复杂一些，主要针对目标方言同源词之间语音相似的程度。我把这个计算称作语音相似指数，该指数表示把一个同源词中的语音单位转换为另一种方言中对应的语音单位需要多少规则形式。这个计算是非对称性的，因为从 A 方言的语素单位到 B 方言所需要的规则数量可能比从 B 方言到 A 方言的规则数量多或少。

15 种方言最好的分类图是从郑先生 1997 年文献中的复合计算中得到的，这个文献是关于汉语方言中 2，770 个常用词的词汇频率计算，由此得到的首辅音+音韵+音调的复合测量结果生成的亲缘关系结构图满意地反映出官话方言和南方方言的分界，除了一个例外：太原方言的归属仍然模棱两可，当然此图也没能反映汉语方言南北两系各自的内部分类。

在**第六章**里，我把在第五章中收集到的客观距离值用作互懂度的预测自变量。我把这些关系计算值和第三、四两章的互懂度值进行相关性计算，通过回归分析，得出以下结果：（1）判断相似值比互懂度（无论是判断互懂度还是功能互懂度）更难从客观语言距离测量中预测出来。判断相似性的决定系数值低于 0.8， 而其他变量的决定系数都高于 0.8。（2）功能互懂度（无论在单词层面还是句子层面）比判断互懂度预测的程度更差。因此，我们很难决定到底是判断互懂度还是功能互懂度更能服从客观距离值的预测。但不管怎样，我们可以得出结论功能法在测试互懂度方面是应该优先推荐的，因为功能测试法生成的树形结构图能够反映传统的汉语方言系谱图。

第七章回顾本研究的主要发现，并回答本论文提出的问题：
（1）我的实验结果表明，方言之间的判断互懂度能和判断语言相似性显著相关，对汉语方言而言，音调信息对互懂度只起次要作用，因为从韵律版和无韵律版测试材料获得的结果基本上很相近。
（2）汉语方言之间总是存在一定的互懂度。互懂度的大小可以通过实验测试建立起来（无论是意见调查法还是在词汇和句子层面上的功能测试法），官话方言对南方方言较之南方方言对官话方言更可懂，官话方言内部各方言之间较之南方方言内部各方言之间更互懂。功能测试法较之意见调查法更清楚地印证

了文献中对汉语各方言之间互懂度非对称性的论断，这些论断只是凭印象得出的。从句子层面的功能实验结果中，最能清楚地看出这种非对称性。

（3）所有的实验结果（意见调查法和功能实验法）都在很大程度上互为相关：判断互懂度与判断相似性相关(r = 0.888, N = 105)，词汇互懂度和句子互懂度相关(r = 0.928.)，词汇互懂度和判断互懂度相关(r = 0.772)，功能互懂度和判断互懂度相关(r = 0.818)。

（4）通过意见调查法和功能测试法确立的汉语方言互懂度在一定程度上能从语言的客观结构距离上预测出来。较之互懂度而言（无论是判断互懂度还是功能互懂度），从语言客观距离预测判断相似性要相对难一些。无论是简单的词汇之间的互懂度还是结构更复杂的句子之间的互懂度，都能从判断互懂度得到很好的预测。

（5）有争议的晋方言（以太原话为代表）的划分问题可以通过本实验研究解决。所有实验结果的树形图表明：太原话是官话方言。因此，我们主张没有直接了当的理由可以把太原话从官话方言中划分出去。

（6）尽管通过主观计算和客观测量结果生成的所有树形图都在一定程度上有效检测了传统汉语语言学家提倡的方言系谱图的基本分界和内部结构，而由实验结果生成的互懂度树形图却在很大程度上反映了官话方言和南方方言的基本分界，有的甚至正确反映了各个分支的内部方言组合。互懂度测试能够用来划分方言分类并用来优先检验传统方言系谱图的论断至少部分地得到证实。最后的结论是功能互懂度测试（功能句子互懂度测试）结果最好地反映了由语言学家提出的传统汉语方言系谱图。

（7）未来的后续研究需要选择更多的方言样本，在更高的语言结构层面上，比如在句法和词素音位的层次上进行更多的实验。

Appendices

Appendix 3.1 Listener information form

No.: ______

Name: _____

Gender: □ Male, □ Female

Age: _____

Date of birth: ___ (year) ___ (month) ____(day)

Nationality: ________

Occupation: ________

Education degree: ________

Standard Chinese Mandarin speaking: □Yes, □No

Standard Chinese Mandarin listening: □Yes, □Most □a bit □No

Places travelled to: ______________________________________

Language(s)/dialect(s): ______________________________________

Home address: ______________________________________

Postal address: ______________________________________

Telephone number: __________________________

Email: __________________________

The language environment at childhood:

Father's name: ___________

Language(s)/dialect(s): ___________

Mother's name: ___________

Language(s)/dialect(s): ___________

MEMO: ________________________

Signature:_________

Appendix 3.2 Proximity matrix generated from Table 3.1 (judged intelligibility based on monotonized speech samples). Horizontal and vertical double lines divide the table into nine non-Mandarin and six Mandarin dialects.

Speaker dialect (down)	Listener dialect (across)														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	0.00	13.62	21.17	18.59	19.63	20.38	19.37	10.97	10.45	10.53	15.21	9.90	12.89	10.00	11.79
Wenzhou	13.62	0.00	15.03	12.73	13.84	14.69	13.41	12.41	11.15	14.42	10.42	13.67	11.14	11.90	10.28
Guangzhou	21.17	15.03	0.00	10.64	12.93	13.26	9.74	18.07	17.97	20.69	13.39	21.49	16.60	20.37	16.38
Xiamen	18.59	12.73	10.64	0.00	10.85	8.11	10.45	15.32	15.17	17.80	10.52	18.55	13.84	17.21	13.34
Fuzhou	19.63	13.84	12.93	10.85	0.00	12.15	11.98	16.83	16.73	19.27	12.54	19.95	15.45	18.67	15.09
Chaozhou	20.38	14.69	13.26	8.11	12.15	0.00	12.30	17.39	17.48	19.81	13.11	20.67	16.06	18.91	15.47
Meixian	19.37	13.41	9.74	10.45	11.98	12.30	0.00	15.83	15.95	18.33	10.08	19.43	14.56	17.82	14.00
Nanchang	10.97	12.41	18.07	15.32	16.83	17.39	15.83	0.00	9.03	7.82	12.30	9.99	12.03	9.18	10.23
Changsha	10.45	11.15	17.97	15.17	16.73	17.48	15.95	9.03	0.00	8.42	12.26	8.83	9.19	7.22	9.32
Taiyuan	10.53	14.42	20.69	17.80	19.27	19.81	18.33	7.82	8.42	0.00	13.94	6.76	12.01	8.33	9.31
Beijing	15.21	10.42	13.39	10.52	12.54	13.11	10.08	12.30	12.26	13.94	0.00	14.48	10.13	13.71	8.64
Jinan	9.90	13.67	21.49	18.55	19.95	20.67	19.43	9.99	8.83	6.76	14.48	0.00	10.58	4.52	7.06
Hankou	12.89	11.14	16.60	13.84	15.45	16.06	14.56	12.03	9.19	12.01	10.13	10.58	0.00	8.99	7.07
Chengdu	10.00	11.90	20.37	17.21	18.67	18.91	17.82	9.18	7.22	8.33	13.71	4.52	8.99	0.00	7.34
Xi'an	11.79	10.28	16.38	13.34	15.09	15.47	14.00	10.23	9.32	9.31	8.64	7.06	7.07	7.34	0.00

Appendix 3.3 Proximity matrix generated from Table 3.2 (judged intelligibility based on intonated speech samples). Horizontal and vertical double lines divide the table into nine non-Mandarin and six Mandarin dialects.

Speaker dialect (down)	Listener dialect (across)														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	0.00	12.84	22.09	17.39	19.44	20.83	20.14	9.95	10.57	10.96	10.08	9.06	12.47	9.05	8.90
Wenzhou	12.84	0.00	17.85	13.65	15.27	16.83	15.40	9.50	10.46	14.29	7.36	12.46	11.33	11.92	12.08
Guangzhou	22.09	17.85	0.00	15.28	13.29	13.85	11.27	18.35	19.86	23.52	16.70	22.55	16.64	21.74	20.98
Xiamen	17.39	13.65	15.28	0.00	13.75	8.20	13.56	12.77	14.61	18.06	11.21	16.96	13.29	16.81	15.78
Fuzhou	19.44	15.27	13.29	13.75	0.00	12.64	12.87	16.07	17.45	20.80	14.17	19.83	14.74	19.06	18.30
Chaozhou	20.83	16.83	13.85	8.20	12.64	0.00	13.64	17.04	18.64	21.78	15.44	20.87	16.17	20.41	19.54
Meixian	20.14	15.40	11.27	13.56	12.87	13.64	0.00	15.40	17.53	21.03	13.89	20.09	14.74	19.48	18.92
Nanchang	9.95	9.50	18.35	12.77	16.07	17.04	15.40	0.00	4.36	8.76	3.77	5.94	7.11	6.19	6.22
Changsha	10.57	10.46	19.86	14.61	17.45	18.64	17.53	4.36	0.00	8.83	6.11	6.73	8.34	6.47	7.81
Taiyuan	10.96	14.29	23.52	18.06	20.80	21.78	21.03	8.76	8.83	0.00	10.18	4.28	12.62	5.12	5.16
Beijing	10.08	7.36	16.70	11.21	14.17	15.44	13.89	3.77	6.11	10.18	0.00	8.14	6.83	7.96	7.47
Jinan	9.06	12.46	22.55	16.96	19.83	20.87	20.09	5.94	6.73	4.28	8.14	0.00	10.80	3.18	3.47
Hankou	12.47	11.33	16.64	13.29	14.74	16.17	14.74	7.11	8.34	12.62	6.83	10.80	0.00	10.30	10.41
Chengdu	9.05	11.92	21.74	16.81	19.06	20.41	19.48	6.19	6.47	5.12	7.96	3.18	10.30	0.00	4.73
Xi'an	8.90	12.08	20.98	15.78	18.30	19.54	18.92	6.22	7.81	5.16	7.47	3.47	10.41	4.73	0.00

Appendix 3.4 Proximity matrix generated from Table 3.3 (judged similarity based on monotonized speech samples). Horizontal and vertical double lines divide the table into nine non-Mandarin and six Mandarin dialects.

Speaker dialect (down)	Listener dialect (across)														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	0.00	11.64	14.09	14.79	14.19	14.78	14.56	16.09	16.95	14.49	15.13	17.65	14.78	17.27	14.99
Wenzhou	11.64	0.00	13.71	13.58	13.68	14.53	13.79	14.11	15.68	13.25	13.10	16.42	14.29	15.89	14.11
Guangzhou	14.09	13.71	0.00	14.31	13.88	14.35	10.22	16.39	17.29	14.72	16.02	18.20	15.00	18.03	15.46
Xiamen	14.79	13.58	14.31	0.00	15.05	9.26	13.94	14.23	14.20	13.36	11.86	14.49	13.35	15.34	12.37
Fuzhou	14.19	13.68	13.88	15.05	0.00	14.34	14.37	17.38	18.20	15.16	16.62	18.83	15.31	18.65	16.10
Chaozhou	14.78	14.53	14.35	9.26	14.34	0.00	14.52	17.48	18.39	15.54	16.78	18.88	15.74	18.95	16.56
Meixian	14.56	13.79	10.22	13.94	14.37	14.52	0.00	15.88	16.73	14.42	15.03	17.39	14.71	17.44	14.99
Nanchang	16.09	14.11	16.39	14.23	17.38	17.48	15.88	0.00	9.56	13.68	9.51	10.03	12.65	11.08	9.69
Changsha	16.95	15.68	17.29	14.20	18.20	18.39	16.73	9.56	0.00	13.38	6.13	8.45	10.89	8.97	7.36
Taiyuan	14.49	13.25	14.72	13.36	15.16	15.54	14.42	13.68	13.38	0.00	11.26	12.18	12.60	12.26	9.72
Beijing	15.13	13.10	16.02	11.86	16.62	16.78	15.03	9.51	6.13	11.26	0.00	5.45	9.41	6.86	3.44
Jinan	17.65	16.42	18.20	14.49	18.83	18.88	17.39	10.03	8.45	12.18	5.45	0.00	11.08	6.88	4.04
Hankou	14.78	14.29	15.00	13.35	15.31	15.74	14.71	12.65	10.89	12.60	9.41	11.08	0.00	10.81	8.47
Chengdu	17.27	15.89	18.03	15.34	18.65	18.95	17.44	11.08	8.97	12.26	6.86	6.88	10.81	0.00	6.37
Xi'an	14.99	14.11	15.46	12.37	16.10	16.56	14.99	9.69	7.36	9.72	3.44	4.04	8.47	6.37	0.00

Appendix 3.5 Proximity matrix generated from Table 3.4 (judged similarity based on intonated speech samples). Horizontal and vertical double lines divide the table into nine non-Mandarin and six Mandarin dialects.

Speaker dialect (down)	Listener dialect (across)														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	0.00	11.64	14.09	14.79	14.19	14.78	14.56	16.09	16.95	14.49	15.13	17.65	14.78	17.27	14.99
Wenzhou	11.64	0.00	13.71	13.58	13.68	14.53	13.79	14.11	15.68	13.25	13.10	16.42	14.29	15.89	14.11
Guangzhou	14.09	13.71	0.00	14.31	13.88	14.35	10.22	16.39	17.29	14.72	16.02	18.20	15.00	18.03	15.46
Xiamen	14.79	13.58	14.31	0.00	15.05	9.26	13.94	14.23	14.20	13.36	11.86	14.49	13.35	15.34	12.37
Fuzhou	14.19	13.68	13.88	15.05	0.00	14.34	14.37	17.38	18.20	15.16	16.62	18.83	15.31	18.65	16.10
Chaozhou	14.78	14.53	14.35	9.26	14.34	0.00	14.52	17.48	18.39	15.54	16.78	18.88	15.74	18.95	16.56
Meixian	14.56	13.79	10.22	13.94	14.37	14.52	0.00	15.88	16.73	14.42	15.03	17.39	14.71	17.44	14.99
Nanchang	16.09	14.11	16.39	14.23	17.38	17.48	15.88	0.00	9.56	13.68	9.51	10.03	12.65	11.08	9.69
Changsha	16.95	15.68	17.29	14.20	18.20	18.39	16.73	9.56	0.00	13.38	6.13	8.45	10.89	8.97	7.36
Taiyuan	14.49	13.25	14.72	13.36	15.16	15.54	14.42	13.68	13.38	0.00	11.26	12.18	12.60	12.26	9.72
Beijing	15.13	13.10	16.02	11.86	16.62	16.78	15.03	9.51	6.13	11.26	0.00	5.45	9.41	6.86	3.44
Jinan	17.65	16.42	18.20	14.49	18.83	18.88	17.39	10.03	8.45	12.18	5.45	0.00	11.08	6.88	4.04
Hankou	14.78	14.29	15.00	13.35	15.31	15.74	14.71	12.65	10.89	12.60	9.41	11.08	0.00	10.81	8.47
Chengdu	17.27	15.89	18.03	15.34	18.65	18.95	17.44	11.08	8.97	12.26	6.86	6.88	10.81	0.00	6.37
Xi'an	14.99	14.11	15.46	12.37	16.10	16.56	14.99	9.69	7.36	9.72	3.44	4.04	8.47	6.37	0.00

Appendix 4.1. Stimulus words used for semantic classification task (10 categories, 15 instantiations per category).

#	English	Mandarin		#	English	Mandarin	
		Char.	Pinyin			Char.	Pinyin
(1) Body parts				**(6) Textiles, articles of clothing, apparel**			
1.	head	头	tou2	76.	blanket	被子	bei4zi
2.	face	脸	lian3	77.	sheet	床单	chuang2dan1
3.	eye	眼	yan3	78.	pillow	枕头	zhen3tou
4.	ear	耳	er3	79.	mosquito net	蚊帐	wen2zhang4
5.	nose	鼻	bi2	80.	thread	线	xian4
6.	mouth	嘴	zui3	81.	yarn	纱	sha1
7.	hand	手	shou3	82.	silk	丝	si1
8.	foot	脚	jiao3	83.	cloth	布	bu4
9.	neck	颈	jing3	84.	skirt	裙子	qun2zi
10.	hair	头发	tou2fa	85.	scarf	围巾	wei2jin1
11.	eyebrow	眉毛	mei2mao	86.	shirt	衬衣	chen4yi1
12.	tongue	舌	she2	87.	shoe	鞋	xie2
13.	tooth	牙	ya2	88.	sock	袜	wa4
14.	shoulder	肩	jian1	89.	earring	耳环	er3 huan2
15.	back	背	bei4	90.	sweater	毛衣	mao3yi1
(2) Plants: Sweet fruits & nuts				**(7) Orientation in time and space**			
16.	apple	苹果	ping2guo3	91.	above	上	shang4
17.	pear	梨	li2	92.	below	下	xia4
18.	banana	香蕉	xiang1jiao1	93.	left	左	zuo3
19.	lichee	荔枝	li4zhi1	94.	right	右	you4
20.	mango	芒果	mang2guo3	95.	front	前	qian2
21.	grape	葡萄	pu2tao	96.	back	后	hou4
22.	watermelon	西瓜	xi1gua	97.	east	东	dong1
23.	peach	桃子	tao2zi	98.	west	南	nan2
24.	apricot	杏	xing4	99.	south	西	xi1
25.	pineapple	菠萝	bo1luo2	100.	north	北	bei3
26.	cherry	樱桃	ying1tao	101.	middle	中	zhong1
27.	strawberry	草莓	cao3mei2	102.	inside	里	li3
28.	date	枣	zao3	103.	outside	外	wai 4
29.	pomegranate	石榴	shi2liu	104.	tomorrow	明天	ming2tian1
30.	walnut	核桃	he2tao	105.	yesterday	昨天	zuo2tian1

Appendix 4.1 continued

#	English	Mandarin		#	English	Mandarin	
		Char.	Pinyin			Char.	Pinyin
(3) Plants: Vegetables				**(8) Natural phenomena**			
31.	celery	芹菜	qin2cai4	106.	sun	太阳	tai4yang
32.	leek	韭菜	jiu3cai4	107.	moon	月亮	yue4liang
33.	eggplant	茄子	qie2zi	108.	star	星星	xing4xing
34.	pumpkin	南瓜	nan2gua	109.	rain	雨	yu3
35.	winter melon	冬瓜	dong1gua	110.	wind	风	feng1
36.	tomato	西红柿	xi1hong2shi4	111.	ice	冰	bing1
37.	potato	土豆	tu3dou4	112.	frost	霜	shuang1
38.	corn	玉米	yu4mi3	113.	snow	雪	xue3
39.	lotus root	莲藕	lian2ou3	114.	fog	雾	wu4
40.	spinach	菠菜	bo1cai4	115.	hail	冰雹	bing1bao4
41.	carrot	胡箩卜	hu2luo2bo	116.	cloud	云	yun2
42.	cucumber	黄瓜	huang2gua	117.	thunder	雷	lei2
43.	pea	豌豆	wan1dou4	118.	lightning	闪电	shan3dian4
44.	string bean	豇豆	jiang1dou4	119.	rainbow	彩虹	cai3hong2
45.	mushroom	磨菇	mo3gu	120.	flood	洪水	hong2shui3
(4) Animals: Four-legged mammals				**(9) Perishables (food/drinks other than fruits and vegetables**			
46.	dog	狗	gou3	121.	beancurd	豆腐	dou4fu
47.	cat	猫	mao1	122.	milk	牛奶	niu2nai3
48.	pig	猪	zhu1	123.	noodle	面条	mian4tiao2
49.	ox	牛	niu2	124.	meat	肉	rou4
50.	goat	羊	yang2	125.	rice	米饭	mi3fan4
51.	tiger	老虎	lao2hu3	126.	soup	汤	Tang1
52.	lion	狮子	shi1zi	127.	wine	酒	jiu3
53.	elephant	大象	da4xiang4	128.	oil	油	you2
54.	horse	马	ma3	129.	salt	盐	yan2
55.	leopard	豹	bao4	130.	soy sauce	酱油	jiang4you2
56.	giraffe	长颈鹿	chang2jing3lu4	131.	vinegar	醋	cu4
57.	bear	熊	xiong2	132.	pepper	胡椒	hu2jiao1
58.	zebra	斑马	ban1ma3	133.	egg	蛋	dan4
59.	wolf	狼	lang2	134.	sausage	香肠	xiang1chang2
60.	fox	狐狸	hu2li	135.	tea	茶	cha2

Appendix 4.1 continued

#	English	Mandarin		#	English	Mandarin	
		Char.	**Pinyin**			**Char.**	**Pinyin**
(5) Animals: other				**(10) Verbs of action/ things people do**			
61.	cock	公鸡	gong1ji1	136.	shake hands	握手	wo4shou3
62.	hen	母鸡	mu2ji1	137.	nod	点头	dian3tou2
63.	duck	鸭	ya1	138.	shake head	摇头	yao2tou2
64.	snake	蛇	she2	139.	laugh	笑	xiao4
65.	swallow	燕子	yan4zi	140.	cry	哭	ku1
66.	magpie	喜鹊	xi2que4	141.	walk	走	zou3
67.	crab	螃蟹	pang2xie4	142.	run	跑	pao3
68.	goose	鹅	e2	143.	jump	跳	tiao4
69.	sparrow	麻雀	ma2que4	144.	stand	站	zhan4
70.	bee	蜜蜂	mi4feng1	145.	sit	坐	zuo4
71.	spider	蜘蛛	zhi1zhu1	146.	sleep	睡	shui4
72.	silk worm	蚕	can2	147.	open	开	kai1
73.	ant	蚂蚁	ma2yi3	148.	close	关	guan1
74.	butterfly	蝴碟	hu2die2	149.	read	读	du2
75.	dragonfly	蜻蜓	qing1ting2	150.	write	写	xie3

Note: digits in Pinyin transcription refer to lexical tones. Tone 1 is the high level tone, Tone 2 is a mid-rising tone, Tone 3 is the low dipping tone and Tone 4 is high falling.

Appendix 4.2 Mandarin SPIN sentences with Chinese characters, Pinyin transliteration and English originals.

#	Mandarin (Characters)	Pinyin	English
1.	他捕鱼用网	ta1 bu4 yu2 yong4 wang3	He caught the fish in his net.
2.	关上窗户,挡住风	guang1 shang4 chuang1 hu4, dang3 zhu4 feng1	Close the window to stop the draft.
3.	我的电视是十二英寸的屏幕	wo3 de dian4 shi4 shi4 shi2 er4 ying1 cun4 de ping2 mu4	My T.V. has a twelve-inch screen.
4.	农民收割庄稼	nong2 min2 shou1 ge2 zhuang1 jia.	The farmer harvested his crop.
5.	舰长指挥舰队	jian4 zhang2 zhi3 hui1 jian4 dui4	The Admiral commands the fleet.
6.	喝啤酒的人举起了酒杯	he1 pi2 jiu3 de ren2 ju3 qi2 le jiu3 bei1	The beer drinkers raised their mugs.
7.	白蚁看起来象蚂蚁	bai2 yi3 kan4 qi3 lai2 xiang4 ma2 yi3	A termite looks like an ant.
8.	他膝盖上的伤口结了一个疤	ta1 xi2 gai4 shang4 de shang1 kou3 jie1 le yi2 ge4 ba1	The cut on his knee formed a scab.
9.	农民堆码干草	nong2 min2 dui1 ma3 gan1 cao3	The farmer baled the hay.
10.	为了你的生日,我做了蛋糕	wei4 le ni3 de sheng1 ri4 wo3 zuo4 le dan4 gao1	For your birthday I baked a cake.
11.	火车脱离了轨道	huo3 che1 tuo1 li2 le gui3 dao4	The railroad train ran off the track.
12.	那只孤独的鸟在找它的同伴	na4 zhi1 gu1 du2 de niao3 zai4 zhao3 ta1 de tong2 ban4	The lonely bird searched for its mate.
13.	他们喝完了一整瓶酒	ta1 men he1 wan3 le yi4 zheng3 ping2 jiu3	They drank a whole bottle of gin.
14.	我们在沙滩上玩沙	wo3 men zai4 sha1 tan1 shang4 wan2 sha1	On the beach we play in the sand.
15.	我们迷路了,所以要看地图	wo3 men mi2 lu4 le,suo3 yi3 yao4 kan4 di4 tu2	We're lost so let's look at the map.
16.	飞机丢下一颗炸弹	fei1 ji1 diu1 xia4 yi4 ke1 zha4 dan4	The airplane dropped a bomb.
17.	把香肠切成条	ba3 xiang1 chang2 qie1 cheng2 tiao2	Cut the bacon into strips.
18.	这把钥匙不配这把锁	zhe4 ba3 yao4 chi2 bu2 pei4 zhe4 ba3 suo3	This key won't fit in the lock.
19.	男孩在踢足球	nan2 hai2 zai4 ti2 zu2 qiu2	The boy gave the football a kick.
20.	为了安全,警察穿了防弹衣	wei4 le an1 quan2 jing3 cha2 chuan1 le fang2 dan4 yi1	The cop wore a bullet-proof vest.

Appendix 4.2 Continued

#	Mandarin (Characters)	Pinyin	English
21.	洗完澡,他穿上睡衣	xi2 wan2 zao3,ta1 chuan1 shang4 shui4 yi1	After his bath he wore a robe.
22.	装汤用碗	zhuang1 Tang1 yong4 wan2	The soup was served in a bowl.
23.	工人们在挖一条水沟	gong1 ren2 men zai4 wa1 yi4 tiao2 shui3 gou1	The workers are digging a ditch.
24.	船长召集他的船员	chuan2 zhang3 zhao1 ji2 ta1 de chuan2 yuan2	The ship's captain summoned his crew.
25.	他们在玩猫捉老鼠的游戏	ta1 men zai4 wan3 mao1 zuo1 lao3 su3 de you2 xi4	They played a game of cat and mouse.
26.	黑腥腥是猿猴	hei1 xing1 xing shi4 yuan2 hou2	A chimpanzee is an ape.
27.	垫子里面塞的是塑料泡沫	dian4 zi li3 mian4 sai1 de shi4 su4 liao4 pao4 mo4	The cushion was filled with foam.
28.	他抛给快淹死的人一根绳	ta1 pao1 gei3 na4 ge4 kuai4 yan1 si2 de ren2 yi4 gen1 sheng2	He tossed the drowning man a rope.
29.	扫地用扫帚	sao3 di4 yong4 sao4 zhou3	To sweep the floor with a broom.
30.	我们听见钟的滴答声	wo3 men ting1 jian4 zhong1 de di1 da1 sheng1	We heard the ticking of the clock.
31.	医生开了处方	yi1 sheng1 kai1 le chu2 fang1	The doctor prescribed the drug.
32.	下棋是一种乐趣	xia4 qi2 shi4 yi4 zhong2 le4 qu4	Playing checkers can be fun.
33.	他早餐喝了一些牛奶	ta1 zao3 can1 he1 le yi4 xie1 niu2 nai3	At breakfast, he drank some milk.
34.	国王戴的金制的皇冠	guo2 wang2 dai4 de jin1 zhi4 de huang2 guan1	The king wore a golden crown.
35.	沙粒堆成了山	sha1 li4 dui1 cheng2 le shan1	The sand was heaped in a pile.
36.	为放木材,他搭了一个棚	wei4 fang4 mu4 cai2,ta1 da1 le yi2 ge4 peng2	To store his wood, he built a shed.
37.	狮子发出一声怒吼	shi1 zi fa1 chu1 yi4 sheng1 nu4 hou3	The lion gave an angry roar.
38.	高速公路有六条车道	gao1 su4 gong1 lu4 you3 liu4 tiao2 che1 dao4	The super highway has six lanes.
39.	汽车滚下了悬崖	qi4 che1 gun3 xia4 le xuan2 ya2	The car drove off the steep cliff.
40.	扔掉那些无用的垃圾	reng1 diao4 na4 xie1 wu2 yong4 de la1 ji1	Throw out all this useless junk.

Appendix 4.2 Continued

#	Mandarin (Characters)	Pinyin	English
41.	她给他做了一吨丰盛的饭菜	ta1 gei3 ta1 zuo4 le yi2 dun4 feng1 sheng4 de fan4 cai4	She cooked him a hearty meal.
42.	房东提高了房租	fang2 dong1 ti2 gao1 le fang2 zu1	The landlord raised the rent.
43.	我们的座位在第二排	wo3 men de zuo4 wei4 zai4 di4 er4 pai2	Our seats were in the second row.
44.	我们跟着狮子找到了它的窝	wo3 men gen1 zhe shi1 zi zhao3 dao4 le ta1 de wo1	We tracked the lion to his den.
45.	她给她自己倒了一杯茶	ta1 gei2 ta1 zi4 ji2 dao4 le yi4 bei1 cha2	She poured herself a cup of tea.
46.	大雨引起了洪灾	da4 yu3 yin3 qi3 le hong2 zai1	The heavy rains caused a flood.
47.	警察寻找线索	jing2 cha2 xun2 zhao3 xian4 suo3	The police searched for a clue.
48.	洗地板用抹布	xi2 di4 ban3 yong4 mo2 bu4	Wash the floor with a mop.
49.	小鸡啄玉米用嘴	xiao3 ji1 zhuo2 yu4 mi3 yong3 zui3	The chicken pecked corn with its beak.
50.	池塘里满是呱呱叫的青蛙	chi2 Tang2 li3 man4 shi4 gua1 gua jiao4 de qing1 wa1	The pond was full of croaking frogs.
51.	游泳的人跳进了游泳池	you2 yong3 de ren2 diao4 jin4 le you2 yong3 chi2	The swimmer dove into the pool.
52.	牧羊人看着他的羊群	mu4 yang2 ren2 kan1 zhe ta1 de yang2 qun2	The shepherd watched his flock of sheep.
53.	把肉切成小块	ba3 rou4 qie1 cheng2 xiao3 kuai4	Cut the meat into small chunks.
54.	西瓜有很多籽	xi1 gua you3 hen3 duo1 zi3	Watermelons have a lot of seeds.
55.	新娘穿着白色的婚纱	xin1 niang3 chuan1 zhe bai3 se4 de hun1 sha1	The bride wore a white gown.
56.	生病的孩子吞下了药片	sheng1 bing4 de hai2 zi tun1 xia4 le yao4 pian4	The sick child swallowed the pill.
57.	自行车有两个轮子	zi4 xing2 che1 you3 liang2 ge4 lun2 zi	A bicycle has two wheels.
58.	她有一条玻璃珠的项链	ta1 you3 yi4 tiao2 bo1 li2 zhu1 de xiang4 lian4	She had a necklace of glass beads.
59.	船驶出了港湾	chuan3 shi2 chu1 le gang2 wan1	The boat sailed across the bay.
60.	奶牛生了牛犊	nai3 niu2 sheng1 le niu2 du2	The cow gave birth to a calf.

Note: digits in Pinyin transcription refer to lexical tones. Tone 1 is the high level tone, Tone 2 is a mid-rising tone, Tone 3 is the low dipping tone and Tone 4 is high falling.

Appendix 5.1a. Lexical affinity index (LAI, proportion of cognates shared) for all pairs of listener dialects (across) and speaker dialects (down). Values have been copied from Cheng (1991: 96).[66] Note that no lexical affinity was available for Taiyuan and Hankou. Horizontal and vertical double lines divide the table into nine non-Mandarin and six Mandarin dialects.

Speaker Dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	1.000														
Wenzhou	0.313	1.000													
Guangzhou	0.184	0.195	1.000												
Xiamen	0.080	0.102	0.171	1.000											
Fuzhou	0.123	0.141	0.165	0.280	1.000										
Chaozhou	0.097	0.101	0.212	0.338	0.246	1.000									
Meixian	0.182	0.190	0.302	0.166	0.141	0.186	1.000								
Nanchang	0.376	0.282	0.246	0.133	0.184	0.150	0.272	1.000							
Changsha	0.345	0.261	0.228	0.120	0.160	0.135	0.226	0.555	1.000						
Taiyuan															
Beijing	0.289	0.218	0.240	0.199	0.269	0.214	0.215	0.443	0.461		1.000				
Jinan	0.310	0.231	0.222	0.164	0.218	0.174	0.212	0.455	0.487		0.672	1.000			
Hankou															
Chengdu	0.295	0.212	0.172	0.089	0.140	0.098	0.166	0.423	0.485		0.448	0.453		1.000	
Xi'an	0.317	0.221	0.209	0.133	0.201	0.140	0.201	0.448	0.484		0.611	0.608		0.487	1.000

[66] The table is based on Cheng 1991: 96, with one decimal less and Cheng 1997: 61. In both publications the index numbers were called 'correlation coefficients'. In the earlier version, Cheng explains that the index is a *phi* coefficient of association. In the later publication Cheng (1997: 53) rejects the *phi* coefficient, and explains that he actually used a different measure for lexical affinity, namely the LAI (lexical affinity index) as defined in our text above.

Appendix 5.1b Proximity matrix derived from Appendix 5.1a (LAI, 13 dialects, Hankou and Taiyuan are not available).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.000**	.994	1.204	1.427	1.322	1.395	1.205	.961	1.013		1.154	1.125		1.050	1.099
Wenzhou	.994	**.000**	1.161	1.346	1.263	1.336	1.165	1.155	1.200		1.316	1.301		1.220	1.292
Guangzhou	1.204	1.161	**.000**	1.218	1.205	1.157	.988	1.229	1.276		1.311	1.333		1.302	1.334
Xiamen	1.427	1.346	1.218	**.000**	1.034	.939	1.230	1.493	1.530		1.480	1.524		1.511	1.542
Fuzhou	1.322	1.263	1.205	1.034	**.000**	1.077	1.238	1.361	1.403		1.321	1.377		1.380	1.386
Chaozhou	1.395	1.336	1.157	.939	1.077	**.000**	1.193	1.455	1.494		1.447	1.494		1.484	1.515
Meixian	1.205	1.165	.988	1.230	1.238	1.193	**.000**	1.208	1.280		1.345	1.353		1.311	1.351
Nanchang	.961	1.155	1.229	1.493	1.361	1.455	1.208	**.000**	.638		.857	.832		.840	.828
Changsha	1.013	1.200	1.276	1.530	1.403	1.494	1.280	.638	**.000**		.821	.778		.750	.766
Taiyuan															
Beijing	1.154	1.316	1.311	1.480	1.321	1.447	1.345	.857	.821		**.000**	.472		.850	.570
Jinan	1.125	1.301	1.333	1.524	1.377	1.494	1.353	.832	.778		.472	**.000**		.828	.562
Hankou															
Chengdu	1.050	1.220	1.302	1.511	1.380	1.484	1.311	.840	.750		.850	.828		**.000**	.766
Xi'an	1.099	1.292	1.334	1.542	1.386	1.515	1.351	.828	.766		.570	.562		.766	**.000**

Appendix 5.2a Occurrence of initials (onset consonants) in the phoneme inventories of 15 dialects (transcription in X-SAMPA). ‘1’: occurs, ‘0’: does not occur.

number	X-SAMPA	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi’an
1.	p	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2.	pf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3.	t	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4.	ts	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5.	ts`	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1
6.	ts\	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1
7.	k	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8.	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9.	p_h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10.	pf_h	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
11.	t_h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12.	ts_h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13.	ts`_h	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1
14.	ts_h	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1
15.	k_h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16.	b	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0
17.	d	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
18.	dz	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
19.	dz`	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
20.	g	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0
21.	m	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
22.	n	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
23.	s	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
24.	s`	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1
25.	J	1	1	0	0	0	0	0	1	1	0	0	1	0	1	0
26.	N	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
27.	f	1	1	1	0	0	0	1	1	0	1	1	1	0	1	1
28.	l	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1
29.	z	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0
30.	s\	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1
31.	z`	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1
32.	h	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
33.	v	1	1	1	0	0	0	1	0	1	1	0	1	0	0	1
34.	x	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
35.	G	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
36.	j	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
37.	h\	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 5.2b Proximity matrix derived from Appendix 5.2a (initials in phoneme inventory).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	1.41	3.16	3.16	3.46	3.00	3.16	3.16	3.74	3.74	3.87	3.46	3.46	3.00	3.87
Wenzhou	1.41	**0.00**	3.46	3.46	3.74	3.32	3.46	3.46	4.00	4.00	4.12	3.74	3.74	3.32	4.12
Guangzhou	3.16	3.46	**0.00**	2.00	1.41	2.24	0.00	2.83	3.46	3.46	3.32	3.16	2.83	3.00	3.32
Xiamen	3.16	3.46	2.00	**0.00**	1.41	1.00	2.00	2.83	3.74	4.00	3.61	3.74	2.83	3.32	3.87
Fuzhou	3.46	3.74	1.41	1.41	**0.00**	1.73	1.41	2.83	3.46	3.74	3.32	3.46	2.45	3.00	3.61
Chaozhou	3.00	3.32	2.24	1.00	1.73	**0.00**	2.24	3.00	3.87	3.87	3.74	3.87	3.00	3.16	4.00
Meixian	3.16	3.46	0.00	2.00	1.41	2.24	**0.00**	2.83	3.46	3.46	3.32	3.16	2.83	3.00	3.32
Nanchang	3.16	3.46	2.83	2.83	2.83	3.00	2.83	**0.00**	3.46	4.00	3.61	3.46	3.16	3.00	3.87
Changsha	3.74	4.00	3.46	3.74	3.46	3.87	3.46	3.46	**0.00**	2.83	2.65	2.00	2.45	2.65	2.65
Taiyuan	3.74	4.00	3.46	4.00	3.74	3.87	3.46	4.00	2.83	**0.00**	1.73	2.00	3.16	3.00	2.24
Beijing	3.87	4.12	3.32	3.61	3.32	3.74	3.32	3.61	2.65	1.73	**0.00**	1.73	2.65	2.83	2.00
Jinan	3.46	3.74	3.16	3.74	3.46	3.87	3.16	3.46	2.00	2.00	1.73	**0.00**	2.83	2.65	1.73
Hankou	3.46	3.74	2.83	2.83	2.45	3.00	2.83	3.16	2.45	3.16	2.65	2.83	**0.00**	1.73	3.00
Chengdu	3.00	3.32	3.00	3.32	3.00	3.16	3.00	3.00	2.65	3.00	2.83	2.65	1.73	**0.00**	3.16
Xi'an	3.87	4.12	3.32	3.87	3.61	4.00	3.32	3.87	2.65	2.24	2.00	1.73	3.00	3.16	**0.00**

Appendix 5.3a Occurrence of vocalic nuclei in the phoneme inventories of 15 dialects (transcription in IPA). '1': occurs, '0': does not occur.

#	Ipapan	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
1.	ʅ	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1
2.	ʅ	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1
3.	ə	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1
4.	əi	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5.	əʅ	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
6.	əu	1	0	0	0	0	0	0	1	1	1	0	0	0	1	0
7.	ə̃	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
8.	ɚ	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
9.	ɘ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
10.	ɯ	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1
11.	ɥ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12.	ɑ	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1
13.	ɑ̃	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
14.	ɑi	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
15.	ɑu	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
16.	a	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17.	aŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
18.	ã	1	0	0	1	0	1	0	0	0	1	0	1	0	0	0
19.	ai	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0
20.	au	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1
21.	ãu	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
22.	ãi	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
23.	ɛ	0	1	1	0	1	0	1	1	0	0	1	1	1	1	1
24.	ɐu	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
25.	ɛu	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
26.	ɛi	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
27.	ɛe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
28.	õ	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
29.	ou	0	0	1	0	1	1	0	0	0	0	1	1	1	0	1
30.	oi	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0

Appendix 5.3a (continued)

#	Ipapan	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
31.	õi	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
32.	õu	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
33.	e	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0
34.	ẽ	0	0	0	1	0	1	0	0	1	1	0	1	0	0	1
35.	ei	0	1	1	0	1	0	0	1	1	1	1	1	1	1	1
36.	eu	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
37.	ɤu	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
38.	ɤ	1	0	0	0	0	1	0	0	0	1	1	1	1	0	1
39.	ʏ	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
40.	ʏ̃	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
41.	ø	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
42.	øy	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
43.	ɔ	0	1	1	1	1	0	1	1	0	1	0	1	1	0	0
44.	ɔi	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0
45.	ɔ̃	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
46.	ɔy	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
47.	ɔu	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
48.	ɐi	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
49.	ɐ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
50.	æ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51.	æ̃	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1
52.	œ	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
53.	ɛi	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
54.	eu	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
55.	õu	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
56.	ou	0	0	0	0	0	1	0	0	0	0	1	1	1	0	1
57.	o	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
58.	ɪ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59.	ĩ	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
60.	ɪu	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Appendix 5.3a (continued)

#	Ipapan	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
61.	iu	0	0	1	1	0	1	1	1	0	0	0	0	0	0	0
62.	ɨu	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
63.	i	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
64.	ɨ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
65.	ũ	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
66.	ui	0	0	1	1	0	1	1	1	0	0	0	0	0	0	0
67.	iɛ	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
68.	ɨi	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
69.	ᴇ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
70.	u	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
71.	m̩	1	1	1	1	0	1	1	1	1	1	1	0	1	0	0
72.	n̩	1	1	0	0	0	0	0	1	1	1	0	0	1	0	0
73.	ŋ̍	1	1	1	1	0	1	1	1	0	0	0	0	1	0	0
74.	y	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1
75.	ỹ	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
76.	uĩ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
77.	uʏ	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
78.	uɤ	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0

Appendix 5.3b Proximity matrix derived from Appendix 5.3a (nuclei in phoneme inventory of 15 dialects).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	4.24	4.90	4.90	5.39	5.48	4.24	5.20	4.24	4.24	4.12	4.69	4.58	4.24	5.10
Wenzhou	4.24	**0.00**	3.46	4.47	4.36	5.29	3.46	4.12	4.00	4.00	3.87	4.24	3.32	3.46	4.90
Guangzhou	4.90	3.46	**0.00**	4.24	4.58	5.10	4.00	4.12	4.69	4.69	4.36	4.69	3.87	4.47	5.48
Xiamen	4.90	4.47	4.24	**0.00**	5.39	3.16	3.74	4.58	4.90	4.90	4.58	4.69	4.36	4.69	5.66
Fuzhou	5.39	4.36	4.58	5.39	**0.00**	5.92	4.80	4.90	5.20	5.20	4.90	5.57	4.90	4.80	5.75
Chaozhou	5.48	5.29	5.10	3.16	5.92	**0.00**	4.00	5.00	5.29	5.10	5.00	5.48	4.58	5.10	5.83
Meixian	4.24	3.46	4.00	3.74	4.80	4.00	**0.00**	3.61	3.74	3.74	3.61	4.47	3.32	3.46	5.10
Nanchang	5.20	4.12	4.12	4.58	4.90	5.00	3.61	**0.00**	4.36	4.36	4.47	4.58	3.16	4.12	5.57
Changsha	4.24	4.00	4.69	4.90	5.20	5.29	3.74	4.36	**0.00**	3.16	3.61	4.24	3.61	3.16	4.69
Taiyuan	4.24	4.00	4.69	4.90	5.20	5.10	3.74	4.36	3.16	**0.00**	3.61	4.69	3.32	3.16	4.90
Beijing	4.12	3.87	4.36	4.58	4.90	5.00	3.61	4.47	3.61	3.61	**0.00**	4.12	3.46	3.61	3.87
Jinan	4.69	4.24	4.69	4.69	5.57	5.48	4.47	4.58	4.24	4.69	4.12	**0.00**	4.12	4.47	4.47
Hankou	4.58	3.32	3.87	4.36	4.90	4.58	3.32	3.16	3.61	3.32	3.46	4.12	**0.00**	3.32	4.58
Chengdu	4.24	3.46	4.47	4.69	4.80	5.10	3.46	4.12	3.16	3.16	3.61	4.47	3.32	**0.00**	4.90
Xi'an	5.10	4.90	5.48	5.66	5.75	5.83	5.10	5.57	4.69	4.90	3.87	4.47	4.58	4.90	**0.00**

Appendix 5.4a Occurrence of codas in the phoneme inventories of 15 dialects (transcription in X-SAMPA). '1': occurs, '0': does not occur.

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
1.	/	1	0	0	1	1	1	1	1	0	1	0	0	0	0	0
2.	}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
3.	k	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
4.	l	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5.	m	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
6.	m`	1	1	1	1	0	1	0	1	1	0	0	0	1	0	0
7.	m`/	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
8.	n	1	0	1	1	0	0	1	1	1	0	1	0	1	1	1
9.	N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10.	n`	1	1	0	0	0	0	0	1	1	0	0	0	1	0	0
11.	N?	1	1	1	1	0	1	0	1	0	0	0	0	1	0	0
12.	N?	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
13.	p	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
14.	r	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
15.	t	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0

Appendix 5.4b Proximity matrix derived from Appendix 5.4a (codas in phoneme inventory)

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	1.73	2.65	2.83	2.24	2.83	2.83	1.41	1.73	2.45	2.24	2.65	1.41	2.24	2.45
Wenzhou	1.73	**0.00**	2.45	3.00	2.00	2.65	3.00	1.73	1.41	2.24	2.00	2.00	1.00	2.00	2.24
Guangzhou	2.65	2.45	**0.00**	1.73	2.83	2.24	1.73	2.24	2.45	3.00	2.45	2.83	2.24	2.45	2.65
Xiamen	2.83	3.00	1.73	**0.00**	3.00	1.41	2.00	2.45	3.00	3.16	3.00	3.32	2.83	3.00	3.16
Fuzhou	2.24	2.00	2.83	3.00	**0.00**	2.65	2.24	2.24	2.00	1.00	1.41	1.41	2.24	1.41	1.73
Chaozhou	2.83	2.65	2.24	1.41	2.65	**0.00**	2.45	2.83	3.00	2.83	3.00	3.00	2.83	3.00	3.16
Meixian	2.83	3.00	1.73	2.00	2.24	2.45	**0.00**	2.45	2.65	2.45	2.24	2.65	2.83	2.24	2.45
Nanchang	1.41	1.73	2.24	2.45	2.24	2.83	2.45	**0.00**	1.73	2.45	2.24	2.65	1.41	2.24	2.45
Changsha	1.73	1.41	2.45	3.00	2.00	3.00	2.65	1.73	**0.00**	2.24	1.41	2.00	1.00	1.41	1.73
Taiyuan	2.45	2.24	3.00	3.16	1.00	2.83	2.45	2.45	2.24	**0.00**	1.73	1.73	2.45	1.73	2.00
Beijing	2.24	2.00	2.45	3.00	1.41	3.00	2.24	2.24	1.41	1.73	**0.00**	1.41	1.73	0.00	1.00
Jinan	2.65	2.00	2.83	3.32	1.41	3.00	2.65	2.65	2.00	1.73	1.41	**0.00**	2.24	1.41	1.00
Hankou	1.41	1.00	2.24	2.83	2.24	2.83	2.83	1.41	1.00	2.45	1.73	2.24	**0.00**	1.73	2.00
Chengdu	2.24	2.00	2.45	3.00	1.41	3.00	2.24	2.24	1.41	1.73	0.00	1.41	1.73	**0.00**	1.00
Xi'an	2.45	2.24	2.65	3.16	1.73	3.16	2.45	2.45	1.73	2.00	1.00	1.00	2.00	1.00	**0.00**

Appendix 5.5a Occurrence of word tones in the sound inventories of 15 dialects. Tones are transcribed with maximally 3-digits expressing relative pitch on a scale from 1 (low pitch) to 5 (high pitch). For other cells: '1': occurs, '0': does not occur.

#	Tone type	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
1.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2.	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3.	2	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0
4.	3	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
5.	5	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0
6.	11	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
7.	13	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0
8.	21	0	0	1	1	0	0	0	1	0	0	0	1	0	0	1
9.	22	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
10.	23	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
11.	24	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1
12.	31	1	1	0	0	0	1	1	0	0	0	0	0	0	1	0
13.	32	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
14.	33	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0
15.	35	0	1	1	1	0	1	0	0	0	0	1	0	1	0	0
16.	41	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
17.	42	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0
18.	44	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1
19.	45	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
20.	51	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
21.	53	0	0	1	1	1	1	1	0	0	1	0	0	0	1	1
22.	54	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
23.	55	1	0	0	1	0	1	0	0	1	0	1	1	1	1	0
24.	212	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
25.	213	0	1	0	0	0	1	0	1	0	0	0	1	1	0	0
26.	214	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
27.	242	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
28.	513	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 5.5b Proximity matrix derived from Appendix 5.5a (tones in the inventories of 15 dialects).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	3.32	3.46	3.16	3.46	3.00	3.00	3.46	2.83	3.46	2.65	3.00	3.00	2.24	3.32
Wenzhou	3.32	**0.00**	3.00	3.00	3.61	2.45	3.16	3.00	3.00	3.32	2.83	2.45	2.00	2.83	3.16
Guangzhou	3.46	3.00	**0.00**	2.45	3.16	2.65	3.32	3.16	3.46	3.16	3.32	3.32	3.32	3.32	3.00
Xiamen	3.16	3.00	2.45	**0.00**	2.83	2.65	3.00	3.16	3.16	3.16	2.65	2.65	2.65	2.65	2.65
Fuzhou	3.46	3.61	3.16	2.83	**0.00**	3.32	2.65	3.46	3.46	3.16	3.32	3.32	3.32	3.00	2.65
Chaozhou	3.00	2.45	2.65	2.65	3.32	**0.00**	2.83	3.00	3.00	3.00	2.83	2.83	2.45	2.45	3.16
Meixian	3.00	3.16	3.32	3.00	2.65	2.83	**0.00**	3.32	3.32	2.65	3.16	3.16	3.16	2.45	2.45
Nanchang	3.46	3.00	3.16	3.16	3.46	3.00	3.32	**0.00**	3.16	2.83	3.32	2.24	2.65	3.32	2.65
Changsha	2.83	3.00	3.46	3.16	3.46	3.00	3.32	3.16	**0.00**	3.16	2.65	2.65	2.65	2.24	2.65
Taiyuan	3.46	3.32	3.16	3.16	3.16	3.00	2.65	2.83	3.16	**0.00**	3.00	3.00	3.00	2.65	2.65
Beijing	2.65	2.83	3.32	2.65	3.32	2.83	3.16	3.32	2.65	3.00	**0.00**	2.45	2.00	2.45	2.83
Jinan	3.00	2.45	3.32	2.65	3.32	2.83	3.16	2.24	2.65	3.00	2.45	**0.00**	1.41	2.45	2.45
Hankou	3.00	2.00	3.32	2.65	3.32	2.45	3.16	2.65	2.65	3.00	2.00	1.41	**0.00**	2.45	2.83
Chengdu	2.24	2.83	3.32	2.65	3.00	2.45	2.45	3.32	2.24	2.65	2.45	2.45	2.45	**0.00**	2.45
Xi'an	3.32	3.16	3.00	2.65	2.65	3.16	2.45	2.65	2.65	2.65	2.83	2.45	2.83	2.45	**0.00**

Appendix 5.6a Occurrences of finals in 15 dialects. ‘1’: occurs, ‘0’: does not occur.

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi’an
1.	ɿ	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1
2.	ʅ	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1
3.	əɽ	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
4.	ɘ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
5.	ɯ	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1
6.	ɯʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7.	ɥ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8.	ɑ	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1
9.	ɑʔ	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
10.	ɑ̃	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11.	ɑŋ	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
12.	iɑŋ	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
13.	uɑŋ	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
14.	a	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0
15.	ai	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0
16.	ak	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
17.	ɑu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
18.	au	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0
19.	am	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
20.	an	0	0	1	1	0	0	1	1	1	0	1	0	1	1	0
21.	ap	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
22.	aŋ	0	1	1	1	1	1	1	1	0	0	0	1	1	1	0
23.	at	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0
24.	ã	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0
25.	aʔ	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
26.	ɛ	0	1	1	0	1	0	0	0	0	0	0	1	1	0	1
27.	ə	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0
28.	ən	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
29.	ɚ	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
30.	o	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
31.	e	0	1	0	1	0	1	1	1	0	0	0	0	1	1	0
32.	ei	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1
33.	ek	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
34.	eŋ	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
35.	ɐi	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
36.	ɐu	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
37.	ɐm	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
38.	ɐn	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
39.	ɐp	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
40.	ɐt	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
41.	ɐk	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
42.	ɛu	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
43.	ɛk	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
44.	ɐŋ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
45.	ɛŋ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
46.	ø	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
47.	øn	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
48.	øt	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
49.	øy	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
50.	æ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51.	iæ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52.	ɛe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
53.	œ	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
54.	œk	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
55.	œŋ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
56.	ɤ	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1
57.	ᴇ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58.	əm	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
59.	əu	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0
60.	ɔ	0	0	1	1	1	0	0	0	0	0	0	1	1	0	0

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
61.	ɔ̃	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
62.	ɔk	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
63.	ɔn	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
64.	ɔt	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
65.	ɔŋ	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
66.	ɵ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
67.	ok	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0
68.	ou	0	0	1	0	0	1	0	0	0	0	1	1	1	0	1
69.	oŋ	1	1	1	0	0	1	1	0	1	0	0	0	1	1	0
70.	oʔ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
71.	ui	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
72.	iɪ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
73.	ɤu	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
74.	ɛi	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
75.	eʔ	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
76.	ɨi	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
77.	ẽ	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1
78.	m̩	1	1	1	1	0	1	0	1	1	0	0	0	1	0	0
79.	ŋ	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
80.	iu	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0
81.	i	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
82.	əl	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
83.	əp	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
84.	ɨu	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
85.	õ	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
86.	ĩ	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
87.	ə̃	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
88.	æ̃	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
89.	on	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
90.	əŋ	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
91.	ɛn	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
92.	aʔ	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0
93.	em	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
94.	ɨn	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
95.	iɑ	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1
96.	iɑʔ	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
97.	iɑ̃	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
98.	ɒ̃	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
99.	iɒ̃	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
100.	uɒ̃	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
101.	ɔŋ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
102.	im	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
103.	ip	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
104.	en	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
105.	ik	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
106.	et	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
107.	ot	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
108.	ɔʔ	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0
109.	ɛt	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
110.	ãi	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
111.	ɨt	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
112.	ut	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
113.	əʔ	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
114.	ɔk	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
115.	oʔ	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
116.	uʔ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
117.	aʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
118.	iã	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0
119.	iaʔ	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0
120.	ou	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
121.	iɛ	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
122.	ia	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0
123.	iai	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0
124.	iau	0	1	1	1	0	0	1	0	1	1	1	0	1	1	0
125.	iɑu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
126.	iam	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
127.	iap	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
128.	iaŋ	0	1	1	1	1	1	1	1	0	0	0	1	1	1	0
129.	ian	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0
130.	iũ	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
131.	iak	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
132.	uk	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
133.	auʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
134.	ãu	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
135.	õu	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
136.	m̩ʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
137.	ĩʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
138.	ãʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
139.	ẽʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
140.	õi	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
141.	om	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
142.	ãuʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
143.	oi	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
144.	iek	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
145.	ieŋ	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
146.	iæ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
147.	iɐi	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
148.	iɐu	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
149.	iɐm	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
150.	iɐn	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
151.	iɐp	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
152.	iɛt	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
153.	iɐt	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
154.	iɛk	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
155.	uᴇ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
156.	iəʔ	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
157.	iɛŋ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
158.	iøy	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
159.	io	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0
160.	ioʔ	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
161.	iœk	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
162.	iœŋ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
163.	ioŋ	1	0	1	0	0	1	1	0	1	0	0	0	1	0	0
164.	iø	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
165.	uø	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
166.	iʏ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
167.	ɔi	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
168.	iə	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
169.	iɔ	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
170.	iʔ	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
171.	iɛ	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1
172.	iɛi	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
173.	ieu	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
174.	iɛu	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
175.	ie	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0
176.	n̩	1	1	0	0	0	0	0	1	1	0	0	0	1	0	0
177.	iẽ	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1
178.	iɛn	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0
179.	iəu	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0
180.	iou	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
181.	io	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
182.	in	1	0	1	1	0	0	1	1	1	0	1	0	1	1	1
183.	iŋ	0	0	0	0	1	1	0	0	0	1	1	1	0	0	1
184.	iuŋ	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
185.	iui	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
186.	iæ̃	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
187.	$ɑ^{ŋ}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
188.	$iɑ^{ŋ}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
189.	$uɑ^{ŋ}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
190.	iaʔ	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0
191.	iok	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0
192.	ɑi	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
193.	ɑu	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
194.	iɔŋ	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
195.	iøn	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
196.	ieʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
197.	ep	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
198.	iøt	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
199.	œʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
200.	ət	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
201.	iat	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
202.	ien	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
203.	iet	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
204.	iɔk	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
205.	uɔi	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
206.	it	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0
207.	iɔʔ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
208.	ion	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
209.	iuʔ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
210.	iut	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
211.	iuk	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
212.	iou	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
213.	iun	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
214.	iuʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
215.	iauʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
216.	iãu	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
217.	iãuʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
218.	iãʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
219.	ɔ̃ʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
220.	ŋ̍ʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
221.	iõu	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
222.	ouŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
223.	uai	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
224.	uɑ	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1
225.	uɔ	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
226.	uɑ̃	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
227.	uã	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0
228.	ua	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0
229.	uiɛ	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
230.	uɛ	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1
231.	op	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
232.	u	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
233.	uʔ	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
234.	uat	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
235.	uəʔ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
236.	uek	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
237.	uĩ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
238.	ueŋ	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
239.	uɐi	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
240.	uɐn	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
241.	uɐt	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
242.	uɐk	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
243.	uɐŋ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
244.	ŋ̍	1	1	1	1	0	1	0	1	0	0	0	0	1	0	0
245.	ue	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0
246.	uə	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
247.	uɤ	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
248.	uɛe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
249.	uẽ	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1
250.	uɔk	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
251.	uo	0	0	0	0	1	0	1	1	0	0	1	0	0	0	1
252.	uaʔ	1	0	0	1	1	1	0	1	0	1	0	0	0	0	0
253.	uɔŋ	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0
254.	uən	1	0	0	0	0	0	0	0	1	0	1	0	1	1	1
255.	uæ̃	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
256.	uəŋ	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
257.	uŋ	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1
258.	yaʔ	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
259.	eu	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
260.	uai	0	0	0	1	1	1	1	1	1	1	1	0	1	1	0
261.	uak	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0
262.	uei	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
263.	y	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1
264.	uã	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
265.	ui	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0
266.	iu	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
267.	uɑi	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
268.	ya	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
269.	uəʔ	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
270.	uaŋ	0	1	1	0	1	1	1	1	0	0	0	1	1	1	0

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
271.	uɑʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
272.	uɑŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
273.	uɔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
274.	uɔʔ	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
275.	uan	0	0	1	1	0	0	1	1	1	0	0	0	1	1	0
276.	uɑn	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
277.	uon	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
278.	uoi	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
279.	uen	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
280.	uet	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
281.	un	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0
282.	ueʔ	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
283.	uoʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
284.	yən	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
285.	ouʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
286.	uk	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
287.	uat	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
288.	uot	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
289.	uɛt	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
290.	ut	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0
291.	uoŋ	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
292.	uok	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
293.	uĩ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
294.	uãi	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
295.	uam	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
296.	yə	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
297.	ɤŋ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
298.	yɛ	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
299.	yã	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
300.	uiʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
301.	uaiʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
302.	yai	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
303.	yei	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
304.	uap	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
305.	yan	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0
306.	uẽʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
307.	uãiʔ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
308.	yæ̃	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
309.	ye	0	0	0	0	0	0	0	1	1	1	0	0	1	1	0
310.	yn	0	0	1	0	0	0	0	1	1	0	1	0	1	1	1
311.	yɛn	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
312.	yẽ	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1
313.	yŋ	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1
314.	yʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
315.	yai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
316.	yon	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
317.	yəʔ	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
318.	yot	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
319.	yt	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
320.	yɔ	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
321.	ɔy	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
322.	yɔʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
323.	yɔŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
324.	yo	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1
325.	yoʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
326.	yoŋ	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0
327.	ieŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
328.	ɤk	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
329.	iɛŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
330.	iuʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
331.	aiʔ	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
332.	oiʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
333.	iouʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
334.	auʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
335.	m̩ʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
336.	ŋ̍ʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
337.	ĩʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
338.	ẽʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
339.	iũʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
340.	ãiʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
341.	uãiʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
342.	ãuʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
343.	iõuʔ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
344.	ɛiŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
345.	aiŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
346.	ɔyŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
347.	ɔuŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
348.	øyŋ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
349.	ieʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
350.	iɛʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
351.	ɛʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
352.	ɛiʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
353.	ɔyʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
354.	ɔuʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
355.	øyʔ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
356.	ər	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
357.	ə̃r	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
358.	iər	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
359.	iə̃r	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
360.	uər	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1

Appendix 5.6a (continued)

#	IPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi’an
361.	uə̃r	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
362.	yər	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
363.	ar	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
364.	ãr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
365.	iar	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
366.	iãr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
367.	ɑr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
368.	iɑr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
369.	uar	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
370.	uãr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
371.	uɑr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
372.	ur	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
373.	yɑr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
374.	ɛr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
375.	iɛr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
376.	uɛr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
377.	yɛr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
378.	ier	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
379.	er	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
380.	uer	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
381.	yer	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
382.	ɤr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
383.	ɔr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
384.	iɔr	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
385.	or	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
386.	uor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
387.	yor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
388.	ɑur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
389.	iɑur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
390.	our	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
391.	iour	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1

Appendix 5.6b Proximity matrix derived from Appendix 5.6a (inventory of finals in 15 dialects).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.0**	7.9	11.3	10.3	10.3	10.6	10.1	9.8	7.8	8.3	8.1	9.9	8.0	8.1	9.5
Wenzhou	7.9	**0.0**	9.1	9.3	8.7	9.6	8.6	8.1	6.6	7.0	7.4	8.5	6.0	5.8	9.1
Guangzhou	11.3	9.1	**0.0**	10.8	11.9	11.0	10.0	10.5	10.1	10.5	10.4	11.4	9.4	9.8	11.8
Xiamen	10.3	9.3	10.8	**0.0**	11.3	9.5	9.6	9.4	9.5	9.7	10.1	11.0	9.2	9.1	11.5
Fuzhou	10.3	8.7	11.9	11.3	**0.0**	11.4	10.9	10.0	9.9	9.2	9.4	10.7	9.5	9.2	10.8
Chaozhou	10.6	9.6	11.0	9.5	11.4	**0.0**	10.2	10.6	10.2	10.1	10.5	11.0	9.6	9.8	11.5
Meixian	10.1	8.6	10.0	9.6	10.9	10.2	**0.0**	9.1	8.5	9.1	8.9	10.7	8.3	8.1	10.9
Nanchang	9.8	8.1	10.5	9.4	10.0	10.6	9.1	**0.0**	8.3	8.3	8.6	10.3	7.6	7.5	10.5
Changsha	7.8	6.6	10.1	9.5	9.9	10.2	8.5	8.3	**0.0**	6.6	6.6	8.8	5.5	5.1	8.7
Taiyuan	8.3	7.0	10.5	9.7	9.2	10.1	9.1	8.3	6.6	**0.0**	6.9	8.6	6.7	6.3	8.5
Beijing	8.1	7.4	10.4	10.1	9.4	10.5	8.9	8.6	6.6	6.9	**0.0**	8.7	6.3	6.2	7.2
Jinan	9.9	8.5	11.4	11.0	10.7	11.0	10.7	10.3	8.8	8.6	8.7	**0.0**	8.4	8.7	8.3
Hankou	8.0	6.0	9.4	9.2	9.5	9.6	8.3	7.6	5.5	6.7	6.3	8.4	**0.0**	4.9	8.5
Chengdu	8.1	5.8	9.8	9.1	9.2	9.8	8.1	7.5	5.1	6.3	6.2	8.7	4.9	**0.0**	8.6
Xi'an	9.5	9.1	11.8	11.5	10.8	11.5	10.9	10.5	8.7	8.5	7.2	8.3	8.5	8.6	**0.0**

Appendix 5.7a Union of occurrences of initials and codas in 15 dialects. This is the concatenation of Appendices 5.2a and 5.4a. These tables are not reproduced here.

Appendix 5.7b Proximity matrix derived Appendix 5.7a (union of initials and codas).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	2.24	4.12	4.24	4.12	4.12	4.24	3.46	4.12	4.47	4.47	4.36	3.74	3.74	4.58
Wenzhou	2.24	**0.00**	4.24	4.58	4.24	4.24	4.58	3.87	4.24	4.58	4.58	4.24	3.87	3.87	4.69
Guangzhou	4.12	4.24	**0.00**	2.65	3.16	3.16	1.73	3.61	4.24	4.58	4.12	4.24	3.61	3.87	4.24
Xiamen	4.24	4.58	2.65	**0.00**	3.32	1.73	2.83	3.74	4.80	5.10	4.69	5.00	4.00	4.47	5.00
Fuzhou	4.12	4.24	3.16	3.32	**0.00**	3.16	2.65	3.61	4.00	3.87	3.61	3.74	3.32	3.32	4.00
Chaozhou	4.12	4.24	3.16	1.73	3.16	**0.00**	3.32	4.12	4.90	4.80	4.80	4.90	4.12	4.36	5.10
Meixian	4.24	4.58	1.73	2.83	2.65	3.32	**0.00**	3.74	4.36	4.24	4.00	4.12	4.00	3.74	4.12
Nanchang	3.46	3.87	3.61	3.74	3.61	4.12	3.74	**0.00**	3.87	4.69	4.24	4.36	3.46	3.74	4.58
Changsha	4.12	4.24	4.24	4.80	4.00	4.90	4.36	3.87	**0.00**	3.61	3.00	2.83	2.65	3.00	3.16
Taiyuan	4.47	4.58	4.58	5.10	3.87	4.80	4.24	4.69	3.61	**0.00**	2.45	2.65	4.00	3.46	3.00
Beijing	4.47	4.58	4.12	4.69	3.61	4.80	4.00	4.24	3.00	2.45	**0.00**	2.24	3.16	2.83	2.24
Jinan	4.36	4.24	4.24	5.00	3.74	4.90	4.12	4.36	2.83	2.65	2.24	**0.00**	3.61	3.00	2.00
Hankou	3.74	3.87	3.61	4.00	3.32	4.12	4.00	3.46	2.65	4.00	3.16	3.61	**0.00**	2.45	3.61
Chengdu	3.74	3.87	3.87	4.47	3.32	4.36	3.74	3.74	3.00	3.46	2.83	3.00	2.45	**0.00**	3.32
Xi'an	4.58	4.69	4.24	5.00	4.00	5.10	4.12	4.58	3.16	3.00	2.24	2.00	3.61	3.32	**0.00**

Appendix 5.8a Lexical frequencies of initials (onsets) in 15 dialects counted in the CASS database (764 items).

Number	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
1.	-	51	16	10	68	99	70	68	109	91	66	102	75	103	83	79
2.	b	16	16	0	27	0	12	0	0	0	0	0	0	0	0	0
3.	d	18	17	0	0	0	0	0	0	1	0	0	0	0	0	0
4.	dz	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0
5.	dz`	15	25	0	0	0	0	0	0	0	0	0	0	0	0	0
6.	f	9	7	30	0	0	0	43	39	36	21	21	21	21	22	34
7.	g	4	0	0	32	0	7	0	0	0	0	0	0	0	0	0
8.	G	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0
9.	h	15	15	69	105	84	97	59	28	0	0	0	0	0	0	0
10.	h\	73	33	0	0	0	0	0	0	0	0	0	0	0	0	0
11.	j -	0	75	86	0	0	0	0	0	0	0	0	0	0	0	0
12.	k	38	40	72	81	89	80	62	44	42	35	35	36	37	38	35
13.	k_h	20	22	19	31	32	37	46	22	21	21	21	20	21	22	21
14.	l	43	42	42	84	42	47	42	64	54	42	42	53	0	0	44
15.	m	24	25	31	3	22	19	26	23	22	22	22	23	22	22	22
16.	n	13	12	24	5	33	19	16	0	0	31	29	10	86	56	36
17.	N	13	13	17	1	31	22	50	16	12	0	0	8	9	13	9
18.	n`	35	39	0	0	0	0	1	34	27	0	0	15	0	23	0
19.	p	21	20	27	34	41	29	20	20	37	28	28	28	28	28	27
20.	p_h	11	11	21	14	13	21	30	25	12	20	20	20	20	21	21
21.	pf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19
22.	pf_h	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
23.	s	56	50	87	106	100	90	99	75	53	71	22	17	68	74	39
24.	s\	20	30	0	0	0	0	0	45	42	60	59	62	60	56	58
25.	s`	0	0	0	0	0	0	0	0	28	0	47	51	0	0	19
26.	t	16	16	25	57	54	49	19	15	33	25	25	26	26	25	24
27.	t_h	13	13	22	22	25	26	30	30	13	21	22	21	21	21	22
28.	ts	50	34	81	62	61	64	49	47	61	69	23	23	65	69	37
29.	ts\	24	41	0	0	0	0	0	26	53	50	49	50	51	47	49
30.	ts_h	10	16	0	0	0	0	0	39	15	21	21	21	26	20	19
31.	ts_h	32	24	64	32	38	50	79	63	21	54	13	13	52	50	27
32.	ts`	0	0	0	0	0	0	0	0	29	0	48	47	0	0	14
33.	ts`_h	0	0	0	0	0	0	0	0	9	0	44	43	0	0	15
34.	v	16	36	0	0	0	0	25	0	0	27	0	21	0	0	16
35.	w	0	0	37	0	0	0	0	0	0	0	0	0	0	0	0
36.	x	0	0	0	0	0	0	0	0	37	48	46	46	48	51	50
37.	z	108	58	0	0	0	25	0	0	0	23	0	0	0	23	0
38.	z`	0	0	0	0	0	0	0	0	15	0	25	14	0	0	14

Appendix 5.8b Proximity matrix derived from Appendix 5.8a. (lexical frequencies of initials in CASS database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.0**	13.0	24.2	23.0	22.3	20.0	21.4	19.1	21.4	19.2	25.0	23.5	22.8	19.9	22.5
Wenzhou	13.0	**0.0**	22.1	25.5	24.8	23.0	23.5	21.1	22.8	20.4	25.8	23.7	24.4	21.7	23.3
Guangzhou	24.2	22.1	**0.0**	19.0	16.3	16.4	15.5	20.5	24.3	22.6	27.3	27.4	23.5	23.3	24.5
Xiamen	23.0	25.5	19.0	**0.0**	12.1	10.3	16.8	19.9	23.2	22.8	26.3	26.6	24.5	24.5	24.8
Fuzhou	22.3	24.8	16.3	12.1	**0.0**	7.9	12.0	17.8	21.6	20.7	24.4	25.2	20.3	20.6	22.3
Chaozhou	20.0	23.0	16.4	10.3	7.9	**0.0**	12.8	18.3	22.5	20.3	25.4	25.7	21.9	20.8	23.2
Meixian	21.4	23.5	15.5	16.8	12.0	12.8	**0.0**	15.6	21.6	19.2	24.9	24.2	20.9	20.7	21.1
Nanchang	19.1	21.1	20.5	19.9	17.8	18.3	15.6	**0.0**	14.9	16.2	20.5	19.6	17.1	16.1	17.8
Changsha	21.4	22.8	24.3	23.2	21.6	22.5	21.6	14.9	**0.0**	15.8	12.4	11.2	16.8	15.7	12.1
Taiyuan	19.2	20.4	22.6	22.8	20.7	20.3	19.2	16.2	15.8	**0.0**	17.5	17.5	12.0	11.1	13.1
Beijing	25.0	25.8	27.3	26.3	24.4	25.4	24.9	20.5	12.4	17.5	**0.0**	7.9	17.6	18.7	11.2
Jinan	23.5	23.7	27.4	26.6	25.2	25.7	24.2	19.6	11.2	17.5	7.9	**0.0**	19.6	19.2	11.4
Hankou	22.8	24.4	23.5	24.5	20.3	21.9	20.9	17.1	16.8	12.0	17.6	19.6	**0.0**	7.5	14.4
Chengdu	19.9	21.7	23.3	24.5	20.6	20.8	20.7	16.1	15.7	11.1	18.7	19.2	7.5	**0.0**	15.4
Xi'an	22.5	23.3	24.5	24.8	22.3	23.2	21.1	17.8	12.1	13.1	11.2	11.4	14.4	15.4	**0.0**

Appendix 5.9a Lexical frequencies of finals (rhymes) in 15 dialects counted in the CASS database (764 items).

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
1.	-	0	0	0	1	2	0	6	29	1	0	0	0	0	0	0
2.	@	0	0	0	0	0	0	0	0	26	0	0	36	0	0	0
3.	@?	37	0	0	0	0	0	0	0	0	22	0	0	0	0	0
4.	@_~	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
5.	@i_-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
6.	@l	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.	@m	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
8.	@n	51	0	0	0	0	0	11	0	58	0	22	0	52	54	0
9.	@N	0	0	0	0	0	0	0	0	0	57	32	33	0	0	37
10.	@p	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
11.	@r	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1
12.	@t	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
13.	@u	25	0	0	0	0	0	0	0	45	19	0	0	43	17	0
14.	_i@u	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
15.	{	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.	{_~	0	0	0	0	0	0	0	0	0	42	0	0	0	0	0
17.	1?	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0
18.	1i	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0
19.	1n	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0
20.	1u	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
21.	2	27	27	0	0	0	0	0	0	0	0	0	0	0	0	0
22.	2n	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0
23.	2t	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
24.	2y	0	10	30	0	9	0	0	0	0	0	0	0	0	0	0
25.	2y?	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
26.	2yN	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
27.	3	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
28.	6i	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0
29.	6k	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0
30.	6m	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0
31.	6n	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0
32.	6N	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0
33.	6p	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0
34.	6t	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0
35.	6u	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 5.9a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
36.	7	1	1	0	0	0	0	0	0	0	12	28	0	27	0	13
37.	7u	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0
38.	8	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
39.	8?	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
40.	8n	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0
41.	9	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
42.	9k	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0
43.	9N	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0
44.	a	0	98	20	11	11	7	21	1	28	7	24	28	24	25	29
45.	A	21	0	0	0	2	0	0	16	0	0	0	0	0	0	0
46.	a?	22	0	0	1	9	6	0	23	0	47	0	0	0	0	0
47.	A?	10	0	0	0	15	0	0	7	0	0	0	0	0	0	0
48.	a_~	14	0	0	0	0	3	0	0	0	0	0	42	0	0	46
49.	A_~	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50.	a_~i	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
51.	a_~u	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
52.	ai	0	43	22	33	12	21	23	33	34	33	35	0	32	33	34
53.	Ai	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0
54.	ai?	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0
55.	aiN	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
56.	ak	0	0	9	6	0	20	9	0	0	0	0	0	0	0	0
57.	am	0	0	15	22	0	14	22	0	0	0	0	0	0	0	0
58.	an	0	0	25	18	0	0	21	30	62	0	39	0	42	43	0
59.	aN	0	56	10	5	28	34	16	0	0	0	0	22	18	19	26
60.	AN	0	0	0	0	8	0	0	11	0	0	17	0	0	0	0
61.	ao	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
62.	ap	0	0	14	15	0	8	14	0	0	0	0	0	0	0	0
63.	at	0	0	13	9	0	0	15	0	0	0	0	0	0	0	0
64.	au	0	17	8	10	8	18	17	0	18	18	1	0	18	19	18
65.	Au	0	0	0	0	10	0	0	16	0	0	19	0	0	0	0
66.	aun	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
67.	e	60	23	0	9	0	8	4	0	0	0	0	0	0	26	0
68.	E	0	0	7	0	10	0	0	0	0	0	0	29	0	0	0
69.	e?	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
70.	E?	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0

Appendix 5.9a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
71.	e_~	0	0	0	0	0	11	0	0	0	0	0	21	0	0	28
72.	ei	0	34	17	0	0	0	0	0	30	15	16	24	24	8	27
73.	Ei	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0
74.	Ei?	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0
75.	EiN	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0
76.	ek	0	0	12	0	0	26	0	0	0	0	0	0	0	0	0
77.	Ek	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
78.	em	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
79.	en	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0
80.	eN	0	23	28	0	0	44	0	0	0	0	0	0	0	0	0
81.	En	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0
82.	EN	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0
83.	ep	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
84.	et	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0
85.	eu	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0
86.	Eu	0	0	0	0	4	0	0	14	0	0	0	0	0	0	0
87.	H	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
88.	i	36	94	25	35	16	36	52	33	54	29	51	50	48	43	57
89.	i?	0	0	0	0	19	1	0	20	0	0	0	0	0	0	0
90.	i@	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0
91.	i@?	50	0	0	0	0	0	0	0	0	49	0	0	0	0	0
92.	i@N	0	0	0	0	0	0	0	0	0	43	0	0	0	0	0
93.	i@u	0	0	0	0	0	0	0	0	14	12	0	0	13	12	0
94.	i_-	0	0	0	0	0	0	0	0	20	0	24	24	0	0	12
95.	i_~	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0
96.	i_~u	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
97.	i_+	10	45	0	0	0	0	23	20	9	21	5	5	28	29	16
98.	i{	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
99.	i2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100.	ia	0	0	0	8	4	8	5	0	11	7	13	11	12	11	9
101.	iA	6	0	0	0	3	0	0	4	0	0	0	0	0	0	0
102.	ia?	1	0	0	0	0	7	0	0	0	4	0	0	0	0	0
103.	iA?	5	0	0	0	0	0	0	3	0	0	0	0	0	0	0
104.	ia_~	9	0	0	0	0	8	0	0	0	0	0	42	0	0	39
105.	iA_~	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 5.9a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
106.	iai	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1
107.	iak	0	0	0	0	0	18	2	0	0	0	0	0	0	0	0
108.	iam	0	0	0	13	0	17	13	0	0	0	0	0	0	0	0
109.	ian	0	0	0	23	0	0	15	0	10	0	42	0	0	38	0
110.	iaN	0	0	0	0	3	26	3	0	0	0	0	11	10	10	10
111.	iAN	0	0	0	0	0	0	0	8	0	0	11	0	0	0	0
112.	iap	0	0	0	12	0	13	10	0	0	0	0	0	0	0	0
113.	iat	0	0	0	17	0	0	13	0	0	0	0	0	0	0	0
114.	iau	0	0	0	13	0	14	14	0	12	13	0	0	13	11	13
115.	iAu	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0
116.	iau?	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
117.	ie	0	0	0	0	7	0	0	0	30	46	0	0	30	32	0
118.	iE	0	30	0	0	11	0	0	0	0	0	29	6	0	0	31
119.	ie?	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0
120.	iE?	0	0	0	0	17	0	0	24	0	0	0	0	0	0	0
121.	ie_~	0	0	0	0	0	0	0	0	35	0	0	15	0	0	16
122.	ien	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0
123.	ieN	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0
124.	iEn	0	0	0	0	0	0	0	37	0	0	0	0	41	0	0
125.	iEN	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0
126.	iet	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
127.	ieu	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0
128.	iEu	0	0	0	0	8	0	0	13	0	0	0	0	0	0	0
129.	iI	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0
130.	iin	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
131.	ik	0	0	0	37	0	11	0	0	0	0	0	0	0	0	0
132.	im	0	0	14	13	0	11	8	0	0	0	0	0	0	0	0
133.	in	45	0	22	13	0	0	26	36	44	0	15	0	43	40	0
134.	iN	0	0	0	63	46	17	0	0	0	0	29	29	0	0	28
135.	io	1	0	0	0	0	3	3	0	7	0	0	0	8	0	0
136.	iO	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0
137.	io?	9	0	0	0	0	6	0	0	0	0	0	0	0	0	0
138.	iO?	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
139.	io_~	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
140.	iok	0	0	0	1	0	2	6	0	0	0	0	0	0	0	0

Appendix 5.9a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
141.	iOk	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0
142.	ion	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
143.	ioN	10	0	0	0	0	7	10	0	8	0	0	0	8	0	0
144.	iON	0	0	0	32	0	0	0	10	0	0	0	0	0	0	0
145.	iou	0	0	0	0	0	0	0	0	0	0	12	12	0	0	13
146.	ip	0	0	13	8	0	7	6	0	0	0	0	0	0	0	0
147.	iQ_~	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0
148.	it	0	0	15	8	0	0	18	0	0	0	0	0	0	0	0
149.	iu	0	0	15	18	0	12	12	9	0	0	0	0	0	0	0
150.	iu?	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0
151.	iuk	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0
152.	iun	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0
153.	iuN	0	0	0	0	0	0	15	9	0	0	0	0	0	0	0
154.	iY	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
155.	M	0	0	0	0	0	13	0	0	0	0	0	0	2	0	0
156.	MN	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0
157.	N	0	3	0	0	0	2	0	0	0	0	0	0	0	0	0
158.	o	22	55	0	20	12	18	15	13	47	0	5	0	5	41	17
159.	O	0	0	18	29	3	0	0	0	0	0	0	18	0	0	0
160.	o?	37	0	0	0	4	3	0	0	0	0	0	0	0	0	0
161.	O?	0	0	0	0	4	0	0	20	0	0	0	0	0	0	0
162.	o_~	0	0	0	0	0	1	0	0	9	0	0	0	0	0	0
163.	o_~i	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
164.	o_~u	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
165.	oi	0	0	0	0	0	8	15	0	0	0	0	0	0	0	0
166.	Oi	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0
167.	ok	0	0	27	1	0	18	20	0	0	0	0	0	0	0	0
168.	Ok	0	0	18	24	0	0	0	0	0	0	0	0	0	0	0
169.	on	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0
170.	oN	30	47	34	0	0	21	27	0	32	0	0	0	33	37	0
171.	On	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
172.	ON	0	0	27	42	0	0	0	26	0	0	0	0	0	0	0
173.	ot	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0
174.	Ot	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
175.	ou	0	1	17	0	9	10	0	0	0	0	20	18	0	0	23

Appendix 5.9a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
176.	ou?	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0
177.	Ou?	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
178.	ouN	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0
179.	OuN	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
180.	Oy	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0
181.	Oy?	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
182.	OyN	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
183.	Q_~	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0
184.	tM	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
185.	u	7	24	6	30	10	15	25	18	17	26	51	50	21	51	44
186.	u?	0	0	0	0	8	0	0	15	0	0	0	0	0	0	0
187.	u@	0	0	0	0	0	0	0	0	1	0	0	27	0	0	0
188.	u@?	7	0	0	0	0	0	0	0	0	20	0	0	0	0	0
189.	u@n	4	0	0	0	0	0	0	0	4	0	17	0	7	14	0
190.	u@N	0	0	0	0	0	0	0	0	0	41	1	0	0	0	0
191.	u_~i	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
192.	u{_~	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0
193.	u1?	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
194.	u2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
195.	u7	0	0	0	0	0	0	0	0	0	7	0	0	2	0	0
196.	u8?	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
197.	u8n	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
198.	ua	0	0	0	8	4	11	3	0	7	7	13	10	13	13	10
199.	uA	2	0	0	0	4	0	0	5	0	0	0	0	0	0	0
200.	ua?	3	0	0	0	4	6	0	4	0	16	0	0	0	0	0
201.	uA?	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
202.	ua_~	2	0	0	0	0	7	0	0	0	0	0	14	0	0	14
203.	uA_~	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
204.	uai	0	0	0	3	2	4	2	4	3	6	6	0	7	7	4
205.	uAi	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
206.	uak	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0
207.	uan	0	0	0	26	0	0	1	5	10	0	17	0	13	18	0
208.	uaN	0	0	0	0	12	29	0	0	0	0	0	10	15	14	7
209.	uAN	0	0	0	0	9	0	0	1	0	0	15	0	0	0	0
210.	uat	0	0	0	18	0	0	1	0	0	0	0	0	0	0	0

Appendix 5.9a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
211.	ue	19	0	0	13	0	16	0	0	0	0	0	0	0	5	0
212.	uE	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
213.	ue?	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
214.	uE?	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
215.	ue_~	0	0	0	0	0	2	0	0	0	0	0	16	0	0	10
216.	ue_~?	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
217.	uei	0	0	0	0	0	0	0	0	9	27	29	29	19	34	29
218.	uet	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
219.	ui	0	0	4	29	0	26	23	28	0	0	0	0	0	0	0
220.	uk	0	0	0	0	0	10	13	0	0	0	0	0	0	0	0
221.	un	0	0	6	33	0	0	21	13	0	0	0	0	0	0	0
222.	uN	0	0	0	0	24	26	21	27	0	0	29	28	0	0	25
223.	uo	0	1	0	0	2	0	0	2	0	0	29	0	35	0	23
224.	uO	0	29	0	0	5	0	0	0	0	0	0	0	0	0	0
225.	uo?	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0
226.	uO?	0	0	0	0	10	0	0	1	0	0	0	0	0	0	0
227.	uoi	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0
228.	uOi	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0
229.	uok	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
230.	uon	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
231.	uoN	0	0	0	0	20	0	1	0	0	0	0	0	0	0	0
232.	uON	0	0	0	0	5	0	0	7	0	0	0	0	0	0	0
233.	uQ_~	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
234.	ut	0	0	4	11	0	0	11	0	0	0	0	0	0	0	0
235.	xiEn	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
236.	y	7	43	9	0	15	0	0	13	22	11	21	23	30	15	22
237.	Y	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0
238.	y?	0	0	0	0	5	0	0	3	0	0	0	0	0	0	0
239.	y@	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0
240.	y@?	8	0	0	0	0	0	0	0	0	27	0	0	0	0	0
241.	y@n	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
242.	y@N	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0
243.	y8?	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0
244.	y8n	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
245.	ya	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0

Appendix 5.9a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
246.	ya_~	0	0	0	0	0	0	0	0	0	0	0	4	0	0	3
247.	yai	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
248.	yan	0	0	0	0	0	0	0	0	7	0	4	0	0	3	0
249.	ye	0	0	0	0	0	0	0	1	6	6	0	0	6	9	0
250.	yE	0	0	0	0	0	0	0	0	0	0	15	0	0	0	7
251.	ye?	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
252.	ye_~	0	0	0	0	0	0	0	0	7	0	0	8	0	0	5
253.	yei	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
254.	yEn	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0
255.	yn	0	0	16	0	0	0	0	7	12	0	6	0	14	10	0
256.	yN	0	0	0	0	23	0	0	0	0	0	7	8	0	0	9
257.	yo	0	0	0	0	0	0	0	0	0	0	0	0	0	16	7
258.	yO	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0
259.	yo?	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
260.	yO?	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
261.	yoN	0	0	0	0	6	0	0	0	0	0	0	0	0	6	0
262.	yON	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
263.	yt	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 5.9b Proximity matrix derived from Appendix 5.9a. (lexical frequency of Finals)

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.0**	33.0	36.3	35.9	36.4	34.3	34.1	34.0	31.1	32.8	35.3	36.1	31.7	30.4	35.3
Wenzhou	33.0	**0.0**	31.8	33.6	33.4	31.3	30.7	34.2	28.7	33.4	31.1	32.0	28.6	27.0	27.6
Guangzhou	36.3	31.8	**0.0**	29.3	35.6	30.9	29.3	33.8	32.1	36.2	33.6	34.7	31.9	32.2	34.3
Xiamen	35.9	33.6	29.3	**0.0**	34.4	28.4	25.4	31.4	31.6	35.1	30.8	34.1	32.4	29.7	32.4
Fuzhou	36.4	33.4	35.6	34.4	**0.0**	32.8	34.9	31.8	35.2	35.6	31.9	32.4	34.2	33.9	31.2
Chaozhou	34.3	31.3	30.9	28.4	32.8	**0.0**	26.6	34.3	33.2	34.7	33.6	31.7	31.8	30.9	29.2
Meixian	34.1	30.7	29.3	25.4	34.9	26.6	**0.0**	31.1	29.0	34.6	31.2	34.3	28.9	27.5	31.6
Nanchang	34.0	34.2	33.8	31.4	31.8	34.3	31.1	**0.0**	32.2	34.7	31.1	35.9	31.1	32.0	34.0
Changsha	31.1	28.7	32.1	31.6	35.2	33.2	29.0	32.2	**0.0**	30.6	26.6	31.2	18.7	18.2	29.2
Taiyuan	32.8	33.4	36.2	35.1	35.6	34.7	34.6	34.7	30.6	**0.0**	31.5	33.3	29.1	29.8	30.3
Beijing	35.3	31.1	33.6	30.8	31.9	33.6	31.2	31.1	26.6	31.5	**0.0**	27.4	25.6	25.3	22.6
Jinan	36.1	32.0	34.7	34.1	32.4	31.7	34.3	35.9	31.2	33.3	27.4	**0.0**	32.0	31.8	18.5
Hankou	31.7	28.6	31.9	32.4	34.2	31.8	28.9	31.1	18.7	29.1	25.6	32.0	**0.0**	17.3	27.5
Chengdu	30.4	27.0	32.2	29.7	33.9	30.9	27.5	32.0	18.2	29.8	25.3	31.8	17.3	**0.0**	28.0
Xi'an	35.3	27.6	34.3	32.4	31.2	29.2	31.6	34.0	29.2	30.3	22.6	18.5	27.5	28.0	**0.0**

Appendix 5.10a Lexical frequencies of codas in 15 dialects counted in the CASS database (764 items).

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
1.	0	429	635	279	269	271	314	273	258	515	422	460	622	461	457	621
2.	?	189	0	0	1	188	43	0	182	0	185	0	0	0	0	0
3.	k	0	0	90	92	0	115	64	0	0	0	0	0	0	0	0
4.	l	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5.	m	0	0	44	48	0	42	46	0	0	0	0	0	0	0	0
6.	n	105	0	117	113	0	0	160	196	208	0	162	0	219	220	0
7.	N	40	129	143	142	303	222	93	99	40	156	141	141	84	86	142
8.	n`	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
9.	p	0	0	30	35	0	28	32	0	0	0	0	0	0	0	0
10.	r	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1
11.	t	0	0	61	63	0	0	89	0	0	0	0	0	0	0	0

Appendix 5.10b Proximity matrix derived from Appendix 5.10a. (lexical frequency of Codas)

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.0**	19.5	22.5	22.8	18.5	23.0	22.1	9.8	15.2	13.3	16.1	19.7	15.7	15.8	19.7
Wenzhou	19.5	**0.0**	21.9	22.4	20.1	19.0	23.3	23.3	16.4	15.2	13.8	1.3	16.0	16.1	1.3
Guangzhou	22.5	21.9	**0.0**	1.4	24.2	15.6	5.3	20.9	19.5	23.6	16.7	21.8	17.8	17.7	21.7
Xiamen	22.8	22.4	1.4	**0.0**	24.4	15.5	5.5	21.1	20.0	23.8	17.3	22.2	18.4	18.4	22.2
Fuzhou	18.5	20.1	24.2	24.4	**0.0**	17.1	26.3	17.3	26.0	9.0	21.4	19.6	24.1	24.0	19.6
Chaozhou	23.0	19.0	15.6	15.5	17.1	**0.0**	19.1	22.8	24.7	17.4	20.9	18.7	23.2	23.1	18.6
Meixian	22.1	23.3	5.3	5.5	26.3	19.1	**0.0**	20.4	18.3	25.0	17.0	23.2	17.0	17.0	23.2
Nanchang	9.8	23.3	20.9	21.1	17.3	22.8	20.4	**0.0**	17.1	15.8	15.8	23.2	15.5	15.5	23.2
Changsha	15.2	16.4	19.5	20.0	26.0	24.7	18.3	17.1	**0.0**	21.7	8.1	16.6	4.3	4.6	16.6
Taiyuan	13.3	15.2	23.6	23.8	9.0	17.4	25.0	15.8	21.7	**0.0**	18.7	15.0	20.7	20.7	15.0
Beijing	16.1	13.8	16.7	17.3	21.4	20.9	17.0	15.8	8.1	18.7	**0.0**	13.6	4.9	4.7	13.6
Jinan	19.7	1.3	21.8	22.2	19.6	18.7	23.2	23.2	16.6	15.0	13.6	**0.0**	16.1	16.1	0.1
Hankou	15.7	16.0	17.8	18.4	24.1	23.2	17.0	15.5	4.3	20.7	4.9	16.1	**0.0**	1.0	16.1
Chengdu	15.8	16.1	17.7	18.4	24.0	23.1	17.0	15.5	4.6	20.7	4.7	16.1	1.0	**0.0**	16.1
Xi'an	19.7	1.3	21.7	22.2	19.6	18.6	23.2	23.2	16.6	15.0	13.6	0.1	16.1	16.1	**0.0**

Appendix 5.11a Lexical frequencies of tones in 15 dialects counted in the CASS database (764 items).

#	Tones	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
1.	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2
2.	1	0	0	0	0	0	0	109	0	0	0	0	0	0	0	0
3.	11	0	0	0	3	0	0	153	0	5	316	0	0	0	0	0
4.	13	149	0	0	0	0	0	0	0	149	0	0	0	0	188	0
5.	2	0	0	81	0	0	104	0	34	0	151	0	1	0	0	0
6.	21	0	0	151	67	0	0	0	115	0	0	0	233	0	0	300
7.	212	0	0	0	0	74	0	0	0	0	0	0	0	0	0	0
8.	213	0	191	0	0	0	78	0	100	0	0	0	251	337	0	0
9.	214	0	0	0	0	0	0	0	0	0	0	95	0	0	0	0
10.	22	0	86	105	116	0	0	0	0	0	0	0	0	0	0	0
11.	23	0	0	37	0	104	0	0	0	0	0	0	0	0	0	0
12.	24	0	0	0	0	0	0	0	51	188	0	0	0	0	0	192
13.	242	0	0	0	0	115	0	0	0	0	0	0	0	0	0	0
14.	3	84	0	66	0	0	0	0	0	0	0	0	0	0	0	0
15.	31	138	150	0	0	0	41	67	0	0	0	0	0	0	332	0
16.	32	0	0	0	102	78	0	0	0	0	0	0	0	0	0	0
17.	33	0	158	64	0	0	157	0	0	159	0	0	0	0	0	0
18.	35	0	112	61	152	0	72	0	0	0	0	207	0	184	0	0
19.	41	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0
20.	42	0	67	0	0	0	0	0	156	0	0	0	194	87	0	0
21.	44	0	0	0	0	158	0	180	0	0	0	0	0	0	0	179
22.	45	0	0	0	0	0	0	0	130	0	182	0	0	0	0	0
23.	5	105	0	42	86	84	83	77	148	0	0	0	0	0	0	0
24.	51	57	0	0	0	0	0	0	0	0	0	256	0	0	0	0
25.	513	68	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26.	53	0	0	147	82	149	77	172	0	0	81	0	0	0	83	91
27.	54	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0
28.	55	163	0	10	155	0	152	0	0	178	0	206	85	156	161	0

Appendix 5.11b Proximity matrix derived from Appendix 5.11a (lexical frequency of tones).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.0**	35.2	35.0	31.9	36.6	30.0	34.1	35.4	29.8	39.1	31.3	36.1	34.8	21.9	39.1
Wenzhou	35.2	**0.0**	31.6	32.7	39.1	26.2	36.6	32.4	34.8	39.1	35.2	29.9	24.7	33.4	39.1
Guangzhou	35.0	31.6	**0.0**	24.2	31.8	26.1	33.1	31.8	36.1	33.3	36.1	33.5	36.2	35.8	30.0
Xiamen	31.9	32.7	24.2	**0.0**	31.1	26.4	33.5	33.3	34.5	36.8	28.7	33.2	29.7	32.3	33.7
Fuzhou	36.6	39.1	31.8	31.1	**0.0**	34.0	26.5	35.8	39.1	36.3	39.1	39.1	39.1	36.2	31.1
Chaozhou	30.0	26.2	26.1	26.4	34.0	**0.0**	32.4	31.8	29.7	33.5	31.1	32.7	27.6	30.2	36.9
Meixian	34.1	36.6	33.1	33.5	26.5	32.4	**0.0**	35.9	38.8	29.8	39.0	39.0	39.0	32.8	30.4
Nanchang	35.4	32.4	31.8	33.3	35.8	31.8	35.9	**0.0**	36.6	32.9	38.7	23.6	31.1	38.7	31.7
Changsha	29.8	34.8	36.1	34.5	39.1	29.7	38.8	36.6	**0.0**	38.8	33.8	36.0	34.6	29.3	33.9
Taiyuan	39.1	39.1	33.3	36.8	36.3	33.5	29.8	32.9	38.8	**0.0**	39.1	39.0	39.1	36.9	36.8
Beijing	31.3	35.2	36.1	28.7	39.1	31.1	39.0	38.7	33.8	39.1	**0.0**	35.9	28.0	34.2	39.1
Jinan	36.1	29.9	33.5	33.2	39.1	32.7	39.0	23.6	36.0	39.0	35.9	**0.0**	22.2	36.1	31.7
Hankou	34.8	24.7	36.2	29.7	39.1	27.6	39.0	31.1	34.6	39.1	28.0	22.2	**0.0**	34.8	39.1
Chengdu	21.9	33.4	35.8	32.3	36.2	30.2	32.8	38.7	29.3	36.9	34.2	36.1	34.8	**0.0**	36.8
Xi'an	39.1	39.1	30.0	33.7	31.1	36.9	30.4	31.7	33.9	36.8	39.1	31.7	39.1	36.8	**0.0**

Appendix 5.12a Lexical frequencies of vocalic nuclei in 15 dialects counted in the CASS database (764 items).

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
1.	@	89	0	0	0	0	0	18	0	84	79	54	69	52	54	37
2.	@_~	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
3.	@i_-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
4.	@r	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1
5.	@u	25	0	0	0	0	0	0	0	45	19	0	0	43	17	0
6.	_~	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0
7.	_i@u	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
8.	{	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.	{_~	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0
10.	1	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0
11.	1i	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0
12.	1u	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
13.	2	27	27	18	0	0	0	0	0	0	0	0	0	0	0	0
14.	2y	0	10	30	0	27	0	0	0	0	0	0	0	0	0	0
15.	3	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
16.	6	0	0	97	0	0	0	0	0	0	0	0	0	0	0	0
17.	6i	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0
18.	6u	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0
19.	7	1	1	0	0	0	0	0	0	0	12	28	0	27	0	13
20.	7u	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0
21.	8	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0
22.	9	0	0	26	0	1	0	0	0	0	0	0	0	0	0	0
23.	a	22	154	106	87	48	89	118	54	90	54	63	50	84	87	55
24.	A	31	0	0	0	25	0	0	34	0	0	17	0	0	0	0
25.	a_~	14	0	0	0	0	3	0	0	0	0	0	42	0	0	46
26.	A_~	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27.	a_~i	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
28.	a_~u	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
29.	ai	0	43	22	33	35	21	23	33	34	33	35	0	32	33	34
30.	Ai	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0
31.	ao	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
32.	au	0	17	8	10	8	18	17	0	19	18	1	0	18	19	18
33.	Au	0	0	0	0	10	0	0	16	0	0	19	0	0	0	0
34.	e	60	46	40	9	0	84	44	0	0	0	0	0	0	26	0
35.	E	0	0	22	0	10	0	0	34	0	0	0	29	0	0	0

Appendix 5.12a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
36.	e_~	0	0	0	0	0	11	0	0	0	0	0	21	0	0	28
37.	ei	0	34	17	0	0	0	0	0	30	15	16	24	24	8	27
38.	Ei	0	0	0	0	59	0	0	0	0	0	0	0	0	0	0
39.	eu	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0
40.	Eu	0	0	0	0	4	0	0	14	0	0	0	0	0	0	0
41.	H	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42.	i	81	94	89	177	81	83	110	89	98	29	95	79	91	83	85
43.	i@	50	0	0	0	0	0	0	0	0	92	0	26	0	0	0
44.	i@u	0	0	0	0	0	0	0	0	14	12	0	0	13	12	0
45.	i_-	0	0	0	0	0	0	0	0	20	0	24	24	0	0	12
46.	i_~	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0
47.	i_~u	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
48.	i_+	10	45	0	0	0	0	23	20	9	21	5	5	28	29	16
49.	i{	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50.	i2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51.	ia	1	0	0	73	7	89	61	0	21	11	55	22	22	59	19
52.	iA	11	0	0	0	3	0	0	15	0	0	11	0	0	0	0
53.	ia_~	9	0	0	0	0	8	0	0	0	0	0	42	0	0	39
54.	iA_~	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
55.	iai	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1
56.	iau	0	0	0	13	0	15	14	0	12	13	0	0	13	11	13
57.	iAu	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0
58.	ie	0	0	0	0	41	0	19	0	30	46	0	0	30	32	0
59.	iE	0	30	0	0	35	0	0	61	0	0	29	6	41	0	31
60.	ie_~	0	0	0	0	0	0	0	0	35	0	0	15	0	0	16
61.	ieu	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0
62.	iEu	0	0	0	0	8	0	0	13	0	0	0	0	0	0	0
63.	ii	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
64.	iI	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65.	io	20	0	0	1	0	18	20	0	15	0	0	0	16	0	0
66.	iO	0	0	0	55	0	0	0	14	0	0	0	13	0	0	0
67.	io_~	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
68.	iou	0	0	0	0	0	0	0	0	0	0	12	12	0	0	13
69.	iQ_~	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0
70.	iu	0	0	15	18	0	12	52	29	0	0	0	0	0	0	0

Appendix 9.12a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
71.	iY	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
72.	M	0	0	0	0	0	30	0	0	0	0	0	0	2	0	0
73.	o	89	102	61	21	16	60	82	13	79	0	5	0	38	78	17
74.	O	0	0	70	95	7	0	0	46	0	0	0	18	0	0	0
75.	o_~	0	0	0	0	0	1	0	0	9	0	0	0	0	0	0
76.	o_~i	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
77.	o_~u	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
78.	oi	0	0	0	0	0	8	15	0	0	0	0	0	0	0	0
79.	Oi	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0
80.	ou	0	1	17	0	49	10	0	0	0	0	20	18	0	0	23
81.	Ou	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0
82.	Oy	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0
83.	Q_~	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0
84.	u	7	24	16	74	42	51	91	73	17	26	80	78	21	51	69
85.	u@	11	0	0	0	0	0	0	0	5	61	18	27	7	14	0
86.	u_~i	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
87.	u{_~	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0
88.	u1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
89.	u2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90.	u7	0	0	0	0	0	0	0	0	0	7	0	0	2	0	0
91.	u8	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
92.	ua	3	0	0	52	20	56	5	9	17	23	30	20	41	45	17
93.	uA	2	0	0	0	17	0	0	6	0	0	15	0	0	0	0
94.	ua_~	2	0	0	0	0	7	0	0	0	0	0	14	0	0	14
95.	uA_~	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
96.	uai	0	0	0	3	2	4	2	4	3	6	6	0	7	7	4
97.	uAi	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
98.	ue	19	0	0	13	0	22	1	0	0	0	0	0	0	5	0
99.	uE	0	0	0	0	0	0	0	1	0	0	0	6	0	0	0
100.	ue_~	0	0	0	0	0	3	0	0	0	0	0	16	0	0	10
101.	uei	0	0	0	0	0	0	0	0	9	27	29	29	19	34	29
102.	ui	0	0	4	29	0	26	23	28	0	0	0	0	0	0	0
103.	uo	0	1	0	0	34	0	3	2	0	0	29	0	35	0	23
104.	uO	0	29	0	0	20	0	0	8	0	0	0	0	0	0	0
105.	uoi	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0

Appendix 5.12a (continued)

#	X-SAMPA	dialects														
		Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
106.	uOi	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0
107.	uQ_~	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
108.	xiE	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
109.	y	7	43	34	0	43	0	0	23	34	11	34	31	44	25	31
110.	Y	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111.	y@	13	0	0	0	0	0	0	0	0	42	0	15	0	0	0
112.	y8	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0
113.	ya	0	0	0	0	0	0	0	0	9	0	4	0	0	3	0
114.	ya_~	0	0	0	0	0	0	0	0	0	0	0	4	0	0	3
115.	yai	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
116.	ye	0	0	0	0	0	0	0	2	6	6	0	0	6	9	0
117.	yE	0	0	0	0	0	0	0	0	0	0	15	0	5	0	7
118.	ye_~	0	0	0	0	0	0	0	0	7	0	0	8	0	0	5
119.	yei	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
120.	yo	0	0	0	0	8	0	0	0	0	0	0	0	0	22	7
121.	yO	0	16	0	0	6	0	0	0	0	0	0	0	0	0	0

Appendix 5.12b Proximity matrix derived from Appendix 5.12a (lexical frequency of nuclei).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.0**	29.1	31.0	32.6	33.4	29.3	28.6	32.3	27.4	29.6	30.9	29.5	29.0	27.4	31.1
Wenzhou	29.1	**0.0**	23.2	28.0	27.1	26.3	23.3	27.1	24.6	31.3	27.8	31.0	23.9	23.4	26.3
Guangzhou	31.0	23.2	**0.0**	26.0	28.6	27.4	26.4	27.5	28.1	33.4	30.3	30.4	29.0	27.6	29.7
Xiamen	32.6	28.0	26.0	**0.0**	28.9	20.2	20.4	24.0	27.0	31.5	26.1	28.3	26.6	23.4	27.3
Fuzhou	33.4	27.1	28.6	28.9	**0.0**	29.9	29.0	24.8	28.7	31.6	24.4	30.0	25.7	27.5	26.4
Chaozhou	29.3	26.3	27.4	20.2	29.9	**0.0**	17.8	29.1	26.5	31.7	27.4	29.4	26.4	22.0	26.2
Meixian	28.6	23.3	26.4	20.4	29.0	17.8	**0.0**	26.4	22.9	29.5	26.2	29.6	23.5	19.0	25.9
Nanchang	32.3	27.1	27.5	24.0	24.8	29.1	26.4	**0.0**	29.9	32.4	25.4	28.6	27.6	28.7	27.7
Changsha	27.4	24.6	28.1	27.0	28.7	26.5	22.9	29.9	**0.0**	24.2	24.0	25.7	16.0	16.4	22.9
Taiyuan	29.6	31.3	33.4	31.5	31.6	31.7	29.5	32.4	24.2	**0.0**	26.8	26.4	22.7	22.7	27.6
Beijing	30.9	27.8	30.3	26.1	24.4	27.4	26.2	25.4	24.0	26.8	**0.0**	22.5	19.8	22.0	18.7
Jinan	29.5	31.0	30.4	28.3	30.0	29.4	29.6	28.6	25.7	26.4	22.5	**0.0**	26.9	26.0	17.7
Hankou	29.0	23.9	29.0	26.6	25.7	26.4	23.5	27.6	16.0	22.7	19.8	26.9	**0.0**	16.2	20.3
Chengdu	27.4	23.4	27.6	23.4	27.5	22.0	19.0	28.7	16.4	22.7	22.0	26.0	16.2	**0.0**	22.6
Xi'an	31.1	26.3	29.7	27.3	26.4	26.2	25.9	27.7	22.9	27.6	18.7	17.7	20.3	22.6	**0.0**

Appendix 5.13a Lexical frequencies of union of initials and finals in 15 dialects counted in the CASS database (764 items). This is the concatenation of Appendices 5.8a and 5.9a. These tables are not reproduced here.

Appendix 5.13b Proximity matrix derived from Appendix 5.13a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.0**	35.4	43.6	42.6	42.6	39.7	40.1	38.4	37.7	38.0	43.2	43.1	39.0	36.3	41.9
Wenzhou	35.4	**0.0**	38.7	42.2	41.5	38.8	38.5	39.5	36.6	39.2	40.4	39.8	37.6	34.7	36.1
Guangzhou	43.6	38.7	**0.0**	34.9	39.1	35.0	33.0	38.8	40.2	42.6	43.3	44.2	39.6	39.7	42.2
Xiamen	42.6	42.2	34.9	**0.0**	36.4	30.2	30.4	36.7	39.2	41.8	40.5	43.2	40.6	38.5	40.8
Fuzhou	42.6	41.5	39.1	36.4	**0.0**	33.7	36.9	36.1	41.3	41.1	40.1	41.0	39.8	39.6	38.3
Chaozhou	39.7	38.8	35.0	30.2	33.7	**0.0**	29.4	38.4	40.1	40.2	42.1	40.9	38.6	37.3	37.3
Meixian	40.1	38.5	33.0	30.4	36.9	29.4	**0.0**	34.4	36.1	39.5	39.9	41.9	35.7	34.4	37.9
Nanchang	38.4	39.5	38.8	36.7	36.1	38.4	34.4	**0.0**	35.1	37.8	36.9	40.5	35.1	35.3	38.0
Changsha	37.7	36.6	40.2	39.2	41.3	40.1	36.1	35.1	**0.0**	34.4	29.4	33.2	25.1	24.0	31.6
Taiyuan	38.0	39.2	42.6	41.8	41.1	40.2	39.5	37.8	34.4	**0.0**	36.0	37.6	31.5	31.8	33.0
Beijing	43.2	40.4	43.3	40.5	40.1	42.1	39.9	36.9	29.4	36.0	**0.0**	28.5	31.1	31.4	25.3
Jinan	43.1	39.8	44.2	43.2	41.0	40.9	41.9	40.5	33.2	37.6	28.5	**0.0**	37.5	37.1	21.7
Hankou	39.0	37.6	39.6	40.6	39.8	38.6	35.7	35.1	25.1	31.5	31.1	37.5	**0.0**	18.9	31.0
Chengdu	36.3	34.7	39.7	38.5	39.6	37.3	34.4	35.3	24.0	31.8	31.4	37.1	18.9	**0.0**	32.0
Xi'an	41.9	36.1	42.2	40.8	38.3	37.3	37.9	38.0	31.6	33.0	25.3	21.7	31.0	32.0	**0.0**

Appendix 5.14a Lexical frequencies of union of initials, finals and tones in 15 dialects counted in the CASS database (764 items). This is the concatenation of Appendices 5.8a, 5.9a and 5.11a. These tables are not reproduced here.

Appendix 5.14b Proximity matrix derived from Appendix 5.14a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.0**	50.0	55.9	53.2	56.2	49.8	52.7	52.2	48.1	54.5	53.3	56.2	52.3	42.4	57.3
Wenzhou	50.0	**0.0**	50.0	53.4	57.0	46.9	53.1	51.0	50.5	55.3	53.6	49.8	45.0	48.1	53.2
Guangzhou	55.9	50.0	**0.0**	42.5	50.4	43.7	46.7	50.1	54.1	54.1	56.4	55.5	53.6	53.5	51.8
Xiamen	53.2	53.4	42.5	**0.0**	47.9	40.1	45.3	49.6	52.2	55.7	49.6	54.5	50.3	50.3	52.9
Fuzhou	56.2	57.0	50.4	47.9	**0.0**	47.9	45.5	50.8	56.8	54.8	56.0	56.6	55.8	53.7	49.3
Chaozhou	49.8	46.9	43.7	40.1	47.9	**0.0**	43.7	49.8	49.9	52.4	52.3	52.3	47.4	48.0	52.4
Meixian	52.7	53.1	46.7	45.3	45.5	43.7	**0.0**	49.7	53.0	49.5	55.8	57.3	52.9	47.5	48.6
Nanchang	52.2	51.0	50.1	49.6	50.8	49.8	49.7	**0.0**	50.7	50.1	53.5	46.8	46.9	52.4	49.4
Changsha	48.1	50.5	54.1	52.2	56.8	49.9	53.0	50.7	**0.0**	51.9	44.8	49.0	42.7	37.9	46.3
Taiyuan	54.5	55.3	54.1	55.7	54.8	52.4	49.5	50.1	51.9	**0.0**	53.1	54.2	50.2	48.7	49.4
Beijing	53.3	53.6	56.4	49.6	56.0	52.3	55.8	53.5	44.8	53.1	**0.0**	45.8	41.8	46.4	46.5
Jinan	56.2	49.8	55.5	54.5	56.6	52.3	57.3	46.8	49.0	54.2	45.8	**0.0**	43.6	51.8	38.4
Hankou	52.3	45.0	53.6	50.3	55.8	47.4	52.9	46.9	42.7	50.2	41.8	43.6	**0.0**	39.6	49.9
Chengdu	42.4	48.1	53.5	50.3	53.7	48.0	47.5	52.4	37.9	48.7	46.4	51.8	39.6	**0.0**	48.7
Xi'an	57.3	53.2	51.8	52.9	49.3	52.4	48.6	49.4	46.3	49.4	46.5	38.4	49.9	48.7	**0.0**

Appendix 5.15a Vowel feature table for LO4. This table specifies all 28 IPA reference vowels with unique feature combinations. Vowels may differ in backness (three degrees: i.e. front, central, back), in rounding (two degrees: spread, rounded) and in height (seven degrees, which are more refined than the traditional four-level division into high, high-mid, low-mid, low). Note that the 1 and 0 specifications are incrementally ordered (unlike the standard use of features which allows any combination of '0' and '1' specifications). This is most clearly seen in the height specification, where seven ones code high vowels, six ones the next degree and so on, until all six height features are 0, indicating a fully low vowel.

		#	Vowels		L04 configuration features								
			IPA	X-SAMPA	frontA	frontB	closeA	closeB	closeC	closeD	closeE	closeF	round
Front	hi	1.	i	i	1	1	1	1	1	1	1	1	0
		2.	y	y	1	1	1	1	1	1	1	1	1
	hi mid	3.	ɪ	I	1	1	1	1	1	1	1	0	0
		4.	ʏ	Y	1	1	1	1	1	1	1	0	1
		5.	e	e	1	1	1	1	1	1	0	0	0
		6.	ø	2	1	1	1	1	1	1	0	0	1
	lo mid	7.	ɛ	E	1	1	1	1	0	0	0	0	0
		8.	œ	9	1	1	1	1	0	0	0	0	1
		9.	æ	{	1	1	1	0	0	0	0	0	0
	lo	10.	a	a	1	1	0	0	0	0	0	0	0
		11.	ɶ	&	1	1	0	0	0	0	0	0	1
Central	hi	12.	ɨ	1	1	0	1	1	1	1	1	1	0
		13.	ʉ	}	1	0	1	1	1	1	1	1	1
	hi mid	14.	ɵ	8	1	0	1	1	1	1	0	0	1
		15.	ə	@	1	0	1	1	1	0	0	0	0
		16.	ɘ	@\	1	0	1	1	1	1	0	0	0
	lo mid	17.	ɜ	3	1	0	1	1	0	0	0	0	0
		18.	ɞ	3\	1	0	1	1	0	0	0	0	1
		19.	ɐ	6	1	0	1	0	0	0	0	0	0
Back	hi	20.	ɯ	M	0	0	1	1	1	1	1	1	0
		21.	u	u	0	0	1	1	1	1	1	1	1
	hi mid	22.	ʊ	U	0	0	1	1	1	1	1	0	0
		23.	ɤ	7	0	0	1	1	1	1	0	0	0
		24.	o	o	0	0	1	1	1	1	0	0	1
	lo mid	25.	ʌ	V	0	0	1	1	0	0	0	0	0
		26.	ɔ	O	0	0	1	1	0	0	0	0	1
	lo	27.	ɑ	A	0	0	0	0	0	0	0	0	0
		28.	ɒ	Q	0	0	0	0	0	0	0	0	1

Appendix 5.15b Consonant feature table used in L04 as example configuration file. The table specifies 74 IPA consonants (seven retroflex consonants were added, as were entries for affricate, increasing the number of manners to ten).

				L04 configuration features					
#	IPA	X-SAMPA	Place	PlaceA	PlaceB	PlaceC	Pulmonic	Voiceless	Manner
1.	p	p	bilabial	1	1	1	1	1	plosive
2.	b	b	bilabial	1	1	1	1	0	plosive
3.	m	m	bilabial	1	1	1	1	0	nasal
4.	ʙ	B\	bilabial	1	1	1	1	0	trill
5.	ɸ	p\	bilabial	1	1	1	1	1	fricative
6.	β	B	bilabial	1	1	1	1	0	fricative
7.	ʘ	O\	bilabial	1	1	1	0	0	click
8.	ɱ	F	labio-dental	1	1	1	1	0	nasal
9.	f	f	labio-dental	1	1	1	1	1	fricative
10.	v	v	labio-dental	1	1	1	1	0	fricative
11.	ʋ	v\	labio-dental	1	1	1	1	0	approximant
12.	t	t	alveolar	1	1	0	1	1	plosive
13.	ts	ts	alveolar	1	1	0	1	1	affricate
14.	d	d	alveolar	1	1	0	1	0	plosive
15.	dz	dz	alveolar	1	1	0	1	0	affricate
16.	n	n	alveolar	1	1	0	1	0	nasal
17.	r	r	alveolar	1	1	0	1	0	trill
18.	ɾ	4	alveolar	1	1	0	1	0	tapflap
19.	s	s	alveolar	1	1	0	1	1	fricative
20.	z	z	alveolar	1	1	0	1	0	fricative
21.	ɬ	K	alveolar	1	1	0	1	1	lateral-fricative
22.	ɮ	K\	alveolar	1	1	0	1	0	lateral-fricative
23.	ɹ	r\	alveolar	1	1	0	1	0	approximant
24.	l	l	alveolar	1	1	0	1	0	approximant
25.	ɖ	d`	retroflex	1	0	0	1	0	plosive
26.	ɖʐ	dz`	retroflex	1	0	0	1	0	affricate
27.	ɳ	n`	retroflex	1	0	0	1	0	nasal
28.	ʈ	t`	retroflex	1	0	0	1	1	plosive
29.	ʈʂ	ts`	retroflex	1	0	0	1	1	affricate
30.	ɻ	r`	retroflex	1	0	0	1	0	trill
31.	ʂ	s`	retroflex	1	0	0	1	1	fricative
32.	ʐ	z`	retroflex	1	0	0	1	0	fricative
33.	ɽ	r\`	retroflex	1	0	0	1	0	approximant
34.	ɭ	l`	retroflex	1	0	0	1	0	approximant
35.	θ	T	dental	1	1	1	1	1	fricative
36.	ð	D	dental	1	1	1	1	0	Fricative

Appendix 5.15b (continued)

#	IPA	X-SAMPA	Place	L04 configuration features PlaceA	PlaceB	PlaceC	Pulmonic	Voiceless	Manner
37.	ǀ	|\	dental	1	1	1	0	0	click
38.	ʃ	S	post-alveolar	1	1	0	1	1	fricative
39.	ʒ	Z	post-alveolar	1	1	0	1	1	fricative
40.	tʃ	tS	post-alveolar	1	1	0	1	1	affricate
41.	dʒ	dZ	post-alveolar	1	1	0	1	0	affricate
42.	!	!\	post-alveolar	1	1	0	0	0	click
43.	c	c	palatal	1	0	0	1	1	plosive
44.	ɟ	J\	palatal	1	0	0	1	0	plosive
45.	ɲ	J	palatal	1	0	0	1	0	nasal
46.	ç	C	palatal	1	0	0	1	1	fricative
47.	ʝ	j\	palatal	1	0	0	1	0	fricative
48.	j	j	palatal	1	0	0	1	0	approximant
49.	ʎ	L	palatal	1	0	0	1	0	lateral-approximant
50.	k	k	velar	0	0	0	1	1	plosive
51.	ɡ	g	velar	0	0	0	1	0	plosive
52.	ŋ	N	velar	0	0	0	1	0	nasal
53.	x	x	velar	0	0	0	1	1	fricative
54.	ɣ	G	velar	0	0	0	1	0	fricative
55.	ɰ	M\	velar	0	0	0	1	0	approximant
56.	ʟ	L\	velar	0	0	0	1	0	lateral-approximant
57.	q	q	uvular	0	0	0	1	1	plosive
58.	ɢ	G\	uvular	0	0	0	1	0	plosive
59.	ɴ	N\	uvular	0	0	0	1	0	nasal
60.	ʀ	R\	uvular	0	0	0	1	0	trill
61.	χ	X	uvular	0	0	0	1	1	fricative
62.	ʁ	R	uvular	0	0	0	1	0	fricative
63.	ħ	X\	pharyngeal	0	0	0	1	1	fricative
64.	ʕ	?\	pharyngeal	0	0	0	1	0	fricative
65.	ʔ	?	glottal	0	0	0	1	1	plosive
66.	h	h	glottal	0	0	0	1	1	fricative
67.	ɦ	h\	glottal	0	0	0	1	0	fricative
68.	ɥ	H	labial-palatal	1	0	1	1	0	approximant
69.	ɕ	s\	alveolo-palatal	1	2	0	1	1	fricative
70.	ʑ	z\	alveolo-palatal	1	2	0	1	0	fricative
71.	tɕ	ts\	alveolo-palatal	1	2	0	1	1	affricate
72.	dʑ	dz\	alveolo-palatal	1	2	0	1	0	affricate
73.	ʍ	W	labial-velar	0	1	2	1	1	fricative
74.	w	w	labial-velar	0	1	2	1	0	approximant

Appendix 5.15c Segmental Levenshtein distance, unweighed, between all pairs of 15 dialects, computed on the CASS database.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.000**	.644	.767	.703	.740	.802	.542	.737	.809	.666	.538	.748	.766	.573	.355
Wenzhou	.644	**.000**	.501	.711	.762	.459	.444	.593	.508	.604	.743	.567	.546	.557	.612
Guangzhou	.767	.501	**.000**	.401	.649	.696	.660	.642	.702	.670	.541	.482	.617	.777	.675
Xiamen	.703	.711	.401	**.000**	.665	.589	.628	.590	.510	.541	.741	.761	.649	.625	.657
Fuzhou	.740	.762	.649	.665	**.000**	.666	.559	.488	.394	.460	.757	.760	.745	.698	.710
Chaozhou	.802	.459	.696	.589	.666	**.000**	.699	.656	.631	.487	.485	.263	.681	.741	.610
Meixian	.542	.444	.660	.628	.559	.699	**.000**	.625	.669	.647	.519	.438	.257	.471	.315
Nanchang	.737	.593	.642	.590	.488	.656	.625	**.000**	.479	.675	.713	.619	.604	.673	.614
Changsha	.809	.508	.702	.510	.394	.631	.669	.479	**.000**	.488	.437	.279	.457	.306	.437
Taiyuan	.666	.604	.670	.541	.460	.487	.647	.675	.488	**.000**	.154	.772	.742	.673	.633
Beijing	.538	.743	.541	.741	.757	.485	.519	.713	.437	.154	**.000**	.646	.623	.567	.546
Jinan	.748	.567	.482	.761	.760	.263	.438	.619	.279	.772	.646	**.000**	.408	.434	.251
Hankou	.766	.546	.617	.649	.745	.681	.257	.604	.457	.742	.623	.408	**.000**	.242	.376
Chengdu	.573	.557	.777	.625	.698	.741	.471	.673	.306	.673	.567	.434	.242	**.000**	.348
Xi'an	.355	.612	.675	.657	.710	.610	.315	.614	.437	.633	.546	.251	.376	.348	**.000**

Appendix 5.15d Proximity matrix derived from Appendix 5.15c.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	1.16	1.26	1.21	1.27	1.33	1.04	1.21	1.43	1.21	1.10	1.41	1.30	1.15	0.98
Wenzhou	1.16	**0.00**	0.94	1.07	1.19	0.90	0.90	0.96	1.07	1.17	1.17	0.97	0.96	1.03	1.06
Guangzhou	1.26	0.94	**0.00**	0.76	1.14	1.05	1.16	1.01	1.22	1.15	1.15	1.18	1.21	1.32	1.29
Xiamen	1.21	1.07	0.76	**0.00**	1.00	1.10	1.21	0.92	1.13	1.04	1.17	1.31	1.22	1.22	1.28
Fuzhou	1.27	1.19	1.14	1.00	**0.00**	1.20	1.23	0.79	1.03	0.96	1.21	1.39	1.35	1.30	1.36
Chaozhou	1.33	0.90	1.05	1.10	1.20	**0.00**	1.21	1.08	1.09	0.98	1.02	0.91	1.26	1.29	1.20
Meixian	1.04	0.90	1.16	1.21	1.23	1.21	**0.00**	1.10	1.09	1.26	1.13	0.97	0.59	0.79	0.64
Nanchang	1.21	0.96	1.01	0.92	0.79	1.08	1.10	**0.00**	0.95	1.14	1.23	1.17	1.11	1.15	1.17
Changsha	1.43	1.07	1.22	1.13	1.03	1.09	1.09	0.95	**0.00**	1.07	1.08	0.89	0.94	0.76	0.95
Taiyuan	1.21	1.17	1.15	1.04	0.96	0.98	1.26	1.14	1.07	**0.00**	0.55	1.44	1.42	1.32	1.33
Beijing	1.10	1.17	1.15	1.17	1.21	1.02	1.13	1.23	1.08	0.55	**0.00**	1.24	1.24	1.16	1.11
Jinan	1.41	0.97	1.18	1.31	1.39	0.91	0.97	1.17	0.89	1.44	1.24	**0.00**	0.81	0.90	0.73
Hankou	1.30	0.96	1.21	1.22	1.35	1.26	0.59	1.11	0.94	1.42	1.24	0.81	**0.00**	0.52	0.72
Chengdu	1.15	1.03	1.32	1.22	1.30	1.29	0.79	1.15	0.76	1.32	1.16	0.90	0.52	**0.00**	0.65
Xi'an	0.98	1.06	1.29	1.28	1.36	1.20	0.64	1.17	0.95	1.33	1.11	0.73	0.72	0.65	**0.00**

Appendix 5.16a Segmental Levenshtein distance, feature-weighed, between all pairs of 15 dialects, computed on the CASS database.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.000**	.459	.499	.465	.469	.570	.360	.508	.594	.438	.352	.462	.545	.375	.199
Wenzhou	.459	**.000**	.338	.444	.540	.291	.273	.372	.312	.378	.525	.327	.304	.333	.337
Guangzhou	.499	.338	**.000**	.202	.441	.457	.405	.410	.452	.422	.338	.282	.377	.534	.444
Xiamen	.465	.444	.202	**.000**	.425	.409	.407	.378	.326	.358	.496	.519	.429	.399	.423
Fuzhou	.469	.540	.441	.425	**.000**	.415	.349	.297	.215	.313	.482	.490	.510	.470	.503
Chaozhou	.570	.291	.457	.409	.415	**.000**	.471	.444	.418	.304	.322	.178	.457	.505	.411
Meixian	.360	.273	.405	.407	.349	.471	**.000**	.395	.425	.407	.337	.274	.134	.336	.146
Nanchang	.508	.372	.410	.378	.297	.444	.395	**.000**	.285	.463	.491	.408	.381	.431	.391
Changsha	.594	.312	.452	.326	.215	.418	.425	.285	**.000**	.318	.267	.151	.320	.137	.258
Taiyuan	.438	.378	.422	.358	.313	.304	.407	.463	.318	**.000**	.098	.494	.511	.478	.438
Beijing	.352	.525	.338	.496	.482	.322	.337	.491	.267	.098	**.000**	.467	.437	.398	.370
Jinan	.462	.327	.282	.519	.490	.178	.274	.408	.151	.494	.467	**.000**	.256	.305	.173
Hankou	.545	.304	.377	.429	.510	.457	.134	.381	.320	.511	.437	.256	**.000**	.156	.230
Chengdu	.375	.333	.534	.399	.470	.505	.336	.431	.137	.478	.398	.305	.156	**.000**	.215
Xi'an	.199	.337	.444	.423	.503	.411	.146	.391	.258	.438	.370	.173	.230	.215	**.000**

Appendix 5.16b Proximity matrix derived from Appendix 5.16a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	0.95	1.00	0.96	1.00	0.95	0.84	1.01	1.20	0.92	0.82	1.01	1.09	0.92	0.72
Wenzhou	0.95	**0.00**	0.66	0.71	0.86	0.57	0.57	0.63	0.74	0.81	0.85	0.55	0.54	0.62	0.68
Guangzhou	1.00	0.66	**0.00**	0.47	0.81	0.55	0.78	0.69	0.86	0.74	0.78	0.73	0.80	0.92	0.93
Xiamen	0.96	0.71	0.47	**0.00**	0.69	0.68	0.83	0.62	0.82	0.71	0.80	0.86	0.84	0.83	0.93
Fuzhou	1.00	0.86	0.81	0.69	**0.00**	0.69	0.87	0.53	0.75	0.65	0.79	0.90	0.96	0.90	1.01
Chaozhou	0.95	0.57	0.55	0.68	0.69	**0.00**	0.74	0.61	0.66	0.56	0.68	0.64	0.76	0.79	0.86
Meixian	0.84	0.57	0.78	0.83	0.87	0.74	**0.00**	0.75	0.79	0.85	0.80	0.58	0.45	0.60	0.43
Nanchang	1.01	0.63	0.69	0.62	0.53	0.61	0.75	**0.00**	0.66	0.79	0.87	0.73	0.74	0.76	0.86
Changsha	1.20	0.74	0.86	0.82	0.75	0.66	0.79	0.66	**0.00**	0.79	0.84	0.61	0.70	0.58	0.82
Taiyuan	0.92	0.81	0.74	0.71	0.65	0.56	0.85	0.79	0.79	**0.00**	0.37	0.96	1.00	0.92	0.97
Beijing	0.82	0.85	0.78	0.80	0.79	0.68	0.80	0.87	0.84	0.37	**0.00**	0.87	0.93	0.84	0.84
Jinan	1.01	0.55	0.73	0.86	0.90	0.64	0.58	0.73	0.61	0.96	0.87	**0.00**	0.49	0.55	0.53
Hankou	1.09	0.54	0.80	0.84	0.96	0.76	0.45	0.74	0.70	1.00	0.93	0.49	**0.00**	0.47	0.61
Chengdu	0.92	0.62	0.92	0.83	0.90	0.79	0.60	0.76	0.58	0.92	0.84	0.55	0.47	**0.00**	0.49
Xi'an	0.72	0.68	0.93	0.93	1.01	0.86	0.43	0.86	0.82	0.97	0.84	0.53	0.61	0.49	**0.00**

Appendix 5.17a Levenshtein distance between all pairs of 15 Chinese dialects based on lexical frequency of 3-digit tone transcriptions (CASS database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.000**	.863	.775	.695	.789	.795	.793	.706	.674	.834	.550	.762	.465	.563	.879
Wenzhou	.863	**.000**	.551	.732	.735	.650	.727	.776	.730	.765	.832	.648	.675	.759	.909
Guangzhou	.775	.551	**.000**	.655	.837	.819	.812	.716	.782	.804	.775	.691	.727	.625	.764
Xiamen	.695	.732	.655	**.000**	.796	.642	.856	.680	.962	.828	.458	.695	.698	.501	.859
Fuzhou	.789	.735	.837	.796	**.000**	.608	.582	.689	.759	.767	.871	.819	.701	.893	.806
Chaozhou	.795	.650	.819	.642	.608	**.000**	.843	.726	.609	.672	.734	.718	.746	.752	.756
Meixian	.793	.727	.812	.856	.582	.843	**.000**	.771	.673	.754	.755	.837	.749	.721	.906
Nanchang	.706	.776	.716	.680	.689	.726	.771	**.000**	.825	.820	.630	.624	.799	.824	.719
Changsha	.674	.730	.782	.962	.759	.609	.673	.825	**.000**	.616	.847	.836	.622	.988	.845
Taiyuan	.834	.765	.804	.828	.767	.672	.754	.820	.616	**.000**	.918	.752	.710	.747	.521
Beijing	.550	.832	.775	.458	.871	.734	.755	.630	.847	.918	**.000**	.760	.691	.585	.928
Jinan	.762	.648	.691	.695	.819	.718	.837	.624	.836	.752	.760	**.000**	.784	.855	.823
Hankou	.465	.675	.727	.698	.701	.746	.749	.799	.622	.710	.691	.784	**.000**	.716	.920
Chengdu	.563	.759	.625	.501	.893	.752	.721	.824	.988	.747	.585	.855	.716	**.000**	.804
Xi'an	.879	.909	.764	.859	.806	.756	.906	.719	.845	.521	.928	.823	.920	.804	**.000**

Appendix 5.17b Proximity matrix derived from Appendix 5.17a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	1.33	1.21	1.09	1.28	1.24	1.23	1.13	1.19	1.35	0.88	1.22	0.74	0.96	1.46
Wenzhou	1.33	**0.00**	0.85	1.19	1.12	1.01	1.12	1.17	1.16	1.20	1.32	0.98	1.08	1.23	1.39
Guangzhou	1.21	0.85	**0.00**	1.05	1.29	1.21	1.24	1.09	1.29	1.24	1.20	1.03	1.14	1.02	1.25
Xiamen	1.09	1.19	1.05	**0.00**	1.33	1.10	1.36	1.07	1.54	1.40	0.73	1.11	1.15	0.80	1.44
Fuzhou	1.28	1.12	1.29	1.33	**0.00**	0.96	0.90	1.08	1.13	1.17	1.38	1.22	1.13	1.40	1.27
Chaozhou	1.24	1.01	1.21	1.10	0.96	**0.00**	1.23	1.10	1.02	1.04	1.21	1.09	1.14	1.25	1.21
Meixian	1.23	1.12	1.24	1.36	0.90	1.23	**0.00**	1.18	1.07	1.18	1.25	1.26	1.14	1.23	1.38
Nanchang	1.13	1.17	1.09	1.07	1.08	1.10	1.18	**0.00**	1.30	1.26	1.02	0.93	1.22	1.25	1.18
Changsha	1.19	1.16	1.29	1.54	1.13	1.02	1.07	1.30	**0.00**	1.00	1.44	1.28	1.02	1.53	1.32
Taiyuan	1.35	1.20	1.24	1.40	1.17	1.04	1.18	1.26	1.00	**0.00**	1.49	1.19	1.18	1.30	0.85
Beijing	0.88	1.32	1.20	0.73	1.38	1.21	1.25	1.02	1.44	1.49	**0.00**	1.20	1.10	0.91	1.52
Jinan	1.22	0.98	1.03	1.11	1.22	1.09	1.26	0.93	1.28	1.19	1.20	**0.00**	1.21	1.29	1.26
Hankou	0.74	1.08	1.14	1.15	1.13	1.14	1.14	1.22	1.02	1.18	1.10	1.21	**0.00**	1.14	1.46
Chengdu	0.96	1.23	1.02	0.80	1.40	1.25	1.23	1.25	1.53	1.30	0.91	1.29	1.14	**0.00**	1.36
Xi'an	1.46	1.39	1.25	1.44	1.27	1.21	1.38	1.18	1.32	0.85	1.52	1.26	1.46	1.36	0.00

Appendix 5.18a Levenshtein distance between all pairs of 15 Chinese dialects based on lexical frequency of staring pitch plus contour tone transcriptions (CASS database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.000**	.863	.775	.695	.789	.795	.793	.706	.674	.834	.550	.762	.465	.563	.879
Wenzhou	.863	**.000**	.551	.732	.735	.650	.727	.776	.730	.765	.832	.648	.675	.759	.909
Guangzhou	.775	.551	**.000**	.655	.837	.819	.812	.716	.782	.804	.775	.691	.727	.625	.764
Xiamen	.695	.732	.655	**.000**	.796	.642	.856	.680	.962	.828	.458	.695	.698	.501	.859
Fuzhou	.789	.735	.837	.796	**.000**	.608	.582	.689	.759	.767	.871	.819	.701	.893	.806
Chaozhou	.795	.650	.819	.642	.608	**.000**	.843	.726	.609	.672	.734	.718	.746	.752	.756
Meixian	.793	.727	.812	.856	.582	.843	**.000**	.771	.673	.754	.755	.837	.749	.721	.906
Nanchang	.706	.776	.716	.680	.689	.726	.771	**.000**	.825	.820	.630	.624	.799	.824	.719
Changsha	.674	.730	.782	.962	.759	.609	.673	.825	**.000**	.616	.847	.836	.622	.988	.845
Taiyuan	.834	.765	.804	.828	.767	.672	.754	.820	.616	**.000**	.918	.752	.710	.747	.521
Beijing	.550	.832	.775	.458	.871	.734	.755	.630	.847	.918	**.000**	.760	.691	.585	.928
Jinan	.762	.648	.691	.695	.819	.718	.837	.624	.836	.752	.760	**.000**	.784	.855	.823
Hankou	.465	.675	.727	.698	.701	.746	.749	.799	.622	.710	.691	.784	**.000**	.716	.920
Chengdu	.563	.759	.625	.501	.893	.752	.721	.824	.988	.747	.585	.855	.716	**.000**	.804
Xi'an	.879	.909	.764	.859	.806	.756	.906	.719	.845	.521	.928	.823	.920	.804	**.000**

Appendix 5.18b Proximity matrix derived from Appendix 5.18a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	1.21	1.09	1.06	1.09	0.97	0.70	0.90	0.99	0.94	0.93	1.33	1.01	1.28	1.03
Wenzhou	1.21	**0.00**	1.09	1.05	1.03	1.21	1.18	1.25	1.20	1.20	1.11	1.09	1.33	1.14	1.24
Guangzhou	1.09	1.09	**0.00**	1.07	1.01	1.29	1.14	0.90	1.11	1.26	1.21	1.17	1.09	1.04	1.02
Xiamen	1.06	1.05	1.07	**0.00**	0.95	1.17	1.12	1.01	1.05	1.25	1.15	1.19	1.28	0.88	1.16
Fuzhou	1.09	1.03	1.01	0.95	**0.00**	0.90	1.25	1.10	1.24	1.21	1.23	0.98	1.17	0.81	1.39
Chaozhou	0.97	1.21	1.29	1.17	0.90	**0.00**	1.01	1.26	1.20	1.01	1.20	1.01	1.01	1.10	1.34
Meixian	0.70	1.18	1.14	1.12	1.25	1.01	**0.00**	1.13	0.85	0.73	0.96	1.32	1.00	1.26	1.02
Nanchang	0.90	1.25	0.90	1.01	1.10	1.26	1.13	**0.00**	1.24	1.38	0.80	1.30	1.29	1.14	1.18
Changsha	0.99	1.20	1.11	1.05	1.24	1.20	0.85	1.24	**0.00**	0.73	1.13	1.39	0.98	1.32	0.60
Taiyuan	0.94	1.20	1.26	1.25	1.21	1.01	0.73	1.38	0.73	**0.00**	1.22	1.23	0.81	1.26	0.88
Beijing	0.93	1.11	1.21	1.15	1.23	1.20	0.96	0.80	1.13	1.22	**0.00**	1.29	1.45	1.26	1.17
Jinan	1.33	1.09	1.17	1.19	0.98	1.01	1.32	1.30	1.39	1.23	1.29	**0.00**	1.37	1.02	1.32
Hankou	1.01	1.33	1.09	1.28	1.17	1.01	1.00	1.29	0.98	0.81	1.45	1.37	**0.00**	1.20	1.10
Chengdu	1.28	1.14	1.04	0.88	0.81	1.10	1.26	1.14	1.32	1.26	1.26	1.02	1.20	**0.00**	1.42
Xi'an	1.03	1.24	1.02	1.16	1.39	1.34	1.02	1.18	0.60	0.88	1.17	1.32	1.10	1.42	**0.00**

Appendix 5.19a Distance between all pairs of 15 Chinese dialects based on lexical frequency of feature-weighed tones (CASS database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.000**	.454	.284	.265	.314	.358	.365	.239	.334	.454	.309	.416	.400	.346	.426
Wenzhou	.454	**.000**	.262	.397	.435	.358	.376	.451	.294	.325	.475	.303	.280	.229	.345
Guangzhou	.284	.262	**.000**	.304	.394	.392	.284	.320	.390	.382	.433	.416	.414	.248	.401
Xiamen	.265	.397	.304	**.000**	.222	.265	.370	.333	.363	.453	.275	.411	.443	.303	.378
Fuzhou	.314	.435	.394	.222	**.000**	.242	.315	.382	.312	.438	.344	.447	.381	.350	.415
Chaozhou	.358	.358	.392	.265	.242	**.000**	.294	.318	.399	.262	.415	.397	.479	.350	.415
Meixian	.365	.376	.284	.370	.315	.294	**.000**	.399	.376	.241	.409	.483	.353	.273	.384
Nanchang	.239	.451	.320	.333	.382	.318	.399	**.000**	.377	.446	.320	.428	.510	.456	.431
Changsha	.334	.294	.390	.363	.312	.399	.376	.377	**.000**	.348	.357	.357	.270	.400	.234
Taiyuan	.454	.325	.382	.453	.438	.262	.241	.446	.348	**.000**	.508	.372	.300	.286	.278
Beijing	.309	.475	.433	.275	.344	.415	.409	.320	.357	.508	**.000**	.525	.418	.347	.450
Jinan	.416	.303	.416	.411	.447	.397	.483	.428	.357	.372	.525	**.000**	.451	.400	.299
Hankou	.400	.280	.414	.443	.381	.479	.353	.510	.270	.300	.418	.451	**.000**	.233	.336
Chengdu	.346	.229	.248	.303	.350	.350	.273	.456	.400	.286	.347	.400	.233	**.000**	.309
Xi'an	.426	.345	.401	.378	.415	.415	.384	.431	.234	.278	.450	.299	.336	.309	**.000**

Appendix 5.19b Proximity matrix derived from Appendix 5.19a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.000**	.454	.284	.265	.314	.358	.365	.239	.334	.454	.309	.416	.400	.346	.426
Wenzhou	.454	**.000**	.262	.397	.435	.358	.376	.451	.294	.325	.475	.303	.280	.229	.345
Guangzhou	.284	.262	**.000**	.304	.394	.392	.284	.320	.390	.382	.433	.416	.414	.248	.401
Xiamen	.265	.397	.304	**.000**	.222	.265	.370	.333	.363	.453	.275	.411	.443	.303	.378
Fuzhou	.314	.435	.394	.222	**.000**	.242	.315	.382	.312	.438	.344	.447	.381	.350	.415
Chaozhou	.358	.358	.392	.265	.242	**.000**	.294	.318	.399	.262	.415	.397	.479	.350	.415
Meixian	.365	.376	.284	.370	.315	.294	**.000**	.399	.376	.241	.409	.483	.353	.273	.384
Nanchang	.239	.451	.320	.333	.382	.318	.399	**.000**	.377	.446	.320	.428	.510	.456	.431
Changsha	.334	.294	.390	.363	.312	.399	.376	.377	**.000**	.348	.357	.357	.270	.400	.234
Taiyuan	.454	.325	.382	.453	.438	.262	.241	.446	.348	**.000**	.508	.372	.300	.286	.278
Beijing	.309	.475	.433	.275	.344	.415	.409	.320	.357	.508	**.000**	.525	.418	.347	.450
Jinan	.416	.303	.416	.411	.447	.397	.483	.428	.357	.372	.525	**.000**	.451	.400	.299
Hankou	.400	.280	.414	.443	.381	.479	.353	.510	.270	.300	.418	.451	**.000**	.233	.336
Chengdu	.346	.229	.248	.303	.350	.350	.273	.456	.400	.286	.347	.400	.233	**.000**	.309
Xi'an	.426	.345	.401	.378	.415	.415	.384	.431	.234	.278	.450	.299	.336	.309	**.000**

Appendix 5.20a Phonological affinity (correlation coefficients) based on lexical frequencies of initials (DOC database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**1**														
Wenzhou	.911	**1**													
Guangzhou	.511	.530	**1**												
Xiamen	.709	.662	.659	**1**											
Fuzhou	.714	.698	.638	.889	**1**										
Chaozhou	.744	.697	.694	.973	.902	**1**									
Meixian	.690	.680	.730	.817	.781	.868	**1**								
Nanchang	.733	.713	.603	.754	.739	.802	.805	**1**							
Changsha	.571	.539	.407	.609	.672	.617	.520	.565	**1**						
Taiyuan	.739	.717	.533	.686	.768	.723	.696	.855	.664	**1**					
Beijing	.652	.655	.530	.610	.686	.634	.581	.752	.736	.861	**1**				
Jinan	.649	.652	.530	.604	.687	.630	.583	.761	.731	.861	.994	**1**			
Hankou	.583	.558	.403	.575	.654	.605	.555	.702	.855	.803	.712	.706	**1**		
Chengdu	.591	.556	.383	.563	.632	.599	.553	.707	.842	.816	.698	.697	.986	**1**	
Xi'an	.702	.700	.550	.652	.738	.681	.638	.815	.729	.928	.962	.963	.759	.755	**1**

Appendix 5.20b Proximity matrix derived from Appendix 5.20a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.000**	.113	.676	.473	.417	.448	.476	.389	.713	.471	.604	.607	.727	.729	.546
Wenzhou	.113	**.000**	.659	.508	.447	.482	.489	.408	.740	.487	.602	.604	.755	.758	.549
Guangzhou	.676	.659	**.000**	.473	.529	.460	.380	.604	.946	.776	.830	.829	.984	1.000	.804
Xiamen	.473	.508	.473	**.000**	.203	.060	.248	.437	.772	.619	.745	.747	.819	.833	.693
Fuzhou	.417	.447	.529	.203	**.000**	.193	.325	.357	.637	.470	.598	.601	.684	.700	.542
Chaozhou	.448	.482	.460	.060	.193	**.000**	.204	.399	.769	.594	.730	.731	.807	.819	.674
Meixian	.476	.489	.380	.248	.325	.204	**.000**	.393	.838	.620	.761	.761	.850	.860	.704
Nanchang	.389	.408	.604	.437	.357	.399	.393	**.000**	.643	.285	.467	.465	.602	.607	.389
Changsha	.713	.740	.946	.772	.637	.769	.838	.643	**.000**	.485	.470	.477	.242	.266	.466
Taiyuan	.471	.487	.776	.619	.470	.594	.620	.285	.485	**.000**	.273	.273	.427	.431	.175
Beijing	.604	.602	.830	.745	.598	.730	.761	.467	.470	.273	**.000**	.013	.496	.507	.103
Jinan	.607	.604	.829	.747	.601	.731	.761	.465	.477	.273	.013	**.000**	.501	.512	.103
Hankou	.727	.755	.984	.819	.684	.807	.850	.602	.242	.427	.496	.501	**.000**	.038	.463
Chengdu	.729	.758	1.000	.833	.700	.819	.860	.607	.266	.431	.507	.512	.038	**.000**	.472
Xi'an	.546	.549	.804	.693	.542	.674	.704	.389	.466	.175	.103	.103	.463	.472	**.000**

Appendix 5.21a Phonological affinity (correlation coefficients) based on lexical frequencies of finals (DOC database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**1**														
Wenzhou	.159	**1**													
Guangzhou	.028	.047	**1**												
Xiamen	.078	.077	.162	**1**											
Fuzhou	.059	.053	.196	.117	**1**										
Chaozhou	.183	.162	.149	.404	.294	**1**									
Meixian	.139	.128	.262	.430	.241	.360	**1**								
Nanchang	.177	.142	.234	.351	.274	.364	.782	**1**							
Changsha	.250	.122	.071	.194	.194	.283	.386	.446	**1**						
Taiyuan	.106	.108	.135	.154	.286	.317	.430	.445	.365	**1**					
Beijing	.201	.084	.133	.278	.256	.356	.476	.502	.504	.625	**1**				
Jinan	.149	.089	.122	.103	.278	.230	.244	.290	.212	.463	.607	**1**			
Hankou	.321	.140	.119	.268	.239	.416	.422	.477	.646	.496	.807	.431	**1**		
Chengdu	.327	.179	.075	.296	.225	.415	.465	.515	.632	.453	.762	.399	.843	**1**	
Xi'an	.217	.131	.103	.169	.191	.394	.320	.326	.401	.524	.682	.732	.661	.582	**1**

Appendix 5.21b Proximity matrix derived from Appendix 5.21a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**0.00**	1.23	1.45	1.38	1.39	1.31	1.53	1.52	1.30	1.51	1.68	1.45	1.56	1.52	1.52
Wenzhou	1.23	**0.00**	1.38	1.41	1.41	1.43	1.65	1.67	1.60	1.65	1.96	1.63	1.91	1.84	1.77
Guangzhou	1.45	1.38	**0.00**	1.28	1.20	1.42	1.46	1.53	1.65	1.59	1.91	1.59	1.93	1.92	1.79
Xiamen	1.38	1.41	1.28	**0.00**	1.29	0.95	1.06	1.17	1.34	1.40	1.57	1.51	1.56	1.49	1.55
Fuzhou	1.39	1.41	1.20	1.29	**0.00**	1.12	1.32	1.32	1.36	1.23	1.56	1.26	1.59	1.57	1.47
Chaozhou	1.31	1.43	1.42	0.95	1.12	**0.00**	1.03	1.04	1.11	1.08	1.24	1.23	1.18	1.15	1.15
Meixian	1.53	1.65	1.46	1.06	1.32	1.03	**0.00**	0.34	1.04	0.99	1.11	1.33	1.15	1.07	1.26
Nanchang	1.52	1.67	1.53	1.17	1.32	1.04	0.34	**0.00**	0.95	0.95	1.02	1.28	1.05	0.97	1.21
Changsha	1.30	1.60	1.65	1.34	1.36	1.11	1.04	0.95	**0.00**	1.00	0.93	1.24	0.75	0.71	1.04
Taiyuan	1.51	1.65	1.59	1.40	1.23	1.08	0.99	0.95	1.00	**0.00**	0.76	0.85	0.93	0.96	0.79
Beijing	1.68	1.96	1.91	1.57	1.56	1.24	1.11	1.02	0.93	0.76	**0.00**	0.92	0.42	0.50	0.61
Jinan	1.45	1.63	1.59	1.51	1.26	1.23	1.33	1.28	1.24	0.85	0.92	**0.00**	1.11	1.15	0.57
Hankou	1.56	1.91	1.93	1.56	1.59	1.18	1.15	1.05	0.75	0.93	0.42	1.11	**0.00**	0.26	0.72
Chengdu	1.52	1.84	1.92	1.49	1.57	1.15	1.07	0.97	0.71	0.96	0.50	1.15	0.26	**0.00**	0.80
Xi'an	1.52	1.77	1.79	1.55	1.47	1.15	1.26	1.21	1.04	0.79	0.61	0.57	0.72	0.80	**0.00**

Appendix 5.22a Phonological affinity (correlation coefficients) based on lexical frequencies of initials and finals in 15 dialects (DOC database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**1**	.861	.852	.992	.992	.838	.763	.809	.879	.113	.684	.674	.668	.668	.675
Wenzhou	.861	**1**	.980	.859	.855	.975	.639	.661	.748	–.004	.556	.549	.543	.541	.549
Guangzhou	.852	.980	**1**	.847	.845	.952	.616	.678	.765	–.038	.562	.555	.549	.547	.555
Xiamen	.992	.859	.847	**1**	.999	.837	.758	.808	.882	.128	.677	.666	.661	.660	.668
Fuzhou	.992	.855	.845	.999	**1**	.833	.766	.820	.892	.127	.694	.683	.678	.677	.685
Chaozhou	.838	.975	.952	.837	.833	**1**	.653	.637	.742	–.007	.568	.561	.555	.553	.561
Meixian	.763	.639	.616	.758	.766	.653	**1**	.617	.677	.348	.938	.928	.917	.915	.925
Nanchang	.809	.661	.678	.808	.820	.637	.617	**1**	.888	.054	.620	.612	.608	.607	.613
Changsha	.879	.748	.765	.882	.892	.742	.677	.888	**1**	.046	.682	.671	.666	.665	.673
Taiyuan	.113	–.004	–.038	.128	.127	–.007	.348	.054	.046	**1**	.303	.298	.294	.293	.295
Beijing	.684	.556	.562	.677	.694	.568	.938	.620	.682	.303	**1**	.992	.969	.967	.983
Jinan	.674	.549	.555	.666	.683	.561	.928	.612	.671	.298	.992	**1**	.950	.947	.994
Hankou	.668	.543	.549	.661	.678	.555	.917	.608	.666	.294	.969	.950	**1**	.999	.946
Chengdu	.668	.541	.547	.660	.677	.553	.915	.607	.665	.293	.967	.947	.999	**1**	.943
Xi'an	.675	.549	.555	.668	.685	.561	.925	.613	.673	.295	.983	.994	.946	.943	**1**

Appendix 5.22b Proximity matrix derived from Appendix 5.22a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.00**	.51	.51	.03	.04	.51	.90	.54	.31	2.73	1.07	1.08	1.09	1.09	1.08
Wenzhou	.51	**.00**	.06	.51	.54	.07	1.17	.70	.60	2.63	1.33	1.33	1.33	1.33	1.32
Guangzhou	.51	.06	**.00**	.51	.54	.10	1.17	.67	.58	2.63	1.32	1.32	1.32	1.32	1.31
Xiamen	.03	.51	.51	**.00**	.04	.51	.91	.54	.32	2.72	1.08	1.09	1.10	1.10	1.08
Fuzhou	.04	.54	.54	.04	**.00**	.54	.88	.54	.31	2.74	1.06	1.07	1.07	1.07	1.06
Chaozhou	.51	.07	.10	.51	.54	**.00**	1.13	.69	.60	2.60	1.29	1.29	1.29	1.29	1.28
Meixian	.90	1.17	1.17	.91	.88	1.13	**.00**	.96	.88	2.46	.24	.26	.27	.28	.25
Nanchang	.54	.70	.67	.54	.54	.69	.96	**.00**	.30	2.42	1.06	1.05	1.05	1.05	1.05
Changsha	.31	.60	.58	.32	.31	.60	.88	.30	**.00**	2.62	1.01	1.01	1.02	1.01	1.01
Taiyuan	2.73	2.63	2.63	2.72	2.74	2.60	2.46	2.42	2.62	**.00**	2.42	2.40	2.38	2.37	2.39
Beijing	1.07	1.33	1.32	1.08	1.06	1.29	.24	1.06	1.01	2.42	**.00**	.04	.09	.10	.05
Jinan	1.08	1.33	1.32	1.09	1.07	1.29	.26	1.05	1.01	2.40	.04	**.00**	.10	.11	.01
Hankou	1.09	1.33	1.32	1.10	1.07	1.29	.27	1.05	1.02	2.38	.09	.10	**.00**	.01	.11
Chengdu	1.09	1.33	1.32	1.10	1.07	1.29	.28	1.05	1.01	2.37	.10	.11	.01	**.00**	.11
Xi'an	1.08	1.32	1.31	1.08	1.06	1.28	.25	1.05	1.01	2.39	.05	.01	.11	.11	**.00**

Appendix 5.23a Phonological affinity (correlation coefficients) based on lexical frequencies of initials and finals in 15 dialects (DOC database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**1**	.726	.404	.578	.582	.631	.574	.597	.505	.583	.549	.533	.530	.535	.586
Wenzhou	.726	**1**	.417	.535	.560	.584	.557	.568	.445	.561	.515	.513	.461	.468	.555
Guangzhou	.404	.417	**1**	.550	.545	.575	.622	.514	.336	.439	.438	.434	.346	.318	.441
Xiamen	.578	.535	.550	**1**	.736	.857	.739	.661	.522	.565	.543	.495	.514	.510	.544
Fuzhou	.582	.560	.545	.736	**1**	.786	.675	.634	.570	.658	.595	.599	.566	.543	.613
Chaozhou	.631	.584	.575	.857	.786	**1**	.765	.701	.551	.633	.580	.547	.573	.566	.621
Meixian	.574	.557	.622	.739	.675	.765	**1**	.804	.503	.638	.569	.512	.537	.545	.569
Nanchang	.597	.568	.514	.661	.634	.701	.804	**1**	.548	.744	.692	.637	.650	.663	.683
Changsha	.505	.445	.336	.522	.570	.551	.503	.548	**1**	.593	.682	.598	.803	.789	.648
Taiyuan	.583	.561	.439	.565	.658	.633	.638	.744	.593	**1**	.800	.754	.724	.718	.816
Beijing	.549	.515	.438	.543	.595	.580	.569	.692	.682	.800	**1**	.887	.750	.728	.885
Jinan	.533	.513	.434	.495	.599	.547	.512	.637	.598	.754	.887	**1**	.641	.623	.899
Hankou	.530	.461	.346	.514	.566	.573	.537	.650	.803	.724	.750	.641	**1**	.945	.741
Chengdu	.535	.468	.318	.510	.543	.566	.545	.663	.789	.718	.728	.623	.945	**1**	.715
Xi'an	.586	.555	.441	.544	.613	.621	.569	.683	.648	.816	.885	.899	.741	.715	**1**

Appendix 5.23b Proximity matrix derived from Appendix 5.23a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.00**	.41	1.00	.72	.69	.71	.73	.73	.87	.81	.92	.88	.94	.92	.88
Wenzhou	.41	**.00**	.94	.79	.76	.80	.79	.82	.97	.90	1.02	.95	1.06	1.04	.98
Guangzhou	1.00	.94	**.00**	.86	.92	.95	.84	1.06	1.22	1.22	1.30	1.21	1.35	1.35	1.30
Xiamen	.72	.79	.86	**.00**	.43	.27	.43	.64	.95	.87	1.01	1.00	1.03	1.03	.99
Fuzhou	.69	.76	.92	.43	**.00**	.36	.51	.59	.84	.69	.83	.82	.90	.91	.81
Chaozhou	.71	.80	.95	.27	.36	**.00**	.40	.55	.93	.77	.92	.94	.96	.97	.89
Meixian	.73	.79	.84	.43	.51	.40	**.00**	.43	.96	.77	.94	.95	.99	.98	.92
Nanchang	.73	.82	1.06	.64	.59	.55	.43	**.00**	.82	.48	.65	.71	.74	.74	.64
Changsha	.87	.97	1.22	.95	.84	.93	.96	.82	**.00**	.72	.66	.73	.38	.39	.71
Taiyuan	.81	.90	1.22	.87	.69	.77	.77	.48	.72	**.00**	.36	.45	.57	.58	.33
Beijing	.92	1.02	1.30	1.01	.83	.92	.94	.65	.66	.36	**.00**	.26	.54	.57	.18
Jinan	.88	.95	1.21	1.00	.82	.94	.95	.71	.73	.45	.26	**.00**	.68	.70	.25
Hankou	.94	1.06	1.35	1.03	.90	.96	.99	.74	.38	.57	.54	.68	**.00**	.10	.58
Chengdu	.92	1.04	1.35	1.03	.91	.97	.98	.74	.39	.58	.57	.70	.10	**.00**	.61
Xi'an	.88	.98	1.30	.99	.81	.89	.92	.64	.71	.33	.18	.25	.58	.61	**.00**

Appendix 5.24a Phonological affinity (correlation coefficients) based on lexical frequency of initials, finals and tones (DOC database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**1**	.834	.738	.900	.887	.789	.722	.763	.792	.266	.666	.655	.650	.651	.669
Wenzhou	.834	**1**	.831	.796	.787	.872	.633	.654	.687	.176	.567	.562	.545	.545	.572
Guangzhou	.738	.831	**1**	.781	.770	.847	.636	.648	.662	.131	.549	.543	.516	.507	.545
Xiamen	.900	.796	.781	**1**	.936	.843	.760	.781	.810	.261	.663	.645	.646	.645	.658
Fuzhou	.887	.787	.770	.936	**1**	.828	.752	.779	.815	.297	.683	.676	.664	.658	.682
Chaozhou	.789	.872	.847	.843	.828	**1**	.702	.673	.704	.203	.592	.578	.580	.577	.597
Meixian	.722	.633	.636	.760	.752	.702	**1**	.690	.645	.452	.841	.818	.818	.818	.833
Nanchang	.763	.654	.648	.781	.779	.673	.690	**1**	.805	.272	.658	.638	.638	.641	.649
Changsha	.792	.687	.662	.810	.815	.704	.645	.805	**1**	.221	.698	.669	.715	.712	.683
Taiyuan	.266	.176	.131	.261	.297	.203	.452	.272	.221	**1**	.452	.438	.426	.425	.447
Beijing	.666	.567	.549	.663	.683	.592	.841	.658	.698	.452	**1**	.967	.918	.910	.960
Jinan	.655	.562	.543	.645	.676	.578	.818	.638	.669	.438	.967	**1**	.877	.870	.971
Hankou	.650	.545	.516	.646	.664	.580	.818	.638	.715	.426	.918	.877	**1**	.987	.900
Chengdu	.651	.545	.507	.645	.658	.577	.818	.641	.712	.425	.910	.870	.987	**1**	.891
Xi'an	.669	.572	.545	.658	.682	.597	.833	.649	.683	.447	.960	.971	.900	.891	**1**

Appendix 5.24b Proximity matrix derived from Appendix 5.24a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.00**	.42	.55	.18	.19	.40	.67	.44	.39	1.97	.93	.92	.93	.93	.92
Wenzhou	.42	**.00**	.27	.46	.50	.22	.89	.62	.61	1.90	1.14	1.12	1.14	1.13	1.12
Guangzhou	.55	.27	**.00**	.54	.58	.29	.93	.65	.67	1.87	1.18	1.16	1.18	1.17	1.17
Xiamen	.18	.46	.54	**.00**	.11	.37	.69	.46	.41	2.01	.96	.96	.97	.97	.95
Fuzhou	.19	.50	.58	.11	**.00**	.41	.65	.45	.39	1.99	.91	.91	.92	.92	.90
Chaozhou	.40	.22	.29	.37	.41	**.00**	.81	.58	.57	1.93	1.08	1.07	1.08	1.08	1.07
Meixian	.67	.89	.93	.69	.65	.81	**.00**	.65	.65	1.76	.39	.40	.42	.42	.39
Nanchang	.44	.62	.65	.46	.45	.58	.65	**.00**	.32	1.77	.84	.84	.84	.83	.84
Changsha	.39	.61	.67	.41	.39	.57	.65	.32	**.00**	1.87	.80	.80	.78	.77	.80
Taiyuan	1.97	1.90	1.87	2.01	1.99	1.93	1.76	1.77	1.87	**.00**	1.76	1.72	1.72	1.71	1.75
Beijing	.93	1.14	1.18	.96	.91	1.08	.39	.84	.80	1.76	**.00**	.09	.19	.20	.07
Jinan	.92	1.12	1.16	.96	.91	1.07	.40	.84	.80	1.72	.09	**.00**	.23	.24	.06
Hankou	.93	1.14	1.18	.97	.92	1.08	.42	.84	.78	1.72	.19	.23	**.00**	.03	.21
Chengdu	.93	1.13	1.17	.97	.92	1.08	.42	.83	.77	1.71	.20	.24	.03	**.00**	.22
Xi'an	.92	1.12	1.17	.95	.90	1.07	.39	.84	.80	1.75	.07	.06	.21	.22	**.00**

Appendix 5.25a Cheng's Phonological Correspondence Index (PCI) for all pairs of 15 dialects (DOC database).

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**1**	.492	.483	.525	.511	.499	.572	.561	.517	.568	.510	.523	.587	.592	.546
Wenzhou	.534	**1**	.473	.455	.499	.489	.485	.468	.514	.485	.407	.452	.467	.483	.464
Guangzhou	.484	.469	**1**	.515	.503	.474	.567	.522	.454	.455	.487	.480	.477	.458	.479
Xiamen	.461	.341	.434	**1**	.498	.510	.511	.489	.388	.468	.457	.421	.486	.449	.453
Fuzhou	.457	.405	.435	.534	**1**	.555	.557	.538	.443	.544	.490	.473	.483	.505	.487
Chaozhou	.439	.402	.396	.498	.545	**1**	.491	.477	.412	.517	.413	.417	.444	.463	.465
Meixian	.480	.418	.528	.535	.540	.504	**1**	.658	.516	.535	.504	.438	.549	.565	.465
Nanchang	.519	.376	.469	.537	.547	.514	.655	**1**	.524	.568	.577	.499	.583	.614	.536
Changsha	.534	.439	.412	.448	.492	.479	.532	.563	**1**	.520	.610	.572	.689	.688	.608
Taiyuan	.549	.400	.437	.476	.539	.516	.557	.560	.529	**1**	.609	.603	.590	.633	.612
Beijing	.489	.382	.464	.503	.536	.473	.553	.587	.608	.608	**1**	.713	.728	.730	.656
Jinan	.500	.404	.429	.458	.451	.414	.492	.497	.541	.612	.725	**1**	.594	.646	.765
Hankou	.512	.378	.463	.529	.481	.492	.576	.622	.663	.574	.727	.582	**1**	.799	.627
Chengdu	.498	.399	.450	.506	.524	.536	.580	.622	.632	.599	.722	.669	.791	**1**	.697
Xi'an	.551	.418	.431	.490	.476	.465	.516	.530	.579	.617	.715	.771	.643	.690	**1**

Appendix 5.25b Proximity matrix derived from Appendix 5.25a.

Speaker dialect	Listener dialect														
	Suzhou	Wenzhou	Guangzhou	Xiamen	Fuzhou	Chaozhou	Meixian	Nanchang	Changsha	Taiyuan	Beijing	Jinan	Hankou	Chengdu	Xi'an
Suzhou	**.00**	.78	.77	.74	.75	.78	.72	.71	.72	.68	.84	.79	.78	.81	.76
Wenzhou	.78	**.00**	.78	.93	.88	.87	.95	1.00	.91	.97	1.14	1.04	1.11	1.14	1.05
Guangzhou	.77	.78	**.00**	.77	.79	.83	.75	.82	.89	.89	.98	.95	.98	1.03	.96
Xiamen	.74	.93	.77	**.00**	.70	.71	.71	.74	.87	.80	.90	.91	.86	.92	.88
Fuzhou	.75	.88	.79	.70	**.00**	.64	.67	.70	.82	.70	.87	.88	.87	.87	.86
Chaozhou	.78	.87	.83	.71	.64	**.00**	.75	.77	.86	.76	.96	.95	.91	.92	.90
Meixian	.72	.95	.75	.71	.67	.75	**.00**	.50	.74	.68	.79	.86	.73	.74	.81
Nanchang	.71	1.00	.82	.74	.70	.77	.50	**.00**	.69	.65	.71	.80	.65	.66	.75
Changsha	.72	.91	.89	.87	.82	.86	.74	.69	**.00**	.71	.63	.68	.53	.57	.64
Taiyuan	.68	.97	.89	.80	.70	.76	.68	.65	.71	**.00**	.63	.62	.66	.64	.60
Beijing	.84	1.14	.98	.90	.87	.96	.79	.71	.63	.63	**.00**	.45	.44	.44	.47
Jinan	.79	1.04	.95	.91	.88	.95	.86	.80	.68	.62	.45	**.00**	.64	.59	.35
Hankou	.78	1.11	.98	.86	.87	.91	.73	.65	.53	.66	.44	.64	**.00**	.31	.57
Chengdu	.81	1.14	1.03	.92	.87	.92	.74	.66	.57	.64	.44	.59	.31	**.00**	.51
Xi'an	.76	1.05	.96	.88	.86	.90	.81	.75	.64	.60	.47	.35	.57	.51	**.00**

Appendix 6.1 Correlation matrix between subjective and objective measures. CC: data from Cheng (1997), Inv: sound inventories of Chinese dialects, CA: lexical frequencies based on the CASS database. Lev: Levenshtein distance, JS: judged similarity, JI: judged intelligibility, F: functional testing. Bolded coefficients are significant ($p < .01$).

Variables	CC_init	CC_final	CC_init_itfin	CC_in_fin_tone	CC_tone	CC_LAI	CC_PCI	Inv_init	Inv_nuc	Inv_coda	Inv_tone	Inv_final	Inv_init_coda	Inv_init_final	Inv_in_fin_tone
CC_final	**0.52**														
CC_init_final	**0.95**	**0.76**													
CC_in_fin_tone	0.19	0.23	0.26												
CC_tone	0.03	0.10	0.09	0.99											
CC_LAI	**0.45**	**0.60**	0.53	0.26	0.14										
CC_PCI	**0.58**	**0.86**	0.74	0.29	0.17	**0.76**									
Inv_init	**-0.50**	**-0.34**	-0.52	-0.37	-0.29	-0.27	**-0.44**								
Inv_nuc	-0.05	**-0.38**	-0.16	0.03	0.06	-0.30	**-0.36**	0.23							
Inv_coda	**-0.34**	**-0.43**	-0.40	-0.22	-0.16	**-0.64**	**-0.56**	**0.29**	**0.36**						
Inv_tone	-0.20	**-0.39**	-0.29	-0.22	-0.17	-0.29	**-0.33**	0.13	0.22	0.22					
Inv_final	-0.22	**-0.51**	-0.33	0.05	0.11	**-0.50**	**-0.53**	0.04	**0.75**	**0.59**	**0.29**				
Inv_init_coda	**-0.55**	**-0.48**	-0.60	-0.39	-0.30	**-0.51**	**-0.62**	**0.90**	**0.36**	**0.67**	0.21	**0.32**			
Inv_init_final	**-0.29**	**-0.56**	-0.40	-0.01	0.06	**-0.54**	**-0.59**	0.18	**0.76**	**0.62**	**0.31**	**0.99**	**0.44**		
Inv_in_fin_tone	**-0.30**	**-0.58**	-0.42	-0.03	0.04	**-0.55**	**-0.60**	0.18	**0.76**	**0.62**	**0.37**	**0.99**	**0.44**	**1.00**	
CA_init	**-0.76**	**-0.59**	-0.79	-0.22	-0.09	**-0.59**	**-0.68**	**0.69**	**0.26**	**0.40**	0.22	**0.33**	**0.73**	**0.42**	**0.43**
CA_final	**-0.36**	**-0.66**	-0.51	-0.32	-0.24	**-0.45**	**-0.60**	**0.34**	**0.49**	**0.50**	**0.40**	**0.62**	**0.51**	**0.66**	**0.67**
CA_coda	**-0.29**	**-0.40**	-0.35	-0.14	-0.08	**-0.41**	**-0.48**	0.25	**0.33**	**0.57**	**0.28**	**0.49**	**0.45**	**0.51**	**0.52**
CA_tone	0.05	0.08	0.05	-0.18	-0.19	0.15	0.15	0.12	-0.02	-0.10	**0.69**	-0.15	0.04	-0.14	-0.08
CA_nuc	**-0.32**	**-0.63**	-0.47	-0.38	-0.31	**-0.31**	**-0.53**	**0.37**	**0.51**	**0.38**	**0.42**	**0.52**	**0.48**	**0.57**	**0.58**
CA_init_fin	**-0.62**	**-0.75**	-0.74	-0.33	-0.21	**-0.62**	**-0.76**	**0.57**	**0.47**	**0.54**	**0.38**	**0.58**	**0.71**	**0.65**	**0.67**
CA_ons_fin_tone	**-0.45**	**-0.53**	-0.54	-0.36	-0.28	**-0.37**	**-0.49**	**0.52**	**0.35**	**0.36**	**0.72**	**0.35**	**0.58**	**0.41**	**0.46**
Lev_unweighed	-0.15	**-0.34**	-0.23	-0.16	-0.13	**-0.32**	**-0.33**	0.14	0.23	0.18	**0.28**	**0.27**	0.21	**0.29**	**0.30**
Lev_weighed	-0.09	**-0.27**	-0.16	-0.13	-0.11	-0.27	**-0.25**	0.11	0.20	0.15	0.24	0.22	0.17	0.24	0.25
Lev_tone	0.18	0.06	0.15	0.06	0.04	0.21	0.08	-0.07	0.04	-0.10	0.13	0.00	-0.11	-0.01	0.00
Lev_tone_change	0.01	0.02	0.01	0.12	0.12	0.17	0.02	0.03	0.02	-0.12	**-0.41**	-0.06	-0.03	-0.05	-0.08
Tne_weighed	-0.03	0.04	-0.03	-0.10	-0.09	0.15	-0.05	0.18	0.06	0.06	0.04	0.07	0.14	0.08	0.08
JS_monotonized	**0.54**	**0.67**	0.63	0.26	0.15	**0.85**	**0.76**	**-0.42**	**-0.35**	**-0.54**	**-0.45**	**-0.50**	**-0.59**	**-0.56**	**-0.58**
JS_intonated	**0.52**	**0.67**	0.62	0.30	0.20	**0.86**	**0.74**	**-0.40**	**-0.37**	**-0.56**	**-0.45**	**-0.51**	**-0.59**	**-0.56**	**-0.58**
JI_monotonized	**0.49**	**0.62**	0.57	0.13	0.04	**0.85**	**0.73**	**-0.28**	**-0.36**	**-0.56**	**-0.44**	**-0.61**	**-0.50**	**-0.64**	**-0.66**
JI_intonated	**0.50**	**0.62**	0.57	0.12	0.02	**0.87**	**0.71**	-0.24	**-0.38**	**-0.58**	**-0.44**	**-0.63**	**-0.48**	**-0.66**	**-0.67**
F_word	**0.35**	**0.73**	0.50	0.08	0.00	**0.77**	**0.77**	-0.25	**-0.40**	**-0.48**	**-0.36**	**-0.50**	**-0.44**	**-0.54**	**-0.55**
F_sentsence	**0.44**	**0.71**	0.56	0.06	-0.04	**0.74**	**0.77**	**-0.26**	**-0.33**	**-0.45**	**-0.44**	**-0.48**	**-0.43**	**-0.52**	**-0.54**

Appendix 6.1 (continued)

Variables	CA_init	CA_final	CA_coda	CA_tone	CA_nuc	CA_init_final	CA_in_fin_tone	Lev_unw	Lev_weight	Lev_tone	Lev_tone_change	Tone_weighed
CA_final	**0.44**											
CA_coda	**0.27**	**0.69**										
CA_tone	-0.02	0.02	0.04									
CA_nuc	**0.43**	**0.88**	**0.41**	0.04								
CA_init_fin	**0.80**	**0.89**	**0.60**	0.01	**0.81**							
CA_init_fin_tone	**0.61**	**0.70**	**0.49**	**0.63**	**0.65**	**0.78**						
Lev_unweighed	0.21	**0.38**	**0.27**	0.00	**0.31**	**0.36**	**0.27**					
Lev_weighed	0.16	**0.34**	**0.26**	0.00	**0.26**	**0.31**	0.23	**0.98**				
Lev_tone	-0.14	-0.08	-0.02	**0.31**	-0.04	-0.14	0.09	-0.22	-0.24			
Lev_tone_change	0.05	-0.09	-0.16	**-0.37**	-0.04	-0.03	**-0.26**	-0.10	-0.10	0.06		
Tone_weighed	0.21	0.20	0.15	0.13	0.20	0.23	**0.26**	0.00	0.03	**0.35**	**0.31**	
JS_monotonized	**-0.60**	**-0.55**	**-0.47**	0.03	**-0.47**	**-0.68**	**-0.50**	**-0.34**	**-0.27**	0.15	0.23	0.06
JS_intonated	**-0.58**	**-0.57**	**-0.47**	0.02	**-0.51**	**-0.68**	**-0.51**	**-0.33**	**-0.25**	0.12	0.23	0.06
JI_monotonized	**-0.54**	**-0.48**	**-0.50**	0.06	**-0.38**	**-0.60**	**-0.42**	**-0.29**	-0.22	0.05	0.19	0.02
JI_intonated	**-0.53**	**-0.49**	**-0.48**	0.08	**-0.40**	**-0.59**	**-0.40**	**-0.26**	-0.20	0.03	0.16	0.04
F_word	**-0.40**	**-0.44**	**-0.37**	0.22	**-0.39**	**-0.50**	-0.23	**-0.32**	**-0.26**	0.11	0.04	0.17
F_sentence	**-0.43**	**-0.42**	**-0.36**	0.15	**-0.39**	**-0.50**	**-0.27**	**-0.27**	-0.19	0.10	0.15	0.14

Variables	JS_monotonized	JS_intonated	JI_monotonized	JI_intonated	F_word
JS_intonated	**0.93**				
JI_monotonized	**0.90**	**0.82**			
JI_intonated	**0.86**	**0.89**	**0.94**		
F_word	**0.73**	**0.74**	**0.74**	**0.77**	
F_sentence	**0.78**	**0.78**	**0.80**	**0.82**	**0.93**

Curriculum Vitae

TANG Chaoju was born on January 28, 1968, in Chongqing, the People's Republic of China. She graduated from senior Middle School in 1987 and was admitted to the English Department, Chongqing Normal University, majoring in English as First Foreign Language and Japanese as Second Foreign Language. In 1991, she obtained her Bachelor's degree and worked as an assistant teacher of English in the South-Western Technology University in Sichuan Province, China; she was promoted to lecturer in 1996. From 1997 onwards, she studied English Linguistics and Literature as a Master candidate in the Foreign Languages Institute of South-Western Normal University, Chongqing. In 2000 she got her Master's degree. She then worked in the Foreign Languages Institute of Chongqing Jiaotong University and was promoted to associate Professor in 2002. In 2004, she was selected as an M.Phil candidate in the Leiden University Centre for Linguistics (LUCL), the Netherlands, and was supported by a grant from the China Scholarship Council. In 2005, she was accepted as a PhD candidate and began to do experiments on testing the mutual intelligibility between Chinese dialects. The present dissertation is the results of this research. Meanwhile she has resumed her job at Chongqing Jiaotong University.

Chaoju Tang: Mutual intelligibility of Chinese dialects: An experimental approach

This study examines the mutual intelligibility between all 225 pairs of 15 Chinese dialects, in two main branches, i.e., six Mandarin dialects and nine non-Mandarin (Southern) dialects. The dialects (often distinct languages by western standards) differ in the richness of their lexical tone inventories, ranging between four (in most Mandarin dialects) to as many as nine (in Guangzhou/Cantonese). Judgment (how well do listeners *think* they understand the speaker?) and functional (how well do speakers *actually* understand the speaker?) intelligibility tests were used. A methodological question was whether (fast and efficient) judgment testing may serve as a viable substitute for (laborious) functional intelligibility testing. Dialect fragments were also monotonized in order to estimate the importance of pitch variation for intelligibility in tone languages. Also, a large number of objective linguistic distance measures were collected, either copied from the literature or computed by the author on existing language resources. A systematic attempt is made to determine how well the judgment and functional intelligibility scores can be predicted from each other and from (combinations of) objective linguistics distance measures.

Mutual intelligibility testing affords a single dimension along which the degree of difference between language varieties can be expressed. The hypothesis is tested that the agglomeration trees generated from mutual intelligibility scores correlate strongly with linguistic taxonomies expressing family relationships among languages and dialects.

This study should be of interest to linguists, more specifically dialectologists, dialectometrists and phoneticians.