Central Asian and Vernacular Yanbian Korean in Comparative Perspective

Simon George Barnes-Sadler

Thesis submitted for the degree of PhD

2016

Department of the Languages and Cultures of Japan and Korea

SOAS, University of London

Declaration for SOAS PhD thesis

I have read and understood regulation 17.9 of the Regulations for students of the SOAS, University of London concerning plagiarism. I undertake that all the material presented for examination is my own work and has not been written for me, in whole or in part, by any other person. I also undertake that any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in the work which I present for examination.

Signed: _____   Date: _____

Abstract

This thesis describes and compares the synchronic linguistic features of the Koryo Mar (KM) and Vernacular Yanbian Korean (VYK): varieties of Korean spoken in Central Asia and Yanbian, PRC respectively. In addition, it provides socio-historical backgrounds of the transplanted communities and a thorough, critical review of prior linguistic research on their varieties.

Its main contribution, though, is to the empirical understanding of linguistic variation between the forms of Korean in transplanted contexts. Its principal outcomes are twofold.

Primarily on the basis of two collections of KM and VYK speech gathered from diverse samples of consultants, parallel descriptions of the transplanted varieties were created following the framework of traditional Korean grammar. This allowed for their systematic, granular comparison using principles adapted from Contrastive Analysis. Such an analysis revealed a high degree of equivalence in form and function between the majority of the over 100 comparative concepts analysed in addition to concrete differences in phonology (e.g. KM /v/, VYK /w/), morphology (e.g. VYK morphological honorification), word order (e.g. divergent patterns of negation) and vocabulary (e.g. VYK Sino-Korean coinings).

The degree to which VYK and KM speech can be distinguished and which features are most salient in this were established using the quantitative techniques of clustering and Principal Component Analysis. The results of these statistical tests demonstrated that KM and VYK may be distinguished on the basis of a selection of their aggregate linguistic features, and allowed for the identification of smaller sets of linguistic features which account for the greater part of the inter-varietal variation.

The main conclusion of this thesis is that varieties of Korean distinct from each other and those on the Korean Peninsula have developed and are further diverging in transplanted contexts. It further advances Korean dialectology by implementing and advocating novel analytical methods for the examination of vernacular speech rather than traditional elicited dialect data.

Table of Contents

List of Tables and Figures

List of Abbreviations

1	First person
2	Second person
3	Third person
ABL	Ablative
ACC	Accusative
ADJ	Adjective
ADN	Adnominaliser
ADV	Adverb
ALL	Allative
CA	Contrastive Analysis
CATPCA	Categorical Principal Components Analysis
CAUS	Causative
CCP	Chinese Communist Party
CIS	Commonwealth of Independent States
CLASS	Classifier
CNJ	Conjunction
COM	Comitative
COP	Copula
CSK	Contemporary Standard Korean
DAT	Dative
DECL	Declarative
DEM	Demonstrative
DM	Discourse Marker
DPRK	Democratic People's Republic of Korea

F	Feminine
FSU	Former Soviet Union
FUT	Future
GEN	Genitive
HON	Honorific
IMP	Imperative
INF	Informal (speech style)
INST	Instrumental
INT	Interrogative
IS	Indirect Speech
KM	*Koryo Mar*
LOC	Locative
LWC	Language of Wider Communication
M	Masculine
MOD	Modal
MOFAT	Ministry of Foregin Affairs and Trade
NEG	Negative
NEUTR	Neutral (speech style)
NLA	National Language Association
NMZ	Nominalizer
NOM	Nominative
OBJ	Object
OCE	Observed Communicative Event
ORD	Ordinal
PCA	Principal Components Analysis
PL	Plural

POL	Polite (speech style)
POS	Part of Speech
PRC	People's Republic of China
PRF	Perfective
PRN	Pronoun
PROP	Propositive
PRS	Present
PST	Past
REFL	Reflexive
RETR	Retrospective
ROK	Republic of Korea
TOP	Topic
SCE	Staged Communicative Even
SG	Singular
SSR	Soviet Socialist Republic
ST	Sociolinguistic Typology
SUBJ	Subject
USSR	Union of Soviet Socialist Repbulics
VYK	Vernacular Yanbian Korean
YKAP	Yanbian Korean Autonomous Prefecture

Acknowledgements

At some point prior to departing on the first period of fieldwork for this project, I attended a workshop on collaborative research. Over the course of the afternoon the question of authorship was raised and, at the time, I was shocked to discover how great a contribution someone could make to a piece of research and still not be entitled to claim a part in its authorship. From the criteria laid down at this workshop I understand that I am the sole author of this work and take full responsibility for any shortcomings or errors contained in it, but without the following people it could not have been written.

In Almaty and Yanbian I was lucky to be helped in my search for consultants, my initial struggles with transcription and with life in unfamiliar surroundings generally by Ilya Yem, Evgeniya Lee and Ma Won-gol. These invaluable assistants were introduced to me by N.S Pak at KazUIR&WL ih. Abilay Khan and Kim Kwang-su at Yanbian University. Prior to making

contact with these researchers, their work was an inspiration for this thesis and I have been delighted and humbled by the time and effort that they have sacrificed to help with this project. I would also like to extend my sincere gratitude to the other, official members of my supervisory committee, Prof. Peter Austin and Prof. Friederike Lüpke for their insightful comments and critiques, which had a profound impact on the shape of the finished thesis.

In addition to their contributions, it has been my greatest good fortune to have Prof. Yeon Jaehoon as my supervisor. He has served as an example of everything I aspire to as a researcher, teacher and mentor in providing support, guidance, considered and constructive critique as well as unique opportunities pursue my interests. I thank him for all of these things as well as his patience and forbearance during the production of this manuscript.

Throughout that time, I have also benefitted from the professional and emotional support of a wonderful collection of friends and colleagues at SOAS, most especially the regulars in the basement at 52 Gordon Square.

The scope of the support I have received from my parents is immeasurable and I am at a loss for appropriate expressions of gratitude. "Thank you for everything" seems inadequate, but I am especially grateful for their tolerance of my frequent absences to "chase people around with a tape-recorder".

Hanna helped, too. A lot.

1 Introduction

It is an inherent property of language to vary over time and space. The Korean language is no exception to this. Its attestation is conventionally divided into five temporal periods and the contemporary language is divided into between five and seven geographically defined varieties (cf. five varieties Rokuro 1945; six varieties Lee and Ramsey 2000 and King 2006a; seven varieties Kim 1982 and Sohn 1999). This dialect classification, though, only takes into account the varieties of Korean spoken on the Korean peninsula and its surrounding islands, whereas Korean is in fact spoken by a large global diaspora community. The variation of the Korean language in diaspora has received far less scholarly attention in comparison to linguistic variation on the Korean peninsula to date.

Going some way towards redressing this imbalance, this thesis focuses on the use of the varieties of Korean spoken by two of the oldest Korean diaspora communities: the Central Asian Korean community – the *Koryo Saram*, specifically the portion of that community centred on Almaty, Kazakhstan, and the Chinese Korean community, specifically that part of it residing in the Yanbian Korean Autonomous Prefecture (YKAP).

The main contributions of this thesis are to the discipline of Korean Studies and fourfold:

- It provides an empirical basis for the argument which holds that new, global varieties of Korean are developing from the traditional peninsula varieties

- It identifies granular formal and functional differences between the linguistic features of the varieties of Korean used by the Central Asian and Chinese Korean communities, which challenges their earlier, limited treatment in Korean traditional dialectology
- It demonstrates that the varieties of Korean used by the Central Asian and Chinese Korean communities may be differentiated from one another on the basis of their aggregate linguistic features
- It identifies the linguistic features which contribute most to the synchronic dissimilarity between these transplanted varieties of Korean

In terms of the discipline of linguistics, this thesis demonstrates the importance of combining quantitative and qualitative approaches in that its fine-grained, single feature oriented contrastive analysis is incorporated into and built upon by a statistical analysis in order to provide a fuller and more nuanced picture of the variation across and within these transplanted varieties than would be possible using just one of these methods alone.

These contributions are made as a result of the investigations into the central hypothesis of this research: the varieties of Korean spoken by the transplanted communities of Central Asia and the YKAP, despite their similar temporal and geographical origins, are synchronically distinct today. Over the course of the thesis, this central contention is broken down into thee research questions. Although these are developed throughout the chapters which precede the contrastive analysis and supplementary quantitative analysis, we present them here, too, for clarity from the outset:

1. What is the form/are the forms of Korean currently in use in China and Central Asia?
2. Do these forms of these varieties differ from one another and, if so, to what extent?
3. If these transplanted varieties can be shown to be distinct from one another, by what linguistic features may we distinguish them?

We now go on in the remainder of this section to provide a brief introduction to this thesis' objects of inquiry, namely, the Korean language and linguistic variation, before going on to give some background information on the global Korean diaspora, and outline the structure and content of the remainder of the thesis.

1.1 The Korean Language

1.1.1 Overview

Korean is a language of uncertain heritage (although strong arguments have been made for its inclusion in the Altaic family cf. Ramstedt 1949; Miller 1971; Starostin et al. 2003 etc.). The languages of the Korean peninsula are attested, albeit in a limited form, back to the third century of the Common Era (CE) (Lee and Ramsey 2011 p. 34) with the amount of attestation increasing sharply from roughly the fifteenth century at which time its unique, indigenous writing system was also developed. It is currently used by in excess of 80 million people, of whom some 75 million reside on the Korean peninsula in the Republic of Korea (ROK – South Korea) and the Democratic People's Republic of Korean (DPRK – North Korea), where Korean is the official state language and the primary means of communication for the population. While the genetic affiliation of Korean has still not been conclusively established, typologically it may be considered a representative head-final SOV language.

Both of the nation-states of the Korean peninsula maintain their own standard variety of Korean and the majority of contemporary research admits to between five and seven geographically delimited traditional dialect areas on the peninsula and an additional dialect area on the largest Korean island, Jeju, which has more recently also been classified as a separate language of the so-called Koreanic family (UNESCO 2010).

Throughout this thesis reference will be made to Contemporary Standard Korean (CSK). This term is somewhat awkward as Korean is a pluricentric language with multiple standard varieties (one developed in the ROK (known as *pyocwune*), on in the DPRK (known as *munhwae*) and one in the Yanbian Korean Autonomous Prefecture). The standard varieties of the ROK and DPRK are defined by these nations' respective national language organisations (National Language Association 2016; Social Sciences Research Assocation 1992), notionally on the basis of the speech of particular regions, while that of the Yanbian Korean Autonomous Prefecture is defined in prescriptive grammars according to principles discussed in more detail in section 3.4. Here, though, the term CSK is not used with these ideologically motivated standards in mind, rather it refers to the ideal variety which appears in contemporary descriptive grammars of Korean (e.g. Martin 1992 or Yeon and Brown 2011). This variety differs from all three of the official standard varieties of Korean, albeit to varying degrees, and may be considered the contemporary, supra-regional 'common language' or *kongthonge* of Korea rather than a prescriptive 'correct' form of Korean to which inferior 'deviant' forms may be compared.

1.1.2 Transplanted Korean

The Korean speaking population that is not resident on the Korean peninsula is spread throughout the world, with particularly large diaspora communities being found in China (especially the Yanbian Korean Autonomous Prefecture), Japan (especially Osaka), the former Soviet Union (especially Almaty, Kazakhstan, Tashkent, Uzbekistan, and Sakhalin, Russia), and the USA (especially Los Angeles and New York) (cf. Yun 2004). Outside of the Korean peninsula, the language only receives official recognition in the Yanbian Korean Autonomous Prefecture, where it is a co-official language along with Modern Standard Chinese. We note that relatively large Korean communities of over 50,000 people may be found in no fewer than five more countries (MOFAT 2012), however due to migratory churn in these communities or their

relatively young age varieties of regionally specific varieties of transplanted Korean are not thought to have developed there.

The varieties which are the particular focus of this thesis are spoken by the ethnic Korean populations of Central Asia and China, although not universally in either case. Among speakers of *Koryo Mar* (KM) and Vernacular Yanbian Korean (VYK), bilingualism in Russian or Chinese, respectively, is ubiquitous. This fact stands in stark contrast to the situation on the Korean Peninsula, which is frequently conceived of as utterly monolingual and linguistically homogenous. Monolithic conceptions of Korean are spreading, though, as increasing numbers of members of the transplanted communities are gaining fluency or at least knowledge of at least one of the contemporary standard varieties of Korean.

Since the seed communities for these populations largely emigrated from the Korean peninsula prior to the modern standardisation of the language (although the variety spoken in China has undergone some independent standardisation over the late twentieth century), the vernacular usages of both of the contemporary transplanted varieties are formed on the basis of dialect speech, principly of the varieties spoken in the North east of the Korean peninsula. Since their transplantation the degree of contact which these communities have had with the wider Korean speech community has been restricted, therefore they may be considered dialect islands. We further note that the original dialects from which the transplanted varieties developed were geographically substantially removed (given the scale of the Korean peninsula) from the dialects which later came to form the basis of the contemporary standard varieties.

The social and historical backgrounds of the Chinese and Central Asian Korean communities which we focus on in this thesis are examined in more detail in chapter 2, however, here we note that the seed communities of each emigrated from the Korean peninsula towards the end

of the so-called (Early) Modern Korean period (1592 - 1895). This was prior to the commencement of the Contemporary Korean period (1895 – present), which not only brought about the first concerted efforts towards the standardisation of the language on the peninsula mentioned above, but the Korean of that period is also characterised as distinct on linguistic grounds. This fact also contributes to the conception of the varieties of Korean spoken in China and Central Asia as especially removed from CSK, even among the contemporary transplanted varieties.

These transplanted varieties were moreover not considered in many of the early dialect taxonomies of the Korean language (cf. Lee 2003) and research which has been carried out on them to date has focussed on establishing their genetic affinity and degree of synchronic similarity with the varieties of the North East of the Korean peninsula. Consequently, prior research has focussed on establishing the presence of linguistic features associated with the varieties of the Korean peninsula's North Eastern dialect zone rather than establishing the patterns of use of these features in a cross-varietal context. While these studies have certainly provided evidence for the shared dialectological origins of these transplanted varieties, such a narrow focus, combined with the methodological issues of traditional Korean dialectology, has obscured the more nuanced picture of these varieties' characteristics and led to a situation in which they are often conflated rather than simply compared with peninsula varieties. While it will be necessary to touch upon their dialectological heritage, to as great an extent as possible in the current work we attempt to approach the varieties of Korean spoken in Central Asian and China in their own rights rather than making inferences friom them concerning peninsula dialects.

1.2 Linguistic Variation

1.2.1 Overview

Variation over space and time was identified as an inherent property of language at the outset of this section. The diachronic variation of language has been a pre-occupation of modern linguistics from its earliest days (cf. Jones 1786/1967; Rask 1818/1967 etc.) and one which endures to the modern day (cf. Fox 1995; Hale 2007; Ringe 2013). The inception of modern, formal study of synchronic variation in language quickly followed (e.g. the work of Georg Wenker or Jules Giliéron etc.), although an "intuitive and casual" (Chambers and Trudgill 1980 p. 14) awareness of geographically distributed linguistic variation pre-dates such formal study[1]. Early traditional dialectology relied on direct elicitation of particular linguistic items which could then be mapped in order to determine areas in which different variants (lexical, phonological, morphological etc.) were used (for a step-by-step description of programmatic data collection in traditional dialectology, see Petyt 1980 pp. 49-51). Despite ostensibly describing synchronic variation, such early work often retained a pronounced diachronic orientation in that variation was considered an object of interest for what it could reveal about linguistic history, particularly when it came to varieities which appeared to exhibit a high degree of 'archaism'.

The variation captured by traditional dialectological research has generally been characterised by sub-dividing a given geographical area, most usually a nation state, using the conceptual lines known as isoglosses. These lines mark the spatial limits of linguistic variants, often as the

---

[1] In the case of Korean, there is particularly clear evidence that there is a longstanding consciousness of regional variation. Documents contemporary to the promulgation of the Korean alphabet in the fifteenth century, in addition to noting the forms of graphemes for representing the contemporary prestige variety of the language and particular Chinese sounds not found in Korean, recorded graphemes which had the sole purpose of transcribing dialect sounds (c.f. Jiphyeonjeon c. 1446; Ledyard 1998p. 314; Lee and Ramsey 2000 p. 308)

boundary between two distinct variants, and most frequently of lexical or phonological features (although see Brandner 2012 on contemporary efforts to reconcile syntactic variation with more traditional dialectology). Dialects were generally defined as areas separated from those which surrounded them by numerous isoglosses which followed the same or a very similar geographic distribution. This construct was later refined to reflect the non-existence of such categorical boundaries in the real world, incorporating the concept of 'transition zones' between dialect areas. The patterns of variation which these disciplines discovered revealed that linguistic change and variation are not entirely random and again was linked to the early diachronic preoccupations of dialectology; specifically, it was suggested that synchronic, geographical variation was the outcome of diachronic changes taking effect at different rates in different areas. Language change in progress, though, was not thought to be observable.

Subsequent investigation into the synchronic variation of language over the mid-twentieth found that it is not only correlated with geographical distribution, but also with a wide and varied range of extra-linguistic factors pertaining to speaker or speech-community identity (e.g. Labov 1964; Trudgill 1974 etc.). Sociological categories which are especially strongly implicated in the patterning of linguistic variation include gender, social class, and age. Furthermore, the global and immediate social contexts of speech events have been implicated in having a great deal of influence in determining the style, form, or even content of speakers' utterances (cf. Halliday 1978). Thus, while even the language produced by an individual speaker may exhibit a great deal of variation, this too has been demonstrated to be rule governed and exhibiting an "orderly heterogeneity" (Weinreich, Labov and Herzog 1968).

Most recently the individual speaker has received increased interest as the locus of linguistic variation. Thus, intra-personal variation has been investigated more intensively through ideas of 'indexicality' (cf. Eckert 2008). Linguistic variation at the macro- or group level has not been neglected, though, rather the nature of the groups being examined has developed. Instead of

examining variation as it patterns to larger sociological indicators (i.e. gender, social class, ethnicity etc.) notions of 'communities of practice' (cf. Wenger 1998), i.e. a more functional conception of groups of people who come together to make language with a view to accomplishing a certain task or collaborating towards a shared goal, have come to prominence.

In addition, in common with much research in the humanities and the social sciences, research into language variation has undergone something of a quantitative turn. While this has been a tendency of sociolinguistics from its earliest days to the present time (e.g. Fischer 1958; Tagliamonte 2006 etc.), increases in computing power and the wider availability of larger data sources has led to a renewed interest in applying quantitative techniques (cf. Seguy 1973) to the phenomenon of geographically distributed linguistic variation (Nerbonne and Kretzschmar 2003). A notable difference between many of the quantitative, so-called dialectometric, research carried out in the early twenty-first century and the traditional dialectological carried out before is the former's departure from the latter's granular, detailed approach, using aggregated linguistic data to provide a generalised account of linguistic (dis)similarity between two or more varieties.

The common thread running through the approaches to variation outlined above is that their starting point is linguistic data. It is, however, a truism to point out that there is linguistically verifiable difference between a 'language' and a 'dialect'. Furthermore, when it comes to differentiating between dialects the subjective evaluations of the researcher often play a large role (see Song 2001 for a discussion of the grading of isoglosses). An approach to linguistic variation which elevates these subjective judgements is perceptual dialectology (Preston 1989; Preston and Long 1999; Jeon and Cukor-Avila 2015 etc.). As a general rule, in this sub-discipline the opinions of people without linguistic training are collected and from these, evaluative characterisations of varieties and maps showing their perceived geographical extents may be created.

In summary, then the study of synchronic variation developed from the study of diachronic linguistic variation and change. Its primary goals have been the fine-grained and detailed description of languages, capturing their full range of variation and the orderly description of this variation. Over time, it has drawn on developments in linguistic and social theory in order to provide more rigorous and complete descriptions, which have greater explanatory power when it comes to accounting for the variation in its descriptions. One final characteristic to note, though, is that there is an overwhelming tendency in research on linguistic variation to focus on contiguous geographical areas. The phenomenon of the globalisation of language has only recently become a focus for research (e.g. Coupland 2010; Smakman and Heinrich 2015), but with this change, an increased attention has been paid to variation in the global context. We go on to discuss this in the following section.

1.2.2 Transplanted Languages

In section 1.1.2 we discussed the transplanted varieties of Korean, which are to be central to this thesis from the point of view of the historical development of the communities in which they are spoken and traditional Korean dialectology. Here we take the opportunity to introduce the broader field of the study of languages which are conceived of as transplanted or which are spoken in diaspora situations in order to provide additional context and further support for the classification and comparative examination of KM and VYK as 'transplanted varieties of Korean'.

Twentieth century and earlier approaches the spatial investigation of language have been characterised as "sedentary" with speaker movement and migration being incorporated only somewhat recently and, even then, only of interest for the investigations of contact phenomena (Blommaert and Dong 2016). Such approaches tend to explicitly or, more often,

implicity[2] take a simple definition of a transplanted language may simply be one which is spoken outside of its homeland. This definition, however, may be problematized on account of its assumption that 'languages', their 'homelands' and what constitutes 'speaking a language' may be incontrovertibly identified. To give concrete examples of how these terms may be problematic, we shall take examples from World Englishes before going on to examine the frameworks which we have adapted for Korean.

Turning first to the notion of 'language', once we have admitted the possibility of variation within a given language, we must ask whether there are limits to be placed on it. For example, in cases where mutual comprehensibility between varieties is very low, may they be called the same language? To give a concrete if extreme example, this question may be asked of the various creole languages, such as Hawai'i Creole English or Tok Pisin, which are derived in part from English.

Moving on to the concept of 'homeland', we may point out that all World Englishes may ultimately be traced back to languages spoken in the northern portion of contemporary Germany and the southern portion of contemporary Denmark. Despite this, World Englishes are conventionally said to have originated on the British Isles (see Trudgill 2004; Jenkins 2009 etc.), a place which may seem as or even more remote to speakers of many World Englishes as the homelands of the Germanic peoples, who bought the precursor to the English of the British Isles with them from northern continental Europe, seem to speakers of British English who reside in the British Isles.

---

[2] See, for example, a foundational text on contact induced dialect change in which examples of transplanted situations are commonly adduced, but no definition of what constitutes such a situation is provided (Trudgill 1986 pp. 83-100)

Finally, we must interrogate the notion of 'speaking a language'. In the context of World Englishes this has lead to the popularly accepted tripartite distinction between countries in which English is used as a native language, second language, and foreign languages, on the basis of which the categories of the Inner Circle, Outer Circle and Expanding Circle English speaking nations have been established (Kachru 1992). If this model were to be applied to Korean, it would be necessary to ask whether the large number of second language Korean speakers all over South East Asia is also to be considered a population which speaks transplanted Korean?

While the origins and spread of English as an international language is starkly different from that of Korean, these questions are just as relevant if we are to fully address varieties of Korean spoken outside of the Korean peninsula and its surrounding islands.

It fact, there is no simple, universally accepted definition of a transplanted language or variety. This holds not only cross-linguistically, but, to once more take an example from World Englishes, models for identifying and classifying transplanted languages may vary considerably for even a single language (for a recent discussion of the range of models for identifying transplanted varieties of English see Galloway and Rose 2015 pp. 7 – 25).

We note that a general characteristic of definitions of transplanted languages and, indeed, a characteristic of the issues that are often found with them, is that they rely more on social than linguistic characteristics of communities since there are no universal unifying linguistic criteria for identifying transplanted languages. On those grounds we, too, take the social characteristics of speech communities as diagnostic of transplanted languages. Following some of the earliest systematic work on transplanted languages (Price 1973), the existence of another region or state in which the language is the dominant language (although not necessarily an official language) may be identified as a fundamental characteristic. This

definition has been challenged and refined (e.g. Sikma and Gorter 1991), often within the framework of minority language studies. Dexpite this critique, the recognition on the part of the transplanted community of this criterion is perhaps the only one which is universally valid for identifying transplanted languages or dialect, due to the high degree of both geographical diversity (see Edwards 2010 p. 81) and sociological diversity (see Trudgil 2011) in 'transplanted communities'.

Consequently, it is by dint of this criterion, the recognition of community external, geographically removed dominant or canonical varieties of the language, that we consider KM and VYK transplanted varieties of Korean. It is important to emphasise, though, that while both KM and VYK may be identified as transplanted varieties of Korean in general, it is quite possible to distinguish them on the grounds of their divergent geneses, contemporary social structure or geographical and cultural relationship with their notional homeland and current host nation. For example, KM is neither adjoining nor cohesive with the contiguous Korean speaker realm and may thus be likenened to such transplanted languages as German in the USA. VYK, on the other hand, adjoins the contiguous Korean speaking realm, although its degree of cohesiveness with it is debatable. This gives it a sociolinguistic profile more similar to that of German in Western Poland. We revisit the historical processes of transplantation and these distinguishing characteristics in chapter two, where they are examined in far greater detail.

1.2.3 Korean

Unlike diachronic variation and change, synchronic variation is an aspect of Korean which research has neglected. A systematic literature review by Silva (2010) revealed that studies into synchronic linguistic variation in Korean are comparatively rare compared to other

avenues of linguistic enquiry which focussed on Korean and comprised only a maximum of five per cent of Korean linguistics articles available from numerous scholarly databases.

Where research has been carried out on variation in Korean, it has tended to be within the framework of traditional dialectology, implicitly endorsing the maxim "*chaque mot a son histoire*", popularly attributed to Gilliéron, in that it is being comprised as a body of work of atomistic studies of individual words and features (cf. Lee et al. 2008). This is analogous to the state of dialectology near-universally in the early twentieth century (Chambers and Trudgill 1980 pp. 15-21). A further characteristic of early Korean dialectology was Japanese origin. The first investigations widely acknowledged as being carried out in the framework of traditional dialectology covering linguistic variation over the territory of the Korean peninsula were carried out during the Japanese colonial period (1905-1945), very frequently by Japanese fieldworkers (e.g. the work collected in Ogura 2009 [1944]). While the scope of such early dialectology was vast, encompassing items of vocabulary, verb endings and other grammatical forms gathered from the whole Korean peninsula, data collection during this period was somewhat unsystematic. For example, in Ogura Shinpei's 1944 *Cosenepangensacen* (Dictionary of Korean Dialect), which collates the results of his earlier dialect studies, twenty survey sites are identified in Gyeonggi Province, but a cursory examination of the book reveals that data is only regularly collected from a very small proportion of them. This is perhaps the most extreme of the gaps in the data which appears in this highly influential work, which is also beset by problems of over-inclusion of data. For example, several sites which do not appear to have had data gathered from them individually are included in some sweeping statements about particular dialect forms which are used over the whole country. Despite these methodological shortcomings we nevertheless acknowledge the profound importance of this early work, not only as the genesis of the field, but especially in terms of its lasting influence on the study of linguistic variation in Korea.

A further characteristic of these early studies was their pronounced diachronic focus, as has been noted elsewhere (for example Lee 2003 p. 48). While the latter part of the twentieth century saw an increasing number of Korean dialectologists begin to acknowledge dialects as "independent linguistic systems" and investigate their synchronic characteristic, there are particular areas of the study of dialect in Korea which retain as strong diachronic tendency, such as dialect taxonomy. On that subject, before going on to chart the development of Korean dialectology after the Japanese Colonial period, we note that, in the more synchronically based dialect taxonomic studies that have been carried out over the course of the twentieth century and beyond, dialect and sub-dialect areas have been generally been identified on the basis of either single feature studies or those which take only a very small selection of isoglosses (for example Lee Iksop's 1987 taxonomy of Gangwon dialect made solely on the basis of the form of the informal interrogative ending or Lee Sungjae's 2004 taxonomy of the southwestern dialect (in this cast incorporating Jeolla and Chungcheong Provinces) solely on the basis of the operation of vowel harmony in verbal inflection) into account pattern strikingly closely with the changing administrative divisions of the Korean peninsula.

It is following the Japanese Colonial period that we see the appearance of smaller descriptive studies of variation in Korean along with a studies focussing on variation not only in vocabulary, but in phonology, morphology and syntax, too (Jeong 2001). It may further be observed that, with a few notable exceptions (for example Ramsey 1978), research into Korean dialectology in the twentieth century was produced domestically rather than internationally, whereas other areas of Korean linguistics appear to have captured the attention of the international research community.

Within the very small field of research into variation in the Korean language, such studies of geographical variation have predominated to the extent that only in the late twentieth and

early twenty first centuries have other approaches to linguistic variation (i.e. sociolinguistic studies, perceptual dialectology etc.) begun to be seriously pursued. While social and regional variation are often conflated in the Korean term *pangenhak*, it is only in very recent times that a more socially integrated approach to dialectology which does not restrict itself to traditional, rural areas and data, for example the so-called 'urban dialectology' of Lee Sang-Oak (Lee 2006)**.**

It is also worth noting that the transplanted varieties of Korean are conspicuous in all earlier works of Korean dialectology and, indeed, in many more recent works of traditional dialectology only on account of their absence. A particularly striking example of the latter phenomenon would be the non-inclusion of Korean spoken outside of the Korean peninsula in the comprehensive record of Korean dialect forms *Hankwukpangencalyocip* (Korean Dialect Resources Collection 1987-1995).

We may thus observe that the conclusions of traditional Korean dialectology in terms of the taxonomy of varieties or sub-varieties of Korean have only rarely been reflective of the full picture of empirically observable linguistic reality since they are based on a conception of the Korean language which only includes those varieties used on the Korean peninsula and, as we shall see in chapter 3, has frequently resolved this problem only by classifying transplanted varieties together with particular peninsula varieties.

The foregoing provides valuable context for this thesis and also demonstrates clearly that it is situated firmly at the intersection between the broader study the global phenomenon of linguistic variation and the specific study of variation in Korean. Our research questions are predicated on cross-linguistically observed patterns of variation which appear to be at odds with the conclusions of traditional Korean dialectology. This thesis presents a principled case for the use of particular methodological frameworks and concrete techniques for the analysis of the transplanted varieties of Korean briefly described above in order to reconcile these two

disciplines and reveal as accurate a picture as possible of the characteristics of and (dis)similarities between the two transplanted varieties in question.

1.3 Structure

We now outline the structure of the remainder of this thesis. In order to elucidate the socio-historical context of the speech communities under investigation and position this project in relation to prior research, we follow this introduction with first a more detailed examination of the historical development and contemporary situation of the Korean diaspora communities of Central Asia and the PRC. This provides a solid understanding of the commonalities and differences in the transplanted contexts of the varieties of Korean which are under examination in this thesis and provides grounds for the comparative examination of KM and VYK.

This is followed by a comprehensive and critical review of the literature pertaining to global varieties of Korean in general and KM and VYK in particular. We demonstrate that, even in the relatively small amount of scholarship dedicated to them, the transplanted varieties are only exceptionally examined in their own rights as independent linguistic systems, rather than as examples of the varieties of the North East of the Korean peninsula. We go on to explore the empirical findings of prior research on these varieties and question the extent to which they accurately reflect the characteristics of the transplanted varieties in light of the methodological shortcomings of traditional (Korean) dialectology. This review allows us to frame concise and significant research questions. In order to answer these, we also consider the ideologies underlying this research in order to identify both appropriate methodological frameworks and concrete procedures for data gathering an analysis.

The next chapter is devoted to the methodological issues which surround the examination of these varieties: how are speakers of these varieties to be identified; on what basis are

descriptions of these varieties to be formulated; how to undertake a systematic comparison of these two varieties in order to reveal any potential linguistic similarities and differences; which framework should be used to analyse the findings of such a comparison to account for them in terms of linguistically significant generalisations?

Through addressing these questions, we arrive at the conclusion that, of all the linguistic sub-disciplines concerned with comparison, the most appropriate and robust comparison for our purposes would be carried out in the broad framework of Contrastive Analysis (CA) and rely on traditional Korean grammatical description for the provision of underlying comparative concepts. It was further determined that such an analysis, while effective for identifying specific points of difference between VYK and KM and providing a fine-grained description of linguistic variation, should be supplemented by a quantitative analysis of the distribution of linguistic features over the sampled population of consultants in order to decisively conclude whether there is an empirical linguistic basis for considering the transplanted varieties distinct. The place of quantitative analysis in contemporary linguistics is summarised and appropriate statistical tools for answering our concrete research questions are identified.

The discussion in this chapter also allows us to determine the form which the primary data should take in order to be suitable for these two types of comparison, namely, recordings of semi-structured interviews and observed communicative events gathered from small samples[3] of self-identifying speakers of Central Asian and Yanbian Vernacular Korean (twelve and fourteen consultants, respectively) which were subsequently transcribed using ELAN (Brugman and Russel 2004). While the transcriptions were sufficient basis for the CA carried out in chapter 5, the data was coded as described in chapter 6.2 for the quantitative analysis.

---

[3] Although see sections 3.3. and 3.4. While the sample of speakers is small in absolute terms, this is not the case when it is put in the context of prior research on these varieties.

In the next chapter we turn to the parallel descriptions and CA of KM and VYK. This examines over one hundred individual comparative concepts (taken from traditional Korean grammatical description) and presents examples of their forms and functions drawn largely from data gathered from semi-structured interviews with both VYK and KM speaking consultants in addition to some data from select secondary sources. Degrees of equivalence between KM and VYK instantiations of these comparative concepts are established and it is concluded that, while the overall majority of surveyed features are so-called 'near equivalents' (i.e. nearly identical to one another in form and function), a not-insignificant number of features on the phonological and morphological levels stand in relationships of partial or non-equivalence. Features which were found to exhibit a great deal of variance in their use by this granular analysis were selected for inclusion in the quantitative analysis.

The final content chapter of this thesis is a quantitative analysis of the primary data upon which the preceding CA is largely based. It incorporates both the findings of CA and the observations of prior research to construct a catalogue of linguistic features which are suitable for analysis. This analysis co-opts multivariate statistical tests for the purpose of determining whether, in contrast to the CA, VYK and KM may be differentiated 'in general' rather than only through the sole examination of specific single features. To this end, we turn to the relatively well-established quantitative dialectological technique of clustering and, in order to answer the supplementary question of which linguistic features most distinguish these transplanted varieties from one another, we also employ a variant of Principal Components Analysis – categorical Principal Components Analysis. This chapter demonstrates conclusively that the speech of VYK and KM consultants is linguistically distinct and identifies a small set of thirteen linguistic features which account for the greater part of this distinction.

In addition to providing a definitive summary of the findings and content of this thesis, the main purposes of the subsequent concluding chapter are to situate the findings of the

foregoing analysis chapters in the disciplines of Korean dialectology and variationist linguistics

and also to identify promising avenues for further research on the transplanted varieties. In

examining the vast amount of fundamental research which is still to be carried out on KM and

VYK, we also reinforce the urgent necessity of carrying out documentation of these

transplanted varieties in all their diverse forms, especially the critically endangered KM, before

it finally becomes extinct.

1.4 Orthography

We finish this introductory section with a brief note on transliteration. Where standard

Korean is transliterated the Yale System of Romanisation (YSR) is used. Examples of the non-

standard varieties of Korean investigated in this thesis which are derived from recorded

primary data will be presented in a broad transcription using the International Phonetic

Alphabet (IPA), whereas those examples of the transplanted varieties taken from sources

written using either original Hangul or Cyrillic transcription will be given in a modified form of

the YSR (Barnes-Sadler 2015), which explicitly encodes distinctions made in the original

transcription, such as voicing, nasalisation and alveolar/post-alveolar realisation, amongst

other things, which are not normally present in the YSR. Where Russian and Chinese are

transliterated, the International Scholarly System and Hanyu Pinyin are to be used,

respectively.

Exceptions to the foregoing are made wherever a conventionalised, non-systematic

transliteration is available as is often the case for personal names or geographical regions. For

example we refer to Hamgyeong Province rather than *hamkyeng-do*.

2 Background to Transplanted Korean Communities

There are in excess of seven million Koreans living outside of the Korean peninsula in 174

countries amongst which eleven host Korean communities with populations in excess of

50,000 and a further five host communities over 20,000 (MOFAT 2011). The Korean language

is maintained in diverse forms and to varying extents in these communities depending on such

factors as community age, the dialectological origins of the seed community and their

experiences in their host nations. It is only comparatively recently that the varieties of Korean

used in these communities have come to be the subject of research, and then they are dealt

with singly or only in relation to standard varieties of peninsula Korean rather than in terms of

direct comparison with one another, which is the purpose of the current study. Given the

large number of nations which host Korean communities, it is impractical to conduct a

complete contrastive linguistic survey of the Korean diaspora, or even of such "large"

communities as those of over 50,000. Thus, it is necessary to establish criteria narrowing the

scope of our investigation.

In describing the development of transplanted varieties of any language, Trudgill makes an analogy with baking and advocates the study of varieties which:

> … arose from similar mixtures of similar dialects in similar proportions occurring at similar times. If you bake cakes, I suggest, from roughly the same proportions in roughly similar conditions for roughly the same length of time, you will get roughly similar cakes. (Trudgill 2004 p. 20)

This serves to identify the factors which most influence the development of transplanted varieties of specific languages in very general terms. Trudgill's "proportions of ingredients" is analogous with the dialectological background of the original seed community of the transplanted speech community and the ratio of speakers of different dialects in that community. To extend the metaphor, then, his 'baking time' may be taken to mean the age of the community and the 'oven temperature' is the linguistic and social conditions in which the transplanted community finds itself.

Bearing in mind that the focus of our examination is on the synchronic linguistic dissimilarity which has arisen as a result of of post-transplantation linguistic change, i.e. linguistic change occurring in the differing linguistic and social conditions in which diaspora varieties of Korean are found, we must ensure that as many of the factors surrounding their genesis, are as similar as possible. We may be sure that each transplanted variety of Korean is the result of dialect mixture to a greater or lesser extent, however, if we hope to conduct a meaningful comparative study which goes beyond simply recapitulating the earlier descriptive research, discussed in sections 3.3 and 3.4, or that on contact between transplanted dialects, we must take the necessary step of screening the varieties to be compared.

A simple example may be adduced to demonstrate that linguistic differences attributable to the demographics of the seed community tell us little about the relationship between

language change and the diaspora situation. While a comparison the variety of Korean used by a diaspora Korean community drawn largely from the southern part of the Korean peninsula, such as that in Osaka, and a community drawn largely from the northern part of the Korean peninsula, such as that in Yanji, is likely to demonstrate the diversity of transplanted Korean varieties, the fact that the former variety exhibits southern dialect features while the latter variety exhibits northern dialect features does not allow us to comment on the role of the diaspora situation in the formation of new varieties of Korean. Ideally, the input varieties or dialect mixtures should be as similar as possible in order to establish this.

Furthermore, this task is also implicitly historical in nature, examining language change or stasis. Language change requires time, often several generations, to manifest; therefore recently established diaspora communities, for example those with histories stretching as far back only as the mid-20th century, are unsuitable for our purposes and it would certainly be inappropriate to compare such varieties with those spoken by older diaspora communities.

One other practical consideration not mentioned by Trudgill which must be taken into account is the accessibility of the varieties in question. Varieties for which any linguistic resources at all are not available and which are inaccessible for the purposes of primary data gathering must also be discounted as objects of study.

Given these requirements two varieties may be identified as suitable for direct comparison: *Koryo Mar*, the spoken variety of the *koryo saram* multi-national Central Asian Korean diaspora and, *Yanbian ciyeke* also known as *Yanbian cosene* (for our purposes Vernacular Yanbian Korean or VYK), the variety of the *cosencok* Korean community of the Yanji Korean Autonomous Prefecture (YKAP) in the People's Republic of China (PRC). As specified above,

these communities differ largely in respect to their experiences in the diaspora situation, but have similar dialectological origins and temporal origins.

2.1 Socio-historical Backgrounds

As seen above, members of the transplanted Korean communities of Central Asia and China use specific ethnonyms to describe themselves. Such self-identification (admittedly, amongst many other factors) has long been understood to be an influential criterion in the construction of social or ethnic identity (Bashirova and Solnyshkina 2015). While this need not be associated with use or proficiency in a particular language or variety, or even with a specific linguistic identity, it may certainly be regarded as a contributing factor towards the delineation of both varieties of language and the constructions designated as languages from one another (Weinrich 1979). In examining the social backgrounds and histories of the two speech communities identified above, albeit necessarily briefly and incompletely, we shed light on how their independent senses of identity and independent conceptions of the varieties of Korean with which they identify developed. Such an examination of socio-historical backgrounds further enables us to more fully understand the extra-linguistic forces which have acted and continue to act on the transplanted Korean speech communities of China and Central Asia and the roles which they played in the development of their varieties of Korean.

We examine each transplanted community in separate sections. Each section comprises a brief discussion of the origins of the community on the Korean peninsula, and then a description of the Korean experience in their host countries up to the present day before concluding with an examination of the host countries' policies pertaining to ethnic or national minorities with a specific focus on language.

2.2 Socio-historical Background of *Koryo Mar*

Until its dissolution in 1991 the Korean community of the USSR was the third largest in the world after those of China and the USA. Rather than break the contemporary Central Asian Korean community into parts based on residence in the successor states of the USSR, we follow German Kim () in dividing the Korean diaspora the CIS into three parts: the descendants of Korean settlers in the Russian Far East, the descendants of those relocated to southern Sakhalin during the Japanese colonial period, then the Karafuto states, and former North Korean citizens, either former students and contract workers in the Soviet Union or refugees. It is due to the first group, which comprises roughly 80% of the total Korean population of the CIS (Yun 2004 p. 89), that this community is deemed appropriate for inclusion in the current study. We break down the examination of their background into the following sub-sections:

> 1860 – 1937 – The Korean community in the Russian Far East

> 1937 – 1939 – Deportation to Central Asia

> 1939 – 1991 – The Korean community in Soviet Central Asia

> 1991 – Present – The Korean community in the CIS

After which we conclude our discussion the socio-historical background of *Koryo Mar* with an examination of the various nationality and language policies which the community has been subjected to.

2.2.1 The Korean Community in the Russian Far East

Korean settlement in the Russian Far East took place primarily in areas known today as *Primorskij kraj* and *Xabarovskij kraj* – territories which were ceded to Russia by China in the Aigun and Peking Treaties, signed in 1858 and 1860, respectively, as a result of the Second Opium War with Britain and France (Belikov 1991). This area in which the incoming Koreans

found themselves was already settled, if sparsely at the time of its acquisition by Russia, by Chinese, Manchu-Tungus and other aboriginal groups who came to be known as the *Tazy* in Russian, and was consequently a multilingual, multicultural environment, albeit under the auspices of the Russian Empire.

The first Korean settlers recorded in the newly acquired Russian territory arrived in 1863, when thirteen families secretly crossed the border from Joseon Korea (Wada 1987; Belikov 1991). In marked contrast to the other significant immigrant population in the area, the Chinese, whose community overwhelmingly consisted of male migrant labourers the relatively even gender ratios of the Korean population and migration of whole families made the establishment of a long-term diaspora community far more viable (Belikov 1991p. 300). Over the following years the Korean population grew steadily to 1,800 people and rose abruptly when a further 6,500 fled famine on the Korean peninsula in 1869. Such official interest as this early influx of Koreans garnered was largely devoted to methods for ensuring their allegiance to the Russian Empire and the Tsar. Although use of the Korean language was never actively suppressed it initially received no official support; Russian medium schools were one measure put into place for "bringing the Russian spirit into the Korean midst" ([Nasekin 1904] in Kim 1993).

These efforts were proved to be largely successful by the Korean community's comparatively rapid integration with the Russian state. Not only was the Korean community was known for using the local language of wider communication (LWC), but for converting to the local religion, practicing exogamy and participating in Russian national service (Wada 1987). In 1871 they were granted some autonomy in the form of the founding of a Korean village, *Blagoslovennoe* (King 2001).

Despite such official recognition on the part of the state and participation in public life on the part of members of the Korean community, the greater part of the community lacked any kind of official status in Russia and engendered mixed attitudes from their new compatriots. A solution of sorts to "the Korean Problem" came about in the form of an 1891 decree naturalising the Koreans who had been resident in the Russian Far East prior to the establishment of diplomatic relations between Korea and Russia on June 25th 1884. It may be noted that it was only with the establishment of diplomatic ties between the Russian Empire and Joseon Korea that it became possible for Korean migrants to legally become fully naturalised Russian citizens.

The population of Koreans in the Russian Far East continued to grow, brought up to 60,000 by refugees from Korea following the annexation of the peninsula by Japan in 1910, an event which prompted both a strong surge of patriotism for their adopted homeland on the part of the Korean community and petitions for an accelerated process of naturalisation for Koreans. The next census to be carried out in 1923, after the Russian Revolution, revealed that the Korean population had grown again to over 100,000, made up of 34,559 ethnically Korean Russian subject and 72,258 noncitizen residents (Chey 1987 p. 63). While there is a broad consensus that the Korean population of the Soviet Union continued to grow throughout the following decades precise numbers are difficult to ascertain with scholarly speculation ranging from a scant 87,0 00 (Kozlov 1982) to "at least 250, 000" by the mid-1920s (Kolarz 1954).

2.2.2 Deportation to Central Asia

This large community eventually came to be regarded with suspicion, despite well attested loyalty to and representation in the Russian Communist Party (Hara 1987). The 1920s saw reports expressing concerns over Koreans living in the Russian Far East acting as "agents of intrigue against Soviet Russia", an attitude of antipathy which only intensified during the 1930s

as prominent Koreans were purged from the Communist Party by Stalin([Arkihpov 1926;Arkhipov 1929] in Wada 1987). Ultimately, the decision was taken to relocate the Korean population to Soviet Central Asia, allowing those who chose to be repatriated to Japanese held Korea. As a result of "Top Secret Order Number 1428 – 326cc" dated 21st August 1937 between September and November of that year, the vast majority of Korean residents of the Russian Far East were forced to migrate. Although precise numbers are unknown, most estimates suggest that between 170,000 and 190,000 people were deported (Yun 2004 p. 89).

Although Koreans were relocated into all five Central Asian SSRs (Kazakhstan, Uzbekistan, Tajikistan, Kyrgyzstan and Turkmenistan), by far the largest populations were to be found within the borders of present day Kazakhstan and Uzbekistan, although due to the then status of Karakalpakstan at an autonomous *oblast'* of Kazakhstan immediately following the deportation the community in the Kazakh SSR was by far the larger (Kimura 1987). The experience of this new migration closely paralleled that of the Korean settlement in the Russian Far East, as Koreans settling in rural areas cultivated undeveloped land before finding employment on collective farms – either Korean kolkhozes or integrated sovkhozes – both types of Soviet collective farms – as they were established with the consolidation of Soviet power in Central Asia (Kim 2008 p. 157).

The relocation of the Koreans changed not only their material conditions and access to institutions, but the very structure of their communities. Korean diaspora communities almost universally characterised as dense and favouring language preservation (Kim and Lee 2008); the early Korean community in the Russian Far East was concentrated in just two *kraj* – that is, large, peripheral administrative units, however the Korean community of Central Asia in the mid-20th century was spread over 23 *oblast'* – an administrative unit considered equivalent to a *kraj* – in five SSRs, lending truth to German Kim's summary of the new living situation of the

koryo saram that "the modern demography of the Korea (sic) population is characterized by dispersion" (1993 p. 125).

2.2.3 Koreans in Soviet Central Asia

Over the second half of the twentieth century, however, social change was rapid. The population of Soviet Central Asia as a whole became increasingly urbanized (Kho 1987; Yun 2004); in 1926 only 10.5 per cent of the Korean population of the USSR at large were urban dwellers with the rural majority living dispersed over a very large territory. The proportion of urban dwellers, though rose to 78 per cent by 1979 (Kimura 1987 p. 96). More remarkably, the proportion of Koreans living in cities was even higher in the Kazakh SSR, where they made up 84.4 per cent of the local Korean population by 1970 (id.). This urban migration has been attributed to the greater opportunities for educational and social advancement available in the cities, which rural areas lacked (Yun 2004 p. 90). Furthermore, it may have been a contributing factor to the speed and extent of the linguistic and cultural Russification of the *Koryo Saram*. As Chey remarks (1987 p. 69), Koreans living in rural areas largely practice endogamy, but exogamy was much more common among urban *Koryo Saram*. Also, as members of the Korean community took on social roles in spheres other than agriculture including industry, the government and education the socio-economic profile of the community changed dramatically (Kim 2008 p. 161). Korean identity also found expression in the form of the growth in prominence of the Korean theatre and the establishment of the Korean newspaper *Lenin Kichi* (Lenin's Banner). While the theatre was originally founded in Vladivostok in 1932, it moved to Kyzylorda in 1937 with the deportees. The peripatetic theatre relocated to Ushtobe between 1942 and 1959 before returning to Kyzylorda for ten years and final settling in its current location in Almaty in 1968. The paper, on the other hand, was founded after the relocation to Central Asia and is published to this day under the title *Koryo Ilbo*.

In addition to such official manifestations of Korean identity, Soviet ethnographers record retention of Korean foodways, culturally specific festivals or rites of passage (*tol* – first birthday celebration, *hwankap* – sixtieth birthday celebration, reverence for ancestors etc.) and some material culture (*ontol* under-floor heating, clothing etc.). Despite this, Koreans resident in Central Asia continued their partial and pragmatic assimilation to the dominant culture as may be seen from their adoption of Russified given names, including patronymics, and an alternative, Slavic rather than Korean language-based ethnonym : *sovetskii koretsi* – "Soviet Koreans".

2.2.4 The Korean Community in the CIS

While the Korean community of the USSR was one of the largest in the world, the creation of the modern successor states in 1991 has fragmented it. Nevertheless, Korean communities of over 100,000 survive to this day in several CIS nations, including Russia, Uzbekistan and Kazakhstan (MOFAT 2011). The reactions of the *Koryo Saram* to this all across the CIS have been characterized by yet more rural to urban migration, a revival of farming practices which were impossible under Soviet collectivisation and a quick adaptation to the new market economies, perhaps aided by contact with South Korean entrepreneurs (Back 2001). The collapse of the Soviet Union has also lead to the development of closer ties between South Korean businesses and the CIS nations of Central Asia and the commencement of missionary work on the part of South Korean Christian churches.

While both traditional and local Korean culture in the CIS is enjoying something of a renaissance compared to its position in the assimilationist USSR, the language is at present in a near moribund state, apparently conforming to Fishman's (1965) three generational pattern of language loss. This is particularly noteworthy, since this distinctive pattern of language shift

dates from the relocation to Central Asia rather than the foundation of the transplanted community, as would be more conventional.

2.2.5 Policy and the *Koryo Saram*

We now examine the relationship between the *Koryo Saram*, their language and their host communities through the lens of policy, a task complicated by the fact that, more often than not, the national identity of the host country of the *Koryo Saram* has changed around them, leading to a change in the parameters of the relationship between minority communities and the state, as during the Russian Revolution and then the fall of the Soviet Union.

In his examination of *Koryo Saram* historiography, German Kim notes that the life of Koreans in the Russian Far East has been romanticised in the contemporary understanding to such an extent that "a stereotype holds that misfortune only befell the Koreans after the deportation, or shortly before it" (Kim 1993 p. 136). While this may be and overly nostalgic view of the lives of people driven to migration by economic hardship or political repression, who were furthermore denied the opportunity of citizenship in their adoptive country, this positive stance seems wholly tenable when it is applied to the Korean language. Korean was widely and officially condemned if not supressed in Imperial Russia, as were all minority languages in the state whose guiding principles were encapsulated in the slogans "one tsar, one religion, one language" and "autocracy, orthodoxy, nationalism" (Lamont 1945). Nevertheless, it remained a viable and high-profile language of intra-ethnic communication in the Russian Far East to the extent that, despite a good deal of cultural assimilation, one governor-general of the Ussuri Region of Russia is recorded as complaining that "the majority of Koreans do not know the [Russian] language" ([Anosov 1928 p. 8] in Chey 1987). It is possible that this refers only to the standard, literary variety of Russian, though, and a contrary view to the effect that a Russian based pidgin, that is a pidgin with Russian taking on the role of lexifier language, was

the principal means of inter-ethnic communication in the Far East in the early 20th century has been proposed (Bellikov 1991 p. 303).

In the early days of post-revolutionary Russia, Korean had a brief respite from official censure with the implementation of the Soviet policy of *koren'izatsija* (literally "taking root", but frequently translated as "indigenization") whereby minority languages were consciously developed with regard to their own traditions and principles (Ornstein 1968; King 1997). This policy progressed to the extent that books were published in Korean in the USSR prior to the deportation, however, this situation was not to last. A policy favouring the use of Russian over the development of other languages was introduced in the 1930s and maintained through the Soviet period (Ornstein 1968).

In light of these policies and ideologies, it comes as a surprise that Korean medium education, though not officially sanctioned until 1923, was common in the Russian Far East, whereas it was non-existent Central Asia. Instead, post-deportation Korean children received their education from so-called "republic schools" where lessons were carried out in the, generally Turkic, language of the republics, while Russian was taught as a second language. Mainstream, Russian medium schools were also present in Central Asia and the pragmatic concern that knowledge of Russian was a *de facto* pre-requisite for participation in Soviet culture led Koreans to seek education at Russian medium schools. Such Korean language education as was available was largely carried out by unofficial Korean schools on collective farms or, one of the standard varieties of Korean could be studied as a second language up to middle school level in selected fourteen schools across the Soviet Union. While some research was carried out on Korean education in Kazakhstan and Central Asia specifically ([Khan 1981] in Kim 1993) the state cannot be honestly said to have actively contributed to the maintenance of the Korean language in Central Asia.

In the face of this lack of Korean language education and the dispersion of the community, the linguistic Russification of Koreans progressed at an even greater speed than during the period over which an official policy of minority assimilation into the Empire was in place when the Korean community was concentrated in the Russian Far East in the late 19th and early 20th centuries. Comparing the adoption of the Russian language among Koreans up to 1979 with other ethnic groups in Central Asia, Kimura (1987 p. 98) comes to the conclusion that "the linguistic Russification of Koreans is proceeding more rapidly than that of other peoples". This observation is supported and another contributing factor posited by Chey:

> Koreans in the USSR do not form any administrative unit, unlike other ethnic groups who reside in native territories now classified as republics or autonomous regions. This lack of support within the Soviet power structure places them at a disadvantage…The lack of official administrative support hastens the assimilation of Soviet Koreans (Chey 1987 p. 73)

Despite the lip-service paid to minorities in the Soviet Union, the Soviet period was undoubtedly a period of pervasive Russian influence and deliberate Russification over the whole of its territory. Although the so-called "nationalities question" was frequently addressed by policy-makers and in such fundamental policy documents as the "Soviet Communist Party Program", groups other than Russians were rarely mentioned specifically, that is by name, and the degree of social mobility (urbanization, access to higher education, access to communication media etc.) enjoyed by the various ethnic groups of the USSR has been shown to have been strongly correlated with their degree of Russification (Armstrong 1968).

With the fall of the Soviet Union in 1991 came the resurgence of national and ethnic identities along with a blossoming of heritage languages as the newly formed successor states of the

USSR re-aligned their language policies to reflect their new identities. The effects which the diminishing role of Russian on the governments and identities of Central Asian countries has exerted on the Korean community and their language, *Koryo Mar*, remain to be seen. However, survey data reveals that less than one per cent of *Koryo Saram* respondents consider any of the Turkic languages of Central Asia their first language, the vast majority of even first generation deportees consider Russian to be the language they are most comfortable using (Yun 2004).

To close this section we now offer a brief summary of the policies pertaining to language and minorities in the countries hosting the two largest transplanted Korean communities in Central Asia, Uzbekistan and Kazakhstan.

In both nations the primacy of the "titular" language of the new nations, Uzbek and Kazakh, was guaranteed as they were granted official, national language status. They differ from one another in that absolutely no special provision has been made for Russian that has not been granted to other minority languages in Uzbekistan since its independence, whereas Russian was classified the "language of interethnic communication" immediately after Kazakhstan's independence was achieved and later granted "official language" status, allowing it to be employed in state institutions and local self-administrative bodies, although use of Kazakh for official purpose is strongly promoted (Landau and Kellner-Heinkele 2001).

In Uzbekistan linguistic minorities have enjoyed the right to native language medium education since the passing of the 1997 Law on Education. Although there are no Korean medium schools, the Korean language is widely taught as a second language (Landau and Kellner-Heinkele 2012), and a cultural centre funded by the South Korean government has been active in Tashkent since 1992.

With respect to minority languages, in Kazakhstan legislation is in place encouraging the preservation and development of all the languages found in the country and since the formation of the "Assembly of Peoples" in 1995, the state has been concerned with the establishment and running of cultural organisations representing each of the ethnic groups of Kazakhstan. The 1997 Law on Languages guarantees freedom of language in communication and creative activities for all the minority languages of Kazakhstan and non-Kazakh or Russian medium education "when necessary" (Landau and Kellner-Heinkele 2012). Given the advanced state of the language shift among the *Koryo Saram*, though, there is not state-funded provision of Korean medium education in Kazakhstan. In recent times, it is the South Korean government which has been the main provider of Korean medium education since the establishment of the Korean Education Centre in Almaty in 1997, although due to the lack of official standardisation or codification of *Koryo Mar* the variety can be considered heteronomous with the standard language of either South or North Korea.

2.2.6 Language Attitudes and *Koryo Mar*

While it does not appear that any research has been carried out with an explicit focus on the language attitudes of the *Koryo Saram*, it is possible to make some inferences from works on other topics, demographic trends, and observations recorded in the field notes for this project.

Perhaps the most explicit statement of speaker attitudes comes from King's 1992 paper on archaisms and innovations in KM (then called Soviet Korean) the translation of which we quote in full:

> As for our speech, it is a jumbled language, a base tongue,
> and we talk raucously and chaotically. But Seoul speech is a
> noble language, spoken delicately.

This kind of negative evaluation is made all the more concrete when the reality of language shift in Central Asia is examined.
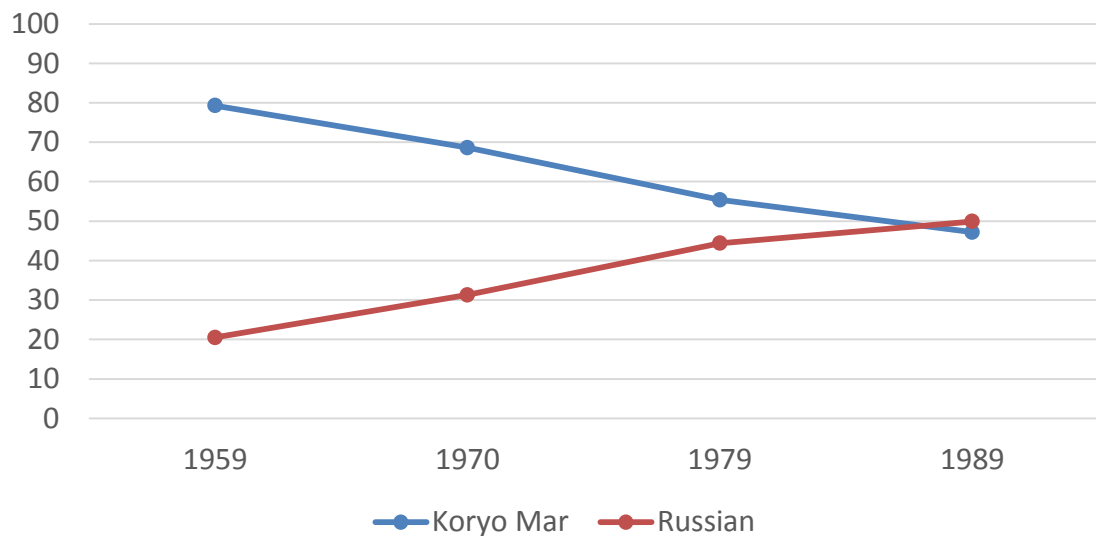
Figure 1: Language Shift from KM to Russian in Central Asia in Percentage of Speakers' Primary Language

The above graph is produced on the basis of data made available by Yun (2004) and illustrates the proportion of the Korean population of the contemporary CIS who consider their language of primary communication to be KM and those who consider their primary language of communication to be Russian. Such a pattern of language shift is highly suggestive that KM is not valued as a heritage language and intergenerational transmission has broken down somewhat, whereas the popularity of the regional language of wider communication has risen. This phenomenon was not explicitly addressed by any of the KM speakings consultants who participated in this study although many younger *Koryo Saram* expressed regret at either being able to communicate with older generations only in Russian or in CSK.

Following the dissolution of the Soviet Union and the new language policies examined above, the role of Russian has decreased. It has been suggested that the post-Soviet language policy of promoting the languages of the titular ethnicity of the successor states of the CIS, for example Uzbek and Kazakh, acts as a 'push factor' in encouraging the return migration of

Koryo Saram to Korea (Jo 2017). It is possible to infer from this that such languages are negatively evaluated by the *Koryo Saram*. Indeed, it is notable that only a very low proportion of them claim any kind of proficiency in these languages.

From other sources, then, we may infer that both Russian and the standard variety of the ROK (no data is available for the standard language of the DPRK) are positively evaluated by *Koryo Saram* while their heritage language KM and other languages used widely in Central Asia are less favourably evaluated. These conclusions are broadly in line with observations recorded during field research for this project. One consultant described Seoul Korean as *chistij* 'clean' and, by implication, superior to her native variety. Furthermore, a commonly expressed attitude was one of shame or embarrassment directed against themselves since they were unable to make themselves understood in Korea using KM rather than, for example, frustration at interlocutors who could not understand them.

Explicit references to multilingualism appear quite infrequent and I recorded no reference from any consultant on the influence of Russian or any other regional language on the pronunciation or grammar of KM. Conversely, consultants seem to consider their use of nonce-loans in KM to be a case of mixing in Russian words. Finally, while observing a group discussion at the Korean Cultural Centre (KCC) in Almaty one participant explicitly connected KM and VYK in the form a complaint that they had great difficulty understanding guest speakers at the KCC, but found communicating with *Cosencok* on a trip to Yanbian much easier.

From the foregoing we may form some impression of language attitudes towards KM in the Central Asian Korean community. It has been negatively evaluated to the point that language shift to Russian is on-going. The loss of KM as a community or heritage language is felt more keenly by younger, non-KM speakers than older KM-speaking *Koryo Saram*. Nevertheless, when these speakers choose to learn a variety of Korean it is almost invariably CSK rather than

KM. In the Central Asian context multilingualism is the norm and consequently not the subject of spontaneous explicit evaluation by *Koryo Saram* although the mixing of loanwords into KM may contribute to its negative evaluation. Although no explicit evaluation of Russian was recorded, it may be inferred from the general trends reported in other sources and the observed use of Russian to the exclusion of KM in all domains of use that it is more positively evaluated, even if only on practical grounds, than KM.

2.3   Socio-historical Background of Yanbian Korean

China and Korea have shared borders since the beginning of their respective recorded histories. Indeed, the borders of older Korean civilisations, such as Goguryeo, extended into the present day PRC, currently home to the world's largest Korean diaspora community. Such Korean population as was settled in North-Eastern China from antiquity until prior to the nineteenth century are said to have assimilated to other regional cultures, such as Han Chinese or Manchurian (Paio 1987 p. 47) and is dwarfed by the scale of subsequent movements of population (Im 2003), therefore we do not examine it further here. Later migrations from Korea to China are conventionally divided into as many as five periods: 1860 to 1881, 1881 until the annexation of the Korean peninsula by Japan in 1910, 1911 to 1920 (the so-called "free migration era"), 1921 to 1931and 1931 to 1945 (the "forced community migration era") (Im 2003 p. 49).   For our purposes here, since settlement in China rather than migration is our main focus, we divide our examination of the socio-historical background of the Korean community in Yanbian into just two parts, the first covering the period between starting with the first modern wave of migration from Korea to China in the 1860s and lasting until the proclamation of the PRC in 1949 and the second looking at the *cosencok* in China since that time to the present day.

2.3.1 Koreans in China 1860 - 1949

The origins of the contemporary *cosencok* community are conventionally said to lie in the mid-19th century (Cui 1990; Piao 1990; Tai 2004; Yun 2004). Similarly to the K*oryo Saram,* this community originated in greater part in North Hamgyeong Province and migrated to new territory out of necessity brought on by famine. Starting from a small community in a settlement known as *Hoylyongpong* in 1862, the community soon expanded to 470 households, roughly two thousand people, spread over 30 villages, largely just to the North of the Tuman and Yalu rivers, neighbouring Northern Hamgyeong Province and Northern Pyeongan Province respectively, by 1870. The Qing administration seemed keen to encourage further migration and opened an area only sparsely settled by Chinese to the North of the Tuman River to migrants in 1885with the result that, as the turn of the century approached, the Korean population of the area had reached roughly 37,000 (Piao 1987 pp. 48-49). The primary occupation of the new settlers was agriculture, either for subsistence on previously uncultivated land or as tenant farmers, initially largely of rice but later expanding into apples.

The Japanese annexation of Korea saw the Korean population of China explode, as it had in the Russian Far East; before 1920 approximately 459,400 Koreans were resident in China. This influx of people differed from that of the nineteenth century in that people came from almost all of the Korean territory were represented in it, although a significant proportion still came from the nearby Northern provinces. Flight from the occupied Korean peninsula for political reasons was not the sole motivating factor for the growth of the Korean population in China during the Japanese colonial period, with contemporary surveys citing economic hardship as the majority's main impetus for migration from Korean prior to the 1930s (Im 2003 p. 65). Japanese plans to use Manchuria as a supply base for the invasion of China were realised by forcibly relocating vast numbers of people from the Korean peninsula for the purposes of

agricultural and industrial labour. Although incomplete, the statistics available for 1945 record

at least two million Koreans in Manchuria alone.

During the Japanese occupation of North Eastern China, the growing population of *Cosencok*,

were still largely occupied in agriculture, but also began the slow process of urbanisation

around this period. It was in more urban areas that the occupation was most strongly felt, by

means of a high degree of surveillance of the Korean population as well as the imposition of

Japanese education. Here, the movement resisting the Japanese occupation flourished,

despite the casualties brought about by its suppression.

Immediately following the end of World War II, however, many people naturally chose to be

repatriated to the Korean peninsula and the population fell once more to between seven and

eight hundred thousand. This reduced community was nevertheless involved in the early

Chinese Communist Party (CCP), and its representatives have played a role in national

government since the formation of the PRC, greatly contributing to the development of the

rights of minority nationalities.

On a point of demography, it must be noted that it is emphasised here that, although the

overwhelming majority of the *Cosencok* seed community, particularly the early settlers in

Hoylyongpong and just inside the Chinese border, had their origins in Northern Hamgyeong

Province and the Yukchin area, the early settlers from Korea did not originate there exclusively.

In addition, the successive waves of migration brought about by the Japanese occupation and

the division of Korea lead to both a rise in the proportion of Korean migrants originating from

other areas of the Korean peninsula, most especially from the Southern provinces of Korea and

a greater diversity of points of origin in general.

2.3.2 The Korean Community in the YKAP

The Korean community grew and flourished in China in the early twentieth century, despite the instability brought about by the Chinese Civil War and Japanese invasion. Moreover, it continued integrating well with its host state. Frequently described of all the minority nationalities as the "model minority" the Korean community has been officially recognised as one of China's minority groups since 1927 (Piao 1990 p. 70), therefore all subsequent legislation and policy pertaining to minorities in general has been applied to Koreans as it has to Tibetans, Mongolians and Zhuang etc.

Following the Revolution of 1949 and the CCP pursued a policy of granting autonomy to minority nationalitie. The practical realisation of this recognition was the establishment of the Yanbian Korean Autonomous Prefecture on the third of September 1952 which conferred certain additional rights including but not limited to the right to establish autonomous regulations on the basis of cultural heritage, the right to manage regional finances and the right to independently manage education, science and culture in the region (Cui 1990 pp. 82-83). This last point is particularly significant for the maintenance of the Korean language, since it allows for the official provision of Korean medium education up to tertiary level and the recognition of Korean alongside Putonghua Chinese as an official language within the Autonomous Prefecture.

We note here, that there are other Korean autonomous regions in China, for example the Changbai (*cangpayk*) Korean Autonomous County (*cachihyen* (the administrative unit smaller than the *cikwu* – "prefecture" or *cachicwu* – "autonomous prefecture" in the PRC), however due to the relatively small populations of these other autonomous areas and the centrality of the YKAP to the *Cosencok* we shall not examine them in any more depth.

The transformation of the situation of Koreans in China following the foundation of the YKAP serves to demonstrate that the nationalities policy of the CCP is more than just 'window dressing'. Since then the social mobility of the *Cosencok* has increased dramatically as industrialisation, the mechanisation of agriculture and access to education increased rapidly. While the initial Korean settlers of the area were almost universally occupied by agriculture, Chinese Koreans are now represented in all walks of life, from industry to the arts to the civil service (Im 2003 p. 305). In addition, various aspects of traditional Korean culture including the retention of traditional dress, foodways, pastimes, festivals and systems of interpersonal relations are maintained in daily life (Zheng 2010 pp. 568-570).

Although Han Chinese now make up the majority of the populations of both Yanbian and its capital city, Yanji, the region retains a high proportion of Koreans, who form an absolute majority of the population of such smaller cities as Tumen and Ryongcheong. In addition to retaining the Korean language both in an official capacity and as a vernacular language of daily communication, it has appeared in print and broadcast media since 1947 (Ma 1994). Other than the language, Korean cultural traditions and festivals, suchas *selnal* and *chwusek*, are maintained by the majority of the Korean community in China (Yun 2004 p. 81) who also enjoy a high level of official support for their language with ts development and maintenance guaranteed by the YKAP national autonomy laws (Heo 2013 p. 162).

These manifestations of culture and identity are not surprising when the Korean population of China as a whole is taken into consideration, since it represents the largest Korean diaspora community in absolute terms in the world with in excess of 2.7 million members, amongst whom knowledge of the Korean language is near universal (MOFAT 2011). It must be acknowledged, though, that the first signs of pressure on the Korean language have been observed as the Korean population of Yanbian has been undergoing a period of slight negative growth since the mid-1990s. This phenomenon has been attributed to multiple causes, highly

educated female population being more selective in choosing their partners (Jeong 2013) and a demographic imbalance in Yanbian whereby ethnic Korean men "of marriageable age" outnumber ethnically Korean women by twenty two to one is leading to a declining birth-rate among the *cosencok* just as the mortality rate among those over sixty is rising. In addition to this, the surge in internal migration within the YKAP from the Korean speaking majority villages to majority sinophone cities such as Yanbian in the 1990s (Lee 2011 p. 310) or, to a more limited extent, population movement out of the YKAP to other regions of China is leading to a slight decline in the use of Korean (Zhou 2004 p. 92-94)

2.3.3 Policy and the *cosencok*

Koreans are officially recognised by the CCP as one of the 56 nationalities of the modern PRC, and thus constitutionally guaranteed equal rights with other nationalities, including the majority Han Chinese, which includes the right to develop a national language and system of writing (He 1990 p. 4; Cui 1990 p. 81). This has been presented as a stark contrast to the treatment of Koreans by pre-revolutionary regimes in China, each of which enacted policies of national suppression and assimilation – Sinification – to a greater or lesser degree (Piao 1990 pp. 68-70).

Given its diversity, the "nationalities question" has long been as prominent in China as the Russian Empire and Former Soviet Union. On a national level, policy is formulated towards all minority nationalities rather than Koreans, Mongols, Manchus etc. specifically; therefore we divide the following discussion of policy into two parts. The first describes the national minority policies of pre-and post-revolutionary China, with a focus of analysis being their relevance for the Korean community, while the second focuses on the language policies and legislation enacted in the YKAP.

Policies towards minority nationalities in the PRC have not been entirely consistent since the country's establishment in 1949. It has been proposed that this vacillation has been due to changing ideology within the CCP as the competing political goals of the accommodation of minority nationalities or their integration into both the CCP and PRC came to the fore, and that the changing minority nationalities policy has been directly reflected in policies affecting minority languages (Zhou 2003 p. 36).

From its inception, the PRC guaranteed in the Common Program of 1949, its provisional constitution, full equality of nationalities and established its support of autonomous governments for minority communities. It further stated that every minority nationality was free to develop its own language and writing system alongside its traditional culture – rights cemented by the1952 "Policy Towards Nationalities of the People's Republic of China" (China 1953). Thus the period immediately following the establishment of the PRC may be characterised as one of linguistic accommodation in which minority languages were promoted. This early position was strongly influenced by, and indeed similar to, early Soviet language policy, which emphasised the possibility of national and socialist unity co-existing with linguistic diversity. As with the case of the USSR, though, this soon gave way to the promotion of a single majority language in the service of nation building. It has been argued that these changes in policy in the PRC were based on the ideologies and attitudes prevalent within the CCP regarding the pace of evolution from socialism to communism (Zhou 2012), however an opposing view that this change in policies was simply based on Han Chinese linguistic nationalism has also been put forward (Harrell 1993).

Contemporary Chinese language policy is broadly covered by two articles in the constitution of the PRC and the specific piece of legislation "Law of the People's Republic of China on Language and Writing Systems in Common Use". In keeping with current accommodationist thinking, these national laws are supplemented by regulations devised by the regional

government of the Yanbian Korean Autonomous Prefecture (YKAP), the most important of these being "Autonomy Regulations for the YKAP", "Korean Education Regulations for the YKAP", and "Korean Language and Writing System Regulations of the YKAP" (Zhang and Li 2007).

1952 saw both the founding of the Yanbian Korean Autonomous Prefecture, and the passing of guidelines for the governance of autonomous regions which had specific relevance to language. While the "Guidelines for Regional Autonomy for Minority Nationalities in the PRC" supported the use of minority languages for official purposes they also required the adoption of another written language for any other minority nationalities resident in autonomous regions who do not use its titular language – a policy which *de facto* guaranteed the continued official use of Chinese alongside Korean and all other minority languages in autonomous regions.

From 1957 the era of 'the Great Leap Forward' began, with the goal of developing communism and national convergence in the CCP. The manifestation of this policy most relevant here is the explicit policy of encouraging minority communities in China to shift from their heritage languages to Chinese since "it was considered a short-cut to language convergence in the fast lane to communism" (Zhou 2003 p. 63). This had little impact on the vitality of Korean in China, though. It was not until 1978 that the era of integrationist minority and linguistic policy came to an end in the PRC.

As mentioned above, language education has fallen within the remit of the YKAP regional government since 1952, although it may be noted that Korean language education also enjoys a long unofficial tradition in China dating back to 1906 and the founding of Yanbian's *sejenseswuk* in1906 (Heo 2013) and in addition has the distinction of having had China's first minority nationality university, Yanbian University, established in the region in 1949. While Korean medium instruction has been a great tool of language maintenance, the time devoted

to Chinese language studies even in ethnic Korean schools has been steadily growing over the later part of the twentieth century. Indeed, some scholars have expressed concerns that the bilingual education policies which have guaranteed a place in the classroom for Korean amongst other minority languages alongside Chinese are merely transitional and serve to promote the ultimate goal of having minority communities shift from their heritage languages to Chinese (Zhou 2012). Survey data, though, indicate that there need be little concern on that account, since four fifths of the *cosencok* population consider themselves fluent and literate in Korean (Yun 2004 p. 80).

2.3.4 Language Attitudes and Vernacular Yanbian Korean

In contrast with the *Koryo Saram* questions relating to language and identity which comprise language attitude have been explicitly investigated with regard to the *Cosencok*. Most recently this has taken the form of a large scale survey carried into the realities of the use of Korean in China (Yun et al 2012). Roughly eight nine per cent of respondents to this survey with *Cosencok* identity identified some form of Korean as their primary language, with eight per cent of respondents claiming to speak the regional variety of Korean. These respondents were less positive about the future of this variety, though; the vast majority believed that Chinese would become the primary language of the *Cosencok* over time with over half of them believing that VYK would either fall out of use or be replaced by CSK. This appears somewhat contradictory, since we may infer a positive attitude towards the language and its continuing transmission and maintenance from such high speaker rates along with observation of VYK in the home domain as part of this project and self report of the use of VYK in intergenerationally (Yun et al. 2012 p. 195). The positive evaluation of VYK may also be inferred from other sources indirectly, for example the on-going standardisation of a variety independent from the standards of either the DPRK or the ROK bespeaks a separate linguistic identity and one which is regarded as equal to the prestige varieties of those countries. We note further that in

explicit comparison with VYK, negative evaluations of CSK were provided, most expecially as giving the impression of "cunning" and "weakness".

In common with the situation of *Koryo Mar*, multilingualism appears to be assumed to the extent that it is rarely explicitly evaluated. Younger consultants in particular expressed satisfaction at having the communicative resources of multiple languages at their disposal, both for engaging with a wider range of Chinese and Korean media and to use among themselves. While reference was made to Chinese in terms of its utility by several consultants I have no record of it being evaluated as superior to any variety of Korean. The attitudes revealed by the foregoing stand in marked contrast to those in section 2.2.6 pertaining to the evaluation of KM by *Koryo Saram* as somehow inferior to CSK and their local language of wider communication.

2.4 Conclusion

We have demonstrated, then, that despite their similar temporal and geographical origins and similar early experiences immediately following their initial migrations, over the course of the twentieth century the Korean communities of Central Asia and China have been subject to very different conditions in their host nations. The most significant of these differences in the diaspora experience of these groups is the secondary transplantation of the *Koryo Saram* from the Russian Far East to Central Asia, compared to the relatively settled expansion of the Korean community in the North East of China. Moreover the *Cosencok* and *Koryo Saram* were granted wildly dissimilar levels of autonomy and official support and recognition in their host nations. Not only were the relationships between host nations and transplanted communities dissimilar, but these communities also maintained ties with the Korean peninsula to differing extents.

There are also characteristics held in common by these communities other than their origins as well as characteristics which find parallels in the experiences of other transplanted communities. We deal with these characteristics in turn. Despite pressure to assimilate into a dominant culture, both transplanted Korean communities have to a very great extent have retained Korean identity, alongside their identities as *Cosencok* and *Koryo Saram*, and continue to perform many Korean cultural practices pertaining to the life cycle, agricultural calendar and interpersonal relations. A further aspect of Korean identity retained by these communities to varying degrees is the Korean language although, as a result of their diaspora situations, these communities are now made up overwhelmingly of bilingual or multi-lingual individuals.

Turning to the broader picture of transplanted languages in general, their migratory trajectories are similar to many others. Edwards' 2010 taxonomy classifies minority communites and their languages with regard to their several of their geographical and social characterstics. On its basis, we may draw direct parallels between *Koryo Saram* and "non-unique, non-adjoining, non-cohesive…immigrant minorities" and *Cosencok* and the so-called "non-unique, adjoining, cohesive…immigrant minorities", since the former community originated in Korean, is currently living geographically remote from the community there has minimal direct contact with it, while the latter community is originated in Korean, but is currently living in a contiguous state with which it enjoys some degree of direct contact. Examples of other communities similar to the *Koryo Saram* include "scattered immigrants of European origin in 'new world' receiving countries" (Edwards 2010 p. 81) such as the Greek community in Australia and the German community in the USA, while communities similar to the *Cosencok* are characterised as "scattered immigrants in neighbouring states" (Edwards 2010 p, 81), for example the Mexican community in the USA or the Algerian community in France. In the most general of social terms, we may classify these varieties of Korean substrates to the regionally dominant LWCs with which they are in contact, Russian and

Chinese, which take on the role of the superstrate. In contrast with situations in which pidgins

and creoles developed from contact between superstrate and substrate languages, though, we

observe that in the case of the varieties of Korean it is the substrate rather than the

superstrate language which is being used in the transplanted situation. In conclusion, it may

be noted that these broad taxonomies do not capture the full social histories or contemporary

situations of the transplanted Korean communities under discussion here as adequately as the

foregoing overview.

We now go on to review research carried out on the varieties of Korean used in China and

Central Asia.

3   Literature Review

3.1   Introduction

In the previous section, we introduced the Korean diaspora in general, before providing a more

detailed socio-historical background to two of the older transplanted Korean communities,

those of Central Asia and China. A number of commonalities between these communities

were identified; most notably the temporal and geographical origins of the founding seed

communities, however, the historical experiences of these two communities were markedly

different. While the twentieth century saw the *Koryo Saram* subject to internal exile and

official indifference in the Soviet Union, the *Cosencok* have remained geographically and

socially stable, while generally prospering under the PRC's minority nationalities policies.

Although the current living conditions of the Korean communities in Central Asia and China

differ greatly in many respects, both communities retain elements of Korean identity. This has

been reflected in the retention of certain elements of traditional Korean culture, foodways and,

of course, language. In the following chapter, we shall critically examine the research which

has been carried out on the varieties of Korean used in Central Asia and China in order to both situate this comparative study within the current research and to refine the research questions which it shall address by identifying the specific issues overlooked by earlier scholarship.

Thus, we shall divide the body of the chapter into four sections. First, we briefly summarise the position of transplanted varieties of Korean and their source dialects in traditional Korean dialectology before going to critically evaluate such linguistic research as has been carried out on *Koryo Mar* and Yanbian Korean in turn.

3.2 Traditional Korean Dialectological Perspectives

3.2.1 The Transplanted Varieties

Little can be said of the dialectological position of transplanted varieties of Korean. A systematic review (Silva 2010 p. 315) of academic databases revealed that dialectological variation is a relatively little studied area in Korean linguistics. This situation is compounded in the case of transplanted varieties of Korean which have been largely or entirely overlooked from the first appearance of classificatory systems of Korean dialects in the early 20th century to the present day.

The earliest Korean dialect taxonomies, for example that of Lee Kuk-lo (1932) and the famous, influential dialect classification of Ogura Shinpei (1940), only concern themselves with the territory of the Korean peninsula and its outlying islands. This attitude prevailed in Korean dialectology until the 1980s when scholarly attention was finally turned to Korean diaspora communities. Even then, the discussion surrounding these transplanted varieties was dominated by peninsula dialects, with which they were often conflated:

> The spoken Korean of Yanbian students has three attributes –
> (sic) a strong North Hamkyong accent, North Korean terms
> and expressions, and Korean transliterations of Chinese

characters. While Koreans in Liaoning Province tend to have
a Pyongan accent, many in Heilongjiang Province speak
Kyongsang dialect. (Lee 1986 pp. 147 - 148).

This excerpt very well demonstrates that transplanted varieties of Korean have received only cursory attention from the field of Korean dialectology but, in the rare cases where they are directly addressed, there has been a strong tendency to explain them only in view of the seed communities' geographical origins and, thus, conflate them with the contemporary descendants of their input dialects.

Such an analysis of Yanbian Korean still finds a certain amount of support in the contemporary research (Sohn 1999 p. 59), whereas, while the descent of *Koryo Mar* is emphasised, its divergence from contemporary peninsula varieties is noted too, albeit not to the extent of categorically and unambiguously considering it wholly separate from the notional group of dialects forming a continuum in the North East of the Korean peninsula (Yeon 2012 p. 168).

Contrary to the position laid out above, other research has strongly argued that the varieties of Korean spoken in Central Asia and China are separate varieties from those spoken on the peninsula (Pak 2005; Jeong 2010; Jin 2013 etc.), with King and Yeon emphatically maintaining that "*Koryo Mar* is not simply Hamkyeong dialect transplanted into foreign soil" (1992 pp. 100-101).

It must be noted here that the idea that the geographically contiguous variety of Yanbian may be amalgamated with its neighbouring dialect on the Korean peninsula is particularly persistent, due to both the origins of its seed community and its geographical position. This concept is accepted to the extent that when dialect data is being collected for the purpose of describing the language of the whole traditional North Eastern dialect zone of the Korean speaking area, consultants may be drawn from the Korean population of China, even if their family have been resident in China for several generations (Choi et al. 2002). While it is

67

absolutely not to be suggested that the input or source forms of the dialects from which

transplanted varieties of Korean descend are not reflected to some extent in their

contemporary forms, we assert that these varieties ought not be classified as the same dialect

as their non-transplanted sibling varieties as doing so ignores several generations'

development in conditions drastically different from those which obtain on the Korean

peninsula. Furthermore, with specific regard to the case of China, it cannot honestly be

claimed that the Koreans resident in the YKAP form one homogenous speech community with

the citizens of the DPRK residing in Northern Hamgyeong Province. Contact and interaction is

essential for the maintenance of coherence within a speech community and, while they remain

geographically proximate and political borders rarely represent a border between speakers of

the same language in real terms, relations between China and Japanese held Korea, then

subsequently the DPRK have precluded all but trace levels of face-to-face contact between

Korean speakers in these countries for over one hundred years.

It is also problematic that there is no absolutely universally agreed upon classificatory system

for the Korean peninsula dialects, so conflation with a given dialect has no consistent meaning

and varies depending on the definitions and nomenclature used by specific authors. In other

words, the degree of similarity the transplanted dialects are said to have or even the particular

linguistic features which they share with peninsula dialects, which in turn lead to their

classification as the same dialect or part of the same group of dialects, varies by author.

Although not strictly within the tradition of Korean dialectology, it is worth noting that the

Korean communities of both Central Asia and China have their own nomenclature for their

varieties of Korean, alongside their own ethnonyms, which are not necessarily the same as

those imposed by scholarship carried out in the Korean tradition. The very fact that Central

Asian Koreans hold their language so distinct from Korean as to refer to it not as a dialect, but

as a full language (*Koryo Mar* – Koryo language) is suggestive of its linguistic distance from

even non-standard peninsula varieties of Korean. Also of particular note is the fact that

Korean is known generally as "*cosene*" or "*cosenmal*" in Yanbian, which could be interpreted in

one of three ways: either it may be an act of identification of the variety used in China with the

standard language of North Korea, where the term *cosene* is preferred over *hankwuke*, with an

earlier, stage of Korean history, prior to the division of the peninsula when the word *hankwuk*

was coined or finally it may be an act of identification with the PRC since *cosene* is simply a

Korean reading of the official Chinese designation for the language (朝鮮語 - cháo xiǎn yǔ).

Given the confused situation of Korean dialect research, it does not come as a surprise when

some researchers conclude that the methods behind traditional Korean dialect classification

may "have some basis in the characterising features of the language, but they are also to a

certain extent constructed simply for convenience of description" (Lee and Ramsey 2000 p.

313). We argue here that an examination of the linguistic features of transplanted varieties of

Korean leads to the more nuanced position that the varieties of Korean to be found outside of

the Koreas have undergone innovation and retained archaism differently from those on the

peninsula and constitute separate varieties of Korean, although some features of the dialects

upon which they were originally based are still present in their contemporary forms, since one

hundred and fifty years is a comparatively brief period of independence in historical-linguistic

terms. Therefore, some knowledge of these dialects is necessary for understanding the

contemporary transplanted varieties. In sections 2.2 and 2.3 we identified the geographical

origins of the seed communities of the Chinese and Central Asian Korean communities, here

we go on to examine the research surrounding the varieties used in these areas – the

transplanted varieties' input dialects – and the issues which arise from it.

3.2.2 Transplanted Varieties' Input Varieties

The lack of a universally accepted taxonomy of Korean dialects naturally complicates the discussion of the source dialects of the transplanted varieties of Korean spoken in China and Central Asia. The designations "Hamgyeong dialect", "Northern Hamgyeong dialect", "North-Eastern dialect(s)", "Myengchen-Kilcwu dialect", "Saeup dialect" (cf. Kim 1986; Kho 1987 p. 102), "Yukchin dialect(s)" and "Yukeup dialect(s)" are all used to describe the main source dialects of the seed communities from which *Koryo Mar* and Yanbian Korean developed.

Clearly this abundant terminology has the potential to lead to some confusion. The names of the administrative divisions of Korea to describe dialect areas have been used since Ogura Shinpei's 1940 dialect taxonomy, whereas compass point based designations were popularised by Kono Rokuro (1945). Although often used interchangeably, these terms cover slightly different geographical areas. Furthermore, since the Yukchin/Yukeup dialect area is located within the borders of Hamgyeong province, specifically within the borders of Northern Hamgyeong province, in the North East of the Korean peninsula, it is not clear whether dialectologists who use terms based on compass points or administrative divisions include the Yukchin variety of Korean in their definition of the Hamgyeong/North-Eastern dialect. This is particularly problematic, since the independent character of these dialects is a foundational assumption for much of the recent research on Central Asian Korean.

It is not for this thesis to be the final arbiter of dialectological nomenclature and decide whether there are grounds for a dialectological boundary between North and South Hamgyeong Province or the geographical extent of the dialect areas within the DPRK. We shall use the term "North Eastern dialect zone" since it makes the rough location of the dialect zone on the Korean peninsula clear and it clarifies that it is an area delineated by the linguistic characteristics of the speech community living there, rather than a geographical area, the

North East of the Korean peninsula, or an administrative area, Hamgyeong Province, both of which it are largely, but crucially not entirely coincident with the North Eastern dialect zone. Numerous different researchers have suggested different schema of isoglosses by which this variety may be distinguished from others which we summarise below.

There are a small number of linguistic features the status of which is agreed upon as being characteristic of the North Eastern dialect zone. The deletion of the nasals /n/ and /ng/ in intervocalic positions and nasalisation of the preceding vowel is the most striking of these. The North Eastern dialect zone is also delineated by retaining a system of contrastive tones (pitch accent), similar to earlier forms of Korean, as opposed the vowel length based system which has is recognised as part of the ROK and DPRK standard varieties and which has also developed in many other dialects. The extent to which the consonants /k/, /kh/, /t/, /th/, /s/ and /h/ are palatalised before high front vowels and /j/ on-glides (both historically and by means of a synchronic phonological process) is a means by which Korean dialects are often distinguished from one another. These consonants have undergone historical palatalization in the North Eastern dialect zone and are expressed with palatalised allophones before high font vowels and /j/ on-glides in contemporary speech, although this process is suppressed across morpheme boundaries (Kwak 2001a p. 102). Also, the following sentence final endings are considered characteristic of the North Eastern dialect zone:

	Declarative	Interrogative	Imperative	Propositive
Polite Style	*-up/supmeyta*	*-sup/supmeyta*	*-up/supsosey, -up/supso, --wu/swuta*	*-up/supseyta, -ciota*
Intermediate Style	*-um/summey, -o/so*	*-sup/supmey, -o/so*	*-o/so,*	*-up/supsey, -kio*
Informal Style	*-um/sum, -nta*	*-ni/-nya*	*-ulai, -nala,-kala, a/ela*	*-ca*

Table 1: North Eastern Dialect Zone Sentence Final Endings

It may be noted that these endings differ markedly from the endings most popularly associated with the North Eastern dialect zone and the transplanted varieties of Korean which are considered to be characteristics of the Yukchin dialect rather than the North Eastern dialect zone more generally (Kwak 2001b p. 284).

In addition to these unanimously agreed upon features, most researchers also present various combinations of other isoglosses as representative of the North Eastern dialect zone. These commonly include, but are not limited to: the retention of pre-Late Middle Korean intervocalic stops, the apocope of the final /l/ of the accusative particle and a vocalic system which is comprised of ten simple vowels among others. Rather than being characteristic of a particular dialect zone, though, these features are spread widely over the Korean peninsula and appear in the descriptions of many, varied dialects. Due to this we conclude that they do not distinguish the North Eastern dialect zone from other varieties of Korean, but are more likely either fast-speech phenomena (such as the phonetic reduction of final consonants) or non-territorially restricted features of vernacular Korean.

On the basis of the isoglosses laid out above, we cannot account for all the features of the transplanted varieties of Korean thought to be archaisms carried over from the Korean peninsula, therefore we follow the Korean Dialect Research Association (2001) amongst other researchers (e.g. Lee 2005) in holding the Yukchin variety of the far North Eastern corner of the Korean peninsula separate. The terms "Yukchin" (六鎭) and "Yukeup" (六邑) are entirely interchangeable and refer to a particular geographical region which comprises the six (Korean: *yuk*) fortresses (Korean: *cin*) or villages (Korean: *up*) of Chongseong (鍾城 - *Congseng*), Onseong (穩城 - *Onseng*), Hweryeong (會寧 - *Hoylyeng*), Gyeongwon (慶源 - *Gyengwen*), Gyeonghung (慶興 - *Gyenghung*) and Buryeong (富寧 - *Pwulyeng*). These geographical

boundaries are arbitrary, but mark an area roughly coincident with that in which a variety of Korean is spoken, which may be distinguished from the surrounding North Eastern dialect zone just identified on the basis of several systematic linguistic criteria.

The most striking of these are the absence of palatalization and the dialect's unique inflectional system for sentence final predicates (see Table 2, below).

	Declarative	Interrogative	Imperative	Propositive
Polite Style	-up/supkkwuma, -up/supkweni,	-um/sumtwung	-up/supsyo	-keypso
Intermediate Style	-o/so	-o/so	-o/so	-kio, keyo
Informal Style	-ta	-ni/-nya	-a,-ela, -nala,-kala,	-cya

Table 2: Yukchin Variety Sentence Final Endings

A further distinctive feature of this variety is the appearance of the consonants /l/ and /n/ in word initial position before the high front vowel /i/ and the glide /y/. That is, the so-called *twuumpepchik* (head sound rule) does not apply as it does in all other Korean dialects save that of Pyeongan Province in the far North West of the peninsula in which words have also not undergone historical palatalization and are not affected by the synchronic phonological process.

Furthermore, the Yukchin variety is further differentiated from the rest of the North Eastern dialect zone by a vocalic system of eight simple vowels, its realisation of the consonants /c/, /ch/, and /cc/ and its regular inflection of CSK's "s" and "p" irregular verbs.

In addition to the foregoing differences, it must be acknowledged that there are also many similarities between the varieties of the Yukchin area and the North Eastern dialect zone.

These include the retention of distinctive vocalic tone, similar treatment of complex syllable final clusters, lack of reinforcement of consonants following nasals, the absence of the phonologically conditioned allomorph of the subject particle –*ka*, similar formation of passive and causative verbs with a more extensive use of the allomorph -*kwu*- than is seen in many other varieties and similar patterns of negation which have not been attested in other varieties (Kwak 2001b).

We have established, then, that while the geographical areas from which the transplanted Korean communities of China and Central Asia originated is clear, the dialectological situation of these areas is less so. Earlier work on Korean dialect classification was not primarily based on linguistic reality, so it is only in more recent scholarship that the heterogeneity of the North Eastern region of the Korean peninsula has been fully acknowledged. Given what is known about the origins of the original seed communities of the *Cosencok* and the *Koryo Saram*, we propose that their original transplanted varieties of Korean were, likewise, not homogenous and combined varieties originating in the North Eastern dialect zone and the Yukchin area. Consequently, the forms of the contemporary varieties of *Koryo Mar* and Yanbian Korean may be expected to show signs of this mixed heritage to varying degrees.

3.3   Koryo Mar

This section examines research carried out in the Korean, Russian and Western traditions sequentially and separately. We then conclude the section by identifying the areas which earlier research has neglected.

As demonstrated above, it is the geographically static descendants of the source dialects of transplanted varieties of Korean which have received the most attention in Korean scholarship. Nevertheless, the body of research devoted to Central Asian Korean has been growing considerably in recent years.

While we have established that overseas varieties have been ignored by traditional Korean

dialectology, they have been an object of study in their own right for Korean academics since

the late 1980s, the field being dominated by Kwak Chung-gu and Kwon Jae-il. From a starting

point of producing sketch descriptions and attempts at reconciling the overseas varieties with

the dialects of the Korean peninsula (e.g. Kwak 1987), study of *Koryo Mar* has advanced to the

point that a full descriptive grammar has been published (Kwon 2010).

Korean scholarship has recently been supplemented by the publication of Kwak Chung-gu's

abundant transcriptions (Kwak 2008a; Kwak 2009; Kwak 2011) and descriptive materials (Kwak

2010a; Kwak 2010b; Kwak 2012) along with his analyses (Kwak 2007). These offer a wealth of

new information to the researcher of Central Asian Korean and give those without direct

access to linguistic consultants some insight into the data upon which the already widely

known characteristics of the variety are based. These materials are not, however, without

their limitations. Foremost among them is the Hangul transcription of dialect speech used

throughout by Kwak. While some attempt to reflect the characteristics of *Koryo Mar* which are

not conventionally rendered in Hangul transcription is made, for example an innovative

diacritic representing vowel nasalisation, several phonetic features are obscured by the use of

a Hangul based transcription including the realization of <ㅂ> (which may be [p], [b], [v] or [β])

and whether the grapheme <ㄹ> represents [l], [ɾ] or *Koryo Mar's* characteristic trill [r]. It is

further not made clear whether the phonological changes associated with conventional Hangul

orthography, for example nasal assimilation of stop consonants preceding nasal consonants,

take place (Kwak 2008 pp. 25 – 28). Preference for a non-Hangul transcription is ubiquitous in

research produced outside of Korea (Pak 2005; King 2006b p. 137).

The remaining major issue with these works is sampling. In each location (Almaty, Bishkek,

Tashkent and Taldikurgan), the speech of one consultant was taken as representative of the

area as a whole. In the case of languages as critically endangered as *Koryo Mar*, where

stratified sampling may not be possible due to patterns of language shift, it is often thought

necessary to employ judgement sampling in order to ensure the validity of one's data (King

1987; Pak 2005; Kwon2010; Kwak 2010b etc.). It must be acknowledged, though, that

conclusions drawn from such a small group of consultants with such similar sociolinguistic

profiles are not necessarily generalizable to the whole Korean-speaking population of Central

Asia.

Finally with regard to Kwak's work, we note the issue of the use of dialect surveys, particularly

the *Hankwuk pangen cosa cilmunci* (Korean Dialect Survey Questionnaire) (NLA 2006), which is

addressed more fully in the discussion in section 4.2 on the current methodology underlying

the data collection for the current work. Here we observe only that this dialect survey was

employed to gather data in this case, therefore the quality of the data is debatable.

Nevertheless, despite this critique, we stand by our initial assessment of these materials as a

great resource to be consulted by researchers working on Central Asian Korean. Having

acknowledged the limitations of these materials, we do not hesitate in including this data in

our analysis of the variety's lexis and grammar, since these aspects of the language should be

minimally affected by the data gathering procedures and representation in Hangul

transcription.

In addition to the descriptive works relying on single consultants a further dialect survey of

Central Asian Korean in the Korean tradition was carried out by Yi Kigap and his colleagues (Yi

et al. 1999) which incorporates data from twenty four consultants drawn from two research

sites: Almaty and Ushtobe, both in Kazakhstan. While there are minor issues with consultant

selection, such as gender balance and lack of age stratification, the Hangul transcription used

in this paper and the questionable ecological validity of dialect surveys, it offers a wealth of

examples of the features of spoken KM. Although there is no conclusion to this description of

the variety which explicitly relates KM back to the peninsula varieties, the paper implicitly

supports the suggestion that the language of the *Koryo Saram* is simply an example of the speech of the North Eastern dialect zone in two ways. Firstly, the term *Koryo Mar* is not used in the paper, the authors preferring the term *cwungangasia haninuy hankwuke* – "the Korean of Central Asian Koreans". Also, the authors explicitly state that "the dialect of Northern Hamgyeong province is divided into the Yukchin dialect and the Myengchen-Kilcwu dialect, one of which is used by each individual *Koryo Saram* depending on their background" (Yi et al. 6-7). Such a point of view is also incompatible with the idea that a new or koine form of these dialects has developed in Central Asia.

The beginnings of the Russian tradition of scholarship have been covered by King in his PhD thesis and subsequent articles (1989; 1991a) on early Russian sources documenting the Korean language. Chiefly, these comprise a dictionary (Pustsillo 1874) and a combined phrasebook and wordlist (Matveev 1900), both of which were compiled by amateur linguists from information obtained directly from Korean speakers whose language exhibited some characteristics of the North Eastern dialect zone. These early materials, though, were purely descriptive and constituted a far from rigorous account of the language of Koreans resident in the Far Eastern regions of the Russian Empire. Soviet era scholarship devoted little time to the language of its Korean population, with such minimal research as was carried out on them being part of the anthropological tradition, for example the works of Dzharylgasinova (1980 etc.).

It is only very recently that researchers working in the Russian tradition have once again focussed on Central Asian Korean, notably the native-speaker linguist N.S. Pak. Much of her research is informed by an awareness of how severely endangered *Koryo Mar* is, consequently there is a strong descriptive component to it. In addition to this, the focus on the genesis of the variety from dialects of the North East of the Korean peninsula which may be seen in Korean dialectology is also evident here, Pak being a proponent of the theory that *Koryo Mar* is

a koine formed due to the intense, multi-generational contact between speakers originally of the Yukchin dialect and Myenchen-Kilcwu variety of the North Eastern dialect zone (Pak 1991; Pak 2005).

English language scholarship on the *Koryo Saram* to date, which a 2001 bibliography by Kim and King reveals to be surprisingly extensive, has largely ignored their language, focussing instead on their history. It is not unknown for some mention of the language of Soviet Koreans to be made in these works, however these cases are usually restricted to mentioning particular culture bound concepts and words, so the language is not scrutinised closely. A notable exception is the first English language publication on the *Koryo Saram*, Kho Songmoo's monograph of 1987. While the focus of this work is indeed the history of the Central Asian Korean community and a description of their society, some attention is paid to the variety of Korean used in Central Asia. The input Yukchin and Myengchen-Kilcwu varieties ("Yugup" and "Saup" in Kho's terminology) are characterised, but by neglecting to discuss the realisation of their various features in Central Asian Korean, Kho implicitly conflates the transplanted variety with these peninsula varieties. Indeed, it is notable that while the ethnonym *KoryoSaram* is used without compunction throughout this monograph, the term *Koryo Mar* does not appear. Wordlists are also provided in a very broad transcription. In contrast to other research, Kho pays specific attention to the non-lexical linguistic outcomes of contact between Korean and the LWCs of Central Asia and offers relatively extensive examples of intra-sentential code-switching. Unfortunately little to no analysis accompanies the presentation of these linguistic features.

Even taking this work into account, *Koryo Mar* has not been widely studied in the West and is known there almost exclusively through the work of Ross King and his collaborators (King 1987a; King 1989; King 1991a; King 1991b; King and Yeon 1992; King 1992; King 1994; King 1996c; King 1997b; King 2001a; Kim and King 2001; King 2006b).

Two main strands of research may be identified in King's work: historical and descriptive. The relevance of the historical research into early Russian language records of Korean to the current research is that it enables us to identify the original input varieties of contemporary *Koryo Mar* and establish which of its features developed under the conditions of diaspora and which have been maintained or lost since the migration of Koreans into the Russian Empire. While King acknowledges the limitations of working with historical sources such as the unreliability of transcription and unrecorded methods of elicitation, taken together the picture that his analyses of the major early Russian sources of Korean (most notably from pre-revolutionary times, the dictionary compiled by Putsillo (1874), Matveev's short Russian Korean phrase-book(1900) and the so-called "Kazan Materials" (1901 - 1904)) demonstrate that Koreans resident in Russia spoke one of the dialects of Northern Hamgyeong Province, with strong Yukchin representation. The most salient linguistic features recorded which allow us to draw this conclusion are the frequent lack of palatalization of historic /t/, the palatalization of then contemporary /k/, the dropping of /n/ and /ng/ before high front vowels and occasional attempts to transcribe tone (King 1989; King 1991a; King 1991b).

The descriptive work is imperfect by the standards of today's documentary and descriptive linguistics. This may be attributed in large part to the challenge of a carrying out research in the USSR for a foreign scholar and the restricted access to consultants. Consequently, while it is hard to say that the materials upon which King's description of the *Koryo Mar* is representative of the variety of large as it was actually used, it nevertheless represents the earliest attempt at a modern, systematic description and certainly gives a more complete picture than the earlier Soviet and pre-Soviet materials, therefore it merits a closer examination.

The first English language description of the *Koryo Mar* (King 1987a) was made on the basis of one three hour elicitation session with a "middle aged" female native speaking consultant,

supplemented with "some data" from a male consultant of unknown age. These interviews were not recorded, but rather transcribed in a mixture of close phonetic transcription and broad Yale Romanisation transcription. Thus, the question of whether these data are applicable to the variety as a whole only represent a description of an idiolect is once again raised. Nevertheless, this description is remarkably complete on the basis of so little primary data and the majority of its findings went on to be corroborated by later investigations of the variety. These include the retention of dialect features noted in King's historical work such as extensive palatalization, which extended to velar stops, and /n/ and /ng/ deletion in addition to archaisms, such as the retention of 'k-stem' nouns (*kalki, nayngi* etc.), lack of the variant form of the subject particle –*ka*, and innovations, such as the appearance of /v/ in non-etymological positions, a characteristic verbal inflection and a wealth of borrowed Russian vocabulary. The subsequent descriptive paper (King and Yeon 1992) offers more than a sketch description and goes so far as to identify the phonological and what they describe as "grammatical" isoglosses which may be used to distinguish this variety from others, as a result of their main conclusions, which are as follows:

> 1. *Koryo Mar* is not simply Hamgyeong dialect transplanted into foreign soil.
> 2. *Koryo Mar* began to form newly and independently in the Russian Far East in the later part of the 18th century. Between 1860 and 1900 speakers of the Yukup dialect made up the absolute majority [of Korean speakers in that area]. From 1900 to 1920, due to the continuing migration of people from the Myengchen-Kilcwu region to the Primorskiy Kray, the dialects mixed and it may be observed that a kind of koine formed. (King and Yeon 1992 p. 100).

The characteristic dialectological features of *Koryo Mar* identified in this paper are divided into twelve phonological and eight "grammatical" features as follows:

Feature	Examples

Non-standard realisation of the phoneme /l/	KM: [hɨrgi] SK: [hɨk] (earth), KM: [murkogi] SK: [mulgogi] (fish), KM: [targari] SK: [talgjal] (egg) etc.
Phonemically distinctive pitch accent	KM: súri/surí SK: swutkalak/swul (spoon/alcohol), KM mógi/mogí SK: moki/mok (mosquito/neck) etc.
"Regular" inflection of /p/ and /s/ irregular verbs	KM: chipun SK: chwuwun (to be cold), KM: cisumwu SK: ciumyen (if you build) etc.
Nasal weakening	KM: ai SK ani (no/not), KM sidzhay SK: sicang (market) etc.
Retention of LMK /su/, /cu/ and /chu/	KM: sure hadzhi SK: silhehaci (to dislike), KM: tsunthori SK: cinphel (swamp), KM: achumi SK: achim (morning)
Intervocalic voicing of /s/	KM:yuzhikhan mari SK: yusikhan mal (refined speech), KM: kazumi, SK: kasum (chest) etc.
Loss of initial /j/ glide when /ye/ follows a consonant	KM: masera SK: masyela (drink!), KM: peri SK: pyel (star) etc.
Loss of /w/ on glide or /w/ > /v/	KM: hanggaybi SK: hwangkap (sixtieth birthday), KM: vekhirabay SK: oyhalapeci (maternal grandfather)
Deletion of /l/ before coronal stops or affricates	KM: kida SK: kilta (to be long), KM: sadzhi SK: salci (it lives)
Non-standard palatalisation	KM (Myengchen-Kilcwu): tsiri SK: kil (road), KM (Yukup): tsaysari kinaygo SK: ceysalul cinayko (after the memorial service had been carried out)
Umlaut	KM: ormay SK: elma (how much?), KM: guksi SK: kwukswu (noodles) etc.
Insertion of epenthetic /u/ between morpheme boundaries	KM: ciphudzhay aykkuma SK: kiphci anhsupnita (it is not deep), KM: kathugi SK: kathkey (together)

Table 3: Phonological Features of *Koryo Mar* (King and Yeon 1992)

Many of these features may broadly be said to be shared with the North-eastern variety of

Korean spoken on the peninsula, for example umlaut and non-standard palatalization, and

thus support the classification of *Koryo Mar* within this variety. Conversely, the breadth of

these features may serve to obscure differences between *Koryo Mar* and the peninsula

varieties. While the transplanted variety's non-standard patterns of palatalization may be

shown to be identical to contemporary peninsula varieties, it is not known whether other

features, such as pitch accent, manifest in exactly the same way on the peninsula as outside of

it. There are also examples of features entirely unique to K*oryo Mar* such as development of

/v/ in non-etymological positions and the non-standard realisations of the phoneme /l/. We

now present the distinguishing morph-syntactic characteristics of *Koryo Mar*:

Feature	Examples
Absence of subject particle /-ka/	KM: tubi SK: tupu (tofu),
Retention of "k" declining nouns	KM: kalki SK: kalu (flour), KM nayngi SK: namu (tree) etc.
Volition expressed with –*kisiphta* and phonologically reduced forms derived from it	KM: kakisiphta SK: kakosiphta (I want to go), KM: kakeyphumdu SK: kakosiphsupnikka (Do you want to go?), etc.
More frequent use of analytic causative and passive constructions and formation of same with –*ki-*, –*kwu-* and –*wu-* over standard allopmorphs	KM: takkinda SK: tathinta (to close something/to be closed), KM: torgwunda SK: tollinta (to turn), KM: paywunda SK: poita (to show/to be visible) etc.
The connective –*a/eya* is replaced by –*kwuya*	KM: kore yeca mandur**kwuya** masisso SK: kolye yecaka mantul**eya** masi issso (It is only tasty is a *koryo* woman makes it)
Characteristic particles	-tula (-tolok/-kkaci), -ma (-pota) etc.
Characteristic negation patterns	KM: mege **mo** pwasso SK: **mos** meke pwasso (I haven't tried eating it)
Characteristic verbal inflection	-kkuma (-supnita), -mdu (supnikka) etc.

Table 4: Grammatical Features of *Koryo Mar* (King and Yeon 1992)

In contrast with many of the phonological features identified by in this paper, these so-called

"grammatical" features are less widely distributed in peninsula varieties than the phonological

features. Also, the formation and phonetic reduction of –*ki siphta* to express volition is unique to the Central Asian variety of Korean.

Strangely, features common to the CSK recorded in the speech of *koryo saram* prior to their deportation to Central Asia, for example the polite neutral speech style ending with –*yo* (King 1989 p. 306), appear to have fallen out of use. This ending is also absent from modern accounts of the varieties of the North east of the Korean peninsula (cf. Kwak 2001; Lee 2005), i.e. the 'ancestral homeland' of the Central Asian Korean community. Thus, in addition to their shared distinctive characteristics, the contemporary absence of such widely spread features as this may be considered among the peculiarities which mark such varieties and their transplanted descendants as separate from other varieties of peninsula Korean. Furthermore, some research suggests that widespread Russian bilingualism may have an influence on word order in *Koryo Mar* (King 1989 p. 312), however, this feature is not given much consideration in later works and the influence of the LWCs with which *Koryo Mar* has found itself in contact with over the years is generally treated as if it were entirely restricted to vocabulary.

We have demonstrated that *Koryo Mar* has been a subject of research in the scholarly traditions of Korea, the Former Soviet Union and the West. In each of these areas the focus of research has been the description of the variety, however, the research tradition of each area informed how this description was carried out. In the vast majority of cases, though, the synchronic forms and phenomena found in KM speech were explained largely with reference to the traditional dialects of the Korean peninsula or in terms of how they differ from CSK and without reference to other transplanted varieties. Although most research into *Koryo Mar* examines the spoken language, it nevertheless treats the variety as somewhat homogeneous and does not approach questions of intra-variety variation, save for Kwon Jae-il's examination of the variation between spoken and written Korean in Central Asia (Kwon 2012) We turn our attention in the next section to the variety of Korean spoken in the YKAP.

3.4 Yanbian Korean

Even compared to *Koryo Mar* the language of the Korean community of Yanbian garnered relatively little scholarly attention outside of China, either in the Anglosphere or in Korea, over the latter half of the twentieth century. A very complete bibliography of all research published between 1950 and 2012 related to the Korean language as it pertains to China's Korean minority has been compiled (Ji et al. 2012). This work casts its net wide in including newspaper editorials and school language textbooks in the bibliography, but the role of this inclusion in giving a good overview of research carried out on the language of the *Cosencok* over this period must be acknowledged. From its focus on language pedagogy and standardisation in the early and sparse research of the 1950s to 1970s the field grew in size and diversity, however, it is clear that dialectology or the description of the Korean language as it is actually spoken in China is not currently and never has been a major focus of research even within the field of Korean linguistics in the PRC.

This overview also serves, however, to reveal the regrettable fact that a sizeable proportion of the scant, relevant research which has been carried out into this variety is not accessible to the present study since it is published exclusively in Chinese. On a more optimistic note, this bibliography emphasises that Yanbian Korean is receiving more scholarly attention now than ever before. We nevertheless find that the amount of accessible research is insufficient for sub-dividing this section according to scholarly tradition; therefore we shall instead review earlier work according to discipline, occasionally focussing on particularly prolific researchers.

A comparatively great deal of research has been devoted to examining this variety through experimental phonetics, notably the questions of tone and the realisation of particular unstable vowels (Ito and Kenstowicz 2009; Jin 2013). Research into other elements of Yanbian Korean has been largely carried out as part of its on-going description.

While the variation of sounds present in Yanbian Korean has been noted very little can be said about its phonology. In some cases, this can be attributed, again, to the use of Hangul as a means of transcription and the acceptance of Contemporary Korean conventions of reading which this forces on the reader (for example, see Jeong 2010). Nevertheless orthography alone does not determine the reliability of transcription or interpretations: one study which does use the IPA as a transcription device records an independent, contrastive phoneme /f/ in VYK (Park 2003p. 17), but closer examination of this claim, though, reveals that it appears only in nonce-loans from Chinese and therefore is not to be considered part of the VYK phonological inventory.

Within this purely linguistic research, the forms of particles and sentence final endings have also been the subject of much scrutiny, being subject of Jeong Hyang Ran's 2010 monograph, as well as multiple papers (e.g. Choi 2002; Kim 2013).

A certain amount of descriptive work has also been carried out. In addition to the materials covering Central Asia, Kwak Chung Gu has also published a Hangul transcription of his interviews with a Yanbian Korean speaker (Kwak 2008b) and presents Yanbian lexis alongside that of other diaspora varieties in his lexical dialectological survey of transplanted varieties of Korean. While this collection of dialect materials has many of the same issues as those recording the speech of Central Asian Koreans mentioned above (Hangul transcription and sample size of one), it is nevertheless a useful source of supplementary language data for the analysis of Yanbian Korean grammar and vocabulary.

A wide variety of research covering numerous disciplines pertaining to Yanbian Korean has been produced by native-speaker researcher Kim Kwang Su. He has addressed the question of the development of the Korean language in China (2009; 2012a), covering largely the period from the initial migration of Koreans into China up to the liberation of Korea from Japan in

1945, and has written numerous papers on subjects as diverse as the linguistic landscape of Yanji (2012b), the procedures necessary for the compilation of a corpus of Yanbian Korean (2012c) and the influence of the two competing standard peninsula varieties of Korean on the language of the *cosencok* (2012d; 2012e).

In-keeping with earlier research, his description of the linguistic situation of Korea in China prior to 1930 is couched purely in terms of the peninsula varieties. The distribution of migrants and their descendants from the various regions of the Korean peninsula is examined, and then a brief description of the dialect of each peninsula dialect area except Jeju-do is given exclusively in terms of certain aspects of phonology: the presence or absence of lexical tone, number of simple vowels and the degree to which the high front vowels /ø/ and /y/ had undergone diphthongization. Kim goes on to characterise the situation in China as being one in which peninsula dialect distinctions were retained long-term due to lack of contact between rural labourers with different dialectological backgrounds (Kim 2009 p. 43). Although a little more detail is given in the examination of the Yanbian Korean of the independence period (roughly the mid-1940s) , it is not made clear by what mechanism the diverse dialect situation presented as obtaining some twenty years previously is transformed into one in which the consonantal and ten vowel vocalic system resembles that of CSK and in which the following phonological processes occur: intervocalic voicing of lax consonants, non-released consonants in patchim position, regular nasal assimilation, dropping of the /j/ glide (especially for /yay/ and /yey/) following consonants and historic palatalization. The only phonological processes which are not associated with CSK mentioned are the synchronic productive palatalization of /k/, /t/ and /h/ both across morpheme boundaries and within them and umlaut – a common feature of many varieties of spoken vernacular Korean (Kim 2009 pp. 203 – 207). Furthermore, the Hangul transcription obscures the phonetic details of the phonological description described. For example, it remains unclear whether <ㅈ> is realised as a post-alveolar affricate,

i.e. [t͡ʃ], or and alveolar affricate, i.e. [t͡s]. The description of the morpho-syntactic

characteristics of Yanbian Korean is also brief and strikingly similar to those of the standard

varieties of peninsula Korean.

Moving on to other papers, though, here Kim emphasises the separation of Yanbian Korean as

something apart from the varieties of both South and North Korea in both spoken (Kim 2012a)

and written (Kim 2012b) forms, although these studies focus almost exclusively on vocabulary.

These differences with peninsula Korean are attributed to the contact with Chinese which the

Yanbian Korean speech community undergoes generally and, more specifically, to the practice

of exogamy and the universal provision of at least some Chinese language education to the

Cosencok.

In contrast with *Koryo Mar*, the patterns of use of Yanbian Korean are frequently examined.

The precarious position of the former dialect and its limited use has been noted from its first

appearance in modern research due to its traditional speaker community being in the last

stages of shifting language into Russian. In order to avoid a similar fate, the usage of Korean in

China is closely monitored by researchers. Taken together, this research paints a mixed

picture of the situation in Yanbian. While some researchers are pessimistic and note, for

example, reductions in the number of Korean medium schools and domains in which Korean is

used by default (Pung 2000 p. 220), others emphasise the current vitality of Yanbian Korean

and its viable intergenerational transmission (Park 2002) or its unique character in the face of

pressure from peninsula varieties of Korean and Chinese (Kang 2003 p. 22).

Standardising and "developing" the Korean language in China appears to be a pre-occupation

even of research which does not explicitly focus on that topic (e.g. Kim 2012b). Unlike *Koryo*

Mar a great deal of work has been put into the creation of a standard variety of Korean in

China and research has been carried out not only for achieving this goal, but on this process

(Tai 2004). This research is especially valuable for charting the influence of the standard varieties of Korean developed in the Korean peninsula on the variety spoken in Yanbian in the mid-twentieth century and the emergence of an identity and linguistic consciousness separate from those of the Korean peninsula which coincided with the beginning of wholly independent standardisation of Chinese Korean in the mid-1970s. The standard language which has been arrived at today is still somewhat informed by both of the standard varieties of the Korean peninsula and there are even some who argue that from the perspective of lexicography at least Chinese Korean may be identified with with *munhwae* – the standard language of North Korea (Choi et al. 1994). This position is supported by examining the use of technical vocabulary in Chinese Korean publications the fields of linguistics and the life sciences, both of which rely heavily on standard North Korean for the provision of new coined words and technical vocabulary (Kim 2012d; 2012e). Prescriptive works on the standard Chinese Korean language include numerous examples of common features of spoken Chinese Korean being classified as non-standard in favour of the variants of standard peninsula Korean varieties. These include the omission of the characteristic verbal inflection and idiosyncratic particles of Chinese Korean as well as a lack of specification of whether vowels contrast in length rather than tone (Kim et al. 2013). This may be attributed to the fact that current thought on the standardisation of Korean in China revolves around the idea of constructing a standard variety which is representative of the language of the whole Korean ethnos, rather than a standard language based on a particular variety as the standards of South and North Korea are (Kim, et al. 2013 pp. 20-22). Given this fact, it is notable then that the standard variety's vocabulary includes many items which may be considered either nonce-borrowings from Chinese or native Korean words only found in the spoken language, the latter being a feature shared with the standard language of North Korea (Choi et al. 1994 p. 316).

A further focus of research has been language pedagogy. In addition to research into the proportion of *Cosencok* students attending Korean or Chinese language medium schools (Park 2010) and the controversial issue of the position of Chinese in Korean medium schools in Yanbian, comparative studies with peninsula Korean have also been carried out which highlight differences in curriculum and cultural content (Kim and Jeong 2011).

We have seen, then, that the Korean language in China has been at least superficially considered from a wide variety of perspectives. While it has been well established that some form of Korean is used in a wide variety of domains in Yanbian, the on-going conflation of the language of the *Cosencok* with peninsula varieties of Korean, both standard and non-standard, has resulted in a situation in which little can be said of the form of the vernacular language, since little attempt is made to gather data from representative speakers. Less still has been said about the underlying linguistic or extra-linguistic causes of these contemporary linguistic forms. The preceding does establish, though, that there are several similarities between the contemporary linguistic forms of Yanbian Korean and K*oryo Mar* which arise from their common heritage.

3.5   Conclusions

Silva's systematic review of academic databases revealed that dialectology and linguistic variation are two habitually overlooked areas within the field of Korean language studies (Silva 2010). We have demonstrated here that even within these specific subject areas, there is relatively little research has been carried out on transplanted varieties of Korean compared to peninsula varieties.

In our examination of the research which has been carried out, we observed an emphasis on description over explanation, a reliance on traditional Korean dialectology in this description and less ecologically valid data gathering procedures are issues common to earlier works

concerning Central Asian and Yanbian Korean. A proportion of this descriptive work, though, may be incorporated into the current examination of overseas varieties of Korean, particularly with regard to questions of lexis and syntax. Moreover, we note that extra-linguistic causes are popularly adduced in an *ad hoc* manner as causal factors with regard to both general and specific instances of language change or maintenance. To date, though, there has been no systematic examination of the relationship between the social and other extra-linguistic conditions of transplanted communities of Koreans and language change, therefore earlier appeals to these as causal factors have the character of "explanations of last resort". In addition to this gap in the research, we also note that there has been no work produced which undertakes a direct comparison of these transplanted varieties.

Despite its relatively brief history, methodological issues and relative paucity compared to other fields of research into the Korean language, the prior research into transplanted varieties of Korean constitutes a foundational body of knowledge which necessarily informs the direction of future research.

3.5.1 Impact on Research Questions

On the basis of the preceding sections, we have identified the following issues with the research:

1. The synchronic linguistic basis for the dialectological position of the transplanted varieties of Korean in China and Central Asia is not clear.

2. Whether these varieties are linguistically distinct from one another is not clear.

3. The precise linguistic features which distinguish KM and VYK have not been identified or investigated.

These issues can be re-formulated into the following research questions:

4. What is the form/are the forms of Korean currently in use in China and Central Asia?

5. Do these forms of these varieties differ from one another and, if so, to what extent?

6. If these transplanted varieties can be shown to be distinct from one another, by what linguistic features may we distinguish them?

These questions are something of a hierarchy. Once a clearer picture of the varieties of Korean spoken in China and Central Asia is established, it will be possible to directly compare their forms to establish correspondences and differences. Once as many points as possible of convergence, divergence, common inheritance and independent innovation identified, we will be in a position to determine more precisely the synchronic relationship between these varieties. In the next section, we discuss how these questions may be approached and the methodological and the practical considerations of doing so.

4 Methodology

In the preceding literature review, we established that research into overseas varieties of Korean has concentrated on description, but has been beset by methodological issues. On the basis of these descriptions, the supposed character and dialectological characteristics of overseas varieties of Korean have been established, often as a result of implicit comparison with one of the standard varieties of Korean. Due to this, a broad consensus that the *Koryo Saram* and *Cosencok* varieties of Korean differ from peninsula standard varieties (i.e. the ROK's *phyocwune* and the DPRK's *munhwae*) has arisen, primarily on account of their shared signs of influence from their North Hamgyeong heritage and respective main contact languages. However, it must also be noted that these varieties have long been implicitly conflated with one another and the varieties of the North East of the Korean peninsula. This thesis seeks to verify the empirical basis for this claim specifically by contrasting KM and VYK directly to one another and answering the research questions laid out at the close of chapter 3's literature review.

In order to determine the extent of the similarities and differences between these varieties, we propose to undertake a systematic comparison of primarily the phonological, morphophonological and morphological characteristics of Vernacular Yanbian Korean and *Koryo Mar*[4]. In the section of this chapter which follows this introduction directly, we discuss principled approaches to undertaking the synchronic comparison of two linguistic systems.

The section following our discussion of comparison is concerned with the theoretical framework of quantitative linguistics and the application of statistical techniques to the examination of linguistic variation. We shall further demonstrate how these two approaches

---

[4] In order to provide additional context and information for the variables discussed in chapter 6's quantitative analysis we provide descriptions of apparent lexical and grammatical peculiarities of KM and VYK, however we acknowledge that the description contained in this thesis is heavily weighted towards the phonological and morphoponological.

complement each other and detail how we shall employ them to obtain a particularly nuanced and robust understanding of the degree of synchronic linguistic (dis)similarity between the transplanted varieties of the YKAP and Central Asia.

4.1   Comparison of Linguistic Systems

Comparison is a concept at the heart of numerous sub-disciplines of linguistics, including comparative historical linguistics, linguistic typology and microvariation (comparative dialectology). Indeed, it has been argued that all linguistic research is implicitly comparative in scope, since the analytical apparatus of any linguistic inquiry should be applicable to any given language and allow comparisons to be made across languages (Willems et al 2003). We turn to the sub-disciplines of linguistics most closely concerned with the comparison of linguistic systems here in order to establish an appropriate framework in which to carry out our systematic comparison of overseas varieties of Korean

Despite the diachronic tendency inherent in much traditional dialectology, our comparison is synchronic in nature and not oriented towards reconstructing an earlier unrecorded form of Korean, therefore we rule out comparative historical linguistics as a framework for this investigation. Classical typological approaches may be considered inappropriate, too, on the grounds that they more commonly focus on particular linguistic phenomena as they manifest in a wide variety of languages.

As microvariation is also known as comparative dialectology, it appears on the basis of nomenclature alone to be the perfect framework for comparing two non-standard varieties of Korean. Upon closer examination of the framework, this appears not to be the case on the grounds that the theoretical underpinning of studies in microvariation is that patterns of variation within a language may be examined in analogy to the interlinguistic variation analysed by typologists (König 2012). While there is much to be gained from such an approach,

there are practical and principled barriers to pursuing it here. Analysing microvariation entails the mass comparison of a large number of varieties, generally in order to analyse the scope and patterns of variation of a single feature (for example, the papers collected in Barbiers, Cornips and van der Kleij 2002). The practical difficulties include the fact that Korean as a whole is only questionably suitable for such investigations since the precise number of varieties into which it may be divided is a matter for debate and, furthermore, microvariation studies tend to rely on the wide availability of extensive dialect corpora which are accessible only for very few languages among which Korean does not number. In addition, our investigation does not focus exclusively on a single feature or share the particular syntactic focus of much of the contemporary linguistic sub-dicipline of microvariation (e.g. doubling, negation etc.), but rather aims to present a detailed description of a wide range of features of only two varieties.

Such a process of elimination demonstrates that the frameworks of comparative historical linguistics, typology and microvariation are unsuitable for our comparison of KM and VYK. Instead, we propose that formal structural similarities between these varieties may be established using a framework adapted from Contrastive Analysis (CA). Although CA has often been characterised as a limited form of language typology, we consider it distinct and more suitable for our purposes than the 'classical' typological approach. Since the purpose of our enquiry is to identify and account for similarities and differences between two specific varieties, we prefer the approach which emphasises the fine-grained comparison of multiple characteristics of a limited number of linguistic systems. A starting point for its definition and underlying goals taken from James' (1980 p. 3) is as follows "CA is a linguistic enterprise aimed at producing inverted (i.e. contrastive, not comparative) two-valued typologies (a CA is always concerned with a *pair* of languages), and founded on the assumption that languages can be compared".

We now examine the CA framework in more detail and, in summarising its development and characteristics, outline the contribution it may make specifically towards the question of whether the varieties of Korean used in China and Central Asia differ, and address any possible outstanding issues.

The modern understanding of CA is somewhat expanded from that presented above and well summed up by Gast's (2013) brief definition, which states that "contrastive analysis investigates the differences between pairs (or small sets) of languages against the background of similarities". Traditionally, though, the goals of the discipline were intimately connected with language pedagogy and the identification of areas of potential difficulty (points of structural difference that may lead to interference or 'negative transfer') for language learners.

For a period of roughly twenty years following its inception in 1949 the discipline was entirely subsumed within applied linguistics due to the demands of a so-called 'objective of applicability', at the time considered a requirement of investigations carried out in the sub-discipline. Thus, a typical a requirement of traditional CA is that the languages to be contrasted are "socio-culturally linked", a criterion canonically taken to mean that languages subject to CA have been selected on the basis of natural or most especially instructed bilingualism.

Over the 1970s and 1980s the existence of a more theoretical CA was acknowledged (James 1980; Rein 1983) and the framework began to be used to examine more general questions about why languages differ in the ways they do. A typology of contrastive studies was proposed by Fisiak (1981) in which a characterisation of the field strikingly similar to Gast's more contemporary definition, presented above, was proposed for so-called "theoretical contrastive studies"; that is, those which are not conducted with traditional CA's focus on identifying potential cases of interference between L1 and L2, and which were created solely

to "give an exhaustive account of the differences similarities between two or more languages" (Fisiak 1981 p. 9). From this time CA came to be characterised as being "a special case of linguistic typology and is distinguished from other types of typological approaches by a small sample size and a high degree of granularity" (Gast 2012).

The characteristics of the 'conservative' CA framework and how it differs from that developed in the later 20th century is summarised below (adapted and expanded from Hellinger and Ammon 1996):

Traditional CA	Modern CA
Basic unit of study is the "language"	Basic unit of study is "the linguistic system"
Necessarily interlingual or cross-linguistic	Incorporating the concept of the linguistic system, may be applied to both distinct languages and different varieties of a single language
Focuses on the differences rather than similarities between two languages	Describes both differences and similarities of two or more linguistic systems
Unidirectional – takes one of the two languages being contrasted as the frame of reference	May take a *tertium comparationis* as the frame of reference for the linguistic systems being compared

Table 5: Summary of Traditional and Modern CA

The notions that CA should only concern itself with the comparison of standardised, codified languages and is obligatorily cross-linguistic in scope are persistent but not universally accepted. The contrary opinion, that the methods and techniques of CA may be fruitfully applied in other comparative sub-disciplines of linguistics, including dialectology, has also arisen among contrastivists (e.g. König 2012). It is particularly through the application of the concepts of modern CA, outlined above, to the transplanted varieties of Korean under investigation here that we see that this framework is appropriate for the investigation of VYK and KM for the following reasons:

- The transplanted varieties are considered distinct linguistic systems

- Although these systems are not widely used together in a parallel, balanced bilingual

  situation, their common dialectological heritage originating in the North East of the

  Korean peninsula along with their myriad points of common cultural heritage and

  histories in transplanted situations (sociological characteristics) are similar enough for

  these varieties to be considered socio-culturally linked.

- Although there is no standardised descriptive terminology particular to KM or VYK in

  which a description of these varieties may be written, by incorporating the concept of

  tertium comparationis from modern CA we are able to set both varieties against the

  common backgrounds of supra-varietal comparative concepts which, here, we propose

  drawing from traditional Korean grammatical description.

Despite the long-acknowledged existence of non-applied, that is non-pedagogically oriented,

CA, the fact nevertheless remains that the discipline is still primarily associated with applied

linguistics and translation studies (for a recent example see Hjulmand and Schwarz 2009).

Where we use the terms 'constrastive analysis' and 'contrastive linguistics' here, we do not use

them of the comparison of the KM and VYK in that sense. Rather, since a particular focus of CA

research has been developing methods for identifying structural differences between two

linguistic systems, we co-opt those methods and much of the underlying theory for our

investigation of overseas varieties of Korean.

We have demonstrated, then, that the fundamental goals and concepts of CA are

commensurate with our own with regard to the investigation of qualitative differences

between the transplanted varieties of Korean of China and Central Asia, so we now go on to

examine the methods and techniques of this framework in more depth,

4.1.1 Carrying out CA

There have been several suggestions as to what constitutes the precise method for carrying out research in the CA framework:

> Traditionally, contrastive methodology starts with a description of selected data. James (1980: 63) sees two basic processes: description and comparison. In his examples he uses a four-step algorithm: (1) assemble the data, (2) formulate the description (3) supplement the data as required, (4) formulate the contrasts. Krzeszowski (1990: 35) sets up three steps for "classical" contrastive studies: (1) description, (2) juxtaposition, (3) comparison. Additionally, of course, the descriptions must use the same theoretical model. The model chosen will then naturally determine how the contrasts are formulated. (Chesterman 1998 p. 52)

It is the intention of this study to compare the forms of transplanted varieties of Korean on multiple levels of linguistic analysis and undertake a comparison of the distribution of these forms. We discuss these in turn, in order to establish appropriate *tertia comparationis* at each level of analysis and determine more concrete methods for going about performing such an analysis.

The precise method followed in this thesis broadly conforms to the process attributed to James in the quotation above, although the order of the workflow is somewhat different. The data consist of field recordings created specifically for this project and have been supplemented (James' third step) by earlier linguistic documentation of the two varieties identified as suitable for undergoing CA in the previous section. They are described in parallel using terminology and comparative concepts taken from descriptions of CSK in chapter 5 and contrasts of form are to be made alongside the descriptions. Such formal contrasts represent qualitative distinctions drawn between the transplanted varieties of Korean.

The steps involved in comparing *Koryo Mar* and Vernacular Yanbian Korean, as they are outlined above, are somewhat nebulous. Therefore, we devote the following sub-sections to

making them more concrete. First, we examine the nature of the data and their collection,

before moving on to the production of comparable descriptions. Finally, we examine the

processes involved in the formulation of contrasts of form, necessarily taking in such concepts

as "comparability" and "equivalence", before concluding this chapter with an examination of

quantitative linguistics and its contribution to the current study.

4.2 The Data

Before embarking on an analysis of KM and VYK, it is necessary to establish the adequacy of

the primary and supplementary data upon which examination of these varieties is to be based.

Here, the term primary data is taken to mean recordings of speech and observations on

utterances recorded in contemporaneous field notes gathered from research sites in and

around Almaty, Kazakhstan and in and around Yanji, PRC. We now go on to discuss how the

form of this primary data was arrived at and the selection of the specific linguistic consultants

who provided it.

The criteria by which the varieties to be compared were selected were outlined in section 2.1

above. To recap, VYK and KM were identified as suitable for comparison with one another on

the grounds that they are varieties of Korean that have been used in different countries by

transplanted Korean communities with similar temporal and geographical origins. Certain

common dialectological features have also been recorded. Similarities or perceived similarities

between linguistic systems in addition to "socio-cultural linking" have been identified as

peerless comparability criteria for identifying objects for CA. It is precisely the similarities

which make these varieties suitable objects for comparison.

In order to carry out formal CA the KM and VYK data must meet the following criteria: it must

be suitable for hypothesis checking claims made about the organisation of these linguistic

systems in earlier descriptions, and it must be suitable for comparing the features of these varieties in use.

These criteria are arrived at on the basis that the ideal case for carrying out CA involves the comparison of similarly formatted descriptions, usually of invariable or minimally variable standard languages. Since there are no such extant descriptions available for either transplanted variety of Korean[5], we instead rely on the description of recorded primary spoken data described in terms of comparative concepts drawn from traditional Korean grammatical description supplemented by reformulations of prior sketch descriptions and thirs parts documentary materials (secondary data) for the creation of the parallel grammars which are the basis of the formal CA. While the compilation of full and exhaustive documentations and descriptions of these varieties on the basis of wholly new primary data would be preferable, it is not practical within the timescale of the current study. The supplementary quantitative analysis, though, may only be based on observation of these linguistic systems in use: the primary recorded data and documentary materials. Thus, the primary data gathering focused on collecting examples of connected speech for this purpose, rather than attempting to elicit examples of particular linguistic forms.

We go on in the next section to discuss the methodology underlying the specifics of the collection of primary data.

---

[5] Here we do not seek to imply that there are no descriptions available of KM or VYK since we refer to several in sections 3.3 and 3.4. Rather, the extant descriptions are incongruous in their data gathering methods or the frameworks within which the varieties in question are described. Therefore, attempting to directly compare exclusively these prior descriptions would be likely to reflect more on the researchers' differing methodologies than the actual contemporary, synchronic linguistic differences between the transplanted varieties.

4.2.1 Primary Data Gathering Methods

The primary data gathering methods used in this research are informed by the necessity of formulating abstracted descriptions on the basis of or informed by the primary data for the CA of the linguistic forms of each variety, along with the requirement that connected vernacular speech be collected for the supplementary quantitative analysis of these transplanted varieties. Key issues to consider include the selection of linguistic consultants and the form in which data shall be gathered. We first examine the selection of consultants before going to give precise details of the data gathered from members of the Central Asian and Chinese Korean speech communities and the additional linguistic resources consulted for the purposes of compiling our descriptions.

4.2.2 Consultant Selection

We noted in section 3.2 that the common classification of Korean dialects has been made more on the basis of descriptive convenience than linguistic reality. Bearing this in mind, we must ask what the actual object of earlier research into transplanted varieties of Korean has been. In other words, it is a matter for debate whether KM and VYK are specific varieties of Korean with pre-determined characteristics related to their dialectological heritage and characteristics, or simply the names given to the varieties used by the Korean speaking members remaining in the Korean diaspora communities of Central Asian and China, respectively. Resolving this issue for the purposes of the current study enables us to establish criteria for the selection of consultants from whom primary data may be gathered.

On the basis of prior research, particularly given the recurring problem in its limited sampling of consultants and its extreme judgement sampling (Pak 2005; Kwon 2010), it seems that the prevalent attitude among the researchers who have already engaged with the transplanted varieties is that KM and VYK are specific, linguistically defined objects of study rather than

designations for geographically restricted varieties of Korean. In selecting consultants for elicitation, researchers have shown a bias towards sampling older speakers of KM, which may be argued to be only the expected outcome of the breakdown of the intergenerational transmission of this variety. The same tendency, however, may be observed in studies on Yanbian Korean (cf. Jeong 2010; Kim 2013). This particularly harks back to the early days of traditional dialectology when NORMs (Non-mobile Older Rural Males) were the preferred consultants, due to their perceived 'authenticity' and assumed linguistic conservatism. Rather than repeat such judgment sampling, the position taken for this study closely mirrors that of Berge (2010 p. 60), that "adequacy in documentation should not be limited to what is considered the purest of most traditional form of the language, nor to a particular speaker population. Speakers of all ages, socioeconomic statuses, etc., should be included".

A further issue which must be resolved in order to establish the scope of eligible consultants is the near universal bilingualism of the Korean speech communities of both China and Central Asia. On the grounds that a monolingual consultant is an anomalous rarity in these speech communities, the decision was taken to include multilingual consultants. Indeed, the primary data for this study was gathered exclusively from multilingual consultants.

Of course, the competence of multilingual speakers in the various languages which they speak varies. Among the *Cosencok* it is rare to find individuals who consider one of their languages "dominant" since both Korean and Chinese are widely used in Yanbian in a variety of domains, whereas the use of *Koryo Mar* among the *Koryo Saram* is much more restricted, given the degree to which the Korean speech community has shifted to Russian over the latter half of the twentieth century, and older speakers may be affected by a certain degree of language attrition. Rather than have consultants take linguistic tests to demonstrate their competence or establish whether the variety they speak could be considered 'authentic', self-identification

as a speaker of Vernacular Yanbian Korean or *Koryo Mar* was sufficient for inclusion in our

sample on the grounds that:

> The speech community is not defined by any marked agreement in
> the use of language elements, so much as by participation in a set of
> shared norms. These norms may be observed in overt types of
> evaluative behaviour, and by the uniformity of abstract patterns of
> variation which are invariant in respect to particular levels of usage.
> (Labov 1972 pp. 120-121)

Consequently, potential consultants who appeared to be speakers of transplanted varieties,

but considered their experience of other varieties of Korean too great to speak 'real' VYK or

KM were excluded from this sample, as were 'rememberers', a group which was particularly

large in the case of KM, who believed that they had lost all communicative competence in the

variety despite demonstrable awareness and proficient use of particular culturally bound

vocabulary, for example greetings, and the names of foodstuffs, festivals, or rites of passage,

as well as often a high degree of passive understanding of the language.

4.3 Data Gathering Techniques

The question of what constitutes adequate data is largely dependent upon the research

questions being investigated and the purposes to which it is to be put. However data is

gathered, though, it can only ever be a small sample of a linguistic system's output since it is

impossible to record all the possible utterances of a language or variety made by the every

individual speaker under all possible circumstances. Since creating a record and analysis of

every possible linguistic form and every possible use of each linguistic form would be the only

way to truly exhaustively describe a linguistic system, we must examine other, more practical

approaches to the collection of primary linguistic data. We now turn to sub-disciplines of

linguistics which rely upon the empirical analysis of data in order to better clarify what would

constitute adequate data for our purposes.

Although it has already been noted that completely exhaustive documentations of given languages are impossible, we may ask whether it is possible to compile a large amount of primary linguistic data, for example in the form of large scale corpora, for comparison. Would it perhaps be possible, say, to compile corpora of transplanted varieties of Korean after the model of the so-called 'Brown family' of corpora[6] used in the comparison of global varieties of English? While such corpora provide an excellent basis for comparison, the restricted domains of use of KM do not allow for the compilation of a genre spanning corpus of primarily written language, as are those of the 'Brown family'. The primacy of written sources in such corpora also presents problems since the written forms of Korean used in both the YKAP and Central Asia owe more to codified varieties of Korean than the local vernacular language, with some even going so far as to call the situation "diglossic" (Kwon 2013).

Nevertheless, given that the researcher conducting this project lacks native competence in either KM or VYK, our descriptions must necessarily have an empirical basis. Since the creation of comparable large scale corpora consisting of a wide variety of published texts drawn from different genres has been demonstrated above to be an unattainable goal, given the advanced shift from *Koryo Mar* in Central Asia and unrepresentative nature of the codified variety of Korean used in much writing in Yanbian, and would furthermore not give the strongest linguistic signal of the varieties which we hope to investigate, we must instead propose an alternative. In our case this is a smaller linguistic database consisting of attestation of the spoken language, such as the primary and supplementary linguistic materials detailed above, which may be considered suitable for informing the descriptions to be created and contrasted.

---

[6] The name derives from the corpus of American English created in 1961 by Francis and Kučera at Brown University in 1961. It was not only among the first generation of computer readable corpora, but its considered, principled approach to sampling and data coding directly led to the creation of other corpora constructed after the same principles, for example the Lancaser-Oslo/Bergen Corpus of British English.

Furthermore, these smaller corpora may be employed very effectively if they are used for contrasting the observed linguistic features of VYK and KM with relation to the comparative concepts of CSK, rather than trying to build self contained grammatical description of these varieties from scratch. Such smaller corpora or textual data-bases may be found in the sub-disciplines of traditional dialectology, sociolinguistics and documentary and descriptive linguistics, the approaches of which we now examine.

When the explicit examination of a linguistic system is embarked upon, attitudes towards and perceptions of that system are hugely influential in its evaluation (e.g. Inoue 1999). Throughout this work we use, where possible, the neutral term "variety" to refer to both standard and non-standard forms of languages since it is felt that use of the term dialect connotes a mildly negative evaluation of non-standard varieties. A further reason for avoiding this term is to distance the current study from the methods of traditional dialectology and the data gathering practices associated with it.

While the field of traditional dialectology has been hugely influential in the development of methods for gathering primary linguistic data, it is predicated on the notions that a linguistic consultant is both willing and able to provide accurate information on the language or variety of which they consider themselves a speaker, and that a fieldworker may accurately and regularly transcribe this information in a relationship analogous to that of the ideal speaker and hearer. Consultants' metalinguistic knowledge, however, often falls short of this ideal, as do the abilities of fieldworkers (Trudgill 1982). Furthermore, overt questioning and surveys are suited to gathering only very particular types of data without shedding any light on their actual use, for example elicited vocabulary and grammatical paradigms. They are, however, quite practical for the main goal of traditional dialectology which has been defined as the "fine grained as possible inventory of the (mostly phonological) characteristics of geographically and historically related language variants" (Brandner 2012 p. 113). We further note that such

methods rely on the researcher and consultant sharing a language or variety other than that under investigation, the use of which can further prejudice the results of the survey depending on each consultant's attitudes towards the common working language.

A further questionable aspect of traditional dialectology's data gathering practice is its approach to sampling. Rather than seeking to gather data from a representative selection of members of a given speech community, traditional dialectology relies on data gathered from very few "representative speakers", who are generally the most conservative speakers of that community's variety or those whose speech diverges most strongly from a standard variety or perceived norms. When combined with the fact that geographical delineation of varieties is the primary concern of dialectology, so varieties within pre-conceived geographical areas are treated as unitary and homogeneous, a situation results in which a variety of a particular language may be described on the basis of speech data which do not adequately represent real language use in the community from which the data are gathered and questions of variation in that variety based in gender, age, social class or other extralinguistic factors are not considered. Not only would data collection carried out using this method have the issues described above but, since prior descriptions of the transplanted varieties of Korean of Central Asia (Yi et al 2000) and the YKAP (Choi et al 2002; Kim 2013) have largely been carried out within the framework of traditional dialectology, such work would not represent a significant or novel approach to them.

The issue of language heterogeneity ignored by traditional dialectology lies at the heart of the discipline of sociolinguistics and informs its data gathering procedures. Unlike traditional dialectology no single consultant's idiolect may be considered representative of a the language or variety used by a given speech community, therefore a diverse group of speakers must be consulted in order for a fuller picture of the language or variety to be drawn – that is a one which acknowledges the heterogeneous nature of that variety and the heterogeneous makeup

106

of the speech community – in order to establish whether there is any correlation between speakers' sociological characteristics and their language use. The approach to consultant selection outlined above was supplemented by the intention of sampling a diverse range of speakers of the transplanted varieties of Korean under examination. In the case of Central Asian Korean, it was not possible to achieve as equal a balance of gender or as wide a range of ages in the population sampled as in Yanbian. This may be attributed to the breakdown of intergenerational transmission of *Koryo Mar*, and its resultant lack of younger speakers, and the great disparity between the life expectancies of men and women in Kazakhstan (CIA 2013).

The data gathering procedure itself has been an essential focus of research for sociolinguists (Macaulay 2009; Tagliamonte 2012; Schilling 2013 etc.). From their perspective, data gathered using traditional dialectological survey methods is not an accurate reflection of the actual variety used by members of the speech community under examination. Not only do wordlist elicitation and direct questioning rob the items elicited of context in connected speech, precluding the possibility of revealing the full spectrum of intraspeaker variation of each item if it is elicited only once, but the accuracy of such self-reported data remains questionable, and the limited sampling of speech communities associated with traditional dialectology does not allow for correlations between linguistic and social patterns to be drawn. Reflecting on the data gathered for this project, it is clear that it would not have been possible to record the full range of variation of many features as they appear in natural speech, for example, the topic particle, examples of inter and intra-sentential code switching or an accurate impression of how similar the VYK and KM are with regard to their word order had we restricted this study to traditional data gathering methods such as wordlist elicitation or direct questioning. Conversely, there are certain forms which are not attested in the primary data which would certainly have been recorded using more traditional methods, although the quality of such data would be suspect. For the purpose of this study, we chose to prioritise the collection of

language that was ecologically valid to as great a degree as possible both in order to avoid duplicating earlier work and improve the reliability if not the exhaustiveness of our data.

Eliciting natural connected speech or "accessing the vernacular", that is the form of language in which "minimal attention is paid to speech" (Labov 1972 p. 112) is a pre-occupation of the discipline and numerous techniques have been developed to this end. Among these, perhaps the most widely used is the sociolinguistic interview, which was the method ultimately selected for gathering the majority of the primary data in Yanbian and Kazakhstan. The main advantages of this tool are twofold: planning the interviews and topics to be covered in advance allows for greater cross-consultant or cross-varietal comparability, while holding this plan to a more flexible structure allows the data-gatherer discretion to direct the interview to ensure that the consultant talks more freely, thus gathering more naturalistic data and in greater quantity. We may also note here that this method has previously been fruitfully employed by researchers looking into language variation and change in diaspora communities (Boas 2003).

There are, though, several criticisms of sociolinguistic interviews as a data gathering technique which must be acknowledged. These stem from the fact that interviews, perhaps most especially sociolinguistic interviews, are not wholly natural speech events. The presence of the researcher, regardless of their intentions, is liable to have some influence on the speech of the consultant. This phenomenon is known as the observer's paradox after Labov's formulation that "to obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed" (Labov 1972 p. 113). Furthermore the methods of the sociolinguistic interview may seem at odds to consultants with expectations cultivated by experience with other, more widely practiced styles of interview. For example, it not as confrontational or interviewer led as interviews that consulants may have encountered in much broadcast media.

In order to overcome the artificial nature of the sociolinguistic interview as an SCE, we turn to another sub-discipline of linguistics: documentary and descriptive linguistics, which is primarily associated with the recording of endangered languages. The goal of this discipline is the creation diverse and enduring records of these languages before they become extinct, consequently it is a field which has paid much attention to the gathering of linguistic data. As with sociolinguistics, the collection of texts which represent natural or 'ecologically valid' language use is paramount. While such methods as semi-structured interviews may be used towards accomplishing this goal, documentary and descriptive linguists do not consider this method of data gathering to be sufficient in and of itself for two reasons. First, while semi-structured interviews represent a real improvement on traditional dialectology's direct survey method in terms of accessing more natural language use, the context of an interview with a linguist, indeed of all SCEs, is still thought to effect consultants' language use to the point that the quality of the data gathered is compromised. Second, collecting data in only one communicative context, the interview, is considered only a limited reflection of the whole repertoire of the language or variety under investigation (cf. Lüpke 2010).

For the purposes of one particular aim of this study, the comparison of the distribution of KM and VYK's linguistic features, ecologically valid, vernacular speech gathered through the recording of observed communicative events (OCEs) would be considered ideal data. Due to restrictions of time, though, the more active approach of organising SCEs was deemed necessary. Despite its limitations, primary data were gathered using a modified form of Labov's (1972) sociolinguistic interview schedule. Modifications were considered necessary in order to localise the interview schedule and make its content relevant and engaging to the consultants involved in this study, rather than those to whom the interview schedule was originally applied. This method must be acknowledged to be something of a compromise, since it is suitable for eliciting only relatively rather than wholly naturalistic connected speech.

Also, while the interview schedule is standardised, the semi-structured nature of the interview allows the interviewer a great deal of discretion in which questions to ask and which topics to pursue with the goal of eliciting maximally natural language from a consultant who is engaged with the topic under discussion rather than the situation of the interview. One consequence of this is that data gathered using this method is not as uniform as that gathered using, for example, dialect surveys or more rigidly structured interviews.

While gathering data through semi-structured interviews ameliorates a significant number of the problems with data gathered using more traditional SCEs, several remain. For example, the set interview format's question and answer format restricts the range of language which may be used and, furthermore, the usage of the language or variety between the consultant and a third party with whom they are familiar and share a native language is likely to be different from their usage with a non-native speaker researcher in cases where a non-embedded researcher conducts the interview. Various steps were taken to mitigate these aspects of the so-called 'observer effect'. Chief among these was engaging in 'snowball sampling', in order that the researcher could make contact through social groups and gain both an introduction and, ideally, an explicit endorsement from someone known to and trusted by each consultant. In addition, interviews were not conducted 'cold' and were usually preceded by a period of conversation and introduction in order to ensure that the consultant was maximally comfortable with the researcher and, where possible, where possible the interview was conducted over multiple sessions.

These steps, though, do not completely render the data gathered through SCEs entirely naturalistic. In order to gather yet more ecologically valid data uncontrolled interactions between two or more native speakers a certain number of supplementary OCEs were also recorded, either instrumentally where possible (where consent had been obtained from all parties involved) or in the form of contemporaneous field notes (for example the case of

110

particular constructions encountered while living surrounded by the speakers of the varieties being investigated). These give the researcher access to a wider range of registers and contexts than possible through the arrangement of SCEs alone. However, due to the advanced state of the language shift in Central Asia, it was only possible to record OCEs in Yanbian.

Of course, compared to fully structured, direct elicitation, both of these methods have the disadvantage of not necessarily attesting every possible linguistic form. Nevertheless, for the purposes of the current study the collection of connected, ecologically valid, vernacular speech was prioritised, since such attestation of these varieties is remarkably sparse and, should such systematic gaps be revealed upon analysis of the primary data, secondary dialectological or linguistic sources may be consulted.

We now give precise details of the primary data collected for each variety, along with details of the supplementary linguistic resources to be consulted before moving on to a discussion of the methodology underlying the formulation of equivalent or comparable descriptions.

4.3.1 Koryo Mar

Koryo Mar data was gathered over a period of fieldwork in Kazakhstan, specifically in the city of Almaty and its surrounding *oblast'* from the 8th of October 2013 to the 8th of November 2013. Twelve consultants who considered themselves speakers of *Koryo Mar* were included in this study. Initial consultants were introduced through the embedded researcher N.S Pak[7].

---

[7] N.S. Pak linguist and native speaker of *Koryo Mar* has been involved in research and documentation projects surrounding this variety since the revival in scholarly interest which it enjoyed in the 1980s. Her establishment of academic links with international researchers and mediation between the *Koryo Mar* speech community and external linguists has led to the production to a great deal of the research upon which this thesis relies (e.g. Kwak 2008, Kwon 2010, King and Yeon 1992 etc.). It is also necessary to acknowledge the significance of her own contribution to research on the variety, notably Pak 1991; Pak 1997 and Pak 2005. Her role in promoting this severely endangered variety of Korean cannot be overstated. Recently retired from being departmental head of the Korean Studies Department at The Kazakh University of International Relations and World Languages ih. Abilay Khan, she continues there as head of the attached Korean Studies research centre and remains committed to carrying out high

Thereafter further consultants were identified using snowball sampling. Informed consent for

inclusion in this study was received from all consultants. An anonymised list of consultants

and metadata pertaining to them can be found below.

We reiterate here that due to the advanced state of shift from *Koryo Mar* to Russian in Central

Asia, it was not possible to record any OCEs of *Koryo Mar*. While native speakers remain

among the eldest generation of Central Asian Koreans, KM is effectively moribund, even as a

means of intra-community communication. Also due to this language shift, younger speakers

of the variety were not available for consultation. Where possible SCEs in the form of semi-

structured interviews based on Labov's 1972 interview schedule were carried out in which the

interviewer used standard South Korean, somewhat accommodated to the norms of KM, as

the working language while answers were given in *Koryo Mar*. Two among the consultants

were interviewed by the researcher using Russian as the language of elicitation and answered

in *Koryo Mar* due to problems of comprehension with standard South Korean (TL and NOS). All

of these interviews are supplemented by contemporaneous field notes.

Identifier	Gender	D.O.B	Place of Origin	Place of Residence
KP	F	Question refused. Age over 60.	Tashkent, Uzbekistan	Almaty, Kazakhstan
EC	F	1938	Almaty, Kazakhstan	Almaty, Kazakhstan
EN	F	Question refused. Age over 60.	Langar, Uzbekistan	Almaty, Kazakhstan
GKH	F	1940	Almaty,	Almaty,

quality documentation of *Koryo Mar* in Kazakhstand, the wider Central Asian region and all over the
FSU.

			Kazakhstan	Kazakhstan
VC	F	1941	Almaty, Kazakhstan	Almaty, Kazakhstan
VL	M	1940	Almaty, Kazakhstan	Almaty, Kazakhstan
SL	F	1942	Katta-Kurgan, Uzbekistan	Katta-Kurgan, Uzbekistan
NSP	F	1944	Almaty, Kazakhstan	Almaty, Kazakhstan
TL	F	1932	Chabirovsky Kray, Russia	Almaty Oblast', Kazakhstan
VSC	F	1933	Chabirovsky Kray, Russia	Almaty, Kazakhstan
AGH	M	1940	Tashkent, Uzbekistan	Almaty, Kazakhstan
NOS	F	1923	Chabirovsky Kray, Russia	Almaty Oblast', Kazakhstan

Table 6: KM Consultants

4.3.2 Yanbian Korean

Yanbian Korean data was gathered over a period of fieldwork in Yanbian Korean Autonomous

Prefecture, People's Republic of China from the 25th of April 2014 to the 25th of May 2014.

Fourteen consultants who considered themselves speakers of Vernacular Yanbian Korean were

included in this study. Initial consultants were introduced through the embedded researcher

Kim Kwang-su[8] and through personal contacts in the area. Thereafter further consultants were

---

[8] Kim Kwang-su (b. 1965) is a linguist and native speaker of Yanbian Korean whose wide range of
research interests include the history of the Korean language, corpus linguistics, divergence between
standardised varieties of Korean, and the development of the Korean language in China. His notable

identified using snowball sampling. Informed consent for inclusion in this study was received

from all consultants. An anonymised list of consultants and metadata pertaining to them can

be found below.

Consultants participated in semi-structured interview SCEs based on Labov's 1972 interview

schedule in which the interviewer used standard South Korean, again somewhat

accommodated toward the local norms, as the working language while answers were given in

Yanbian Korean. One among the interviews was carried out by a trained, local native speaker

of Yanbian Korean using that variety as the working language (KSN). All of these SCEs are

supplemented by contemporaneous field notes. Moreover, several of these consultants also

provided data in the form of conversational OCEs.

Identifier	Gender	D.O.B.	Place of Origin	Place of Residence
KD	F	1993	Yanji, Yanbian	Yanji, Yanbian
KGL	M	1992	Ryeongjeong, Yanbian	Yanji, Yanbian
KOC	M	1991	Yanji, Yanbian	Yanji, Yanbian
KSN	M	Question refused. Age over 60.	Yanji, Yanbian	Yanji, Yanbian

works pertaining to Korean in Yanbian include research into the linguistic divergence of Yanbian Korean
from peninsula varieties in partuclar semantic fields and its on-going standardisation, which represents a
particularly prolific contribution to the field and provides a great deal of background for this thesis. He
has furthermore been involved in the codification of a standard Yanbian Korean (Kim et al 2008), which
has been significant for drawing attention to the divergence of Yanbian Korean from peninsula varieties
and considerably raising the profile of Korean within the PRC. He continues in his role as Professor of
Korean Language and Linguistics at Yanbian University, where he is actively engaged not only in research,
but teaching at both the undergraduate and post-graduate level.

LWB	M	1992	Ryeongjeong, Yanbian	Yanji, Yanbian
MWG	M	1992	Yanji, Yanbian	Yanji, Yanbian
JHW	M	Question refused. Age over 40.	Hwaryong, Yanbian	Yanji, Yanbian
KSJ	F	1961	Hwaryong, Yanbian	Yanji, Yanbian
YH	M	1991	Yanji, Yanbian	Yanji, Yanbian
HSL	F	1934	Yanji, Yanbian	Yanji, Yanbian
GM	F	Question refused. Age over 60.	Ryeongjeong, Yanbian	Ryeongjeong, Yanbian
GMB	M	1951	Ryeongjeong, Yanbian	Ryeongjeong, Yanbian
LG	M	1993	Yanji, Yanbian	Yanji, Yanbian
HPS	F	1981	Andu, Yanbian	Yanji, Yanbian

Table 7: VYK Consultants

4.4 Supplementary Linguistic Resources Consulted

The recordings of KM were supplemented by transcriptions of notable linguistic features observed outside of recording sessions in field notes, and by the transcripts of Kwak Chung-gu's 2008 interviews with Glaphira Han in Almaty, Kazakhstan. We further consult King and Yeon's sketch description of the variety (1992) and Kwon Jae-il's (2010) grammar of *Koryo Mar* and accompanying transcription of the interviews upon which the grammar was based for the compilation of descriptions of the variety for formal comparison.

The recordings carried out in Yanbian are supplemented by similar transcriptions to those described above in field notes, and transcripts of Kwak Chung-gu's 2007 interviews with Park Nam Seong in Hunchun, Yanbian. For the purposes of compiling the description which forms

the basis of the formal CA, Jeong Hyang Ran's documentation of Yanbian Korean's declension and conjugation (2010) and Choi Myeong Ok's sketch description (2002) were also consulted.

Where linguistic examples are given, the source of the material shall be cited. Primary data shall be presented in broad IPA transcription whereas data taken from other sources shall be presented using a unified transcription system for non-standard varieties of Korean written in Cyrillic or Hangul (Barnes-Sadler 2015). Examples shall be accompanied by a grammatical gloss, following the conventions of the Leipzig Glossing Rules, and an English translation. Where they are drawn from the primary data they shall also be presented with a tag identifying their position in their linguistic database in terms of line number and and indicator of variety.

Having established the constitution of the primary and supplementary linguistic data to be examined here, we now turn to the questions of how to the next step of the CA process, i.e. the creation of comparable parallel descriptions of the transplanted varieties for CA on the basis of linguistic form.

4.5 The Formulation of Contrasts

4.5.1 Description

Having somewhat idealised or abstracted descriptions of the data outlined above for comparison is considered a pre-requisite for carrying out formal linguistic CA, here we address the theory underlying the creation of such descriptions and the precise methods which we shall follow in their construction here. While traditional CA concerned itself largely with standard languages, in the case of transplanted varieties of Korean the question of what constitutes adequate pre-analysis description is particularly relevant, since there is no widely agreed upon codified standard to consult for *Koryo Mar*. While some standardisation and codification has taken place in the YKAP, the officially recognised standard variety of Yanbian is

not representative of the vernacular speech of the area, as established in section 3.4; therefore, a description of specifically vernacular rather than standard Yanbian Korea (hence VYK) is required. The insights into the differences which have arisen between the varieties of Korean used in China and Central Asia since their original transplantation which may be gained from comparison between the primarily oral KM and the Yanbian standard variety are hugely limited.

Despite the absence of codified standards, in contrast with other situations in which linguistic descriptions are required, the varieties in question are not completely undescribed. Furthermore, the sole purpose of this study is not their description; rather similarly formatted descriptions are simply pre-requisite for undertaking a systematic comparison. Our descriptions, then, differ from those compiled under the ambit of documentary linguistics in that the grammatical structures and rules are not solely to be worked out inductively from first principles on the basis of a substantial textual corpus[9]. Rather the descriptions presented in this study represent those compiled in the "hypothesis checking stage" (cf. Dixon 2010) of a full documentation of a given language. We regard earlier sketch descriptions of these varieties and the comparative concepts supplied by grammatical descriptions of CSK as providing hypotheses relating to their grammatical organisation, which may be checked against actual vernacular usage as attested in the primary data described above.

The descriptions created for the formal CA component of this thesis shall not simply be phenomenological descriptions in which each utterance is to be described and taken to be canonical. While we shall be using the terminology borrowed from the description of CSK for

---

[9] The size of a textual corpus upon which a description which is conventionally considered adequate may be based is not firmly established. However it may be noted that the timeframe for the compilation of such a corpus of several nine to twelve month periods of fieldwork carried out "over a period of three or more years" as recommended by Dixon (2010 p. 2) is impractical for the description of two varieties within the timeframe of a PhD.

reasons discussed in the following sub-section, the ultimate classification of each aspect of the

linguistic systems of both varieties relies on structuralist principles and their empirically

observed, system internal behaviour.

Thus, we are aware that the signification of the labels used for describing grammatical

categories may vary between linguistic systems, just as the number and complexity of linguistic

features may, even in cases of linguistic systems commonly understood as 'being part of the

same language'. This is the rationale behind the supplementary quantitative analysis –

establishing whether the linguistic forms identified in the descriptions and described using

similar terminology have different distributions to one another, or whether form alone is the

differentiating factor.

4.5.2 Establishing Equivalence

In this context, establishing equivalence at the basic level is synonymous with determining

which variety particular features are suitable for comparison with one another. While it may

seem intuitive simply to compare formally similar features, this approach has come under

some criticism for its circularity:

> The circularity consists in the following: we compare in order
> to see what is similar and what is different in the compared materials;
> we can only compare items which are in some respect similar, but we
> cannot use similarity as an independent criterion in deciding how to
> match items for comparison since similarity (or difference) is to result
> from the comparison and not motivate it. (Krzeszowski 1990 p. 20)

Rather than rely on surface similarities between the systems being compared, the equivalence

or similarity which justifies such a comparison may be established with reference to an

external *tertium comparationis*. In cases of interlingual comparison, it is necessary to have in

mind broad comparative concepts (after Haspelmath 2010, but see clarification below), which

are not language (or in this case variety) particular.

While it may seem contradictory to the position of identifying VYK and KM as distinct linguistic systems, we propose that the description of CSK provides us with comparative concepts in the form of descriptive categories which, while particular to Korean, are not variety particular to the point of exclusivity and may therefore be held to be analogous to cross-linguistic comparative concepts.

This terminology, and the language particular descriptive categories which it represents, is considered suitable for use in the descriptions of the varieties of Korean which we examine here, and for identifying equivalent features of each variety, even in cases where these features exhibit divergent form to function mapping, for the following reasons:

- It is conventional to use the same terminology for cross-linguistic 'comparative concepts' as for intralingual descriptive categories, even in papers which explicitly consider the two.

- Particularly with regard to grammatical descriptive categories, the intralingual categories of Korean have not been shown to vary significantly between varieties when they are compared. For example, while the precise use of the Korean subject particle may differ between varieties, the fundamental or definitional semantic-conceptual domain which it is understood to mark is similar.

- It is outside of the scope of this thesis to formulate new descriptive categories for the Korean language and there is not as yet a universally agreed upon set of 'comparative concepts' which may simply and efficiently be used in place of the intralingual descriptive categories conventionally used to describe all varieties of Korean.

In short, the intralingual 'descriptive categories' of CSK fulfil the function of 'comparative concepts'[10] or *tertium compartionis*, and allow us to establish the basis for determining the degree of equivalence between the empirically observed forms of VYK and KM for the purposes of phonological and morphological CA.

Having established which features may be compared to one another, we now go on to clarify how the strength of the relationship between comparators, in other words their degree of equivalence may be described. The term "near-equivalence" denotes *de facto* equivalence or near identical form and function (in as much as that may be established from the attestation of the more naturalistic data collected for this project). "Non-equivalence", conversely denotes a situation in which the comparative concept is realised radically differently in the comparator varieties or where the functional category is simply absent from one of the varieties. The final, intermediate degree of equivalence is "partial equivalence". This degree of equivalence may be established in cases where the forms corresponding to the comparative concepts are plainly cognate, but exhibit minor divergence in form, or where the the corresponding instantiations of the comparative concepts in the varieties being contrasted exhibit some degree of functional difference (again, with the caveat that such functional differences may be artefacts of the data collection).

To illustrate these terms, let us take an example from the case of the transplanted varieties of Korean: near equivalence is represented by the labial stop consonants of VYK and KM, which

---

[10] We take this opportunity to emphasise the distinction between Haspelmath's 'comparative concepts' and the use of that term here. In the context of Haspelmath's work, the terms in which we frame our comparison of VYK and KM would still be regarded as language specific 'descriptive categories'. They are not and, indeed, should not be recommended for the typological comparison with which Haspelmath is primarily concerned. It is rather the case that we recognise the necessity of a framework for comparison between varieties commonly conceived as being part of the same language which does not frame one of the varieties being compared in the 'variety particular' terms of a description of the other in a manner analogous (although practically speaking very different from) Haspelmath's more semantic, universally applicable comparative concepts.

are realised near identically in their basic forms and all of their allophonic realisations in both

transplanted varieties. Non-equivalence is exemplified by the forms corresponding to CSK –

pota in VYK and KM in that a similar form is used ubiquitously in VYK whereas the form [ma]

appears in KM. Finally, partial equivalence may be understood through a further phonological

example, that of the phonemes corresponding to CSK /l/. Here we see dissimilar allophony, for

example VYK /l/ (/l/$_{YK}$) havs only the two realisations [l] and [ɾ], whereas KM /l/ (/l/$_{KM}$) may be

realised as [l], [r] or [ɾ]. The actual process by which such relationships are to be established is

discussed in more detail immediately below.

4.5.3 Formulating Formal Contrasts

The process by which formal contrasts may be formulated may most concisely be summarised

diagrammatically (adapted from Gast 2012):
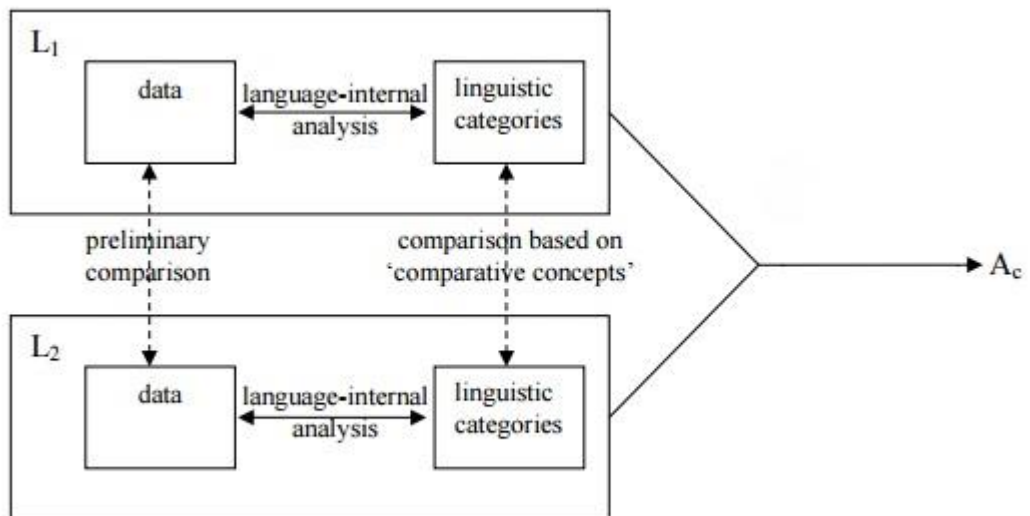
Figure 2: Diagram of the CA Process (Gast 2012)

For theoretical CA, the kind which we undertake in this study, the preliminary comparison

provides the motivation for undertaking further, more detailed analysis of the two linguistic

systems in question. The linguistic features of these systems are then described in terms of

'comparative concepts', here the categories used in grammatical descriptions CSK. The final output "A$_C$" is the contrastive analysis or parallel grammar.

Among the formal contrasts to be ascertained by this process, those of the phonological systems are perhaps the simplest to formulate. The algorithmic process which we shall follow is based on James (1980 p. 74) which consists of four steps: "draw up a phonemic inventory of L1 and L2; equate phonemes interlingually; list the phonemic variants (allophones) for L1 and L2; state the distributional restrictions on the phonemes and allophones of each language". The phonemic inventories of both KM and YK are well documented and easily checked against the field recordings mentioned above. The phonemes may be equated simply in terms of their place and manner of articulation and further analysis the field recordings shall reveal the allophonic variation present in both varieties, as well as the distribution of their allophones.

Similar processes may be followed for the formulation of formal morphological contrasts between the varieties being compared. Establishing equivalence between phonemes, though, is somewhat simpler than establishing equivalence between morphemes. For this purpose, we have already determined that it is necessary to refer to CSK as a *tertium comparationis*. Inventories of grammatical forms with similar functions to those of CSK shall be drawn up, then their variable forms (allomorphs) shall be recorded along with the distributional restrictions of each. These inventories shall be drawn in the first instance from field recordings, and are supplemented where necessary and appropriate with secondary data drawn from prior descriptions of the transplanted varieties.

The foregoing has demonstrated that CA has traditionally required a high level of idealisation of the linguistic systems being compared. In creating comparable descriptions for the CA of *Koryo Mar* and Yanbian Korean, a certain amount of idealisation is an inevitable. This is a characteristic of any description of any linguistic system due to the inter-speaker and intra-

speaker variation inherent to language. Such traditional CA, though, does not fully capture the similarities and differences between two linguistic systems, since it neglects comparison of the systems in use and homogenises to the data. By undertaking CA of the transplanted varieties we will be able to provide a detailed description of the linguistic features of KM and VYK and impressionistically determine their linguistic (dis)similarity. For this reason we propose that the CA be supplemented by a quantitative analysis of the primary data which examines the distribution of linguistic features across the population of consultants in order to determine whether there are any regular patterns in the variance of use of particular linguistic forms which may be considered distinctive of VYK or KM. Before going on on to explain the utility of such an analysis and examine how it may be undertaken, we provide a brief summary of the foregoing sections.

4.5.4 Summary

The purpose of CA is to discover and describe similarities and differences between two linguistic systems. In order to do this, descriptions of the linguistic systems in question must be available and these descriptions must be comparable, that is based on the same descriptive model. Currently, no such comparable descriptions of KM and VYK exist, therefore they must be created for this project.

To create comparable descriptions, it is first necessary to collect linguistic data upon which they must be based. For this purpose, we turn to the concrete data gathering methods of sociolinguistics (the semi-structured interview) and documentary linguistics (the gathering of naturalistic "ecologically valid" speech acts). The current study does not necessarily address the core concerns of these disciplines, i.e. social conditioned language variation and the creation of varied and enduring records of a given linguistic system, however its data collection is greatly influenced by them.

From this primary data, comparable descriptions are to be constructed using CSK grammatical concepts as the *tertium comparationis*. From these descriptions, specific points of similarity and difference between the transplanted varieties will be identified which may be conceptualised into three degrees of equivalence. These evaluations will for the basis for our qualitative comparison of VYK and KM. Furthermore, the specific points of difference found between the transplanted varieties in these descriptions will also inform the quantitative analysis of these varieties, to which we now turn.

4.6 Quantitative Analysis

Here we discuss the application of quantitative methods to linguistic problems in order to demonstrate that the suggested supplementary quantitative analysis combined with the above described CA will conclusively answer the research questions established at the end of chapter 3.

Quantitative methods have widespread application throughout linguistics ranging from describing, interpreting or making predictions about individual variation (e.g. Zipf 1939) or more generalised social variation (e.g Cedegren and Sankoff 1974) when applied to quantified language data, to applied linguistics (Hatch 1982) when applied to quantified language data or observations of learners (e.g. test scores) and the results of linguistic experiments in diverse sub-disciplines. It has been argued that linguistics is undergoing something of a quantitative turn in the early twenty first century (Gries 2009). Consequently, a relatively large number of works disemminating information about the application of general quantitative or statistical techniques to linguistic data are also available (Baayen 2008; Cantos Gomez 2013; Eddington 2015 etc.). They are not used here simply due to their developing prevalence in the field, though. Rather they will give us an alternative, complementary perspective on variation revealed by more qualitative methods, such as CA.

The CA described in section 4.5 above is will give us a detailed impression of the linguistic features and possible areas of (dis)similarity between the transplanted varieties. While this may lead us to draw certain conclusions about the synchronic relationship between KM and VYK, quantitative statistical methods allow us to test the reliability of these intuitively drawn conclusions. Indeed, it has been claimed that in cases where a linguistic researcher is attempting to definitively determine whether two or more groups are distinctively different, the ultimate use of quantitative techniques is "inevitable" (Rasinger 2008 p. 16), and they have, in fact, been serving this purpose for sociolinguists since the mid-twentieth century (e.g. Fischer 1958; Trudgill 1974) and dialectologists since the late twentieth century (cf. Chambers and Trudgill 1980; Davis 1990).

Quantitative examinations of geographically conditioned variation in language has been characterised as the application of set techniques to specially encoded dialect atlas data often subsumed under the term dialectometry (e.g. Goebl 2005). In reality, the applications of quantitative techniques to questions of linguistic variation between what are popularly conceived of as dialects or varieties of 'the same language' are manifold (Nerbonne and Wieling forthcoming). These uses may be subsumed into four tasks which have been summarised as follows:

- Data reduction

- Inference

- Discovery of relationships

- Exploration of probabilistic processes (cf. Johnson 2008 p. 3)

For answering our research questions we are interested primarily in the first and third of these. In other words, we intend on using statistical methods to discover the relationship between VYK and KM and to reduce the data from the many features to be examined in the CA to those

which most significantly account for any dissimilarity between the transplanted varieties.

Before discussing statistical techniques which may be employed to address these questions, we first discuss certain axiomatic principles which underlie all linguistic quantitative analysis and examine the form of our data and how we are to process it to ensure its suitability for being subjected to such analyses.

4.6.1 Principles of Quantitative Linguistics

Perhaps the most important axiomatic concept in quantitative linguistics is the acceptance of variation as a significant linguistic fact. This variation may be examined on the level of idiolect (i.e. the variation of an individual's realisation of a particular linguistic variable) in terms of the language use of a pre-defined group (as we do here with the language of the *Koryo Saram* and the *Cosencok*).

A further implicit assumption of quantitative linguistics, specifically where we seek to use statistical tools not only for their descriptive power, but for their inferential power, is encapsulated in "the principle of quantitative modelling" (cf. Young and Bayley 1996). This consists of the idea that, with adequate data, we are able to "make statements about the likelihood of co-occurrence of a variable form and any one of the contextual factors in which we are interested" (Bayley 2003 p. 85). To relate this directly to our objects of enquiry, it is our belief that it is possible to use the primary data to determine which linguistic features are more likely to be associated with the contextual factor of the speaker's origin in Central Asia or the YKAP.

Having established the fundamental positions required for carrying out our quantitative analysis, we may now turn our attention to our data and discuss its suitability for such an analysis and how it is to be encoded for such a purpose before moving on to examine the statistical tools which are suitable for our addressing our specific questions.

4.6.2 Encoding the Data

The pre-requisite of carrying out a quantitative analysis is quantifiable data. This may take a number of difference forms and be described in a variety of different ways (e.g. nominal, ordinal, interval and ratio). To a very great extent the form of the data determines the analyses to which it may be subjected. The linguistic data which forms the basis for the parallel description in the following chapter, recordings of SCEs and OCEs, is not inherently quantitative, so we must make choices about how we are to quantify it. Specifically we must identify the cases and the variables[11] and then discuss how they are to be measured.

In the CA, we use comparative concepts drawn from CSK to compare the linguistic features of VYK and KM and capture an impression of their inter and intra-varietal variation. Rather than continue to use these concepts for our variables in the quantitative analysis, instead we use particular features identified by the CA which were observed to vary across and between the transplanted varieties. Given the large number of features which are likely to be identified as variable across and within these varieties, our data will be analysed using multivariate statistical tests, as opposed to, for example, the commonly employed and misleadingly named univariate test prevalent in sociolinguistics: multiple regression.

In terms of cases or units of analysis, the CA deals with only two: *Koryo Mar* and Vernacular Yanbian Korean. Rather than deal with the KM and VYK datasets as wholes, for the purposes of our quantitative analysis we instead take our individual consultants as cases. This is done in order to determine whether certain among the similarities between VYK and KM noted in the CA could be accounted for on the basis of the speech of a particular groups of consultants and to test the characterisation of these transplanted varieties provided by earlier studies and

---

[11] A case may be defined as " unit or observation to be analysed" which may be either an individual person or a group (Zedeck 2014 p. 37) while a variable is a "characteristic of an entity, person or object that can take on different categories, levels or values and that can be quantified" (Zedeck 2014 p. 405)

determine whether there are groups of KM and VYK speakers whose use of particular linguistic features is particularly similar despite their geographical separation or whether there will be a clear delineation between VYK and KM speakers.

It must be acknowledged, though, that there are limitations to quantitative analysis and several caveats that must be made. The largest of these raises once more the issue of sampling and whether our sample of consultants may be considered representative of KM and VYK speakers at large. Although more data would undoubtedly improve the representativeness of our sample, we discussed in section 4.3 above the measures taken to recruit a wide range of consultants and to ensure that they produce language that is as naturalistic as possible and, thereby, gather data which is relatively representative, most especially when placed in the context of the research carried out into transplanted varieties of Korean to date.

These measures, though, call into question the comparability of the data, which is another key requirement of quantitative analysis. Whereas directly elicited data has a very low ecological validity and may be subject to a wide variety of observer effects, it is possible to the researcher to ensure the data gathered using such methods is consistent in terms of the amount of data and its precise composition. Relying on primarily on data derived from SCEs such as interviews does not come with the same guarantees. This effects the manner in which we may quantify our data and also the inclusion of our consultants in our quantitative analysis.

Given that not all speakers provided the same amount of data or within their data may have provided considerably more examples of a particular linguistic feature than other consultants, it is clear that we cannot rely on simple absolute frequency counts to represent our consultants' data lest such outliers skew the results unacceptably (Gries 2009 pp. 107-108). Rather, we must normalise this data in some way. Again, the variation in the number of

attestations of each feature presents a strong case against normalising these scores as a relative frequency, for example using percentage scores, since small amounts of variation in the attestation can result in large differences in these values. Instead, while we acknowledge the variation present in our data in practice and address it explicitly in the CA section, we opt to consider the linguistic features categorically in terms of their presence or absence in a given consultant's speech. Thus the data which we shall analyse in chapter 6 may be characterised as dichotomous or binary.

While characterising our data in this way may appear to implicitly endorse a categorical standpoint that runs contrary to the acceptance of variation in language, which was identified above as axiomatic for the practice of quantitative linguistics, in actual fact it simply changes the main focus of the variation that we are examining. Unlike classical sociolinguistic studies which examine the relevance of a number of contextual factors for on the production of a single or very small number of linguistic variables (for example Bell and Holmes 1992), instead we are investigating the relevance of a single contextual factor (geographical origin) for the use of a large number of linguistic variables. An alternative way of clarifying this issue is in contrast with traditional dialectology. Unlike the underlying assumptions of that sub-discipline of linguistics, we take the position here that not every speaker from a given region will exhibit all of the linguistic features identified with it, but using the statistical tools at our disposal we are able to see whether there is a pattern in the variable use of these features which allows us to determine geographical origin and which features are most consistently associated with the areas under investigation.

This dichotomous coding of the data is one way in which we shall deal with the unequal distribution of linguistic features over our naturalistically collected primary data. We must also acknowledge, though, that there may also be extreme cases where particular consultants have not produced several of the linguistic features which we hope to include in our quantitative

analysis. While data drawn from such consultants could contribute to the creation of the parallel descriptions for our CA, when it comes to our quantitative analysis in cases where we have insufficient data, that consultant must be excluded from our analysis.

Having determined that our data will be quantified to take the form of the dichotomous representation of a number of variables to be determined in chapter 6 for as large a number of individual consultants as possible (i.e. who are not excluded for missing data), we now introduce the methods which meet the general goal of quantitative linguistic analysis, namely, to "characterize surface variation in terms of the largest (and most abstract) set of objects that vary in the same way" (Labov 2008 p. 8) in addition to our specific goals of testing the hypothesis that VYK and KM may be distinguished on the basis of their linguistic features and discerning which of these features contributes most to this dissimilarity.

4.6.3 Discovery of Relationships

Here we explore the third use of quantitative techniques identified above: the discovery of relationships. There is a wide array of statistical techniques available for this purpose (e.g. correspondence analysis, linear regression etc.), but the most suitable for our purposes is cluster analysis since it is "a method of multivariate data analysis in which individuals or units are placed into distinct subgroups based on their strong similarity with regard to specific attributes" (Zedeck 2014 p. 47).

Since the primary concern of this work is establishing the extent to which Central Asia and the YKAP may be considered separate varieties, it may be observed that cluster analysis has been described as "a quantitative analog (sic.) to the identification of dialect regions in traditional dialectology" (Grieve et al. 2011: 20).This technique is widely used in contemporary quantitative dialectology and dialectometry (Kretzschmar 1996; Shackleton 2005; Grieve et al. 2011). As is implicit in the quotation above, though, its application, though, tends to be

somewhat more geolinguistic than our proposed use of it here. Specifically, it is used to establish the spatial distribution of aggregate linguistic (dis)similarity on the basis of data drawn from a data gathered from consultants living over a contiguous area. Most commonly this takes the form of linguistic atlas data, although corpus data may also be used. While establishing the spatial limits of KM and VYK or mapping the spatial patterns of variation in Korean in China and Central Asia fall outside of the scope of the current study, our task is in essence the same as that of dialectometrists using cluster analysis to establish dialect areas: namely, we seek to establish what groups our consultants may be divided into on the basis of the linguistic features evident in the data which they have supplied, we simply do not map these findings back onto a geographical space.

There is no consensus on which clustering algorithm is most suitable for linguistic data, since this data may take a wide variety of forms (continuous, categorical, dichotomous etc.). The data here are dichotomous and mark the presence or absence of a particular linguistic feature in a given consultant's speech. While a number of hierarchical, agglomerative clustering methods may be applied to such data, in this case Ward's Method (Ward 1963) is the selected clustering algorithm.

This is not only commensurate with other research in quantitative linguistics (Grieve et al 2011; Wolk 2014), but Ward's Method may be judged as particularly suitable for our purposes since it differs other simple linkage based methods, in that its aim is to 'join cases into clusters such that the variance within a cluster is minimised' (Field 2005 p. 5). To put this in linguistic terms, clusters established using Ward's Method are likely to represent more generally linguistically similar groups than clusters formed using single or complete linkage based methods.

4.6.4 Data Reduction

Although the general characterisation of quantitative research is that it answers questions of 'how much/many' rather than simply 'how', we suggest that it may be combined with the CA and the data described above to establish more precisely how KM and VYK differ from one another. This is to be accomplished using techniques which were designed to achieve the first of the tasks of quantitative methods identified above: data reduction, specifically dimension reduction in our case.

As with the discovery of relationships, there are a large number of techniques which are associated with dimension reduction. At first glance, the application of the commonly used data reduction technique principal components analysis (PCA) to our data is particularly appealing since it consists of an "approach in which a number of independent linear combination of underlying explanatory variables are identified for a larger set of original observed variables" (Zedeck 2014 p. 271).

Principal components analysis, or some analogous statistical technique from the family of factor analysis tests, may also be relatively commonly encountered in contemporary quantitative dialectology (cf. Nerbonne and Wieling forthcoming). These statistical tests have been described as "a quantitative analog (sic.) to the identification of isogloss bundles in traditional dialectology" (Grieve et al 2011: 17). In our case, we employ this technique in an attempt to objectively identify the characteristic features which distinguish KM and VYK from one another.

So-called 'classical' PCA is only applicable to linear or conintuous data and, while certain of the linguistic features in the feature catalogue below may be continuously measured or considered in terms of their frequency of use in the data as opposed to other linguistic forms. In the interests of consistency, and in order to ensure the inclusion of all of the same features in the

132

PCA as the cluster analysis, we instead use non-linear or categorical PCA (CATPCA), which may be used to establish patterns in the variance of dichotomous or binary data (cf. Linting et al. 2007).

We revisit these techniques to provide more detail on the concrete methods followed in their application our primary data in chapter 6.

4.6.5 Summary

In conclusion, we have demonstrated the utility of a quantitative analysis as a supplement to the CA described in the previous sub-section on the grounds that it may definitively demonstrate the impressionistic conclusions drawn from the CA. Furthermore it will give us a more nuanced picture of the variation between KM and VYK since the analyses suggested here take the variation in language use between individual speakers into account. Finally, in addition to statistical tests for the discovery of relationships, we have also determined the suitability of data reduction techniques for helping us identify the most significant features distinguishing the transplanted varieties from one another. In closing this section, we take the opportunity to re-affirm the position that "the inferences to be drawn from quantitative data can be no stronger than our confidence that the data is significant and replicable" (Labov 2008 p. 22) and emphasise the consideration given to the collection of high quality data in sections 4.2 and 4.3 above.

4.7 Limitations of this Study

It must be acknowledged that despite all of the care taken in its collectionthe data upon which this analysis is based is far from perfect. In comparison to a full documentation project, the data are relatively sparse and has several systematic gaps. For example, given the nature of the adapted interview schedule, we must rely on other linguistic sources (e.g. King and Yeon 1992; Yi et al. 2001; Pak 2005; Kwak 2012) derived from explicit elicitation for forms such as a

full range of imperatives or particularly informal language and similar. Furthermore, while the interview process was standardised to as great a degree as possible, the precise frequency with which various linguistic forms arose varied both between speech communities and between speakers. This, however, is moderated by taking the steps described in section 4.6.2 above.

In addition to these shortcomings of the data, the differing social situations of the Korean community in China and Kazakhstan, and the differing linguistic vitalities of *Koryo Mar* and Yanbian Korean also contributed to discrepancies in the constitution of the data gathered in each location. Examples of this include the impossibility of creating the same kind of age stratified sample of *Koryo Mar* speakers as of Yanbian Korean speakers, and more evenly representing the genders in sampled *Koryo Mar* speech. Both of these issues are brought about by the Korean language in Central Asia nearing the completion of its language shift to Russian, only being spoken by the so-called "grandparent" generation, which exhibits a striking gender correlated difference in the age of mortality.

4.8 Conclusion

In this section we have examined the theoretical underpinnings of language description, CA, quantitative linguistics, and our data collection protocols.

Before going on to the parallel descriptions of *Koryo Mar* and Yanbian Korean, we summarise the reasoning behind the use of these frameworks in approaching the research questions laid out in chapter at the close of the literature review in section 3.5.1.

Our first research question revolved around identifying the forms of Korean used in China and Central Asia. In order to answer this question, primary data was gathered from a number of linguistic consultants in both locations. In contrast with prior research, since this primary data came from a wide variety of sources and was gathered using the SCEs described above, it may

be considered representative of the vernacular transplanted Korean varieties to a greater extent than earlier dialectological research. Using this data and recent supplementary linguistic resources we may not only more accurately describe the forms which Korean takes in China and Central Asia, but better reflect its variability.

In order to establish whether the varieties of Korean spoken by the *Cosencok* and the *Koryo Saram* are different from one another or as similar to each other as implied in prior research, we proposed both analysis in the CA framework which will give us a qualitative impression of the formal differences between these varieties. The framework of theoretical CA is uniquely suited to a wide ranging yet granular comparison of a pair of linguistic system due to its methodological rigour. Furthermore, by combining this formal approach with a quantitative analysis, we avoid the complete reduction of transplanted varieties of Korean to abstracted descriptions and take their full range of inter-speaker variability, made available by our unique data, into account in the comparison and provide relatively objective answers to the questions of whether KM and VYK are linguistically distinct and what their distinguishing linguistic features may be.

In the following chapter, we formulate the formal contrasts between *Koryo Mar* and Yanbian Korean.

5 Contrastive Analysis

5.1 Phonology

In order to formulate contrasts between VYK and KM, we must first establish the concepts which underlie their comparison. As specified in the preceding chapter, we rely on descriptions of CSK to provide our comparative concepts. This section is divided into an examination of first consonant, then vowel phonemes, followed by brief discussions of prosody and phonotactics.

For the sub-sections dealing with phonemes, we first identify comparative concepts for before going on to discuss their corresponding phonemes in VYK and KM and assessing the degree of equivalence between the phonological systems of these transplanted varieties.

5.1.1 Consonants

While the *kwuklipkwukewen* advocates a nineteen consonant system for CSK, we follow contemporary phonology in considering the approximants to be part of the consonant rather than vowel inventory (Choi 2004; Ahn 2009; Kang 2011). Thus, the inventory of CSK phonemes which serve as comparative concepts are laid out in the following chart in their Yale transliteration:

		Labial	Alveolar	Palatal	Velar	Glottal
Stop and affricate	**Lax**	p	t	c	k	
	Tense	pp	tt	cc	kk	
	Aspirated	ph	th	ch	kh	
Fricative	**Non-tense**		s			h
	Tense		ss			
Nasal		m	n		ng	
Liquid			l			
Approximant		(w)		Y	w	

Table 8: The Phonological Inventory of CSK

We examine them now in turn according to their manner of articulation.

5.1.1.1 Stops

These phonemes as recorded in the primary data may be observed to have near identical places and manners of articulation[12] as well as similar distributions in CSK and both

---

[12] Unless otherwise noted (for example in section 5.1.3) the phonemes under discussion here are auditorily rather than instrumentally coded. This is considered sufficiently accurate for the vast majority

transplanted varieties of Korean. Cells filled with N/A in the tables below represent non-attested forms.

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
KM2495[pam]	night	KM1771[abai]	uncle	KM267[sɨdip̚]	marriage
KM938[pʰal]	eight	KM477[nampʰjən]	husband	N/A	N/A
KM207[p'ali]	quickly	KM1158[kip'ɨda]	to be happy	N/A	N/A

Table 9: Allophonic Realisations of KM Bilabial Stops

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
KM1313[ta]	all	KM660[adɨri]	son		
KM1827[tʰɨkʰi]	especially	KM1284[katʰɨn]	the same	KM565[pat̚]	field
KM359[t'aɾi]	daughter	KM2038[ət'əkʰe]	how?		

Table 10:Allophonic Realisations of KM Alveolar Stops

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
KM125[korjə]	Korea	KM671[t͡ʃogɨm]	a little	KM[səntʰɛk]	choice
KM1950[kʰɨn]	big	KM925[ʃikʰet͡ʃ'i]	(I/you/(s)he) sent	N/A	N/A
KM928[k'ad͡ʒi]	up to	KM416[kak'aun]	close	N/A	N/A

Table 11: Allophonic Realisations of KM Velar Stops

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
VYK2[pumonim]	parents	VYK1745[pwubun]	proportion	VYK557[pap̚]	food
VYK2941[pʰalʃip̚]	eighty	VYK301[sɨlpʰɨn]	sad	VYK4[jəp̚]	next to
VYK2235[p'uri]	roots	VYKKm807[nap'ɨn]	bad	N/A	N/A

Table 12Allophonic Realisations of VYK Bilabial Stops

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
VYK953[ta]	all	VYK958[ədi]	where?	N/A	N/A
VYK2517[tʰojoil]	Saturday	VYK286[potʰoŋ]	normal	VYK1975[mit̚]	beneath
VYK1546[t'o]	also	VYK1249[it'aga]	later	N/A	N/A

Table 13:Allophonic Realisations of VYK Alveolar Stops

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
VYK362[kat͡sok̚]	family	VYK2320[jəgi]	here	VYK263[kwamok̚]	subject
VYK3654[kʰɨn]	big	VYK2869[t͡ʃokʰa]	niece	N/A	N/A
VYK1184[k'ok̚]	precisely	VYK325[hamnik'a]	do you?	N/A	N/A

Table 14: Allophonic Realisations of VYK Velar Stops

In addition to similarities in their production and distribution, we further note that the stops are affected by similar phonological processes in all varieties of Korean. For example, lax stops undergo intervocalic voicing and all stops (although largely only lax stops are attested in the primary data) undergo word final 'neutralisation'. These processes can also be seen in the examples above. A further phonological process relevant to the stop consonants of the transplanted varieties is the palatalization of the dental and velar stops. This synchronic phonological process may occur when these stops are followed by high front vowels. Examples of this may be found for both alveolar and velar stops in KM:

KM	CSK	English
KM2276[t͡ʃil]	kil	road
KM1538[t͡ʃirɨmi]	kirum	oil/fat (SBJ)
KM696[t͡ʃimt͡ʃʰi]	kimchi	kimchi
KM153[kat͡ʃʰi]	kathi	together

Table 15: Palatalised Velar Stops in KM

Although this process does not operate wholly without exception:

KM	CSK	English
KM871[kiri]	*kil*	road (SBJ)
KM714[kirɨmi]	*kirum*	oil/fat (SBJ)
KM153[k'ɨtʰi]	*kkuth-i*	the end (SBJ)

Table 16: Non-palatalised Velar Stops in KM

Turning to the VYK primary data, though, this process is a little more rare, but is attested for both velar and dental stops:

VYK	CSK	English
VYK1769[t͡ʃirɨm]	*kirum*	oil/fat
VYK751[kat͡ʃʰi]	*kathi*	together

Table 17: Palatalised Alveolar Stops in VYK

We discuss other phonological processes which affect stops where they appear next to other consonants in the section on phonotactics below

Consequently we consider these nine phonemes near equivalents.

5.1.1.2   Fricatives

As with the stops, these phonemes have similar phonetic and distributional characteristics in both transplanted varieties.

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
KM62[saram]	person	KM788[nosɛ]	Russian	KM2031[mot̚]	not
KM1495[s'am]	fight	KM417[vas'o]	I came	N/A	N/A
KM1[hao]	I do	KM514[kohɛ]	church	N/A	N/A

Table 18: Allophonic Realisations of KM Fricatives

Initial Position	Translation	Intervocalic	Translation	Final	Translation

	Position			Position	
VYK6[suksa]	dormitory	VYK1401[t͡sosən]	Korea	VYK1645	similar
				[pisɨt̚]	
VYK1766[s'ɨmda]	they use	VYK121[is'ɨmda]	I am present	N/A	N/A
VYK822[hand͡ʒok̚]	Han Chinese	VYK979[sohak'yo]	primary school	N/A	N/A

Table 19: Allophonic Realisations of VYK Fricatives

We further note that the phonological processes associated with them (word final 'neutralisation', palatalization before high front vowels in the case of phonemes corresponding to CSK /s/ and /ss/ and the aspiration of surrounding lax consonants in the case of /h/) also appear regularly in each transplanted variety.

KM141	[o-ʃip-sam	njən-do-e	hak'o-r	pʰil	ha-go]
	five-ten-three	year-degree-LOC	school-OBJ	graduation	do-DM

I graduated from school in (nineteen) fifty three.

KM684	[han-ʃ'ik̚	hansan]
	one-each	always

(They drank) one each every time.

KM800	[i-ʃip-sam	jən-e	na-kʰo]
	two-ten-three	year-LOC	be born-DM

They were born in (nineteen) twenty three

VYK1048	[han	ʃigan	sa-ʃip	pun	t͡ʃəŋdo]
	one	hour	four-ten	minute	extent

| | It | lasts | for one | hour | and | forty | minutes. |

VYK511   [put̚-kɨlʃ'i]

writing brush-handwriting

calligraphy

VYK2257   [t͡ʃʰiŋgu-dɨl-do   man- kʰo]

friend-PL-ADD   be many-DM

I   also   have   a   lot   of   friends.

A voiced alveolar fricative also appears in the speech of some Koryo Saram. This appears largely in nonce-loans from Russian (e.g. [pozdravlʲaj hago] 'they greet and…'), very occasionally as an allophone of /c_KM/ (e.g. [azu] c.f CSK *acwu*) or even more rarely as a voiced allophone of /s_KM/ in intervocalic position (e.g. KM 233 [ezə] 'in'). This sound does not appear as an allophone of /s_VYK/, but appears very frequently as an allophone of /c_VYK/, as discussed below.

The fricative phonemes of the transplanted varieties may also be classified as near equivalents.

5.1.1.3 Affricates

In both KM and VYK these phonemes share a tri-partite manner distinction with their CSK comparator.

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
VYK1585[t͡ʃuŋguk̚]	China	VYK2545[id͡ʒe]	now	VYK294[nat̚]	day time
VYK2135[t͡ʃʰingu]	friend	VYK28[kot͡ʃʰu]	chilli	VYK397[mjət̚]	how many?
VYK370[t͡ʃ'uk̚]	continuously	VYK1432[ət͡ʃ'i]	how?	N/A	N/A

Table 20: Allophonic Realisations of VYK Affricates

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
KM80[t͡ʃar]	well	KM546[abəd͡ʒi]	father	KM1299[nat͡ʼ]	day time
KM2090[t͡ʃʰakʰham]	kindness	KM1676[at͡ʃʰɨm]	morning	KM2271[mjət͡ʼ]	how many?
KM2411[t͡ʃʼok]	direction	KM1306[ət͡ʃʼɛ]	why?	N/A	N/A

Table 21: Allophonic Realisations of KM Affricates

Furthermore, unconditioned allophonic realisations of both palatal and non-palatal variants of

the lax affricate are present in both varieties.

KM492 [taʃkent saram mar t͡sar ha-nɨnde]

    Tashkent person language well do-DM

    People from Tashkent are good speakers.

KM145 [t͡so-ɨn   sarɨm]

    good-ADN   person

    (They are) good people.

KM571 [kolxoz-ɨ-do   t͡ʃogɨm-t͡ʃogɨm t͡ʃos'-e]

    collective farm-TOP-also a little-a little good.PST-DECL.NEUTR

    Little by little the collective farm got better, too.

VYK1565 [t͡soaha-nɨn   ɨmʃig-ɨnɨ   man-ɨnde]

    like-ADN   food-TOP   be many.PRS-DM

    There is a lot of food that I like.

VYK2444 [t͡seil   t͡ʃoaha-nɨn   gə- nɨn]

most		like-ADN	thing-TOP

Do	you	mean	the	one	that	I	like	the	most?

We note, however, that this is considerably more common in VYK than KM and has developed

to such a point that the lax affricate $/c_{VYK}/$ is often realised as [s] or [z], as below.

VYK1010   [sənsɛŋnim   t͡sazu   gyohwan   hɛs'-ə]

               teacher   often   exchange   do.PST-DECL.NEUTR

               The   teachers   often   changed.

VYK1675   [t͡sosən   minzog-i   tʰɨkt͡ʃ'əm-iga

               Korea   people-SBJ   special characteristic-SBJ

               norɛ   t͡sal   puru-go]

               song   well   call out.PRS-DM

               A special characteristic of the   Korean people is singing well.

Due to these differences in the realisations of $/c_{VYK}/$ and $/c_{KM}/$, we consider them to be only

partial equivalents. Palatal variants of the tense affricates are not attested in the primary data

and attestation of palatal variants of the aspirated affricate (e.g. VYK 1752 [t͡sʰai], KM 498

[t͡sʰɛk̚]) is very rare, thus the phonemes $/cc_{VYK}/$ and $/cc_{KM}/$ are considered near equivalents as

are $/ch_{VYK}/$ and $/ch_{KM}/$.

5.1.1.4 Nasals

Turning first to the phoneme /m/, it may be observed that the VYK and KM realisations of this

phoneme are similar in terms of both their realisations and distributions. Thus, they may be

considered near equivalents. Examples of their distribution realisation may be found below.

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
VYK1066[mabəp̚]	magic	VYK52[əməni]	mother	VYK2550 [t͡ʃigɨm]	a little bit
VYK1699[minzok̚]	people	VYK449 [t͡ʃumal]	weekend	VYK716 [poram]	reward

Table 22: Realisations of VYK Bilabial Nasal

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
KM2500[moɾo]	I do not know	KM436[ədimɛ]	where?	KM76[sənsɛ̃m]	teacher
KM1217[mar]	language	KM150[əmaj]	mother	KM1676[at͡ʃʰɨm]	morning

Table 23: Realisations of KM Bilabial Nasal

The phonemes corresponding to CSK /n/ and /ng/ are not so straightforward in their behaviour. While their basic realisations in CSK, KM and VYK are similar phonemes corresponding to CSK /n/ may appear in initial, intervocalic or final positions in the transplanted varieties as in CSK, while /ng/ may only appear in intervocalic or final positions. Furthermore, there are additional restrictions on the distribution of these phonemes which we discuss below the following examples of their basic forms and distributions.

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
VYK2909[nɨl]	always	VYK878 [halməni]	grandmother	VYK744 [kjəron]	marriage
VYK2468 [nad͡ʒɨŋe]	later	VYK1121[əni]	elder sister	VYK1388 [musɨn]	which?

Table 24: Realisations of VYK Alveolar Nasal

Initial Position	Translation	Intervocalic	Translation	Final	Translation

144

		Position		Position	
KM239[noŋt͡ʃʰən]	village	KM1201[hanɛ]	one	KM2371 [ton]	money
KM1399[nalt͡ʃʼa]	date	KM1493 [minuri]	daughter-in-law	KM1316 [jesun]	sixty

Table 25: Realisations of KM Alveolar Nasal

A phonological rule of CSK is the so-called Head Sound Rule (*twuumpepchik*). This restricts /n/ from appearing in word-initial position before /i/ and the glide /y/[13]. It has been argued elsewhere that the Head Sound Rule does not operate in KM (Pak 2005 p. 73). However, there are no examples of the phoneme /n/ appearing in an otherwise restricted position in the primary data for either transplanted variety. Rather an operation of the Head Sound Rule congruent with CSK is attested in both varieties. One example of this is the word for 'woman', which is regularly realised as [jəd͡ʒa] without an initial [n] without exception in the primary data.

In addition to the operation of the Head Sound Rule, nasal deletion may take place where the phonemes /n/ and /ng/ precede high front vowels in non-initial positions, most especially when followed by the subject particle [i] in KM, although its application is variable. The deletion of the nasal consonant often leads to the nasalisation of the preceding vowel.

Reference	KM	CSK	English
KM1889	[mãi]	*manhi*	many
KM916	[toŋsɛ̃]	*tongsayng-i*	younger sibling
KM699	[kand͡ʒɛ̃]	*kancang-i*	soy sauce
KM1113	[nampʰjə̃i]	*namphyen-i*	husband

---

[13] This rule also restricts /l/ from appearing in word initial position. We discuss this further in the following sub-section.

| KM101 | [kĩjaŋ] | *kunyang* | just |

Table 26: Nasal Deletion in KM

Reference	VYK	CSK	English
VYK105	[mãi]	*manhi*	many
VYK1421	[wand͡ʒə̃i]	*wancenhi*	completely
VYK2863	[halmə̃i]	*halmeni*	grandmother
VYK768	[koŋd͡ʒãe]	*kongcang-ey*	in a factory
VYK748	[kɨrə̃ik'a]	*kulenikka*	therefore

Table 27: Nasal Deletion inVYK

Despite occurring in both varieties, this appears to be a more productive process in KM than VYK. Nevertheless, the alveolar and velar nasal phonemes of the transplanted varieties may also be considered near equivalents.

5.1.1.5   The Liquid

In CSK as in VYK this phoneme may be realised as either an alveolar tap [ɾ] or an alveolar lateral approximant [l] depending on its position in the word and its surroundings.

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation
VYK1139[le]	example	VYK2625[saram]	person	VYK025[t͡ʃumal]	weekend
VYK266 [lɛŋmjən]	cold noodles	VYK477[tarɨn]	different	VYK886[sɛŋhwal]	life

Table 28: Allophonic Realisations of the VYK Liquid

Not all such realisations are common KM, although they are attested. The appearance of the liquid phoneme in initial position is especially limited. This is discussed further below with regard to the Head Sound Rule.

Initial Position	Translation	Intervocalic Position	Translation	Final Position	Translation

KM2025[rodoŋ]	labour	KM1052[saram]	person	KM6[sawəl]	April
N/A	N/A	KM1539[t͡ʃirɨm]	fat/oil	KM2018[nal]	day

Table 29: Allophonic Realisations of the KM Liquid

In addition, this phoneme may be realised as the trill [r] in word final and intervocalic positions in KM as well as (rarely) in consonant clusters.

Reference	KM	CSK	English
KM14	[ir]	*il*	work
KM788	[kɨr]	*kul*	writing/text
KM450	[t͡ʃar]	*cal*	well
KM987	[korjə]	*kolye*	Korean
KM1323	[irgət͡ʃʼi]	*ilkesse*	(I/you/(s)he) read

Table 30: Realisation of the KM Liquid as [r] in Word-Final Position

While this realisation is not unheard of in peninsula varieties and there are many strong arguments against for this feature being an archaism held over from a form of Hamgyeong Korean common to both the YKAP and Central Asia (cf. King 1992 p. 203), we note that this realisation of /l/ is very common in KM and not attested in the primary VYK data. It may thus be suggested that prolonged contact with Russian is a contributory if not causal factor to the (continued?) allophonic realisation of /l/ as a trill since only KM and not VYK has been in such a contact situation.

Although an example of an initial liquid phoneme in KM was provided above, this is exceptional in the primary data. Rather, it appears that there is also a significant difference between these varieties with regard to the distribution of the liquid phoneme. Specifically, as claimed in earlier works (cf. Pak 2005 p. 74) /l_KM/ does not canonically appear in word initial position in KM. As with /n_KM/, discussed above, we see the Head Sound Rule operating as it would in CSK with the vast majority of appearances of word initial liquids accounted for by nonce-borrowing from Russian.

Reference	KM	CSK	English
KM2025	[rodoŋ]	*nodong*	labour
KM846/1303	[jəksa]	*yeksa*	history
KM187	[juk]	*yuk*	six
KM2174	[nɛŋ]	*nayng*	cold
KM598	[nosa]	*lesia*	Russia

Table 31: Operation of the Head Sound Rule in KM

Conversely, the liquid phoneme was frequently observed to appear in word initial position, even before high front vowels in the VYK data.

Reference	VYK	CSK	English
VYK888	[rjəksa]	*yeksa*	history
VYK2057	[rjuhak̚]	*yuhak*	study abroad
VYK2330/1611	[lihɛ/rihɛ]	*ihay*	understanding
VYK94	[lje]	*yey*	example
VYK441	[rjed͡ʒəl]	*yaycel*	manners
VYK286	[rogɨm]	*nokum*	recording
VYK567	[rɛŋmjən]	*nayngmyen*	cold noodles

Table 32: Operation of the Head Sound Rule in VYK

A combination of their differences in realisation and distribution leads us to conclude that the liquid phonemes of VYK and KM are only partially equivalent.

5.1.1.6 Approximants

The position of these sounds in Korean phonology has long been uncertain. It has been common practice since the fifteenth century to consider them part of complex vowel nuclei rather than independent phonemes. Here, though, we follow the practice of contemporary Korean phonology (Choi 2004; Ahn 2009; Kang 2011) in including them in our discussion of the consonants of VYK and KM.

The post-alveolar approximant /y/ appears with a similar realisation, i.e. [j], and distribution, i.e. it may not precede all vowels apart from /i/ or /ɨ/, in VYK and KM. We further note that, in common with CSK, the phonemes /y_VYK/ and /y_KM/ may only appear in initial or intervocalic positions, as below. In careful speech in CSK the canonical diphthong /uy/ may be realised as [ɨj], but in this is not attested in the transplanted varieties.

VYK	English	KM	English
VYK284[jakʰhan]	a little bit	KM2254[jag]	medicine
VYK2102[jenal]	the old days	KM1316[jesun]	sixty
VYK86[sarajo]	(I/you/(s)he) lives	KM1203[ijɛgi]	conversation
VYK2399[ijagi]	conversation	KM1935[nɛjad͡ʒimu]	I have to pay

Table 33: Realisations of the Post-Alveolar Approximant /y/ in VYK and KM

In both varieties the phonemes corresponding to CSK /y/ are subject to irregularly applied glide deletion following consonants, as can be seen in the examples below.

Reference	KM	CSK	English
KM981	[hɛk'o]	*hakkyo*	school
KM125	[hamə̃]	*hamyen*	if I do
KM986	[hak'jo]	*hakkyo*	school
KM1159	[kip'umjə̃]	*kippumyen*	if I were happy

Table 34: Variable Deletion of /y/ in KM

Reference	VYK	CSK	English
VYK409	[ələun]	*elyewun*	difficult
VYK222	[hamə]	*hamyen*	if I do
VYK2133	[əljəum]	*elyewum*	difficulty
VYK277	[hamjən]	*hamyen*	if I do

Table 35: Variable Deletion of /y/ in VYK

Finally, we note that both /y_VYK/ and /y_KM/ may also cause umlaut in surrounding vowels:

Reference	KM/VYK	CSK	English
KM156	[ir hamje]	*il hamyen*	if I work
KM1209	[koɾe]	*kolye*	Korea
VYK2788	[hamjensə]	*hamyense*	while doing
VYK104	[kjeŋhəm]	*kyenghem*	experience

Table 36: Umlaut Caused by /y/ in KM and VYK

Like the post-alveolar approximant, the labio-velar approximant has some restrictions on its distribution in CSK which also occur in VYK and KM, i.e. it may not precede /o/, /wu/ or /u/. However, it has starkly different realisations in VYK and KM. In the case of Yanbian Korean, its realisation is almost identical to CSK:

Initial Position	Translation	Intervocalic Position	Translation
VYK1071[waŋ]	king	VYK2649[ənniwa]	with my sister
VYK1346[wəlɛ]	originally	VYK2155[towa]	(I/you/(s)he) helps and…

Table 37: Realisation of the Labio-Velar Approximant /w/ in VYK

In common with CSK, too, neither /w$_{VYK}$/ nor /w$_{KM}$/ appear in final position.

In KM, however, this phoneme has undergone fortition and is realised frequently as [v]:

Initial Position	CSK	English	Intervocalic Position	CSK	English
KM419[vəl]	*wel*	month	KM2339 [kat'avat͡ʃ'i]	*kasstawasse*	(I/you/(s)he) went and returned
KM905[vəndõ]	*wentong*	the Far East	KM874 [terjəvat͡ʃ'im]	*teylyewasse*	they brought (us) here

Table 38: Realisation of the Labio-Velar Approximant /w/ in KM

A with a similar realisation attested in intervocalic positions in earlier forms of Korean.

Whereas this has undergone lenition and disappeared in CSK and VYK, examples of a corresponding reflex may still be found in KM:

150

KM581	[havund͡ʒa	kɨ	gə	t͡ʃɛmi	əp-t͡ʃ'i]
	alone	that	thing	fun	be absent-PRS.DECL.NEUTR

It is no fun to do that alone.

The appearance of this sound in place of etymological [w], though, we attribute to the influence of Russian, since the labial approximant is not present in the phonemic inventory of that language. Supporting this argument in the fact that many Russian loanwords in which this sound appears are used in KM speech (e.g. [univʲersitʲet] – university, [sovʲetskɨj] – Soviet, [avtobus] – bus etc.) whereas VYK loanwords which taken from words in which this sound appears in source languages (e.g. [ketibɨi] – KTV and [telibi] - television) have generally been incorporated into the existing phonological system of VYK.

Finally, an additional point of similarity between the phonemes corresponding to the CSK labio-velar approximant in the transplanted varieties is that they may both be dropped following consonants as in the examples below:

Reference	KM	CSK	English
KM1442	[pas'o]	*pwasseyo*	I saw
KM406	[kɛnt͡ʃʰank'e]	*kwaynchanhgey*	pleasantly
KM84	[pwas'o]	*pwasseyo*	I saw
KM50	[kjokwasə]	*kyokwase*	textbook

Table 39: Variable Deletion of /w/ in KM

Reference	VYK	CSK	English
VYK2061	[dɛge]	*toygey* [twɛge]	extremely
VYK2529	[kɛnt͡ʃʰansɨmida]	*kwanchansupnita*	it is OK
VYK2258	[dwɛ]	*toy* [twɛ]	it works
VYK1094	[kwɛnt͡ʃʰant'ago]	*kwaynchantago*	he said it was OK

Table 40: Variable Deletion of /w/ in VYK

Due to their entirely different realisations, but similar distributions we consider these phonemes to be partial equivalents.

5.1.2 Summary

A chart outlining the equivalences established between the phonemes of VYK and KM which correspond to specific CSK phonemes may be found below:

Consonant Phoneme	Near Equivalent	Partial Equivalent	Non-Equivalent
/p/	✔		
/ph/	✔		
/pp/	✔		
/t/	✔		
/th/	✔		
/tt/	✔		
/c/		✔	
/ch/	✔		
/cc/	✔		
/k/	✔		
/kh/	✔		
/kk/	✔		
/s/	✔		
/ss/	✔		
/h/	✔		
/m/	✔		
/n/	✔		

/ng/	✔	
/l/		✔
/w/		✔
/y/	✔	

Table 41: Summary of Contrasts Between VYK and KM Consonants

5.1.3 Vowels

Descriptions of CSK are split between those traditional, orthographically based analyses which identify ten so-called 'simple vowels' (*tanmoum* - monopthongs) and eleven 'complex vowels' (*icwungmoum* - diphthongs) (cf. *kwuklipkwukewen* 2016) and those which as few as seven 'simple vowels' (cf. Shin 2015 p. 4) and do not consider 'complex vowels' to be phonological units at all, but rather combinations of approximant consonants and monophthongs. While we hold with the latter view, we follow scholarly consensus (Lee and Ramsey 2000 p. 64; Ahn 2009 p. 43; Kang 2011 p. 117) and use a ten vowel system to provide comparative concepts for our discussion of the vocalic systems of KM and VYK. The traditional 'complex vowels' were discussed above in the section devoted to approximate consonants, consequently our discussion of the vowels is restricted to the ten canonical simple vowels:

/i/, /ey/, /ay/, /a/, /o/, /e/, /u/, /wu/, /oy/, /wi/

Despite being Romanised using digraphs /ey/, /ay/ and /wu/ are monopthongs in spoken vernacular Korean. The Romanisation of /oy/ and /wi/ reflects the fact that they have developed into diphthongs in most spoken varieties of Korean. Indeed, they are largely pronounced as diphthongs in VYK and KM but they are discussed along with the phonemes which correspond to the monopthongs out of adherence to our comparative concepts. Due to their diphthongal realisation in the transplanted varieties, though, these phonemes do not appear on the formant charts presented below.

Unlike the consonants examined above, the contrasts between the monophthongs of VYK and KM are not so striking. That is, the corresponding vowels could all be described similarly in terms of their phonological features (i.e. lip rounding and relative position of the body of the tongue in the vowel space). Therefore, in order to identify any slight differences between the acoustic properties of the monophthongs of KM and VYK formant charts were created.

They were created on the basis of measurments of the values of the first and second formants (F1 and F2) of a sample of vowels taken from the primary data gathered for this thesis. Ten random instances of monophthongs appearing in the same positions as the CSK comparator vowels /i/, /ey/, /ay/, /a/, /o/, /e/, /wu/ and /u/ were identified for ten speakers of both VYK and KM. The first and second formants were measured using Praat (Boersma and Weenink 2013) and then the mean value was taken from the one hundred F1 values and one hundred F2 values of each vowel. These mean values of the F1 and F2 of the monophthongs are displayed on the charts, along the Y and X axes, respectively.

Discussion of the monopthongs follows the formant charts and is in turn followed by discussion of the diphthongs.
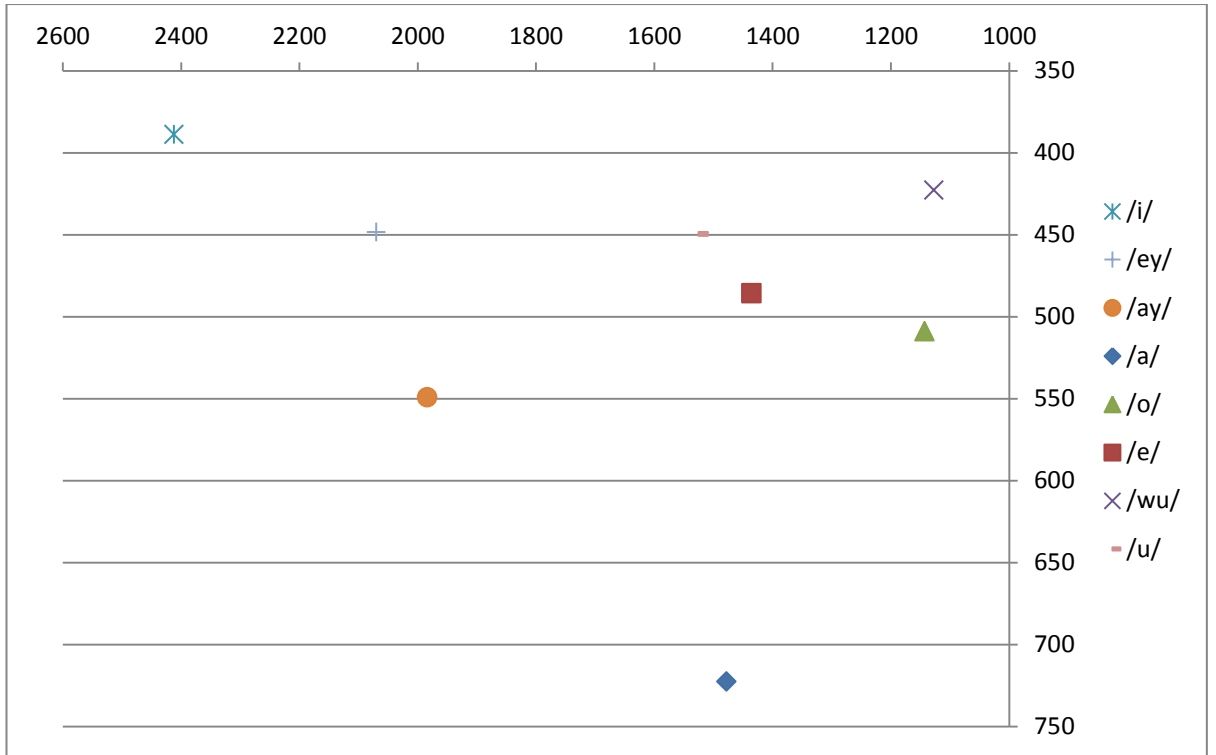
Figure 3: Formant Plot for KM Monopthongs
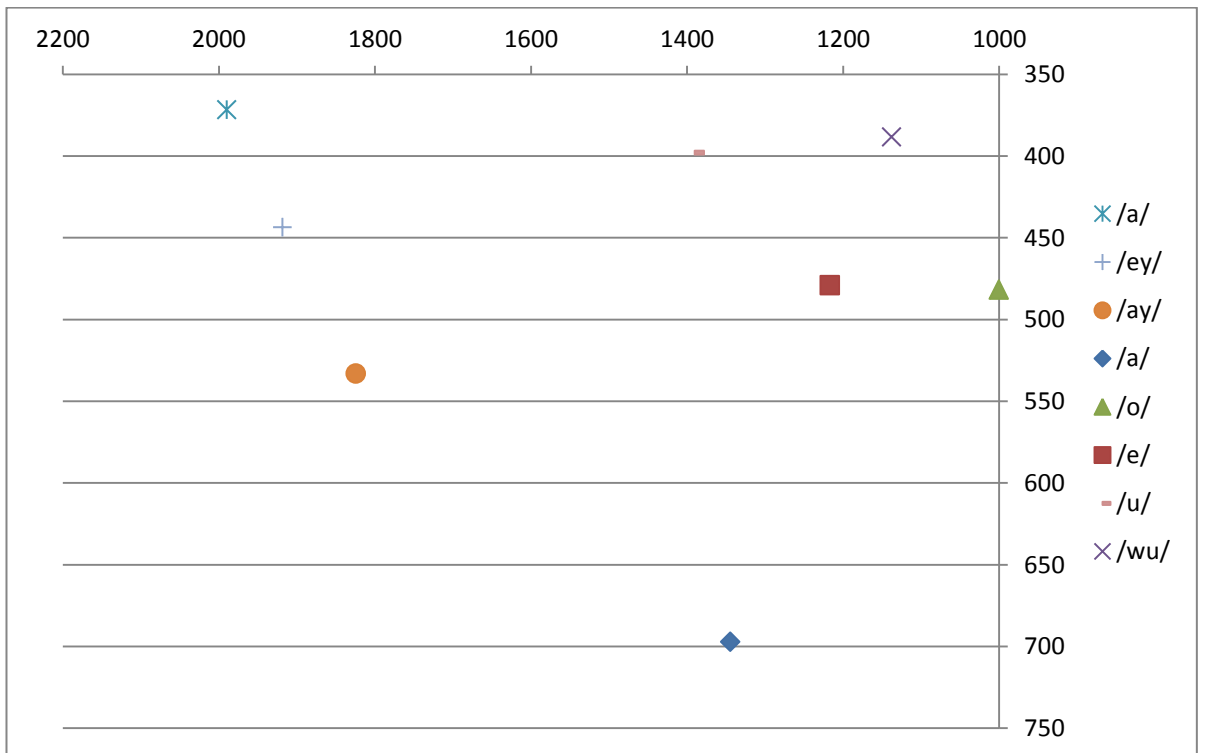
Figure 4: Formant Plot for VYK Monopthongs

The most striking initial difference between these formant plots is the fact that the KM

formant plot is generally higher in pitch, most especially of the F2 (x-axis). This may be

attributed to the fact that only very few male speakers of KM survive, therefore our KM

sample was largely drawn from female speakers whereas the VYK sample was more evenly

divided along gender lines. One point of note, though, is the difference in the relative spacing

of the vowels corresponding to CSK /o/ and /wu/. Specifically, this distinction has been

attributed to the influence of contact with Chinese (Lee 2005). While the charts presented

above support the theory that these vowels are realised differently in VYK and KM, further

investigation into the realisation of the corresponding vowels in Chinese is required on the

part of *Cosencok* consultants before any concrete conclusions can be drawn.

While not conclusive on this particular matter, these charts provide a good overview of the

position of the body of the tonguein the oral cavity for the monophthongs of the transplanted

varieties of Korean, which we now examine in more detail.

5.1.3.1   /i/
The phoneme /i/ in each transplanted variety is realised at a similar height (i.e. the F1 values

are similar) and is the front-most vowel in both KM and VYK. The large difference in F2 values

may be attributed to the unequal gender distribution in the sampled groups. These phonemes

also have a similar distribution:

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
KM606[irəpso]	it is OK	KM2057[t͡ʃip̚]	house	KM299[iŋgi]	here
KM894[irɨmi]	name	KM1358[ʃip̚]	ten	KM62[uɾi]	we

Table 42: Realisations of KM /i/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
VYK2876[irɨm]	name	VYK31	mixed rice	VYK1915[uɾi]	we

156

[pibimbap̚]

VYK2625[irepk'e]   easily   VYK2935[toŋil]   unity   VYK1690[kəgi]   there

Table 43: Realisations of VYK /i/

In addition to the palatalization of preceding dental and apical consonants (e.g. /s/ to [ʃ] as in

example KM 1358 above or /th/ to [t͡ʃʰ], it may cause regressive height assimilation (umlaut) in

preceding syllables and the deletion of the nasals /n/ and /ng/, although both of these

processes are variable. For examples of nasalisation, see section 5.1.1.4. Below we present

examples of umlaut.

Reference	KM/VYK	CSK	English
KM273	[hanɛ]	*hana-ga*	one (of them)
KM556	[kɨrẽjk'a]	*kulenikka*	therefore
KM423	[ijɛgi]	*iyagi*	conversation
KM1030	[kɨred͡ʒi]	*kuleci*	exactly right!
VYK2746	[hanɛ]	*hana-ga*	one (of them)
VYK1833	[ijɛgi]	*iyagi*	conversation
VYK95	[səmjeŋi]	*selmyeng*	explanation
VYK652	[kɨred͡ʒi anɨn]	*kuleci anhun*	not like that

Table 44: Umlaut Cause by /y/ in KM and VYK

5.1.3.2   /ey/ and /ay/

We deal with these similar (possibly merged in CSK) phonemes together. These phonemes are

both produced relatively high and towards the front of the mouth. Again, the relatively large

difference in degree of F2 values may be attributed to the gender imbalance in the KM sample

group. These phonemes remain distinct in both KM and VYK as they are in official description

of CSK, although they have merged in many other spoken varieties (Shin 2015 p. 5). Words

featuring the phonemes corresponding to CSK /ey/ or /ay/ in initial position do not appear in

the primary data, although the sound [ɛ] does very rarely appear in word initial position as a

result of the phonemes corresponding to CSK /a/ undergoing umlaut.

157

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
N/A	N/A	KM1041[kaget'a]	(I/you/(s)he) will go	KM431[ənd͡ʒe]	when?
N/A	N/A	KM319[sesa]	world	KM2211[kit͡ʃʰage]	enormously

Table 45: Realisations of KM /ey/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
N/A	N/A	KM1309[t͡ʃɛmi]	enjoyment	KM137[kundɛ]	army
N/A	N/A	KM660[hɛt͡ʃ'i]	(I/you/(s)he) did	KM917[t'ɛ]	time

Table 46Realisations of KM /ay/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
N/A	N/A	VYK1227[sedɛ]	generation	VYK349[hamk'e]	together
N/A	N/A	VYK2093[hages'ɨmda]	(I/you/(s)he) will do	VYK1709[nɨk'e]	late

Table 47: Realisations of VYK /ey/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
N/A	N/A	VYK1557[sɛŋil]	birthday	VYK813[jəndɛ]	Yanbian University
N/A	N/A	VYK2337[sontʰɛk̚]	choice	VYK51[t'ɛ]	time

Table 48: Realisations of VYK /ay/

Both the pairs of phonemes /ey_VYK/ and /ey_KM/ along with /ay_VYK/ and /ay_KM/ may be

considered nearly equivalent on the basis of their phonetic and distributional similarities.

5.1.3.3   /a/, /e/, /o/ and /wu/

The VYK and KM phonemes which correspond to these CSK phonemes may all be considered

near equivalents. Their realisations, low front unrounded vowel, mid-central unrounded vowel,

mid-back rounded vowel and high back rounded vowel are similar in both transplanted

varieties, despite the sample-induced difference of roughly two hundred Hertz in their F2s as

are their distributions. Examples of this may be seen below.

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
KM2457[adɨri]	son	KM570[sargi]	life/living	KM830[hobund͡ʒa]	alone
KM1877[at͡ʃʰɨm]	morning	KM1224[t'ar]	daughter	KM178[na]	I

Table 49: Realisations of KM /a/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
KM82[əpso]	It is absent	KM2024[t͡ʃən]	before	KM504[kjogwasə]	textbook
KM765[əmə̃i]	mother	KM1462 [kəŋgɛsə]	there	KM1246[putʰə]	from

Table 50: Realisations of KM /e/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
KM2142[onɨr]	today	KM9434[kosa]	worries	KM532[t͡ʃabulo]	by oneself
KM419[ovəl]	May	KM1257[kobɨn]	beautiful	KM46[t'o]	Also

Table 51: Realisations of KM /o/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
KM638[uri]	we	KM1256[kunir]	military work	KM2474[nəmu]	too much
KM1679[us'o]	(I/you/(s)he) laughed	KM309[turi]	two	KM1109[ad͡ʒu]	Very

Table 52: Realisations of KM /wu/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
VYK354[adɨl]	son	VYK[pʰalʃip̚]	eighty	VYK1920[t͡ʃtint͡ʃ'a]	really

VYK588[azu]	very	VYK1974[mal] language	VYK48[ta] all

Table 53: Realisations of VYK /a/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
VYK2103 [əpsɨmda]	It is absent	VYK1612[pən]	time	VYK1466 [t͡ʃosənə]	Korean language
VYK198 [ərjəsʼɨlt'ɛ]	when I was young	VYK1565 [t͡ʃosənd͡ʒok]	Chinese-Koreans	VYK1017 [jərə]	several

Table 54: Realisations of VYK /e/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
VYK2935[oʃi]	clothes	VYK45[hond͡ʒa]	alone	VYK876[pumo]	parents
VYK1996[orɛ]	a long time	VYK186[sohakʼjo]	primary school	VYK513[t'o]	Also

Table 55: Realisations of VYK /o/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
VYK634[undoŋ]	exercise	VYK1814[pukˀ]	drum	VYK2031[t͡ʃʰiŋgu]	friend
VYK2589[uri]	we	VYK1698 [t͡ʃuŋgukˀ]	China	VYK1501[modu]	everything

Table 56: Realisations of VYK /wu/

Furthermore, where these vowels occur next to each other, 'glide insertion' may occur. In other words, the vowels corresponding to CKs /o/ and /wu/ may be either replaced by or followed by an intrusive labio-velar glide when they appear before another vowel. This process is also ubiquitous in CSK. In keeping with the different realisations of /w$_{KM}$ / and /w$_{VYK}$/ as [w] and [v], respectively, where glide insertion occurs this difference in realisation is maintained.

KM2280   [o-n-da]

come-PRS-DECL.NEUTR

/o/ + /n/ + /ta/

I come

KM86   [v-as'-o]

come-PST.DECL-POL

/o/ + /as'/ + /o/

I came

VYK1092   [o-n-da]

Come-PRS-DECL.NEUTR

/o/ + /n/ + /ta/

I come

VYK860   [w-as'-o]

come-PST.DECL-POL

/o/ + /as'/ + /o/

He came

VYK507   [pɛu-da]

learn-DECL.NEUTR

/pɛu/ + /ta/

Learning

VYK518   [pɛu-was'-ɨmda]

learn-PST-DECL.POL

/pɛu/ + /as'/ + /sɨmta/

I learned

KM836   [pɛu-d͡ʒi]

learn-PRS.DECL.NEUTR

/pɛu/ + /t͡ʃi/

I learn

KM348   [pɛ-vas'-ɨmda]

learn-PST-DECL.POL

/pɛu/ + /as'/ + /sɨmta/

I learned

Glide-insertion may also occur word internally, again with the labio-velar glide being realised as [w] in VYK and [v] in KM:

KM2280   [pʰi-va-do]

smoke-PRS-DM

/phiwu/ + /e/ + /to/

Even if one smokes

VYK713   [pɛu-wa-ja]

learn-PRS-DM

/paywu/ + /we/ + /ya/

One has to learn

5.1.3.4   /u/

/u$_{KM}$/ is pronounced considerably lower in the vowel space than /u$_{VYK}$/. This makes it rather more similar in quality to the KM phoneme which corresponds to CSK /e/ and we may tentatively suggest that these phonemes are undergoing a merger. Indeed, contemporary, native-speaker Cyrillic alphabet attestation of KM regularly transcribes sounds appearing in positions corresponding to CSK /u/ with the symbol used to transcribe CSK /e/ (cf. Pak 2005 p. 85).

Nevertheless, there are also instances in which the realisations of the corresponding phonemes in the transplanted varieties are similar to that of CSK and each other in the primary data. We present examples of this below in order to demonstrate the similar distribution of these phonemes.

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
VYK2458[ɨmakˀ]	music	VYK977[t͡ʃigɨm]	now	VYK1697[kɨ]	that
VYK321[ɨmʃik]	food	VYK2011[kɨnjaŋ]	only	VYK165[apʰatʰɨ]	apartment

Table 57: Realisations of VYK /u/

Initial Position	Translation	Medial Position	Translation	Final Position	Translation
KM2275[ɨmakˀ]	music	KM267[sɨdipˀ]	marriage	KM86[kɨ]	that
KM		KM1083[kʰɨn]	big	KM727[ənɨ]	which?

Table 58: Realisations of KM /u/

A phonological process which may be observed in both varieties is the rounding of the phonemes /u$_{KM}$/ and /u$_{VYK}$/ to /wu$_{KM}$/ and /wu$_{VYK}$/ when they are in the environment of labial consonants such as those corresponding to CSK /p/ or /m/. For example:

Reference	VYK	CSK	English
VYK353	[t͡ʃibu]	*cip ul*	the house (OBJ)
VYK1637	[namu]	*nam ul*	others (OBJ)

Table 59: Rounded Realisation of /u/ in VYK

Reference	KM	CSK	English
KM695	[pabu]	*pap ul*	rice (OBJ)
KM1239	[t͡ʃibu]	*cip ul*	the house (OBJ)

Table 60: Rounded Realisation of /u/ in KM

Despite their similar distribution, given the difference in the realisation of the VYK and KM phonemes corresponding to CSK /u/, they are considered only partial equivalents.

5.1.3.5 /oy/ and /wi/

Traditionally, and in, for example, the ROK National Language Association's description of CSK, these phonemes are realised as the high, front rounded vowels [ø] and [y]. In recent years, though, they have undergone a dipthongization not yet officially recognised and are largely pronounced as [wɛ] and [wi]. Jin (2008 p. 129) found that the dipthongisation of /oyᵧₖ/ is almost entirely complete and its realisation as [wɛ] is near universal in the YKAP, while her subsequent research indicates that a high, front rounded pronunciation of /wiᵧₖ/ may still be encountered (Jin 2012 p. 88). This is borne out in the primary data. Very few examples of the monopthongal pronunciation of /wiᵧₖ/ are attested along with the exclusive dipthongal pronunciation of /oyᵥᵧₖ/. Examples of are presented below.

Reference	VYK	CSK	English
VYK2465	[wɛgukˀ]	*oykwuk*	foreign
VYK304	[tɛge]	*toygey*	extremely
VYK943	[wi hɛsə]	*wi hayse*	for
VYK512	[t͡ʃʰimi]	*chwimi*	hobby
VYK2517	[ʃygo]	*swiko*	(I/you/(s)he) rest(s)

and…

Table 61: Realisations of /oy/ and /wi/ in VYK

The examples above also demonstrate that the first segments of these diphthongs may be deleted where they follow obstruents.

Moving on to KM, both /oy$_{KM}$/ and /wi$_{KM}$/ have been reported as having fully undergone dipthongization (Pak 2005 p. 69). This is born out in the primary data and no high, front rounded vowels were recorded. A slight difference in their realisations as 'dipthongs', though, is that /oy$_{KM}$/ is generally pronounced as [vɛ] whereas the initial glide of /wi$_{KM}$/ is still realised as such, rather than as a labio-velar fricative.

Reference	KM	CSK	English
KM483	[vɛgukʼ]	*oykwuk*	abroad
KM1845	[tɛs'o]	*toysseyo*	(I/you/(s)he) became
KM742	[kwiha]	*kwiha*	sir

Table 62: Realisations of /oy/ and /wi/ in KM

Furthermore, as in VYK, the initial segments of the sounds derived from earlier high, front rounded vowels may undergo glide deletion, even in the case of /oy$_{KM}$/, the initial segment of which is not, phonetically speaking, a glide.

While these phonemes share a similar distribution in VYK and KM, their realisations differ, particularly that of the phonemes corresponding to CSK /oy/. Thus, the corresponding transplanted phonemes are only partially equivalent.

5.1.4 Summary

A chart summarising the equivalences between the VYK and KM phonemes established in the foregoing discussion is found below:

Vowel Phoneme	Near Equivalent	Partial Equivalent	Non-Equivalent

/i/	✔	
/ey/	✔	
/ay/	✔	
/a/	✔	
/o/	✔	
/e/	✔	
/wu/	✔	
/u/		✔
/oy/		✔
/wi/		✔

Table 63: Summary of Contrasts Between VYK and KM Vowels

5.1.5 Prosody

Since the primary data gathered for this project is not ideal for making inferences about word level prosody, we rely on the descriptions which appear in secondary data to inform our understanding of the prosody of the transplanted varieties of Korean. Furthermore, we cannot rely on CSK for comparative concepts, since contrastive tone rather than contrastive length is thought to operate in VYK and KM, unlike CSK. It manifests in the form of a pitch accent system; that is, each lexical item contains a syllable which is the locus of a pitch peak (Ito 2008)[14]. This tonal means of distinguishing words is distinct from tone as it is realised Chinese and Thai where each syllable may bear a particular contrastive tone.

---

[14] Although it must be noted that appending a conjugational or inflectional ending to a lexical item, for example a particle or verb ending, may lead to the formation of words with additional tonal peaks (Ramsey 1978; Ito 2008; Jeong 2010)

Turning first to VK, only two tonemes are identified in this variety: low tone and high tone. Choi (2002 p. 21) has observed that while this two tone distinction is similar to that of the North eastern and South eastern dialect zones of the Korean peninsula, the same tonal pattern does not necessarily mark the same lexical item. For example, the CSK word *mal* has the meaning 'horse' when realised with a short vowel and 'language' when realised with a long vowel. In the North eastern dialect zone, the tones become evident when the word is combined with a subject particle, so [márì] denotes 'language' and [màrí] denotes 'horse'. In VYK, however, 'language' is denoted by [màrí], whereas 'horse' is denoted by [márì].

Most analyses of *Koryo Mar* propose that it possesses contrastive tone, similar to the pitch accent of Yanbian Korean and the varieties of the North East of the Korean peninsula. That is, it consists of two tonemes which distinguish high and low tones (King 1987 p. 243; King and Yeon pp. 102 – 103). The individual uses of pitch accent in contrastive function provided by King and Yeon, in contrast to those provided by Choi (2002 p. 21) are identical to those of the North eastern dialect zone, consequently in both *Koryo Mar* and that peninsula variety 'language' is pronounced [márì] in order to distinguish it from 'horse', pronounced [màrí].

Pak's analysis, though, suggests that there is not only contrastive tone in *Koryo Mar*, but also contrastive stress (as in the main contact language Russian), and contrastive vowel length (as in CSK) (2005 pp. 74 – 77). The examples of contrastive stress presented here, though, pattern with the pitch accent so that stressed syllables invariably correlate the high tone. For this reason, we concur with King's conclusion (1992 p. 203) that this high tone is largely perceived as stress by the universally bilingual *Koryo Saram* as a result interference from their (metalinguistic) knowledge of Russian. The case for the development of some contrastive vowel length in KM is a little stronger, though, with examples of a long vowel, albeit without contrastive function, being recorded by King (1987 p. 246), and contrastive length distinguishing *mudenhada* and *mudeenhada* meaning 'generous' and 'exhausted', respectively

(Pak 2005 p. 75). Nevertheless, scholarly consensus suggests that a full system of contrastive vowel length has not developed in KM.

This discussion of prosody in VYK and KM has served to demonstrate that despite the presence of contrastive tone in both transplanted varieties its realisation is sufficiently different that their prosodic systems can be considered only partially equivalent.

5.1.6 Phonotactics

The phonemes examined and described, of course, realised in together with each other rather than in isolation in natural speech. Therefore, we now briefly examine their combination and a selection of associated phonological processes.

As with CSK, the maximal syllable of VYK comprises a consonant followed by a glide, a vowel and a single consonant (CGVC). In writing it may often appear that syllables may end in clusters of up to two consonants, but this is an artefact of Hangul orthography. The minimal VYK syllable is composed of a single vowel. The maximal theoretical syllable of KM is slightly more complex than that of VYK or CSK in that it may be composed of an initial consonant followed by a glide, a vowel and up to two consonants (CGVCC). While this is not attested in the primary data, there are examples of the key distinguishing feature, final consonant clusters in the KM syllable, to be found. These become especially numerous when nonce-borrowings from Russian are taken into account. Its minimal syllable, too, is composed of single vowel. Examples of the VYK maximal syllable and KM syllables with final consonant clusters may be found below:

Reference	VYK	CSK	English
VYK955	[hwakʃin]	*hwaksin*	conviction/certainty
VYK95	[səlmjeŋi]	*selmyeng*	explanation

168

VYK526	[kwind͡ʒaŋi]	*koyngcanghi*	Incredibly

Table 64: Consonant Clusters in VYK

Reference	KM	CSK	English
KM2329	[hɨlkmun]	*hulkmun* [hɨkmun]	earthen door
KM224	[irkdaga]	*ilkdaka* [ikt'aga]	(I/you/(s)he) read(s) and then

Table 65: Consonant Clusters in KM

While a contact origin for the more complex syllable structure of KM may be suggested, it is worth noting that such syllables, or even more complex ones, were admissible in LMK. We are thus not able to rule out that such syllable structure is retained from an earlier stage of Korean rather than arising from a contact situation. Nevertheless, it is plausible that prolonged contact with Russian was influential in the retention of such a feature (if, indeed, it is a retention) and examination of this phenomenon from the point of view of contact linguistics would contribute greatly to a fuller understanding of it.

Numerous phonological processes occur at syllable boundaries in CSK. These are also characteristic of both VYK and KM. They include so-called neutralisation and tensification as well as various kinds of assimilation (e.g. liquid nasal assimilation, nasal assimilation and velar assimilation).

In our examination of the stop, fricative and affricate consonants above, in word final position these have an unreleased pronunciation and are said to have undergone 'neutralisation'. In the transplanted varieties, as in CSK, this process also takes place word internally in syllable final position, i.e. not in intervocalic position (for the general rule see Ahn 2009 p. 31). In these positions the neutralised consonants may also have the effect of tensification on non-sonorant consonants in the initial position of the following syllable (Ahn 2009 p. 35). The examples below demonstrate a particularly common pattern in both transplanted varieties in the

formation of the past and future tenses, namely the neutralisation of a tensed affricate to an

unreleased [t̚] and the tensification of the following consonant:

Reference	KM	CSK	English
KM1221	[kaget'a]	*kakeyssta*	(I/you/(s)he) will go
KM2331	[hɛt͡ʃ'im]	*haysscim*	(I/you/(s)he) did

Table 66: Consonant Neutralisation in KM

Reference	VYK	CSK	English
VYK2041	[kat'a]	*kassta*	(I/you/(s)he) went
VYK780	[wa t͡ʃ'i]	*wassci*	(I/you/(s)he) came

Table 67: Consonant Neutralisation in VYK

Turning to the various kinds of assimilation, we begin with liquid nasal assimilation. This is a

phonological process whereby a /n/ takes on the pronunciation of a neighbouring lateral

approximant (cf. Lee and Ramsey 2000 p. 70). It may be observed in both transplanted

varieties, as below:

Reference	VYK	CSK	English
VYK2810	[kolan]	*konlan*	tough/difficult
VYK1346	[wəlɛ]	*wenlay*	originally
VYK19	[jəlak̚]	*yenlak*	contact

Table 68: Liquid Nasal Assimilation in VYK

Reference	KM	CSK	English
KM2327	[kaljən]	*kanlyen*	accomplice
KM911	[t͡ʃiljən]	*chil nyen*	seven years

Table 69: Liquid Nasal Assimilation in KM

Two further phonological processes which operate in both transplanted varieties are velar and

nasal assimilation. These are assimilations cause lax obstruents (including neutralised

obstruents) to take on a velar pronunciation where they are followed by /k/, /kh/ and /kk/ (cf.

Choi 2002 p. 26) or nasal pronunciations when they followed by /m/, /n/ and /ng/ across a

morpheme boundary (cf. Ahn 209 p. 85).

Reference	KM	CSK	English
KM370	[t͡ʃɛmi ik'e]	*caymi isskey*	enjoyably
KM1799	[ək'o]	*epsko*	It is absent and…

Table 70: Velar Assimilation in KM

Reference	VYK	CSK	English
VYK1800	[ək'o]	*epsko*	It is absent and…
VYK920	[tɨk'o]	*tutko*	(I/you/(s)he) hear(s) and…

Table 71: Velar Assimilation in VYK

Reference	KM	CSK	English
KM417	[t͡ʃaŋnjən]	*caknyen*	last year
KM342	[imnida]	*ipnita*	(I/you/(s)he) am/are/is
KM1441	[panɨnde]	*patnundey*	(I/you/(s)he) receive(s)

Table 72: Nasal Assimilation in KM

Reference	VYK	CSK	English
VYK1538	[maŋnɛ]	*maknay*	the youngest child
VYK141	[mas'ɨmnida]	*macsupnita*	correct
VYK1051	[panɨnde]	*patnundey*	(I/you/(s)he) receive(s)

Table 73: Nasal Assimilation in VYK

This brief review of the phonotactics of KM and VYK has demonstrated that, despite very

slightly different syllable structure, the rules pertaining to the combination of phonemes and

the phonological processes which they undergo when combined are nearly equivalent in both transplanted varieties.

In the next section, we examine the similarities in form and function of the inflectional and conjugational systems of VYK and KM.

5.2 Morphology

In this section we first present comparative inventories of inflectional and conjugational Korean morphemes (respectively, *cosa* – particles and *emi* – word endings) before going on to discuss the contrasts of form and distribution of each as appropriate. For the classification of inflectional or nominal morphology we follow the Korean grammatical tradition (cf. Ko and Nam 1975 pp. 93-107; Lee and Ramsey 2000 p. 139; Lee and Lee 2010 pp. 19-24; Jeong 2010 p. 23) and Yeon and Brown (2011 p. 94) in delineating particles into 'case particles' (*kyekcosa*) and 'special particles' (*thukswucosa* or *pocosa*)[15], both of which may be found in the transplanted varieties of Korean spoken in China and Central Asia. Where the function of particles is discussed, it must be acknowledged that this is done on the basis of severely restricted data. This is due to the limitations inherent in the structure of PhD courses in general and with regard to the data collection for this thesis specifically. Consequently, all evaluations of specifically functional equivalance must be taken as provisional and await full exploration and verification following more extended data collection.

We now go on to examine both case particles and delimiters below, in turn.

---

[15] In the interests of avoiding the term "special particle" which is particular to the traditional Korean grammatical description, we follow Yeon (2003 p. 25) in referring to these particles using the more cross-linguistically valid term "delimiters" since they "delimit the semantico-pragmatic content of the co-occurring constituent without affecting its grammatical relation to other constituents".

5.2.1 Case Particles

Although according to some analyses it is possible to further divide the particles considered here into case particles, delimiters and post-positions on the grounds of their syntactic behaviour in CSK[16], the cross-varietal comparative concepts used in this study make only a bipartite distinction between case particles and delimiters. Following this model we identify the case particles as follows: subject particles, object particles, the genitive particle, particles of movement and location, instrumental particles and comitative particles. We list the CSK forms of these particles below along with a full inventory of their attested allomorphs drawn from the primary data, which forms the basis of this description. We then go on to briefly examine CSK allomorphy in general. More detailed analyses of the behaviour of each particle and discussion of the corresponding particles in VYK and KM follows in dedicated sub-sections.

Particle	CSK Forms	Yanbian Korean Forms	*Koryo Mar* Forms
Subject	*-i, -ka*,	[i], [ga], [iga]	[i], [j]
	-kkeyse	[k'esə]	N/A
Object	*-ul, -lul*	[ɨl], [rɨl], [ɨ], [rɨ], [u], [l]	[ɨl], [ɨ], [rɨ], [rɨ], [r], [ɨr], [u], [ur]. [ul]
Genitive	*-uy*	[e], [i]	[e], [i]
Movement and Location	*-ey*	[e]	[e] i
	-eyta	[eda] /-esə/, /-etaka/, e	[eda] *eytaka*

---

[16] This view, put forward by Yeon (2003 pp. 21 – 25), distinguishes case particles from post-positions due to the fact that case particles may be dropped, may appear outside post-positions where NPs bear two particles, may appear outside the delimiter –*man* (only), may appear outside of a copied plural particle –*tul*, and can appear only once at the end of a conjoined NP rather than at the end of each conjunct. Other systems for classifying particles yielding into five (Chang 1996) and three sub-classes (Kim-Renaud 2009) have been devised, although the basis for such groupings is not clear.

	-eyse	[esə], [sə]	[esə], [ezə], [sə], [zə] isa
	-eykey	[ege] ndey, indey, keyda, bogo	[ege] *indey, key, keyse, eykey*
	-hanthey	[hantʰe]	N/A
Instrumental	-lo, -ulo	[ro], [ɨro], [lo], [ɨlo] ullu, llu, lu (Jeong 44)	[ɨro], [uro], [lo], [ro], [ɨlo] *ulu, ullu, ullo, ullwu, ulla*
Comitative	-wa, -kwa,	[wa], [kwa], [ga] kka, (Jeong (46-47))	[ga]
	-lang, -ilang	[rang], [iraŋ] iramey, ramey	[rang], [irang]
	-hako	[hago]	[hago] *ko, iko*

Table 74: Attested Forms of VYK and KM Case Particles

Much of the allomorphy exhibited in CSK is phonologically conditioned. A stem final consonant or vowel may lead to either wholly distinct allomorphs being appended to NPs (as in the case of the subject particle, for example) or a vowel being inserted between the final consonant of the NP and the particle (for example, the instrumental particles –*lo* and –*ulo*). Some particles also exhibit allomorphy which reflects some degree of honorification (for example, the honorific subject particle –*kkeyse* or the comitative particle –*(i)rang*, which tends to be used in less formal situations than its alternatives). Before commencing our discussion of the individual particles we further observe that the phonological processes examined in the preceding section may lead to a great deal of variation in the realisation of VYK and KM particles, depending on the surrounding phonological environment. Contrasts are noted following our discussion of each particle. The examination of delimiters follows our discussion of the features of the individual case particles.

5.2.1.1 The Subject Particle

As can be seen from the inventory of case particles presented above, the form and distribution

of the subject particle are crucial distinguishing features of Yanbian Korean and *Koryo Mar*.

Canonically, the allomorph /-ka/ is absent from the dialects of the Northeastern dialect zone of

the Korean peninsula. It is certainly the case that this form of the particle is absent from *Koryo*

Mar, notably even from the speech of most consultants who had received some education in

CSK. Rather, where the subject particle appeared it was only in the form /-i/, although in the

speech of numerous speakers it is possible to drop it[17] entirely, or replace it with a delimiter

(for example the topic particle). Given the many productive phonological processes associated

with high front vowels in Central Asian Korean (for example, umlaut, nasal deletion, and glide

insertion), the subject particle's lack of allomorphy in this variety has given rise to variation

between the form of nouns where they are combined with the subject particle and their forms

elsewhere:

KM854 [ʃid͡ʒaŋ t͡ʃənd͡ʒa sond͡ʒɛj-dɨr sonje

now in total grandson.SUBJ-PL granddaughter.SUBJ

jəsɨʃ   i-o]

six   COP-DECL.PRS.POL

I have six grandchildren in total now

KM809 [nɛ arɛ ilt͡ʃəngubɛksaʃipo-jən-do-e na-ɨn

PRN.1.PL.GEN beneath one thousand nine hundred and forty five-year-degree-LOC be born-ADN

---

[17] In CSK the subject particle is also frequently dropped. Its appearance is thought to convey a 'focus' meaning (Yeon 2003; Kim-Renaud 2009).

	My	younger	sibling,	who	was	born	after	me	in	1945,

doŋsɛ̃		sibaŋ		vondo-esə	sa-o]
younger sibling.SBJ		now		far East-LOC	live-DECL.PRS.POL

Is	now	living	in	the	(Russian)	Far	East

In the above examples we can see that where subject marking has led to vowel raising (cf. CSK [sond͡ʒa] and [sonjə] with KM [sond͡ʒɛj] and [sonje]) and nasal deletion in the case of [doŋsɛ̃].

The examples below, though, make it clear that there are those speakers of *Koryo Mar* for whom the subject particle has remained just that, a particle which may be appended to nouns and dropped where the grammatical relations between constituents are clear. In both examples the subject particle has been dropped and neither sound change associated with high front vowels has occurred.

KM1940	[toŋsɛŋ		irəkʰe		and͡ʒ-as'-o]
	younger sibling		thus		sit-PST.DECL-POL

	My	younger	sibling	just	sat	down	like	this.

KM247	[uri		nampʰjən		doɾaga-t͡ʃ'i]
	PRN.1.PL		husband		die.HON-PST.DECL.NEUTR

	My	husband	passed	away.

It is also the case, though, that the appending of this invariant form of the subject particle has been reanalysed by many speakers of *Koryo Mar* as the citation form of the nouns. This appears to be the case in the above example KM854, in which the plural particle [dɨr] is appended to the subject marked noun [sond͡ʒɛj]. Typically in CSK and other varieties of Korean

case particles appear 'outside' of delimiters including the plural particle where they appear together. Indeed, Yeon considers this a defining characteristic of case particles (Yeon 2003 p. 23). There are no examples of other case marking particles appearing 'inside' of plural particles and, furthermore, the appearance of the subject particle inside the plural particle varies both between and within consultants' data. Consequently, we regard it only as a limited case of reanalysis of certain lexemes on the part of certain speakers. Further examples of this reanalysis are presented below:

KM254   [adur-i-nɨ   lʲeningrad-sə   isa-dejhak   pʰirha-go

           son-SBJ-TOP   Leningrad-LOC   medicine-university   complete-CNJ

           t'ar-i-nu   sənsej-dejhak   pirhe-so]

           Daughter-SUBJ-TOP   teacher-university   complete-PRES.DECL

           My son graduated Leningrad medical school and my daughter teacher's college.

KM441   [irəkʰe   tɨrɨva-s'ɨmdu   ʌʃ-i-dur-i]

           like this   enter-INT.PST   parents-SBJ-PL-SBJ

           Is   that   how   your   parents   came   here   (the Russian Far East)?

The above example most clearly demonstrates how the subject particle has been reanalysed in that in the word [ʌʃiduri] we see the subject particle [i] appended to the root [ʌʃ] and also following the plural particle. This reanalysis is not attested as having taken place in the VYK data.

Although it is canonically invariant and thought to appear only as /-i/, the /-ka/ form of the subject particle did appear in the speech of one *Koryo Mar* consultant. This consultant also uniquely used many CSK verb endings. We consider their speech to be the result of a

particularly strong observer effect caused by the interview situation and intense exposure to CSK through formal study.

We now move on to VYK. While some researchers hold that the /-ka/ form of the subject particle is also absent from Yanbian Korean (Jeong 2010 pp. 26 – 31), our data demonstrated that in fact the subject particle in this variety exhibits allomorphy consistent with CSK, that is the phonologically conditioned complementary distribution of the allomorphs /-i/ and /-ka/, with the latter following nouns with a final vowel and the former following nouns ending with a consonant. This assessment holds even for older, more conservative consultants. In the below examples, subject marking is consistent with that of CSK.

VYK2735   [uri-ga   irə-kʰe   ton   t͡ʃogɨm-ʃ'ire   mo-a-gad͡ʒi-go]

           PRN.1.PL-SBJ   thus-ADV   money   a little-each   gather-CNJ-have-DECL.DM

           That's   how   we   saved   up   our   money   little   by   little

VYK116   [t͡ʃe-ga   pu k̚   t͡ʃint͡ʃ'a   t͡ʃal   t͡ʃi-mnida]

          PRN.1.SG-SBJ   *puk*   really   well   play-DECL.PRS. POL

          I   play   *puk*   really   well

VYK690   [t͡ʃosənəmun-ey   huŋt͡ʃʰi-ga   t͡ʃom   də   mana-sɨmnida]

          Korean language and   interest-SBJ   a little   more   more-DECL.PRS. POL
literature-DAT

          I   am   a   little   more   interested   in   studying   Korean   language   and   literature

In addition to these, the form /-ika/ may very occasionally be observed in Yanbian speech (cf. Kim 2013). This does not systematically alternate with the other forms, but seems to represent a reanalysis of the subject particle /-i/ as part of the noun or particle. In other words,

speakers believe that they are producing a word ending with the vowel /i/ and mark it as a subject using the particle /-ka/ as they would for any other noun ending with a vowel.

VYK1505　[kɨredo　tebubun　saram-dɨr-iga　　pe　　pul-i

　　　　　however　majority　person-PL-SUBJ　　stomach　full-ADV

　　　　　məg-ɨl　　　　　su　　　　　　　　　　　ik'-o]

　　　　　eat-ADN　　　　　case　　　　　　　　　　be present-DECL.PRS

　　　　　However　most　people　can　eat　as　much　as　they　need

VYK2178　[se　　　　　　　　ʃid͡ʒaŋ-iga-do　　　　　　ikˈo]

　　　　　new　　　　　　　　market-SBJ-also　　　　　be present-DECL.PRS

　　　　　There　　　　　is　　　　a　　　new　　　market,　too

A final contrast which may be observed between the forms of the subject particle of *Koryo Mar* and Yanbian Korean is their role in subject honorification. While speakers of Yanbian Korean had fluent command of this aspect of honorific language, no honorific subject particle was attested in the primary data of *Koryo Mar*. This mirrors earlier findings about the absence of the honorific subject particle in Central Asian Korean (Pak 2005 p. 197; Kwon 2010 p. 24), and its presence in Chinese Korean (Jeong 2010 p. 31).

VYK90　　[kɨ　　　　　harabəd͡ʒi　　harabəd͡ʒi –keyse　　　　　ama

　　　　　that　　　　grandfather　grandfather-SBJ.HON　　　　perhaps

　　　　　isa　　　　　　　　　　　wa-sˈɨmnikane]

　　　　　move　　　　　　　　　　come-PST.CNJ

　　　　　My　grandfather's　grandfather　probably　moved (here to China from Korea)

The contemporary regional LWCs both exhibit less elaborate means of codifying honorification or politeness grammatically than comparison to Korean. Russian maintains a bipartite T-V distinction of the in its pronominal and verbal systems, in common with many European languages, and, following the establishment of the PRC in 1949 "the elaborate historical Chinese system of honorifics and the denigration and elevation terms of address collapsed" (Pan and Kadar 2011 p. 1533). Thus, we consider it unlikely that contact with regional LWCs led to this divergence in the honorific system of the transplanted varieties. Rather, it is more probable that this feature is found in VYK due to the higher degree of contact with other varieties of Korean, especially standardised varieties, in which this honorific form of the subject particle is used.

The subject particles of *Koryo Mar* and Yanbian Korean, then, may be considered partially equivalent on the grounds that their function and global distribution is similar. We note at this point that partial equivalence as a category is much broader than either near equivalence or non-equivalence. Therefore, while the considerable difference in both the phonologically and pragmatically conditioned realisations of the subject particles of the transplanted varieties are not sufficient grounds for them to be classified as non-equivalent there are certainly other partially equivalent features of these varieties which exhibit a good deal more similarity.

5.2.1.2   The Object Particle

The object particle marks the direct object of transitive verbs, a function which is consistent between VYK, KM and CSK. It is the realisations of this particle, however, that are particularly variable in both of the transplanted varieties of Korean under examination here. CKS admits only allophony conditioned by the NP to which the object particle is attached. When the NP ends with a vowel it takes the form *-lul*, whereas it takes the form *-ul* when the NP ends with a consonant. In addition, where the appearance of *–lul* might be expected following a vowel in

very particular frequently used expressions, it may take the reduced form –*l* as in *ku kel* 'that thing or *nal* 'me' (cf. Yeon and Brown 2011 p. 98).

While forms starting with a consonant invariably follow nouns with final vowels (and vice-versa) the remaining variation between forms of the transplanted object particle does not seem to be so regularly conditioned. Rather, it appears to be due to (generally) irregular phonological reduction in both varieties and the highly variable realisation of the phoneme /l/ in *Koryo Mar*. This variability encompasses both the qualities of the realisations of phonemes which make up the object particle as well as the presence or absence of the final segment.

The full extent of the variation for each variety is demonstrated by the examples given below, drawn from the KM and VYK primary data.

KM804   [adɨr-ɨl   uri   əʃ-i-nɨ   mãj   na-sə]

             son-OBJ   PRN.1.PL   Parents-SBJ-TOP   many   bear-DECL.PST

             My   parents   had   a   lot   of   sons

KM2007   [mun- ɨr   hur   tad-as'-o]

             door-OBJ   forcefully   close-DECL.PST-POL

             I   slammed   the   door

KM531   [kɨrɛ   korjə   saram   i-dejanaj   kudu-rɨ   no-kʰo]

             Therefore   Korean   person   COP-CNJ   under-floor   heating-ACC   put-DECL

             So   because   we   are   Korean   we   put   down   under-floor   heating

KM721   [na   kor ʲejsk-ij   nov-   god-ur   ha-m adɨr-ga   mar   ha-m]

                                         ɨj

| PRN.1.SG | Korean-SG.M | New-SG.M | year.M -OBJ | do-if | son-COM | talk | do-DECL.PRS.POL |

If   I   celebrate   Korean   new   year   I   talk   to   my   son.

KM550   [gɨr-ɨ   mo   irɣ-əs'ə]

writing-ACC   NEG   read-DECL.PST

I   could   not   read

KM327   [kɨrəndɛ   uri   abəd͡ʒi-rɨ   kɨrə-n

However   PRN.1.PL   Father-OBJ   thus-ADN

t͡ʃir-ha-rago   ponɛ-s'ə]

work-do-in order   send-DECL. PST

However, my father was sent to do that kind of work.

KM689   [pabo   da   sul-u   mək-t͡ʃ'i]

Idiot   all   liquor-OBJ   eat-PRES.DECL.POL

Idiots   all   drink   liquor

KM2192   [igi   d͡ʒib-u   pʰaɾ-ə   ka-d͡ʒi-o]

here   house-OBJ   sell-PURP   go-DECL.PRES-POL

I am going to sell the house that I have here

VYK2136 [uri mar-ɨl al-d͡ʒi mosʼ ha-go]

PRN.1.PL language-OBJ know-CNJ NEG do-DM

The do not know our language

VYK44 [kəgi-e ga-sə ʃiksa-rɨl hago]

there-LOC go-CNJ meal-OBJ do-CNJ

We go there and have a meal

VYK1051 [se bən jəŋhwa-rɨ pa-nɨnde]

three time film-OBJ see-DECL.DM

I have seen that film three times.

VYK2581 [uri paŋsoŋ-ɨ t͡ʃad͡ʒu po-go]

PRN.1.PL broadcast-OBJ often see-DM

They often watch our programmes

VYK1637 [na nam-u sərə mana-go]

PRN.1.SG others-OBJ each other meet-DECL.PRS

I meet up with other people.

VYK353 [t͡ʃib-u se doŋne-rɨl isa ha-ge tɛ-nɨnde]

house-OBJ new neighbourhood-OBJ move do-ADV become-CNJ

So we came to move house into a new neighbourhood.

Above we see that object particles conforming to CSK, that is taking the forms [ɨl] and [rɨl] after

a consonant and a noun, respectively, appear in both transplanted varieties, as do forms in

which the final consonant undergoes apocope to leave the object particle realised as [ɨ] or [rɨ].

We also see a demonstration of the variation in the realisation of liquid consonants, namely, their possible realisations as trills in KM. Finally, we observe that the quality of the vowel in the object particle is also variable, and may be realised as more rounded, that is [u] rather than [ɨ], particularly when it is in close proximity to labial consonants.

While many of the same forms appear in both varieties, it may be noted that the apocope of the final /l/ was a significantly more frequently occurring phenomenon in *Koryo Mar* than Yanbian Korean.

Furthermore, as in CSK, the object particle may be dropped entirely in both *Koryo Mar* and Yanbian Korean where it is contextually redundant.

KM305	[kɨ		da:m-e		noŋsa-d͡ʒir	hɛ-s'ə]
	That		after-LOC		farming-work	do-DECL. PST
	After	that	I	did	agricultural	work

VYK71	[undoŋ		t͡ʃint͡ʃ'a		mot	ha-mda]
	exercise		really		NEG	do-PRS.DECL.POL
	I	am	really	bad	at	sports.

In the above examples, an expected object particle was omitted following the direct object of the verbs, i.e. [noŋsad͡ʒir] 'agricultural work' and [undoŋ] 'exercise'. An alternative interpretation to simple contextual redundancy is that the above examples represent the cross-linguistically commonly observed phenomenon of Differential Object Marking (DOM cf. Bossong 1985). We note that in both of the above examples, the objects which are not marked with particles are inanimate, nonspecific nouns phrases and, thus, considered the least likely to be overtly case-marked as per Aissen's generalised prediction on DOM (Aissen 2003

184

pp. 436-437). Whether the apparent DOM observed in KM or VYK can be more fully explained with reference to focus structure, as has been claimed for CSK (Kwon and Zribi-Hertz 2008), or exhibits different conditioning of DOM would be a worthwhile topic for future research.

A final characteristic of the accusative particle is its use to mark the goal of verbs of movement. This occurs in both transplanted varieties (and in CSK), however it is a strikingly common feature of KM speech compared to VYK.

KM2448	[vit͡ʃʰa	t͡ʃib-ul-do	ka-ge-nɨnde]
	Vicha (personal name)	house-OBJ-too	go-MOD-DM

I should go to Vicha's house, too

KM 282	[i-wəl	tar-butʰə	inge-r	taŋgi-mjɛ]
	two-month	month-DAT	here-OBJ	frequent-CNJ

I have been coming here since February.

VYK243	[jəgi-rɨ	nəm-ə	wasʼ-o]
	here-OBJ	cross-CNJ	come.PST.DECL-POL

They crossed (over the border) and came here.

In addition to this we shall demonstrate in section 5.2.1.4 that the object particle has taken on the additional function of marking animate recipients in KM. Nevertheless, the combination of their broadly similarity in functions and the overlap in their hugely diverse phonetic realisations, leads us to conclude that the object particles of KM and VYK are partially equivalent.

5.2.1.3 The Possessive Particle

The CSK possessive particle connects two NPs in a possessor-possessed relationship with the

possessor preceding the possessed, as is typical for SOV languages. This also describes the

function and usage of the possessive particle in both VYK and KM. The canonical form of the

possessive particle, namely [ɨj], can only be found in remarkably careful speech[18], even in CSK

where it is largely pronounced as [e]. The same tendency may be found in both *Koryo Mar* and

Yanbian Korean, perhaps unsurprisingly due to the absence of the phoneme /uy/ in the

transplanted varieties' phonological systems.

VYK2111   [hanguk-e   kasu   aju]

            Korea-GEN   singer   IU

            The   Korean   singer   IU

VYK1587   [i   leŋmjən-iga   hanguk-e   leŋmjən-hago   tɨl-imnida]

            these   noodles-SBJ   Korea-GEN   Noodles-COM   different-DECL.PRS.POL

            These   noodles   are   different   from   Korean   noodles.

KM1866   [nam-e   t͡ʃip̚]

            other-GEN   house

            Someone else's house.

Kwon (2010 p. 25) also records that this form may be realised as [i] in *Koryo Mar*, which would

make it homophonous with the subject particle. This assertion is difficult to generalise to

---

[18] This is called elsewhere "a spelling pronunciation" and "evidence of overcareful speech" (Lee and Ramsey 2000 p. 65). Thus while it may appear in elicitation or reading tasks it is not clear the extent to which this canonical, orthographic form is actually represented in natural speech of any variety of Korean.

Central Asian Korean as a whole, since it is based on data gathered from only a single

consultant, in which a scant four instances of the genitive particle were recorded.

In the KM primary data collected for this project, only one plausible case of a possessive

particle being realised as [i] could be found, as below:

KM1942   [i   daɨm-e-nɨ   ad͡ʒubaj-ga   həŋnim-ɨnɨ   amwɾe

           this   after-LOC-TOP   uncle-COM   elder brother_TOP   somehow

           After   this,   somehow   my   uncle   and   my   elder   brother

           mak̚-toŋsɛ̃j-i   t͡ʃib-ɨro   ka-ya   tɛ-t͡ʃʼi]

           youngest-younger sibling-GEN   house-INST   go-CNJ   become-DECL.PST

           had   to   go   to   my   younger   sibling's   house.

We note, however, that since it is possible to drop the possessive particle an alternative

interpretation of this sentence is that the subject particle has been reanalysed as part of the

noun [dongsɛ̃i] and the possessive particle is simply not marked at all.

The pronunciation of the possessive particle as [i] may be found, although rarely, in Yanbian

Korean as in the following example:

VYK2869   [uri   nampʰjən-i   t͡ʃokʰa-ga   kəgi   isʼə-nɨnde]

           PRN.1.PL   husband-GEN   niece-SBJ   there   reside-DECL.DM

           My   husband's   niece   lived   there

Finally, the primary data reveals that this particle may also be dropped in both KM and VYK (as it may in CSK) with the relationship between possessor and possessed being made clear through context and word order, as in the following two examples:

KM292   [ne   t͡ʃagi   abəd͡ʒi   sɛ̃-i-nɨ   han]

           PRO.1.SG   PRN.REFL   father   surname-SUBJ-   Han
                                                           TOP

           My   father's   surname   was   han

VYK1827   [t͡ʃusək̚   t͡ʃəs̚   nar-e   halməni   t͡ʃib   ga- nɨnde]

           Chuseok   first   day-LOC   grandmother   house   go-DECL.DM

           On   the   first   day   of   Chuseok   we   go   to   my   grandmother's   house.

The genitive particles of KM and VYK may be considered near equivalents. They are functionally identical and have similar realisations, even to the extent of having similar allomorphs. Furthermore, there does not seem to be any factor conditioning the variable realisation of these particles as specific allomorphs, [i] or [e], in either transplanted variety.

5.2.1.4   Particles of Movement and Location
Elsewhere considered together as dative particles (e.g. Sohn 1999 p. 213), we follow Yeon and Brown (2011) in considering these particles in terms of particles of movement and location. Often, both movement and location may be expressed by the same form of a given particle with its meaning only discernable from context. Due to this, we deal with particles in turn on

the basis of the form of the CSK particle which forms the basis for their comparisons, namely -*ey*, *-eyda*, *-eyse*, *-eykey*, and *–hanthey.*[19]

We examine the corresponding particles of VYK and KM in turn in that order.

-ey

This single-form particle may denote either place or movement in CSK. It has been compared to the English prepositions 'to', 'in' and 'at' (Yeon and Brown 2011 p. 102). It has similar form and function to this in both VYK and KM. The examples below demonstrate its use as a particle of movement:

KM 1294	[kɨrənde	pəland-e	ka-tˈa-va-sˈo]				
	Anyway	Poland-DAT	go-PST.DCL-come.PST.DECL.POL				
	Anyway,	I	went	on	a	trip	to	Poland.

VYK425	[abəd͡ʒi-nɨ	kwaŋdoŋ-e	ka-s'o]	
	father-TOP	Guangdong-ALL	go-PST.DECL.POL	
	My	father	went	to	Guangdong.

We further note that it may be dropped where its meaning would be clear from context:

VYK2306	[t͡ʃʰegɨn-e-nɨn	halbin	ga-pwa-s'ɨm]				
	recently-LOC-TOP	Harbin	go-see-PST.DECL.POL				
	I	recently	tried	to	go	to	Harbin.

---

[19] We omit the particles *–eykeyse, -hantheyse* and *–kkey* from our discussion since they do not appear in the primary data for VYK or CSK.

Where it denotes location, this particle may refer to a location in either time or space in both

KM and VYK as below:

KM6   [sa-wəl   dar-e   wa-sˈo]

four-month   month-LOC   come-DECL.PST.POL

I   came   (here)   in   April.

KM2191   [kurgan-e   ɛ̃-jo]

Kurgan-LOC   NEG-PRS.DECL.POL

It   is   not   in   Kurgan.

VYK2530   [han   d͡ʒuir-e   itʰɨl   ʃi-go]

one   week-LOC   two days   rest-DM

I   get   two   days   off   per   week.

VYK1479   [uri   t͡suŋguk-e   munhwa-heŋmi

PRN.1.PL   China-LOC   culture-revolution

ɾa-n   gej   isˈ-ək'ədu]

COP-ADN   thing.SBJ   be present-PST-DM

There   was   this thing in China   calle   the   Cultural   Revolution, you see.

The VYK and KM particles corresponding to CSK –*ey* are near equivalents of one another.

-eyda

This single form particle often replaces –*ey* in CSK and connotes the placement of something in or on a location. The attestation of the single corresponding particle in VYK appears to have the same usage and form as CSK:

VYK1162　　[on　　　　　sesaŋ-eda　　　ərət'a-n　　　mar　　　i-mda]

　　　　　　　whole　　　　world-LOC　　　difficult-ADN　　language　　COP-PRS.DECL.POL

　　　　　　　There　　are　　　difficult　　languages　　all　　over　　the　　world.

A corresponding KM particle was used a little more frequently and, again, seems to retain a similar connotation and a similar form to CSK and VYK:

VYK560　　[jug-ɨ　　　　　　　　pʰur-eda　　　　　　　səg-əse

　　　　　　　meat-OBJ　　　　　　　grass-LOC　　　　　　　mix-CNJ

　　　　　　　kɨrɛ　　　　　　　　　mək'-o　　　　　　　　sar-at͡ʃ'i]

　　　　　　　thus　　　　　　　　　eat-CNJ　　　　　　　　live-PST.DECL.NEUTR

　　　　　　　We　　survived　　by　　mixing　　meat　　into　　the　　grass　　and　　eating　　it.

These particles are near equivalents.

-eyse

Like CSK –*ey*, this particle may also be used to express location or movement. In terms of its use expressing movement, it marks movement away from a place. In terms of its use indicating a location, its most significant distinguishing feature is that it expresses dynamic rather than static location.

KM492　　[ədi-esə　　　　　　　　　ənd͡ʒe　　　　　　　tɛjəna-sˈɨmdu]

　　　　　　　where-LOC　　　　　　　　　when　　　　　　be born-INT.PST.POL

	Where	and	when	were	you	born?

VYK45/46	[t͡ʃə	hond͡ʒa	dɛhak-esə	suksa-esə
	PRO.1.SG	alone	university-LOC	dormitory-LOC

sal-go	is'-umda]
live-CNJ	be present-DECL.PRS.
	POL

I	am	living	alone	in	the	university	dormitory.

One slight difference in form between these particles in KM and VYK is that the intervocalic /s/ may undergo voicing in KM, although this process is not entirely predictable.

KM233	[ros'i-ezə			sa-daga]
	Russia-LOC			live-and then

I	lived	in	Russia	for	a	while.

In both KM and VYK a reduced form of this particle in which the first vowel is dropped, was recorded. This form is also known in colloquial peninsula varieties of Korean (c.f Yeon and Brown 2011 p. 107). Just like the non-reduced form, it may express either dynamic location or movement away from a location.

KM145	[na-nɨ	vondo-sə	saxalin-sə
	PRN.1.PL-TOP	(Russian) Far East-LOC	Sakhalin-LOC

sluʒi-l	ha-go]
serve-PST.SG.M	do-CNJ

I   did   military   service on   Sakhalin   in   the   Russian Far   East.

KM585   [vondo-sə   iŋgi-r   tɨr-o-i]

(Russian) Far East-ABL   here-OBJ   enter-come-DM

I   came   here   from   the   Russian   Far   East.

YK1332   [kɨ   saram-ɨn   ədi-sə   wa-kilɛ]

that   person-TOP   where-ALL   come-DM

It   is   because   of   where   that   person   is   from.

VYK122   [mə-n   de-sə   o-n   saram-dɨr-i   man-kʰo]

distant-ADN   place-ABL   come-ADN   person-PL-SBJ   many-DM

There   are   a   lot   of   people   who   came   from   far   away.

VYK2248   [nɨl   kəgi-sə   sar-a's-ɨmda]

always   there-LOC   live-PST-DECL.POL

I   have   always   lived   there.

As with the non-reduced form, the sibilant may be irregularly voiced in KM, but not in VYK:

KM262   [na-nɨ   sʲevʲer-zə   puk-t͡ʃ'ok-esə   ir-ha-d͡ʒ'i]

PRO.1.SG-TOP   North-LOC   North-direction-LOC   work-do-DECL.PST

I   worked   in   the   North

Although this reduced form is present in both transplanted varieties, we note that there is a difference in the distribution of the reduced forms in VYK and KM. Namely, this form appears freely after both consonants and vowels in KM as seen above and in this example:

193

KM397	[tɛhak-ɨ̃		taʃkʲent-sə	kɨ	gə	pʰir-ha-go]
	university-TOP		Tashkent-LOC	that	thing	graduate-do-DM
	I	graduated	from	a	university	in	Tashkent.

Conversely, this variant appears with only one exception after vowels in the VYK data.

Despite their occasional slight difference in form and the restricted distribution of the Yanbian form [sə], on the grounds of their formal and functional similarity, these particles are considered near equivalents.

-eykey and *-hanthey*

We deal with these two invariant particles together since their functions are so similar. Their basic function in CSK is to mark animate indirect objects as recipients or goals. This also describes their function in VYK.

VYK434	[kɨ	jəŋhjaŋ-i	t͡ʃə-ege
	that	influence-SUBJ	PRN.1.SG-DAT

	ad͡ʒu	d͡ʒotʰa-go	sɛŋgak-ha-mda]
	very	good-IS	think-do-PRS.DECL.POL
	I think that that influence is very good for me.		

VYK2071	[sunzə-bun-dɨl-hantʰe	t͡ʃɛsa	duɾi-l	nal]
	ancestor-person.HON-PL-DAT	offeratory food	give.HON-ADN	Day
	(That is) the day when we perform *chesa* for our ancestors.			

Drawing such conclusions for KM, however, is challenging since a particle corresponding to –*hanthey* does not appear in the primary data and the usage of the very few attested examples

of a particle corresponding to –*eykey* in the KM data is differs starkly from the CSK usage as demonstrated below:

KM217   [i   sarɨm   t'ar-ege   s'a-wə]

            this   person   daughter-?OBJ   fight-PRS.DECL.PLAIN

            They   are   fighting   with   their   daughter.

Examining the KM marking of animate indirect objects, for example with the verb *cwuta*, we see that they are more often marked using a particle corresponding the object particle in or may be left unmarked as in the examples below:

KM20170   [əmej-r   ton-do   t͡ʃak'uma   d͡ʒu-o]

            mother-OBJ   money-too   constantly   give-PRS.DECL.POL

            I constantly give money to my mother.

KM529   [t͡ʃib-ɨ   han-kan-ʃ'i   t͡ʃu-ikadi]

            house-OBJ   one-CLASS-each   give-DM

            They would give each household one handful each.

KM 2073   [ʒenja   d͡ʒu-go   i   nat͡ʃʰan

            Zhenya   give-CNJ   this   Face

            aj   d͡ʒu-n-da]

            NEG   give-PRS-DECL.PLAIN

            I give (some) to Zhenya, but I do not give (any) to that person.

195

KM2481	[t͡ʃir-ɨ		nɛ		aj		d͡ʒu-o]
	work-OBJ		PRN.1.SG		NEG		give-PRS.DECL.POL
	They	do	not	give	me	any	work.

On the basis of the foregoing it appears that the means of expressing the comparative concepts of CSK –*hantey* and –*eykey* in VYK and KM are non-equivalent.

5.2.1.5 The Instrumental Particle

In contrast to CSK, in which the instrumental particle has two shapes which appear after nouns ending with vowels and the phoneme /l/ (-*lo*) or all other consonants (-*ulo*), respectively, it has a particularly large number of allomorphs in the transplanted varieties. As with the object particle, this is due to its variable vowel quality, the variably quality of the phoneme /l/ and the possibility of apocope of the final vowel. Furthermore, these variants do not seem to be entirely reliably predictable on the basis of the final segment of the word to which these particles are appended, i.e. the epenthetic vowel may be omitted following consonants and the intervocalic realisation of the phoneme /l/ may be either [l] or[ɾ] in VYK, or [l], [ɾ] or [r] in KM.

As a result of these factors, most especially the realisations of the KM instrumental and object particles may be very similar or even homophonous. Given that these particles share one similar function, marking movement towards a goal, this renders the classification of some KM particles in context challenging.

KM21	[rosija		puk-t͡ʃʼok-ɨr		ka-sʼo]
	Russia		North-direction-?INST/OBJ		go-PST.DECL
	I	went	to	the	North	of	Russia

KM2007	[mun- ɨr		hur		tad-asʼ-o]

door-OBJ	forcefully	close-DECL.PST-POL
I	slammed	the	door

This situation is complicated by the fact that various theories about the case marking of the goal of verbs of movement in KM abound. The forms *–ru* and *–r* are considered to mark the accusative by King (1986 p.261) and archaic datives by Pak (2005 p. 84). Settling this particular issue decisively is outside the scope of the current research, however for the purposes of our glosses we follow King in classifying these forms as object particles. There can be no doubt, though, that there are cases where the instrumental particle is used to mark movement toward a goal in KM, as in CSK. For examples, see below:

KM1866 [t͡ʃib-ɨro aj ka-get'a]

house-INST NEG go-MOD.PLAIN

I will not go home.

KM2425 [kɨ daɨm-e-n t͡ʃib-uro va-s'-o]

DEM after-LOC-TOP house-INST come-PST.DECL-POL

After that I came home.

KM612 [pastor jəg-lo va-tta]

priest here-INST come-DECL.PST.PLAIN

A priest came here.

We now present less ambiguous examples of the instrumental particle in its uses marking instrument, material, capacity or motion towards in both VYK and KM in order to demonstrate its range of forms.

KM15 [korjo gɨl-lo malo ilg-əs'-o]

Korean writing-INST scarcely read-PST.DECL-POL

I read very little Korean.

KM599 [ise-ro va-s'-o-n korjə

move-INST come-PST.DECL-POL- Korean
TOP

saram-i əps-əs'-o]

person-SUBJ absent-PST.DECL-POL

When we first moved here, there were not any Korean people.

KM1942 [mat͈- toŋsɛŋ d͡ʒib- ɨro ga-ja dɛ-t͡ʃ'i]

youngest-younger sibling house-INST go-CNJ become-PST.DECL.NEUTR

I had to go to my youngest sibling's house.

KM1485 [pjatʲ ʃʲest t͡ʃʲelovʲek

five six Person

uri d͡ʒib-ɨlo o-nola]

PRN.1.PL house-INST come-play.PRS.DECLPLAIN

Five or six people come to celebrate at our house.

In contrast, the forms of the instrumental particle in Yanbian Korean bear a stronger

resemblance to the CSK instrumental particle, as below:

VYK2651 [kunde-ro pone-go]

army-INST   send-CNJ

(My brother)   was   sent   to   do   his   national   service

VYK2462   [uri   t͡ʃigəpt͡ʃək-ɨro   ijagi-ha-mjən]

PRN.1.PL   professional-INST   conversation-do-if

If   we   are   talking   professionally…

VYK44   [ad͡ʒu   əri-l-t'ɛ   jəŋgil-ɨro   wa-s'ɨmda]

very   be young-ADN-time   Yanji-INST   come-
                                                             PST.DECL.POL

I   came   (to live)   in   Yanji   when   I   was   very   young.

VYK254   [jəŋgil-lo   ola-wa-s'ɨmda]

Yanji-INST   raise-come-PST.DECL.POL

I   came   up   (to live)   in   Yanji.

VYK979   [tɛbubun   ta   t͡ʃosən-mal-lo]

majority   all   Korea-language-INST

kaɲi-rɨl   ha-mda]

lecture-OBJ   do-PRS.DECL.POL

The   majority   (of   our   teachers)   lecture   entirely   in   Korean.

One peculiarity in the realisation of the VYK instrumental particle is the variable quality of the liquid sound. While this phoneme is typically realised in intervocalic as a tap, where it appears in the instrumental particle in such positions it may be irregularly realised as [l]:

VYK403	[sɛ	doŋne-lo	isa-ha-n	taɨm-e]
	new	neighbourhood-INST	move house-do-ADN	After-LOC

After moving house to a new neighbourhood…

YK398	[t͡ʃosən-esə	irəkʰe	jənbjən-ɨlo	kənə-wa-t'a-go]
	Korea-LOC	thus	Yanbian-INST	cross-come-PST.DECL.PLAIN-IS

| | They | say | they | came | into | Yanbian | from | Korea | like | that. |

The functions of the transplanted particles corresponding to the CSK instrumental particle are identical and, despite some variation in their realisation, there is considerable overlap. Therefore, these particles may be considered near equivalents.

5.2.1.6 The Comitative Particle

The form of this particle represents another case which may be used to distinguish VYK and KM. Whereas the allomorphs /-wa/ and /-kwa/ appear in CSK in phonologically conditioned complementary distribution (the latter appearing after nouns ending in in vowels, the former after those ending in consonants), neither of these surface forms appears in *Koryo Mar*. The form of the comitative particle in *Koryo Mar* is [ka], the etymological [w] having undergone glide deletion, and it does not exhibit allomorphy conditioned by the NP to which it is attached. It is homophonous with one of the allomorphs of the subject particle in Yanbian Korean, but its use is quite distinct:

KM321	[s'ak	korjə	saram-dur-ga	kat͡ʃiʰ-da]
	entirely	Korean	person-PL-COM	together-COP.DECL.PRS

| | He | is | together | with | all | the | Korean | people | (*Koryo Saram*) |

KM905	[uri	papa-ga	mama-nɨ	kegi-sə	sara-t͡ʃ'imusɨ]

PRN.1.PL	father-COM	mother-TOP	there-LOC	live-PST.DECL.POL
My	father and	mother	lived	there.

KM216   [t'ar-do-ga   s'a-u]

daughter-ADD-COM   fight-PRS.DECL.POL

I   argue   with   my   daughter,   too.

In addition to this characteristic form of the comitative particle, other forms which are identical to their CSK equivalents may also be observed in KM:

KM588   [ja   əməj-raŋ   kəgi-sə   sa-o

PRO.1.SG.SUBJ   mother-COM   there-LOC   live-DECL.PRES.POL

uʃtobʲe-ra-n   dɛ-sə]

Ushtobe-COP-ADN   place-LOC

I   live   there   with   my   mother   in   a   place   called   Ushtobe.

KM702   [kand͡ʒɛ-hago   t͡ʃɛ̃j-hago   s'ak̚]

soy sauce-COM   fermented soy bean paste- COM   Entirely

(We put in)   all of   the soy   sauce   and fermented   soy bean   paste

In the case of VYK, the forms of the comitative particle are even more varied. In addition to forms of the comitative particle identical to those of CSK, we further note that glide deletion is variably applied Yanbian Korean, therefore forms of the comitative particle homophonous with both CSK and *Koryo Mar*, [kwa] and [ka], respectively, number among the possible realisations of the comitative in this variety.

VYK349   [pumonim-gwa   hamk'e   sal-go   i-k'o]

parents-COM   together   live-CNJ   be-DECL.PRS.DM

I   am   living   together   with   my   parents

VYK2263   [jəŋə-wa   hanə-ro   jəŋə-ɾɨl   pɛju-go]

English-COM   Chinese-INST   English-OBJ   learn-DECL

We   learn   English   in   Chinese   and   English.

VYK1517   [kɨ-t'e-nɨ   puk-t͡ʃosən-ga   nam-t͡ʃosən   i-ge

that-time-TOP   North-Korea-COM   South-Korea   this-thing.SUBJ

da   han-ge   nara-i'-t͡ʃ'i]

all   one-CLASS   country-COP-DECL.PST

Then   North   Korea   and   South   Korea   were   all   one   country

Above is example of regular comitative allomorphy in VYK followed by a comitative particle
which has undergone glide deletion. We also present examples of stylistic variants of the
comitative,–*(i)rang* and *–hago*, as they are used in VYK.

VYK2495   [nampʰjən-iraŋ   aj-raŋ   kat͡ʃʰi   ka-gəna]

husband-COM   child-COM   together   go-DM

Do you mean going together with my husband and children or something?

VYK969   [t͡ʃʰin-harabəd͡ʒi-halməni-hago   wɛ-harabəd͡ʒi-halməni

paternal-grandfather-grandmother-COM   maternal-grandfather-grandmother

My   grandparents   on   both   my   father's   and   my   mother's   side

	da	t͡ʃosən-d͡ʒok	i-mda]
	all	Korean-ethnicity	COP-DECL.PRS.POL

are all members of the Korean community of China.

A final characteristic of this particle which has been noted for KM as a distinguishing factor from CSK and Yanbian Korean counterparts are the verbs by which it is governed. As well as the set of verbs governing the comitative case common to Korean (e.g. *kyelhon hata, ssawuta* etc.), verbs of asking also govern this case. In the primary data, though, no examples of verbs of asking combined with the comitative case are to be found. Curiously, two examples of this structure may be found in the primary data of VYK.

VYK89	[jet͡ʃən	gɨ	halməni-ga	murəba-nde]
	previously	DEM	grandmother-COM	ask.DM

I asked my grandmother about this a long time ago.

VYK2030	[i	sanaj-ga	murəba-s'o]
	DEM	man-COM	ask-DECL.PST.POL
	I	asked that	man

Despite not appearing in this project's primary data, this use of the comitative particle in KM is well attested (cf. King 1992 p. 208; Pak 2005). Therefore, we tentatively suggest that this is a relic feature of the antecedent Hamgyeong varieties of both transplanted varieties. The comitative particles of KM and VYK are considered to be near equivalents of one another on the basis of this and their other similarities.

5.2.2 Delimiters

The structure of our examination of the delimiters is to be similar to the above enumeration and subsequent discussion of case particles. We reiterate that, as noted in the introduction to section5.2, the naturalistic data under discussion here does not allow us to draw absolute conclusions about the equivalances of either the case particles discussed above or the delimiters which we discuss below. Thus, the evaluations which follow must be taken as provisional. Delimiters are a class of Korean particles distinct from case particles in that they add meaning to the constituents which bear them rather than simply showing grammatical relations and have a different syntactic distribution from case particles. Specifically, they can be combined with noun phrases, with or without an accompanying case particle, and also with adverbs or inflected forms of predicates (Lee and Ramsey 2000 p.139; Yeon 2003 p. 25). In the following table we present a select inventory of the delimiters of Korean and a catalogue of the realisations of their corresponding particles in VYK and KM:

Particle	CSK Forms	Yanbian Korean Forms	*Koryo Mar* Forms
Topic	/-un/, /-nun/	[ɨn], [nɨn], [ɨ], [n], [nɨ], [ɨnɨ]	[ɨn], [nɨn], [ɨ], [nɨ], [ɨnɨ], [n], [ɨ] -*u, nu*
Plural	/-tul/	[dɨl], [dɨɾ]	[duɾ], [dɨl], [dɨɾ]
Particles of Extent	/-man/	[man], [mã]	[man], [mã]
	/-to/	[do] *twu, to*	[do]-*to*
	/-puthe/	[putʰə]	[putʰə]
	/-k'aci/	[k'ad͡ʒi]	[k'ad͡ʒi]
			-kkaci, kkeci, kkasi
	/-pakk ey/	[pak'e]	[pak'e]
Particles of Frequency	/-mata/	[mada]	[mada]
	/-ssik/	[ʃ'ik], [ʃ'i]	[ʃ'i]

Particles of Comparison	/-cherem/	[t͡ʃʰərəm]	[t͡ʃʰərəm]
	/-taylo/	[dɛro]	[dɛro]
	/-mankhum/	[mankʰɨm]	N/A
	/-pota/	[poda]	[ma]
	/-kathi/	N/A	[gat͡ʃʰi]
Particles of Approximation of Optionality	/-ina/, /na/	[na], [ina]	[na], [ina]
	/-ccum/	[t͡ʃʼɨm]	N/A

Table 75: Attested Forms of VYK and KM Delimiters

More precise explanations of the functions of these particles and, where appropriate, suggested translations are provided in their respective sections.

While it must be conceded that this is not an exhaustive[20] list of the delimiters of Korean, it is representative of those which appear in the primary data upon which this analysis is largely based. Again, as above, those forms which do not appear in the primary data, but which have been elicited by other researchers appear in italics.

5.2.2.1 The Topic Particle

As with the majority of case particles, but in contrast with the majority of other delimiters, the topic particle exhibits phonologically conditioned allomorphy in CSK, appearing as *–un* after consonants and *–nun* after vowels. In addition, it may be further reduced to *–n* in certain contexts (e.g. after first person singular pronouns cf. Yeon and Brown 2011 p. 123). These forms are all attested in KM and VYK with a similar distribution to CSK. In addition to these, forms which undergo apocope of the final consonant and a form which introduces an

---

[20] To put this selection of delimiters in context through comparison with other treatments of similar subjects in CSK, the "fairly exhaustive" list of delimiters presented by Lee and Ramsey (2001 p. 161) only amounts to twelve items. Of the twenty three delimiters identified by Yeon and Brown (2011 pp. 121-149), whose taxonomy is followed for expository convenience in the structure of this section and the grouping of these particles, the majority are included in our selection and those which are not, for example *–(i)yamallo* and *–khenyang* appear only very infrequently in Korean of all varieties.

epenthetic vowel between the final consonant of the noun to which the particle is appended and the particle were recorded. Examples of each form attested in the KM primary data may be found below.

KM 943   [uɾi-nɨn   irɛ   kosa-ha-mje   saɾaga-t͡ʃ'imɨsɨ]

PRN.1.PL-TOP   thus   worry-do-CNJ   surivive-PST.DECL.NEUTR

We have survived this long while living with a lot of worries.

KM561   [aj-dɨɾ-i   man-ɨn   d͡ʒib-ɨn   kit͡ʃage   pap'-əs'-o]

child-PL-SBJ   many-ADN   house-TOP   extremely   busy-PST.DECL-POL

It was extremely difficult for households that had a lot of children.

KM34   [sever-nij   korej   na-n   mor-o]

North-M.SG   Korea.M   PRN.1.SG-TOP   not know-PRS.DECL.POL

I do not know North Korea.

KM1480   [na-nɨ   kɨrən   tɛdap-t͡ʃ'ir   ãj   ɛs'-o]

PRN.1.PL-TOP   that kind of   answer-work   NEG   do.PST.DECL-POL

I was not answering you (the way that you want).

KM169   [kəŋge-sə   korej   saram-dur-ɨ   mantʰa-naikkana]

there-LOC   Korean   person-PL-TOP   many-DM

There are a lot of Korea people there.

KM1477   [kɨ   taɨm-ɨ̃   ət'i   ha-ges'-o]

that   after-top   how   do-MOD-POL

	What	was	I	to	do	after	that?

KM823 [ʃibaŋ   mat-adur-ɨnɨ   mahɨn-sal   i-go]

now   eldest-son-TOP   forty-years   COP-DM

My   eldest   son   is   forty   years   old   now.

As seen in examples KM1477 and KM169 above, the apocope of the final consonant of the topic particle in KM may take place with or without nasalisation of the preceding vowel.

Before moving on to VYK it must be noted that either the restriction which prohibits subject particles and topic particles from co-occurring does not operate in KM or there are frequent cases where the topic particle co-occurs with a subject particle that has been reanalysed as part of the noun stem as in the example below:

KM255   [tʼar-i-nɨ   sənsej   tayhakˀ   pir-hɛsʼ-o]

daughter-?SBJ-TOP   teacher   university   graduate-do.PST.DECL-POL

My   daughter   graduated   from   teacher's   college.

We now present examples of the variation of the forms of the topic particle in VYK.

VYK431   [kɨrən   paŋmjən-e-nɨn   dɛge   maj

that kind of   aspect-LOC-TOP   very   many

With   regard   to   that,   they   (my   parents)

paŋd͡ʒo   hɛ   t͡ʃu-ʃi-go]

help   do.CNJ   give-HON-DM

have   helped   me   a   very   great   deal.

VYK2583	[ohɛ	gatʰ-ɨn	gə-n	əp-k'o]
	misunderstanding	same-ADN	thing-TOP	not present-DM
	There is not anything like miscommunication between us.			

VYK1594	[na-nɨ	han	pən	ka	pas'-o]
	PRN.1.SG-TOP	one	time	go.CNJ	see.PST.DECL-POL
	I have been there once.				

VYK834	[t͡ʃosən	adur-ɨ̃	t͡ʃosən-mal	ha-go
	Korea	son-TOP	Korea-language	do-CNJ
	han͡d͡ʒok̚	adur-ɨ̃	han͡d͡ʒok-mal	ha-go]
	Han Chinese	son-TOP	Han Chinese-language	do-DM
	The Korean boys speak Korean and the Chinese boys speak Chinese.			

VYK733	[kohjaŋ-ɨnɨ	t͡ʃal	moru-d͡ʒi
	home town-TOP	well	not know-PRS.DECL-NEUTR
	I am not sure where their home town is.		

In contrast to KM, the co-occurrence of subject particle and topic particle is exceedingly rare in VYK, attested in the primary data only twice in ambiguous contexts, as below:

VYK759	[nui-n	id͡ʒe	sam	njən
	elder sister-TOP	now	three	Year

$$\widehat{t\int}ən-e \qquad\qquad toraga-go]$$

before-LOC   die.HON-DM

My   elder   sister   passed   away   three   years   ago   now.

Since there appears to be no difference in the function of the topic particle in the transplanted

varieties, and the full complement of identical reduced forms appears in both KM and VYK, we

consider them to be near equivalents.

5.2.2.2   The Plural Particle

The plural particle behaves similarly in CSK, KM and VYK. They may be distinguished only on

the basis of phonetic realisation, i.e. the final liquid of the particle –tul may be realised as a trill

in KM. Examples of the plural particle in KM and VYK are given below.

KM 615   [jesu-l   min-nɨn   saram-dɨl   iŋgi   vas'-o]

Jesus-OBJ   believe-ADN   person-PL   here   come.PST.DECL.-POL

Some   missionaries   came   here.

KM206   [osobʲeno   namd͡ʒa-dɨr   havund͡ʒa   pap'-o]

especially   man-PL   alone   busy-PRS.DECL.POL

It   is   especially   difficult   for   men   to   be   alone.

KM1432   [toŋsẽ-dur-ɨn   is'-o]

younger sibling-PL-TOP   be present-PRS.DECL.POL

I   have   some   younger   siblings.

VYK2303   [t͡ʃigɨm   hangukʾ   ga-ʃi-n   bun-dɨl-do   man-kʰo]

now   Korea   go-HON-ADN   Person.HON-PL-too   many-DM

There are   a lot of   people   who   have   gone   to   Korea   now,   too.

209

VYK1515	[harabəd͡ʒi- dɨr-ɨn	kibonsaŋ-ejsə	puk-t͡ʃʼosən	saram	i-mda]
	grandfather-PL-TOP	basically-LOC	North-Korea	person	COP-PRS.DECL.POL
	My	grandfathers	were	basically	North	Korean.

Since the only point of difference between them is the potential but not invariable realisation of the final consonant of the KM plural particle as a trill, the plural particles of the transplanted varieties may be considered near equivalents.

5.2.2.3 Particles of Extent

Of the nine CSK particles of extent identified by Yeon and Brown (pp. 129-139), only five corresponding particles are attested in the combined primary data of both VYK and KM, therefore we restrict our discussion to those particles, namely those corresponding to *-man, -do, -pwuthe, -kkaci* and *pakk ey*. We present examples of each below in that order and comment on differences in their realisation below each set of examples.

The Delimiting Particle -man

KM268	[kɨ	sɨd͡ʒip- t͡ʃip-ejsə	t͡ʃosən-mar-mã	ha-d͡ʒi]
	that	marriage-house-LOC	Korean-language-only	do-PRS.DECL.NEUTR
	In the	family	that I married into	they only speak Korean.

KM528	[kosa-ha-n	gə-n-man	sɛŋgak-hɛ-t͡ʃʼi]
	worry-do-ADN	thing-TOP-only	thought-do-PRS.DECL.NEUTR
	We	only	thought	about	our	troubles.

VYK2138	[kɨnjaŋ	t͡ʃagi	irɨm-man	mal-ha-nɨn

	just	self	name-only	language-do-ADN

	t͡ʃəŋdo		is'-ɨmda]
	extent		be present-PRS.DECL.POL

They only spoke (Korean) to the extent that they could introduce themselves.

VYK1814	[puŋ-mã		t͡soaha-mida]
	pwuk-only		like-PRS.DECL.POL

I only like the *pwuk* (a kind of traditional drum).

This delimiting particle is identical in function in CSK, KM and VYK. The one case of variable realisation is the dropping of the final /n/ and the accompanying nasalisation of the preceding vowel which may be found in both VYK and KM. Thus the KM and VYK particles are nearly equivalent.

The Additive/Focus Particle -do

This particle exhibits allomorphy in neither of the transplanted varieties, nor in CSK. Although this particle has been recorded with a front rounded vowel in both KM and VYK elsewhere (cf. Pak 2005 p. 169; Jeong 2010 p. 59) this form did not appear in the primary data collected for this project.

KM1964	[uri	vit͡ʃa-do	sur-u	t͡ʃogɨm	t͡ʃoha-hadanaj]
	PRN.1.PL	Vicha-too	alcohol-OBJ	a bit	like-DM

Vicha liked to have a bit of a drink, too.

KM512   [kohε-do   tεŋgi-mdu]

church-too   attend-PRS.INT.POL

Do   you   go   to   church,   too?

VYK21   [kat͡ʃʰi   jəŋhwa-do   po-go]

together   film-too   watch-DM

We   watch   films   together,   too.

VYK2591 [t͡ʃə-do   kamsa-ha-mnida]

PRN.1.SG-too   gratitude-do-PRS.DECL.POL

Thanks   from   me,   too.

An apparent point of difference between these particles, though, is its appearance with object

case particles in KM. This is notable since, typically, where a delimiter such as –*to* is used in

CSK and VYK with a subject or direct object, the case particle is omitted (cf. Yeon 2003 p. 26).

KM58   [nε   arε-ru-do   t͡suku]

PRN.1.SG.GEN   junior-OBJ-too   die-DM

My   younger   sibliings   died,   too

KM2448   [vit͡ʃʰa   t͡ʃib-ul-do   ka-ge-nɨnde]

Vicha (personal name)   house-OBJ-too   go-MOD-DM

I   should   go   to   Vicha's   house,   too

Furthermore, this particles appears combined with other particles, the comitative in the below

example, in a non-canonical order, i.e. with the delimiter appearing 'inside' or closer to the

noun than the other particle.

212

KM216 [t'ar-do-ga   s'a-u]

    daughter-ADD-COM   fight-PRS.DECL.POL

    I argue with my daughter, too.

These characteristics of this particle are not mentioned in any prior research and this does not appear to be attested in secondary sources. Without more extensive examination of this feature in KM, though, it is difficult to conclude whether these examples show a new a feature of the variety or the idiosyncracic habits of a few consultants. Therefore, due to its broadly identical form and function in the YKAP and in Central Asia, these particles are considered nearly equivalent.

The Ablative and Allative Particles -pwuthe and –kkaci

These particles express origin and goal, respectively. They are often translated into English as 'from' and 'up to' and share with those words the possibility of being used to delimit both spatial and temporal extent. Neither particle varies in form either within or across the transplanted varieties.

VYK429 [nɛ   əri-l   t'ɛ-butʰə   kjujug-ɨ

    PRN.1.SG.SBJ be young-ADN time-from   education-OBJ

    dɛge   t͡ʃuŋʃi-hɛ-t'a]

    very much   emphasise-do-DECL.PST.PLAIN

    They emphasised education from when I was young.

VYK1793 [at͡ʃʰim   jəseʔ-ʃi-butʰə   t͡ʃənje   ahop- ʃi k'ad͡ʒi]

    morning   six-hour-from   evening   nine-hour-up to

    From six in the morning until nine in the evening

VYK2389	[kɨrən	il-k'adʒi	ha-go	is'-ɨmnida]
	That kind	work-up to	do-CNJ	be present-PRS.DECL.POL
	I am even doing that kind of work.			

KM843	[ku-wər	t͡ʃəsʲ	nar-putʰə	i	sʲemʲester
	nine-month	first	day-from	this	Term

	t'o	kɨr-ɨ	ir-t͡ʃʼi]
	also	writing-OBJ	read-PST.DECL.NEUTR
	I have been studying this term from the first of September, too.		

KM117	[kəŋgɛ-sə-n	t͡ʃʰil-haŋ-njən- k'adʒi	ilg-əs'-o]
	there-LOC-TOP	seven-study-year-up to	read-PST.DECL-POL
	I studied there until seventh grade.		

It is notable that the particle corresponding to CSK –*kkaci* appears only very seldom in the KM data. This may be attributed to the fact that the accusative particle is more commonly used for its KM function of expressing 'motion towards/into'. We also note that although these particles appear together frequently in CSK and VYK, they do not appear to do so in KM. Nevertheless, we consider the transplanted particles corresponding to –*kkaci* to be near equivalents.

Turning to the particles corresponding to CSK –*pwuthe*, they exhibit near identical form and function in the transplanted varieties, so they may be considered nearly equivalent.

The Exclusive Particle -pakk ey

This invariant complex particle has an exclusive meaning. Its form and function appear to be

identical in both VYK and KM, as in the examples below.

VYK2154   [kɨnjaŋ   t'am   hɨl-ət'a-nun

           just   sweat   flow-PST.IS-AND

           ke-j-pak'-e   kijək   əp-s'ɨmda]

           thing-SBJ-except for   memory   be absent-DECL.PRS.POL

           I   do   not   remember   anything   apart   from   sweating   heavily.

KM2220   [i   gə- pak'-e   moɾ-u]

           this   thing-except for   not know-DECL.PRS.POL

           I   do   not   know   anything   apart   from   this.

The forms and functions of this particle are consistent across the varieties of Korean spoken

both within and outside of the Korean peninsula. Thus, we have a further case of near

equivalence.

5.2.2.4   Particles of Frequency
The Universal Quantifier -mada

This CSK particle exhibits no allomorphy and is equivalent in meaning to English 'each' or

'every'. Its realisations and functions in VYK and KM are identical to each other and to those of

CSK, as seen below:

KM1352   [saram-mada   ta   haŋgab- ʃi   moʃi-nɨn

           person-each   all   sixtieth birthday-party   visit-AND

	saram-dɨr	man-so]
	person-PL	be many-DECL.PRS.POL
	Everyone gets a lot of guests for their sixtieth birthday party.	

VYK1572	[papˀ	məg-ɨl	t'ɛ-mada
	food	eat-ADN	time-each

	kug-u	hɛ	mək'o]
	soup-OBJ	do.CNJ	eat.DM
	Every time we eat food we make a soup (to go with it).		

These particles are considered nearly equivalent on the basis of their formal and functional similarity.

The Distributive Particle -ssik

This particle, also invariant in CSK, marks "regularity or equal distribution" (Yeon and Brown 2011 p. 140). In KM and VYK it may appear with a form identical to the CSK particle, but it appears overwhelmingly more frequently having undergone the apocope of its final consonant as below.

KM651	[t͡ʃogɨm-ʃ'i	adɨr	ta	va-s'-o	na-rã]
	a little-apiece	son	all	come-DECL.PST-POL	PRN.1.SG-COM
	Little by little my sons all came (to be) with me.				

KM125	[han	d͡ʒuil-e	du-bən-ʃ'i]
	one	week-LOC	two-time-apiece

Twice   per   week.

KM684   [modu   han   d͡ʒan-ʃ'ik   maʃi-ra]

all   one   glass-apiece   drink-IMP.INF

Everyone   drink   one   glass   each!

VYK1276   [il   ljən-e   han-bən-ʃ'i   an   mana-go]

one   year-LOC   one-time-apiece   NEG   meet-DM

We   do   not   (even)   meet   once   per   year.

VYK2258   [t͡ʃogɨm-ʃ'i   tarɨ-n   mjən-i   is'-ək'o]

a little-apiece   be different-   aspect-SUBJ   be present-PST.DECL-
                       ADN   DM

There   were   slight   differences   in   some   respects.

These particles may be considered near equivalents.

5.2.2.5   Particles of Comparison

Yeon and Brown identify six CSK particles of contrast (2011 pp. 144-149): *-cherem, -kathi, -mankhum, -pota, -ttala* and *taylo*. Of these, four appear in the primary data, but with an uneven distribution between VYK and KM. Since a particle corresponding to or *–kathi* does not appear anywhere in the primary data, we restrict our discussion to those that do, namely particles corresponding to *-cherem, -mankhum, -pota,-kathi* and *taylo*. The KM primary data, however, lacks attestation of a particle corresponding to *–mankhum* and the VYK primary data lacks attestation of a particle corresponding to *–kathi*.

The Particle of Equivalence -cherem

This one-shape particle is used to express the idea of things being 'alike'. It has identical form and function in CSK, VYK and KM, as can be seen in the examples below. We consider the particles of these varieties to be near equivalents.

KM124   [inostran-ij   jazɨk-t͡ʃʰərəm]

            foreign-ADJ.M.SG.NOM   language.M.SG.NOM-like

            (It   was)   like   a   foreign   language.

VYK1806   [t͡ʃinan   pən-e   mar-ha-n   gə- t͡ʃərəm]

            previous   time-LOC   language-do-   thing-like
                                                      ADN

            It   is   as   I   said   last   time.

The Conformative Particle -taylo

This particle, which has been translated as 'in accordance with' or 'in conformity with' (Yeon and Brown 2011 p. 149) appears in both the VYK and KM data. Its form and function are near identical in each transplanted variety and also in CSK. Thus, the particles of VYK and KM are near equivalents.

KM437   [mar   nao-n-dɛro]

            language   come out-ADN-in accordance with

            It   is   just   as   you   said.

VYK1190   [pʰjənan-ha-n- dɛro   murobo-ʃi]

            comfortable-ADN-in accordance with   ask-IMP.HON

            Please   ask   whatever   you   like.

The Equative Particle -mankhum

This CSK particle "indicates that two things are equal or have reached the same extent" (Yeon and Brown 2011). While the VYK particle, which appears only once in the data, appears to be a near equivalent of that of CSK, no equivalent particle appeared in the KM data.

VYK2441 [himdɨl-d͡ʒiman   kɨ-mankʰɨm   poram-do   i-k'o]

difficult-but   that-as   reward-too   be present-DM

It   is   difficult   but   it   is   rewarding   to   the   same   extent.

The Comparative Particle -pota

In CSK this invariant particle is used in expressions of direct comparison. A formally similar particle appears in the VYK data, whereas no such particle appears in the KM data:

VYK306 [kidɛ-poda-nɨn   t͡ʃɛmi-əps-ɨmda]

expectation-than-TOP   fun-be absent-PRS.DECL.POL

It   was   less   fun   than   I   expected.

VYK2306 [jenal-poda   də   hɨtʰəd͡ʒi-nɨn   saŋhwaŋ i-mnida]

the old days-than   more dispersed-   situation COP-PRS.DECL.POL
                                ADN

The   (living situation of my family) is   more dispersed than   it used to be.

While very few comparative constructions appear in the KM data, those that do are marked with the particle *–ma*, a form unique to this variety, but functionally identical to CSK and VYK *–pota*:

KM161   [mwəd͡ʒi-ma   nosa   t͡ʃən-esə   korjɛ

	what-than	Russia	Before-LOC	Korea

	adɨr-tɨr	kit͡ʃʰi-gɛ		pap'-as'-o]
	son-PL	exceeding-ADV		busy-PST.DECL-POL

Above all, in Russia, before life was exceedingly hard for Koreans.

KM1364	[jenal-ma	t͡ʃogɨm	ta-d͡ʒi
	the old days-than	a little	be different-PRS.DECL.NEUTR

It is a little different from how it was in the old days.

Due to the large discrepancy in form, but similarity in function, VYK –*pota* and KM –*ma* are classified as partially equivalent.

The Analogical Particle -kathi

Used as a particle, CSK –*kathi* denotes similarity in a way similar to the English word 'like'. A homophonous adverb is commonly used in both VYK and KM, however there is no attestation of its use as a particle in the VYK data and only one case where it is used as such in KM, as below:

KM665	[adɨr-i-gat͡ʃʰi-nɨn	tambɛ-do	ãj	pʰju- d͡ʒi]
	son-SBJ-like-TOP	tobacco-even	NEG	smoke-PRS.DECL.NEUTR

Like my son, I do not smoke.

5.2.2.6 Particles of Approximation and Optionality
The Disjunctive Particle -(i)na

The CSK particle -*(i)na* exhibits allomorphy with the vowel being inserted between consonant final phrases to which it is attached and dropped with vowel final phrases. Its meaning is roughly equivalent to English 'or', but it can also be used to express approximation and non-

best option. Its function and variable form is similar in the transplanted varieties which may

be seen in the examples below:

KM1132   [jər-jədep   saɾ-i   jəl-ilgup   saɾ-ina

         ten-eight   years old-SBJ   ten-seven   years old-or

         na   ʃidip-ha-o]

         PRN.1.SG   marriage-do-PRS.DECL-POL

         I   got   married   when I   was   about   seventeen   or   eighteen.

KM503   [moskva-na   taʃkʲent-sə   nao-n   t͡ʃʰɛg   is'-ɨpkuma]

         Moscow-or   Tashkent-LOC   come out-   book   be present-PRS.INT.POL
                                          ADN

         Do you have   any   books that   were   published in Moscow or Tashkent?

VYK635   [ɨmag-ina   kɨrən   paŋmjən-e   də

         music-or   that sort   area-LOC   More

         hɨŋt͡ʃʰi-ga   Is'-əs'-ɨmnida]

         taste-SBJ   be present-PST.DECL-POL

         I   had   more   of   a   taste   for   music   or   that   sort   of   thing.

VYK288   [t͡ʃʰiŋgu-dɨl-gwa   kat͡ʃʰi   maʃ-i-nɨn

         friend-PL-COM   together   taste-be present-ADN

	gə-na	mək'o]
	thing-or	eat.DM

I   (go out to)   eat   something nice   with   my   friends or   something.

VYK348	[toŋsɛj-na	hjəŋ-ɨn	əps-sɨmda]
	younger sibling-or	older brother-TOP	not present-PRS.DECL.POL

I   do   not   have   any   younger   or   older   siblings.

These particles are near equivalents in VYK and KM.

The Approximative Particle -ccum

This particle indicates approximate amount and is most commonly with constructions denoting specific numbers. Its VYK form, presented below, is invariant and similar to the form used in CSK, however no equivalent of this particle appeared in the KM data.

VYK316	[halməni	jəl	mjəs̚-sal-t͡ʃʼɨm	tɛsʼ-ɨl	t'ɛ]
	grandmother	ten	how many-years old-approximately	become.PST-ADN	Time

It was when my grandmother was something-teen years old

VYK2543	[jəhɛŋ-ɨl	han-bən-t͡ʃʼɨm	kjehik̚-ha-go	ik'o]
	travel-OBJ	one-time-approximately	plan-do-and	be present-DM

I have plans to go travelling around once.

It is not possible on the basis of this attestation alone to determine a degree of equivalence between the VYK and KM particles which correspond to this CSK comparative concept.

5.2.3 Summary

In this section we have reviewed the nominal morphology of KM and VYK. The correspondences and dissimilarities between the functions and forms of particles recorded in the primary data of each transplanted variety were catalogued and explained with reference to the particles of CSK, which served as as comparative concepts. We summarise these relationships of equivalence between the particles of the transplanted varieties established in the foregoing section below in tabular format:

Marking	CSK Comparative Concept	Near Equivalents	Partial Equivalents	Non-equivalents
Subject	*-i, -ka, -kkeyse*		✔	
Object	*-ul, -lul*	✔		
Genitive	*-uy*	✔		
Movement and Location	*-ey*	✔		
	-eyta	✔		
	-eyse	✔		
	-eykey	✔		
	-hanthey			✔
Instrumental	*-lo, -ulo*	✔		
Comitative	*-wa, -kwa, -lang, -ilang, -hako*	✔		
Topic	*-un, -nun*	✔		
Plural	*-tul*	✔		
Extent	*-man*	✔		
	-to	✔		

	-puthe	✔		
	-k'aci	✔		
	-pakk ey	✔		
Frequency	*-mata*	✔		
	-ssik	✔		
Comparison	*-cherem*	✔		
	-taylo	✔		
	-mankhum	N/A	N/A	N/A
	-pota			✔
	-kathi	N/A	N/A	N/A
Approximation or Optionality	*-ina, -na*	✔		
	-ccum	N/A	N/A	N/A

Table 76: Summary of Contrasts Between VYK and KM Particles

From this we can see that most of the particles of VYK and KM are partial equivalents in that they have substantial similarities in both form and function. Notable exceptions include the subject particle and the particle of comparison *-pota*. This reiterates the findings of earlier works on these transplanted varieties. The large number of near equivalents is suggestive of a high-degree of similarity between these varieties, however, it is important to emphasise that these equivalences are drawn up on the simple presence or absence of particular forms or uses in the primary data. For example, a contributing factor the evaluation of the KM and VYK topic particles as near equivalents is their commonalities in terms of irregular forms, including the realisation of the topic particle as [ɨnɨ]. However, this form appears with roughly three times more frequently in the VYK data than the KM data. We go on to interrogate such a characterisation of these varieties in chapter six's quantitative analysis, but first we provide a contrastive account of the verbal morphology of these transplanted varieties.

The morphological features of transplanted varieties of Korean discussed in the foregoing section all relate to declension (*kokyong*) in Korean. In the forthcoming sections we describe and contrast the conjugational (*hwalyong*) morphology of Central Asian and Chinese Korean by examining both the final endings and pre-final endings of predicates.   It is necessary to note here that the term 'predicate' covers more than a single part of speech in the description of CSK. Rather, it may refer both to *tongsa* (action verbs) and *hyengyongsa* (stative verbs), as well as nouns combined with the copula[21]. These parts of speech are distinguished not simply by differences in meaning, but by differences in the forms of the plain style final endings and adnominalizing endings which they may take.

These sections follows a similar structure and methodology to the preceding section on nominal morphology. Again, we follow the Korean grammatical tradition in order to identify the underlying concepts which will form the basis for our comparison. This divides the inventory of conjugational morphology into the categories: 'pre-final endings' (*senemalemi*) and 'final endings' (*emalemi*)[22]. We examine each of these categories in turn presenting first an inventory of comparative concepts, then we go on to discuss each morpheme in more detail to establish equivalences before presenting a summary of equivalences and going on to provide some contextual contrastive descriptions of several characteristic features as well as the vocabulary of the transplanted varieties.

---

[21] The status of the copula in CSK is disputed. It has been variously considered it similarly to a particle (c.f. the characterisation of Korean pedagogic grammars in Lee and Ramsey 2000 p. 86, 103), a verb (c.f. Yeon and Brown 2011 pp. 155-156 on "the equational verb" in whose terminology 'verb' covers all inflecting forms, i.e. both descriptive and processive verbs), specifically an adjective/descriptive verb (Chang 1996 p. 32) and a separate and independent part of speech (c.f. Kim-Renaud 2009 p. 63).

[22] These correspond to a larger number of Yeon and Brown's more precise categories including "subject honorifics" and "tense and aspect" endings in the case of pre-final endings and "hearer honorifics and "sentence endings" for final endings.

5.2.4 Pre-Final Endings

These endings have a diverse array of functions and mark tense, modality, evidentiality, and subject honorification.

Pre-final Ending	CSK Forms	Yanbian Korean Forms	*Koryo Mar* Forms
Simple Past Tense	*-ass-, -ess-*	[as'], [əs'], [ɛs']	[as'], [əs'], [ɛs']
Past-Past Tense	*-assess-, -essess-*	[as'əs']	[as'əs']
Observed Past Tense	*-te-*	[də], [sɨpdi]	[də], [dəra], [ɨpdi]
Future/Modal	*-keyss-*	[ges']	[ges']
Subject Honorific	*-(u)si-*	[ʃi]	N/A

Table 77: Attested VYK and KM Pre-Final Endings

Due to the wide range of phonological environments in which these endings may find themselves, there are diverse realisations of each morpheme in the transplanted varieties.

5.2.4.1 Simple Past Tense

The past tense is marked similarly in both transplanted varieties and in CSK, that is with the morphemes corresponding to CSK *–a/ess-*. This is also the case in both transplanted varieties. However, examples of these canonical forms are rare. As noted in section 5.1.6 of our discussion of the phonological systems of the transplanted varieties above, the final consonant of this morpheme frequently appears in positions in which it undergoes final neutralisation and causes the fortition of following initial consonants. In addition, the vowel may elide with the verb stem's final vowel. Examples of simple past tense forms in VYK and KM may be seen below:

VYK330   [kɨ   t'ɛ   sənsɛŋnim   əmgjək̚   hɛs'-ɨmnik'a]

           that   time   teacher   strict   do.PST-INT.POL

           Were   teachers   strict   in   those   days?

VYK723 [kəgi kʰɨ-n koɲd͡ʒaŋ is'-ətʃ'i]

there big-ADN factory be present-PST.DECL.POL

There was a big factory there.

VYK1434 [ilbon saram-dɨr-i t͡ʃuŋgug-e was'-o]

Japan person-PL.SBJ China-ALL come.PST-DECL.POL

Japanese people came to China.

VYK2856 [oʃ-i əps-əs'-o]

clothes-SBJ be absent-PST-DECL.POL

We did not have our own clothes.

KM1899 [mat'oŋsɛ̃ kɨrəkʰe t͡ʃir hɛs'-o]

youngest younger sibling thus work do.PST-DECL.POL

My youngest younger sibling did work like this.

KM462 [is'-əs-ɨmdu]

be present-PST-INT.POL

Did you have one of those?

KM617 [kəŋi vas'-o jərirɨ saram]

there come.PST-DECL.POL many People

A lot of people went there.

KM1448 [jag-i əps-əs'-o]

medicine-SBJ	be absent-PST-DECL.POL
We did not	have any medicine.

Given their formal and functional similarity, these morphemes are near equivalents in VYK and KM.

5.2.4.2 Past-Past Tense

In CSK this tense is formed by simply attaching the past tense pre-final ending to the verb stem twice. As with the past tense, it denotes either an action which takes place prior to the moment of speaking, but connotes that the actions is either fully completed in the past (and thus not relevant to the present) or somehow more temporally remote than actions marked only with the simple past tense pre-final ending. Formally, the vowel harmony rules of CSK apply only to the first syllable of the past-past tense ending, therefore its two possible forms are –*essess*- and –*assess*-. Although not frequently attested, similar forms appear to be used in VYK and KM:

VYK877   [abədʒi-nɨn   imi   toragas'-əs'-ejo]

father-TOP   already   die.HON.PST-PST-DECL.POL

My father has already passed away (long ago).

KM2116   [uri   t͡ʃogɨm-hã   orɛbi   gas'-ət͡ʃ'i]

PRN.1.PL   a little-do.ADN   elder bother   go.PST-DECL.PST.NEUTR

My youngest elder brother had been there.

KM451   [mo   kas'-əs'-ɨmdu]

NEG   go.PST-PST-INT.POL

Were you not able to go there (back then)?

Since this ending appears only once in the VYK primary data and it appears only with verb stems with an inherent [a] in the KM data, it is not possible to ascertain whether vowel harmony operates equivalently in the two transplanted varieties. On the basis of secondary data, though, we are able to confirm that vowel harmony operates in both transplanted varieties just as it does in CSK (cf. Pak 2005 pp. 98-99; Kwon 2010 pp. 56; Jeong 2010 pp. 148-149;). Given their other formal and functional similarities, though, it is reasonable to consider these endings near equivalents.

5.2.4.3   Observed Past Tense

In addition to the endings denoting simple tenses described above, CSK a pre-final ending which is used to denote actions which have been carried out prior to the time of speaking, but which the speaker has observed or has sensory evidence of. It is is also known elsewhere in the literature as the "retrospective" (cf. Lee and Ramsey 2000 p. 179). In CSK its basic form is -*te*- in the plain or informal speech style, but it may also appear as –*tey*- or even –*ti*- in more formal speech styles.

All of these forms appear in the KM primary data with similar meaning to the corresponding CSK pre-final endings:

KM1929	[kɨ	daɨm-e	amu
	that	after-LOC	any

	sol-do	ai	ha-də-ra]
	care-even	NEG	do-RETR-DECL.NEUTR
	After that (I remember that) no-one looked after me.		

KM941   [su   tən   saram   ka-de]

               number   thousand   person   go-RETR.POL

               (I   recall   that)   thousands   of   people   went.

KM2255   [hanjaŋ   ʃʼip   pisʼa-de]

               Korean traditional medicine   vulva (obscenity)   expensive-RETR.POL

               Korean traditional   medicine was bloody   expensive,   to my memory.

KM90   [ut͡ʃitilʲja   nʲemʲet͡s-tɨr   isʼ-ɨpdi]

               teachers.PL   German-PL   be present-RETR.POL

               (I   remember   that)   the   teachers   were   Germans.

KM1661   [banʲja-r   ka-getta-m   iɾ-ep-di]

               public baths-OBJ   go-MOD-if   work-be absent-RETR.POL

               It   was   OK   if they   could   go   to   the   public   baths.

It is a possibility that some KM speakers have reanalysed the observed pasts when it is

combined with the declarative, plain ending to be a simple rather than complex form. This is

demonstrated by the fact that final endings marking politeness may be added to it.

KM2277   [kəŋgi-rɨ   taŋ-dəra-o]

               there-OBJ   go-RETR-DECL.POL

               (I   recall)   that   they   went   there.

This form was observed only once in all of the KM primary data, though, so the extent which

we may generalise this conclusion to KM is questionable.

Turning to VYK we see the appearance of the same forms in the primary data:

VYK111	[mwə				hɛ-də-ra]		
	what				do-RETR-NEUTR		
	What	did	(I	hear	that) they	call	it	again?

VYK2080	[əɾi-l		t'ɛ		mana-de]	
	be young-ADN		time		many-RETR.POL	
	(I remember that)	there were	a lot (of them)	when I	was	younger.

VYK2145	[ta		aɾadɨ-sɨpdi]		
	all		understand-RETR.POL		
	Everyone	understood	(as	far as	I	recall).

These endings, though, appear with far less frequency in the VYK data than the KM data. Since this difference could simply be attributed to sampling or chance and both the forms and functions of the KM and VYK endings are very similar, we consider these endings to be near equivalents.

5.2.4.4 Future/Modal

Although there are many linguistic means for referring to actions which are yet to have taken place in all varieties of Korean, here we restrict our discussion to "the closest thing Korean has to a future tense marker" (Yeon and Brown 2011 p. 203), the invariant pre-final ending –*keyss*-. As can be seen in the examples below, there is a corresponding pre-final ending in both VYK and KM which has a similar form and function:

KM2526	[nɛil-t͡ʃɛ		subota		ət'e		ha-ges'-o]
	tomorrow-ORD		Saturday		how		do-MOD-POL.INT
	What	will	you	do	tomorrow	on	Saturday?

KM1863	[na-n	i	t͡ʃib-esə	aj	ka-get'a]
	PRN.1.SG-TOP	this	house-ALL	NEG	go-MOD.NEUTR
	I	will	not	go	into	this	house.

I will not go into this house.

VYK2093	[ta	səlmjəŋ	mosˀ	ha-get͡ʃ'i]
	all	explanation	NEG	do-MOD.NEUTR

I will not be able to explain it all.

As with the past tense endings examined above, we note that the canonical form of this ending /keys/ is realised comparatively rarely, and instead we see the neutralisation of the final consonant and the tensification of following consonants.

The vowel of this ending in KM is identified in earlier research as [ɛ] (Pak 2005 p. 99) rather than [e], as it is in CSK or VYK. Instrumental examination of the primary data used for this project revealed that the F1 and F2 values of the vowel which occupies this position vary between the average values of those of /ey$_{KM}$/ (that is, [e]) and /ay$_{KM}$/ (i.e. [ɛ]) but there is no observable tendency that /ay$_{KM}$/ is replacing /ey$_{KM}$/ in this KM ending.

This ending is also sometimes called the modal ending, since in addition to futurity it may express a variety of modal meanings including connotations of obligation or tentativeness. This function, too, is common to both transplanted varieties and attested more frequently than the endings' use as a simple future ending in both varieties.

KM1322	[ət'əkˀ	hangap-ʃi	seu-ges'-o]
	how	sixtieth birthday-party	build-MOD-POL.INT

How should one prepare a sixtieth birthday celebration?

KM2151   [t͡ʃislo   moru-ges'-o]

          date   not know-MOD-POL

          I   do   not   know   the   date   (tentative).

VYK1917   [t͡ʃal   moru-ges'-ɨm]

          well   not know-MOD-POL

          I   do   not   know   (tentative).

VYK408   [saram-mada   tarɨ-get͡ʃ'i]

          person-each   different-MOD.NEUTR

          Everyone   is   different   (tentative).

Consequently, these pre-final endings are near equivalents.

5.2.4.5 Subject Honorific

Honorification may be expressed lexically as well as morphologically. Here we discuss only the

morphological marking of honorification, the lexical aspect of which is discussed in section

5.2.8.4. Verbal subject honorificaton is widely used in VYK in a manner analogous to CSK, i.e. it

is marked with the two-form pre-final verbal ending [ʃi]/[ɨʃi].

VYK313   [tarɨ-n   toŋnɛ-esə   sa-ʃi-n   t͡ʃək̚   is'-ɨmik'a]

          different-   neighbourhood-   live-HON-   experience   be present-
          ADN   LOC   ADN   PRS.INT.POL

          Have   you   ever   lived   in   a   different   neighbourhood?

VYK456　　　　[kɨrən　　　　　　　　　　gə-edɛhɛsə　　　　　　　mals'ɨm

　　　　　　　　That sort　　　　　　　　　　thing-about　　　　　　speech.HON

　　　　　　　　tə　　　　　　hɛ　　　　　　　　　　　　　t͡ʃu-ʃi-go]

　　　　　　　　more　　　　　do　　　　　　　　　　　　give-HON-DM

　　　　　　　　Could you please tell me a little more about that sort of thing?

VYK581　　　　[əməni-nɨn　　　　　jori-rɨl　　　　　t͡ʃal　　　ha-ʃi-mnida]

　　　　　　　　mother-TOP　　　　　cookery-OBJ　　　well　　　do-HON-PRS.DECL.POL

　　　　　　　　(My)　　　　　mother　　　is　　　　good　　　at　　　cooking.

VYK2749　　　[tul-t͡ʃ'ɛ　　　　　əni-ga　　　　　jəgi　　ke-ʃi-nɨnde]

　　　　　　　　two-ORD　　　　　elder sister-SBJ　　here　　be present.HON-HON-DM

　　　　　　　　My　　　second　　older　　sister　　is　　(also)　　here　　(in　　　Yanbian).

This ending is absent from earlier accounts of KM and also from the primary data. It is worth noting that in the primary it does not even appear in examples of honorific vocabulary common to both transplanted varieties.

VYK2748　　　[op'a-ga　　　　　　　　toraga-ʃi-go]

　　　　　　　　elder brother-SUBJ　　　　die.HON-HON-DECL.DM

　　　　　　　　My　　　elder　　brother　　has　　passed　　away

KM247　　　　[uri　　　　　　　nampʰjən　　　　　　　toraga-t͡ʃ'i]

　　　　　　　　PRN.1.PL　　　　　husband　　　　　　　die.HON-PST.DECL.NEUTR

　　　　　　　　My　　　husband　　　passed　　　away.

KM1166	[torgovlʲe	institut	pʰil-ha-go	toraga-o]
	trade	FE college	graduate-do-CNJ	die.HON-PRS.DECL.POL

| | They | graduated | om | busines | school | and | then | died. |

Thus, we are led to the conclusion that there is no equivalent of the VYK subject honorific pre-final verbal ending in KM.

Thus, we may conclude this section by saying that the tense, modality, and evidentiality pre-final endings of Yanbian Korean and *Koryo Mar* are near equivalents of one another and those pre-final endings in CSK. The subject honorification marking pre-final ending of Yanbian Korean, though, has no equivalent in *Koryo Mar*, although it does stand in a relationship of near equivalence with its CSK counterpart. We summarise these finding in tabular form below:

Pre-Final Ending	CSK Comparative Concept	Near Equivalents	Partial Equivalents	Non-equivalents
Simple Past	*-a/ess-*	✔		
Past- Past	*-a/essess-*	✔		
Observed Past	*-te-*	✔		
Future/Modal	-keyss-	✔		
Subject Honorific	*-(u)si-*			✔

Table 78: Summary of Contrasts Between VYK and KM Pre-Final Endings

5.2.5 Final Endings

This is a large and diverse set of endings which may be defined according to their morphotactic distribution and their functional characteristics, i.e. they express mood and the relationship between speaker and interlocutor (see Kim-Renaud 2009 p. 85-86). Conventionally, the final endings are further sub-divided into sentence final and non-sentence final endings. Non-

sentence final endings may themselves be sub-categorised along functional lines into

connective endings (*cepsok emi*) and the function converting endings (*censing emi*):

nominalizers (*myengsahwa emi*) and adnominalizers (*kwanhyengsahwa emi*) (Lee and Ramsey

2000 p. 174). Each of the categories identified above shall be contrasted on the basis of CSK

comparative concepts below.

5.2.5.1   Non-Sentence Final Endings

This class of endings covers the vast number of connective endings (over thirty in Lee and

Ramsey 2000 pp. 186-191 and over sixty in Yeon and Brown 2011 pp. 259-326) as well as the

small closed classes of adnominalizing and nominalizing endings. While all adnominalizing and

nominalizing endings can be dealt with individually, limitations of space prevent us from

discussing each connective ending in the same way. Rather, we classify connective endings

according to their broad semantic characteristics (after Yeon and Brown 2011 pp. 259-326) and

examine a representative selection of each. The full inventory of non-sentence final endings to

be examined and their forms in the transplanted varieties is given in the table below:

Non-sentence Final Ending	CSK Forms	Yanbian Korean Forms	*Koryo Mar* Forms
Addition/Sequence	*-ko*	[go], [ko], [k'o]	[go], [ko], [k'o]
	-myense	[mjənsə], [mesə]	[mjənsə], [mesə], [mesəri]
	-ca(maca)	[d͡ʒa mad͡ʒa]	[zamɛsa]
	-nuntey	[nɨnde], [ɨnde], [nde]	[nɨnde], [ɨnde], [nde]
	-(a/e)se	[sə], [asə]	[sə]
Cause	*-(u)nikka*	[nik'a]	[nik'a], [ãik'a], [naik'a], [kadi]
	-killay	[gilɛ]	[gilɛ]
Contrast	*-ciman*	[d͡ʒiman]	???[d͡ʒiman]???

	-(e)to	[do]	[do]
	-kena	[gəna], [k'əna]	[ɨgəna]
Option	*-tunci*	[dɨndʑi]	[dɨndʑi]
	-(u)n/nunci	[nɨndʑi]	[ɨndʑi], [ndʑi], [nɨnga]
Condition	*-(u)myen*	[mjən], [mu], [mə], [m]	[mjən], [mu], [mje], [m]
	-kedun	*-kedeng* (Jeong2010)	*-geden* (Pak 2005)
Causation	*-key*	*-key* (Jeong 2010)	[ge]
	-tolok	[dorok̚]	[torok̚]
Intention	*-le*	[rə]	[la]
	-lyeko	[rjəgo]	N/A
	-(u)n	[ɨn], [n]	[ɨn], [n]
Adnominalizers	*-ten*	[t'ən]	[dən]
	-nun	[nɨn]	[nɨn]
	-(u)l	[ɨl], [l]	[ɨl], [l]
Nominalizers	*-(u)m*	[ɨm], [m]	[ɨm], [m]
	-ki	[gi], [k'i]	[gi]

Table 79: VYK and KM Non-Sentence Final Endings

We divide the following analysis into three sections based on the classification of non-

sentence final endings discussed above: connective endings, adnominalizers and nominalizers.

5.2.5.2 Connective Endings

Connective endings are a large category of predicate final endings which may be divided into

coordinating and subordinating types (Lee and Ramsey 2000 p. 186). Predicates to which they

are attached appear in clause final position. The coordinating type of connective ending

expresses a range of meanings which includes the enumeration of a sequence, contrast, and

choice, whereas subordinating connective endings may indicate a far wider variety of

meanings including reason, concession, condition, and result among others. In the alternative

analysis of the endings of KM, several of Kwon's (2010) 'commonly used endings' fall into this class, for example -ketun.

Addition/Sequence

The CSK non-sentence final endings –ko is a simple connective which links clauses that express either simultaneous or sequential actions as below:

VYK202   [s'irɨm-do   ha-go   kɨrən   gə

                  wrestling-ADD   do-CNJ   that sort   Thing

                  mai   ha-d͡ʒio]

                  much   do-PRS.DECL.POL

                  I   wrestle   and   do   that sort   of thing   a   lot.

VYK925   [undoŋ   ha-go   ir   hɛ-jo]

                  exercise   do-CNJ   work   do-PRS.DECL.POL

                  I   do   exercise   and   work.

KM1413   [mək'o   t͡ʃʰum   t͡ʃʰu-go   nor-go

                  eat.CNJ   dance   dance-CNJ   play-CNJ

                  i   vsʲo   hangap̚   kɨrə-d͡ʒim]

                  and   all   sixtieth birthday   to be thus-DECL.PRS.NEUTR

                  We eat and dance and celebrate and all sixtieth birthdays are like that.

KM1111	[maɨm-i	kop-k'o	t͡ʃo-ɨn	jəd͡ʒa]
	heart-SBJ	beautiful-CNJ	good-ADN	Woman
	(She is) a	good	woman with a	good heart.

It may also appear in an utterance final position in order to "give the feeling that the utterance is adding additional information to what has already been said" (Yeon and Brown 2011 p. 287).

KM1411	[nawa-sə	t'o	pozdravi	ha-go]
	come out-CNJ	again	greet	do-DM
	(And then) they	come out	and greet	you again.

VYK2548	[kugjəŋ-do	t͡ʃom	ha-go]
	sightseeing-ADD	a little	do-DM
	I also did	a little	sightseeing.

Finally, for these connective endings in VYK and KM, we observe that they appear in several near identical construction in each transplanted variety. This is most clearly exemplified by the formation of the present continuous, which follows equivalent patterns to CSK's –ko issta, as below.

VYK46	[suksa-esə	sal-go	is'-ɨmda]
	dormitory-LOC	live-CNJ	be present-PRS.DECL.POL
	I am	(currently) living	in a dormitory.

KM659	[il	ãj	ha-go	i- d͡ʒim
	work	NEG	do-CNJ	be present-PRS.DECL.NEUTR
	I am	not	(currently)	working.

We note, though, that some constructions do not use the same connective endings in VYK and KM. For example, volition is expressed in CSK using the *hako siphta* construction as it is in VYK

VYK59   [uɾi   mal-do   pɛu-go   ʃipʰ-ə]

         PRN.1.PL   language-too learn-CNJ   want-PRS.DECL.NEUTR

         I   also   want   to   learn   about   our   language.

This is not the case in KM and this construction is examined specifically in section 5.2.7.3 below.

Although they are not perfectly identical, on account of their great similarity in forms and in function, the VYK and KM endings corresponding to CSK –*ko* are considered near equivalents.

We now turn to the CSK ending –*myense*, which expresses the simultaneity of actions or states. The equivalent endings are presented in examples below:

VYK2380   [hak'jo- ɾɨl   tani-ge   tɛ-mjənsə   jəŋgil-e   wa-s'o]

          school-OBJ   attend-ADV   become-CNJ   Yanji-ALL   come-
                                                                    PST.DECL.POL

          I   came   to   Yanji   when   I   began   attending   school.

VYK20   [maɾ   ha-mjesə   ədi-ɾɨ   gatʃʰi-do   ka-go]

         language   do-CNJ   where-OBJ   together-too   go-DM

         We   go   somewhere   or   other   together   while   talking.

KM1418   [korjə   mal-lo   ta   ha-mjənsə   tɨ

         Korean   language   all   do-CNJ   PRN.2.SG

	ət'əkʰe	sar-anɨnga]
	how	live-PRS.INT.NEUTR
	How could you live speaking only (all in) Korean?	

KM619	[va	isʼ-ɨmesəri	uri-rɨ
	come.CNJ	be present-CNJ	PRN.1.PL-OBJ

	jesu	mitʰ-ɨra	tɛŋgira-o]
	Jesus	trust-IMP.NEUTR	travel-PRS.DECL.POL
	While they were here they went around telling us to believe in Jesus.		

KM1116	[ir	ha-mjesə	kɨrəkʰe]
	work	do-CNJ	like that
	While doing work it is like that.		

Most strikingly, the KM ending is suffixed by the syllable [ri/ɾi], the origin of which is obscure (King 1992 p. 212). Due this difference in form, the VYK and KM endings are considered partial equivalents.

The CSK ending -ca(maca) denotes immediate succession of action. The VYK and KM endings which correspond to it are unambiguously attested only once each in the primary data. Despite being classified here as connective endings, neither ends the first clause of a two-clause sentence. They nevertheless appear to have similar functions and, while their attested forms are quite different, the fact that only a single attestation appears in the primary data combined with the likelihood that much of this divergence can be accounted for by unconditioned allophonic variation and their attestation in secondary sources are sufficient

reasons for not classifying these endings as any more distant from one another than as near

equivalents for now.

VYK195   [o-d͡ʒa mad͡ʒa]

                 come-DM

                 (It   happened)   as   soon   as   I   came.

KM153   [korjə   mar   ha-zamɛsa-di]

                 Korean   language   do-DM-RETR

                 I recall that it happened as soon as I spoke some Korean.

From the wide range of possible usages of the non-sentence final ending *–(nu)ndey*, for the

purposes of establishing its equivalence across the transplanted we focus on its function in

ending clauses which provide background information for a following clause. Examples of this

function drawn from the KM and VYK primary data may be seen below:

KM2272   [kəɾ-ə   nao-nde   tu   kilomʲetr   i-o]

                 walk-CNJ   come out-CNJ   two   kilometre   COP-PRS.DECL.POL

                 I   was   walking,   you   see,   and   it   is   two   kilometres.

KM807   [t͡ʃənd͡ʒa   ʃid͡ʒakʼ   ha-n

                 war   start   do-AND

                 nar   i-nde   na-t͡ʃʼi]

                 day   COP-CNJ   be born-PST.DECL.NEUTR

                 It   is   the   day   that   the   war   started   that   I   was   born.

KM1452 [kɨrən   gə   murəbo-d͡ʒi an-ɨnde

       that sort   thing   ask-CNJ   NEG-CNJ

       kɨrəm ʃiəməni-n mjə sar   sar-ak'o]

       anyway step-mother-TOP several year   live-PST.DM

       I didn't ask about that sort of thing, but my step-mother lived for several (more) years anyway.

VYK690 [tul   da   t͡ʃoaha-nɨnde kɨ   kaunde

       two   all   like-CNJ   that   among

       t͡ʃosənmun-e   huŋt͡ʃʰi-ga t͡ʃom tə   man-sɨmnida]

       Korean language and   interest a little more much-PRS.DECL.POL
       literature-LOC

       I like both, but between the two I am a little more interested in Korean language and literature.

VYK1931 [t͡ʃontʰoŋ   nori   i-nde   nɛ

       tradition   game   COP-CNJ   PRN.1.SNG

       jut͡ʃʰ nori-do   t͡ʃɛmi   it͡ʃ'i-o]

       Yuch game-ADD   fun   be present.DEC.PRS-POL

       It is a traditional game, you see, and I do enjoy *yuch nori*.

VYK44   [kiog-ɨ   na-d͡ʒi   an-ɨnde   ad͡ʒu

       memory-TOP   rise-CNJ   NEG-CNJ   Very

əri-l	t'ɛ	jəŋgil-ɨro	was'-ɨmda]
be young-ADN	time	Yanji-ALL	come.PST-DECL.POL

I don't remember exactly, but I was very young when we came to Yanji

This ending may also appear in utterance final position as a discourse marker indicating that there are inferences to be drawn from the content of the utterance (cf. Yeon and Brown 2011 p. 301).

KM2051 [madaŋ t͡ʃir ta hɛ-nɨnde]

      yard work all do.PST-DM

      I did all the yard work (so it was hard for me).

KM493 [taʃkʲent saram mar tsar ha-nɨnde]

      Tashkent person language well do-DM

      People from Tashkent talk well (so I expect you to talk well).

VYK1046 [na t͡ʃ'ɨm pa-nɨnde]

      PRN.1.SNG a little see.PST-DM

      I saw some of it (so I can tell you a little about it).

VYK1565 [tsoaha-nɨn imʃig-ɨnɨ man-ɨnde]

      like-ADN food-TOP be many-DM

      There are a lot of foods I like (so it is hard to talk about just one).

Given their similarities in form and function, these endings are considered near equivalents of one another in VYK and KM.

Cause

The main meaning of the CSK connective ending *–(a/e)se* is to join two clauses, the second of which is a direct result of the first. It attaches to the infinitive form of the verb, so its realisation varies depending on rules of vowel harmony. Examples of corresponding endings with this usage in KM and VYK may be found below:

VYK744   [mana-sə   kjəron   ha-go]

              meet-CNJ   marriage   do-DM

              We   met   each   other   and   then   we   got   married.

VYK1271   [toŋnɛ   ta   aɾ-asə   toŋnɛ   mind͡ʒoŋgug

              neigbourhood   all   know-CNJ   neighbourhood   State

              i-ɾago   hɛt'ən-d͡ʒi]

              COP-IS   do.PST.RET-NEUTR

              Everyone in the neighbourhood knew each other so we used to call it The
              Republic of Neighbourhood (lit. a neighbourhood state).

KM133   [pensije   nawa-sə   ir-ɨ   mos'   hɛt͡ʃ'i]

              pension   come out-CNJ   work-OBJ   NEG   do.PST.DECL.NEUTR

              I   could   not   work because   I   had   (already)   taken   my   pension.

KM1411   [daln-i: rodstvʲenik-i   nawa-sə

              distant-PL   relation-PL   come out-CNJ

<table>
<tr><td></td><td>t'o</td><td>pozdravi</td><td>ha-go]</td></tr>
<tr><td></td><td>again</td><td>greet</td><td>do-DM</td></tr>
</table>

Distant relations come out so that they can greet you again.

It is clear from the near identical forms and functions of these endings that they are near equivalents.

The form of the CSK ending *-(u)nikka* is conditioned by whether the verb stem ends with a vowel or a consonant. The VYK ending is almost identical to the CSK ending in form and function.

VYK696   [haŋnjən-i   nopʰad͡ʒi-nik'a   suk͡t͡ʃɛ-do

                year of study-SBJ   become higher-CNJ   homework-too

                ha-n   jogu-do   nopʰad͡ʒi-go]

                do-ADN   demand-too   become higher-DM

                Because my year of study has risen, so has the demand to do homework.

VYK615   [əri-nik'a   kɨ   t'ɛ-nɨn   əməni-ga

                be young-CNJ   that   Time-TOP   mother-SBJ

                tɛrə-da   t͡ʃu-ʃi-go]

                accompany-CNJ   give-HON-DM

                Because I was young my mother dropped me off in those days.

While the KM ending may appear as both identical in form and function, its form is considerably more variable then the VYK ending as the /n/ is frequently deleted under the influence of the following /i/.

KM1249	[kəgi	ədime	kãik'a	tɨsʲat͡ʃ-i-sto	mʲetr-ovɨ]
	there	wherever	go.CNJ	thousand-PL-hundred	metre-GEN.PL

Because I went wherever there (I travelled) hundreds of thousands of metres.

KM244	[a	nãik'a	ne	tebi	toravas'-o]
	child	be born-CNJ	PRN.1.SG.SBJ	again	return.DECL.PST-POL
	I	went	back	again	because	I	had	a	child.

I went back again because I had a child.

There is also some variation on the form of the verb to which equivalent endings may attach. While CSK–*(u)nikka* and the corresponding VYK ending attach to the infinitive form of the verb, an equivalent KM ending may also attach to the 'dictionary form' of the verb. We note also that there are differences in the form of this ending, i.e. an initial [na] and the possibility of dropping the final syllable [k'a], as in the examples below.

KM355	[maʃina	avaria	ha-da-naik'a	sansɛna-go]
	car	accident	do-CNJ-CNJ	die-DM
	He	died	because	he	got	into	a	car	accident.

He died because he got into a car accident.

KM980	[uɾi-nɨ	abəd͡ʒi	əpt'-anai
	PRN.1.PL-TOP	father	be absent-CNJ

	ingan	kosa	ta	hɛs'-o]
	person	worry	all	do.PST-DECL.POL
	Because	my	father wasn't	there	I had	all	sorts	of problems.

Because my father wasn't there I had all sorts of problems.

Furthermore, a wholly novel form of this ending [k'adi], which is held to have an equivalent meaning to CSK –*(u)nikka*, has also been attested in KM secondary materials[23], but not YK materials. It does indeed appear in the KM primary data while being absent from the VYK primary data, but it is dubious the extent to which its function can be considered wholly analogous with that of either the CSK or VYK ending since its attestation is so limited:

KM778   [mãi   t͡ʃug-əs'-o   koŋgi-rɨ   pak'ui-kadi]

         many   die-DECL.PST-POL   air-OBJ   change-CNJ

         Many (of them) died because of the change in the air.

Fort this reason the endings which correspond to CSK -*(u)nikka* are only partially equivalent.

In CSK the ending –*killay* is used to express cause or reason for an action taken by the speaker under their own volition. It appears somewhat rarely in the primary data of both transplanted varieties, especially in canonical connective positions. Where it does appear, though, there do not seem to be any restrictions on the sort or source of the cause being expressed in VYK or KM as there are CSK.

VYK1763   [parɨm-i   kɨrakʰe   dɛ-gilɛ   kɨrɛ-t͡ʃ'i]

         pronunciation-SBJ   thus   become-CNJ   to be thus-PST.DECL.NEUTR

         It was like that because my pronunciation changed in that way.

VYK1293   [kɨ   ɛgi-do   kɨrɛ-gilɛ

         that   child-too   to be thus-CNJ

---

[23] Strangely, this ending is absent from Kwon's inventory of KM connective endings (Kwon 2010 p. 107). This may be attributed to the limited sampling used to compile this description (a single speaker), since elsewhere this ending is identified as a feature particular to speakers of Myengchen-Kilcwu rather than Yukchin dialectoloical heritage (King 1992 p. 212).

	kɨ	gə		musəwas'e-do]
	that	thing		to be scared.PST.DECL-even

Even   though that   child   was scared, too, of that thing on account of that…

KM824	[t͡sazɨk-du		nɨk'e	nas'o		kɨrɛ-gilɛ]
	offspring-ADD		late.ADV	be born.PST-DECL.POL		to be thus-CNJ

It   is   also   because   I   had   my   children   late.

KM567	[kɨ		norma		ta-gilɛ
	that		ration		accept-CNJ

	il		ha-la			tɛŋgi-t͡ʃ'i]
	work		do-CNJ			go-PST.DECL.NEUTR

I   went   to   work   because   I   accepted   that   ration.

These endings are considered near equivalents.

Contrast

The CSK invariant ending –*ciman* marks "a strong opposition between two contrasting or contradictory states of affairs" (Yeon and Brown 2011 p. 273). This usage is evident only in the VYK primary data and the form of this ending is similar, too:

VYK284	[t͡sogɨm	katʰ-ɨn	jənbjən	d͡ʒu	i-d͡ʒiman
	a little	same-ADN	Yanbian	state	COP-but

<table>
<tr><td></td><td>t͡som</td><td>jak-ha-n</td><td>tar-ɨmda]</td></tr>
<tr><td></td><td>a little</td><td>weak-do-ADN</td><td>different-PRS.DECL.POL</td></tr>
</table>

It is somewhat like Yanbian State, but als o a little bit different.

VYK2441   [himdɨl-d͡ʒiman   kɨ-mankʰɨm   poram-do   ik'o]

difficult-but   that-as   reward-ADD   be present-DM

It is difficult, but it is rewarding to the same extent.

There is no corresponding ending used with processive or descriptive verbs in the KM primary data. Rather, something with a similar form and meaning is appended directly to nouns (perhaps with a zero copula).

KM2253   [isa   maud͡ʒɛ-d͡ʒiman-ed   isa   irɛ   ha-o]

doctor   Russian-but-despite   doctor   thus   do-PRS.DECL.POL

The doctor is Russian but (still) does this.

KM1198   [mama-d͡ʒiman]

mother-but

She is my mother but…

Thus, these endings are non-equivalent in the transplanted varieties.

Contrast in terms of concession, as in the English word 'even' or the phrase 'at least' may be expressed in CSK using the non-sentence final ending *–(a/e)to*. Endings with a similar form attach to the infinitive form of the verb in both KM and VYK, too, and stand in a relationship of near equivalence with one another.

KM1740   [nɛ   t͡ʃʰuwə-do   ta   ai   t͡ʃu-go]

          PRN.1.SG.SBJ   cold-CNJ   all   NEG   give-DM

          Even   though I   was   cold   they did not give me all   of   it.

KM1087   [kɨɾɛ-do   na-n   mar   ãi   ha-o]

          to be thus-   PRN.1.SG-   language   NEG   do-PRS.DECL.POL
          CNJ   TOP

          Nevertheless   I   do   not   speak.

VYK1878   [t͡ʃɛ-ga   mal   hɛ-do   aradɨr-ɨd͡ʒi]

          PRN.1.SG.SBJ   language   do-CNJ   understand-PRS.INT.NEUTR

          Do   you   understand   even   when   I   am   talking?

Option

The CSK ending *–kena* marks optionality in a way roughly equivalent to English 'or'. It attaches
to the verb stem and is invariant in form. This form and usage is found in VYK:

VYK592   [t͡ʃɛ-ga   t͡ʃal   mos͈   hɛk'əna

          PRN.1.SG.SBJ   well   NEG   do.CNJ

          kɨɾən   tan-t͡ʃəm   is'-ɨmjən]

          that sort   short-dot   be present-DM

          (Maybe that would happen) if I made a mistake or if there was some
          shortcoming.

It appears very infrequently, however, in the KM data, and in those cases it tends to be in sentence final position.

KM181   [kʰɨ-n   hjən-nim-i   kəŋe   is'-ɨgəna]

         big-ADN   elder brother-HON-   there   be present-DM
                            SBJ

         My   elder   brother   is   there   or   something.

We note that it may also appear in sentence final position in VYK where a list of options is given or similar indefinite meaning is to be inferred:

VYK2458   [ɨmag-ɨl   tɨk'əna   jəŋhwa   po-gəna]

           music-OBJ   listen.CNJ   film   see-DM

           I   listen   to   music   or   watch   a   film   or   something.

On the basis of the available evidence, it is not possible to establish the degree of equivalence of these endings.

The CSK ending *–dunci* has a near identical function to *–kena* and likewise appears with a similar form and usage in both VYK and KM:

VYK2314   [hwarjəŋ   ira-dɨnd͡ʒi   hunt͡ʃʰun   irən   saram]

           Hwaryeong   COP-CNJ   Hunchun   this kind   person

           The   kind   of   person from   Hwaryong   or   Hunchun.

KM565   [kolxoz   iri   pad-ɨnd͡ʒi   ãi   pad-ɨnd͡ʒi]

         collective farm   work.?SBJ   receive-CNJ   NEG   receive-CNJ

         Whether   you   received   work   from   the   collective   farm   or   not…

The apparent difference in the forms of the above examples may be accounted for by the final /t/ consonant of the verb stem therefore an underlying form /tunci/ may be inferred for both transplanted varieties and we may conclude that these endings are near equivalents.

The CSK ending *-nunci* appears most frequently in oblique questions with cognitive verbs (such as *alta* 'to know' or *molwuta* 'to not know') or verbs of telling and asking (such as *kaluchita* 'to tell' and *mutta* 'to ask'). Its form depends on whether it is appended to a processive or stative verb and on the tense of that verb.

The use and form of the corresponding ending in VYK are similar to CSK, as below:

VYK855   [ətəkʼ   hɛ-nɨnd͡ʒi]

           how   do.PST-CNJ

           (I   wonder)   what   was   I   to   do?

VYK2482   [pɛu-rɨl   a-ʃi-nɨnd͡ʒi   moru-gesʼ-əjo]

           actor-OBJ   know-HON-CNJ   not know-MOD-POL

           I   do   not   know   whether   you   would   know   that   actor.

VYK357   [nugi   sa-nɨnd͡ʒi-do   t͡ʃal   moru-go]

           who.SBJ   live-CNJ-even   well   not know-DM

           I   do   not   even   know   who   is   living   there.

This form appears only very rarely in the KM primary data, though, and only once in non-utterance final position, i.e. accompanied by a cognitive verb.

KM772   [ətekʰe   dɨrɨva-nd͡ʒi]

          how   enter.PST-CNJ

          How   did   (we   manage)   to   get   here?

KM2118	[t͡ʃo-ɨnd͡ʒi		mor-u]
	good-CNJ		not know-PRS.DECL.POL
	I do not know whether it is good.		

Far more frequently, oblique questions are asked using the very different form [nɨnga].

KM1842	[əti		ha-ge-nɨnga]
	how		do-MOD-CNJ
	(I wonder) what am I to do		

KM1361	[hangap̚	ətekʰe	se-nɨnga mor-o]
	sixtieth birthday	how	build-CNJ not know-PRS.DECL.POL
	Do you not know how to prepare for a sixtieth birthday party?		

KM2104	[ədime	i-nɨnga	kalxaz-esə	mor-u]
	where	be present-CNJ	collective farm-LOC	not know-PRS.DECL.POL
	Where it is on the collective farm I do not know.			

Due to the existence of an additional form in KM with a surface realisation which does not appear with the same usage in the VYK data, these endings are considered partial equivalents.

Condition

The forms of the endings which correspond to CSK –*myen* are highly variable in both VYK and KM. This variability comes about as a result of variably applied nasal deletion, glide deletion and umlaut. Examples of the full range of forms in each transplanted variety may be found below:

KM42   [kɨrə-mjən-ɨn   əʃi-dɨr   kəgi-sə

to be thus-if-TOP   parent-PL   there-LOC

ros'ija   ta   tɛjənas'-ɨmdu]

Russia   all   be born.PST-DECL.HON

If that is the case were both of your parents born there in Russia?

KM718   [nov-ɨj   god-ɨr   ha-mu   t͡ʃɛbi   t͡ʃib-esə

new-SG.M   year.SG.M-OBJ   do-if   PRN.REFL   house-LOC

nosa   sər-ɨ   se-n   gə   t͡ʃɨrəm]

Russia   New Year-OBJ   build-ADN   ting   Like

If you do something for New Year it is like preparing for Russian New Year in your own house.

KM777   [ai   tɨrɨo-mje   mãi   t͡ʃug-əs'-o]

child   enter-if   many   die-PST.DECL-
POL

If   children   went   there,   many   of   them   died.

KM94   [dʲeŋgi   əps-ɨm   ed͡ʒi-ge-nɨnga-o]

money   be absent-if   how-MOD-INT-POL

If   one   does not   have   any   money,   what   is   one   to   do?

VYK533   [ʃiganpʰjo   po-ʃi-mjən   wəljoir-e   suəb-i

timetable   see-HON-if   Monday-LOC   class-SBJ

	du	ge	ik'o]
	two	CLASS	be present-DM

If you look at the timetable you will see that I have two classes on Monday.

VYK1610   [pukərəwəha-mə   mo   pɛu-n-da]

to be shy-if   NEG   learn-PRS-
                                     DECL.NEUTR

You	cannot	learn	if	you	are	shy.

VYK1540   [mjəŋd͡ʒəl   i-mu   gat͡ʃʰi   mana-sə   nol-go]

festival   COP-if   together   meet-CNJ   play-DM

If	it	is	a	festival	we	get	together	and	celebrate.

VYK848   [t͡ʃagi   mal   tə   ha-m   t͡ʃo-t͡ʃʰi]

PRN.REFL   lanuage   more   do-if   good-PRS.DECL.NEUTR

If one	speaks	one's	own language	more,	it	is	good.

Despite varying frequency of the application of the various phonological processes identified

above across the transplanted varieties, a similar range of forms was attested in the speech of

consultants from both Central Asia and the YKAP. We do note, though, that there was no

attestation of vowel raising in the VYK ending. Nevertheless, these endings are classified as

near equivalents.

The CSK ending –*kedun* is similar in function to CSK –*myen*, but appears in slightly fewer

environments, specifically, it joins clauses of which the second is "usually a command,

proposition or promise…or conjecture" (Yeon and Brown 2011 p. 316). A similar form appears

as a final ending with a significantly different meaning, marking explanation as a sentence final ending. We must rely on secondary sources for the provisional establishment of equivalence of transplanted forms corresponding to CSK *–kedun*. The endings attested in secondary sources (*-keteng* (Jeong 2010 p.132) and *–geden* (Pak 2005 p. 105)) are near identical in form and function, therefore we provisionally classify them as near equivalents.

Causation

There are two frequently used connective endings in CSK which denote a causal relationship between clauses: *-key* and *–tolok*. While the ending [ge] frequently appears in the VYK primary data, no examples of it being combined with a processive verb with causative meaning may be found, in contrast to the many instances where it is combined with stative verbs to give an adverbial meaning or appears with a processive verb as part of a construction analogous to CSK *–key toyta* 'to come to do something'.

A corresponding ending with a similar function does appear, although only once, in the KM primary data:

| KM2041 | [irɛ | ai | ha-ge | toŋmi |
| | thus | NEG | do-CNJ | friend-?GEN |

	adur	sul		məg-əs͈]
	son	alcolhol		eat-PST.DECL.POL
	My	friend's	son	drank	in	order	to	avoid	doing	that.

Due to a lack of comparable data it is not possible to say whether these endings are equivalent in the transplanted varieties.

The endings corresponding to CSK –*tolok*, albeit on the basis of very limited attestation, may

be concluded to be non-equivalent due to their very different functions. The VYK ending has a

similar form to the CSK and similarly denotes causation, somewhat like English 'in order to'.

VYK648   [t͡ʃal   t͡ʃipt͡ʃuŋ   ha-doɾokʼ

                well   attention   do-CNJ

                kjojukʼ   hɛ   d͡ʒu-ʃi-go]

                education   do.CNJ   give-HON-DM

                (My   parents)   taught   me   to   pay   attention   well.

While a similar surface form does appear in the KM data, its function appears to be very

different from the VYK ending with its meaning seeming more similar to CSK –*teylo* 'in

accordance with'. On the basis of this data, these endings must be considered non-equivalent.

KM97   [abud͡ʒi   skaza-l   toɾokʼ]

                father   say.PRF-PST.M   just as

                (I   did)   just   as   my   father   said.

Intention

Forms corresponding to CSK ending of processive verbs –*(u)le* appear in the primary data for

both VYK and KM. As with CSK, it links clauses where an intention is stated in the first and

then where the speaker went to carry out their intended activity is explained in the second. In

contrast to VYK, which exhibits a surface form very similar to CSK, in KM the vowel seems to be

realised exclusively as /a$_{KM}$/ rather than /e$_{KM}$/:

VYK318   [undoŋ   ha-rə-do   tani-go]

           exercise   do-CNJ-even   go-DM

           I   even   go   to   do   exercise.

VYK2426   [əma   apa-gat͡ʃʰi   ka-gəna   hogɨn

           mother   father-COM   go-CNJ   or

           t͡ʃib-esə   ʃy-gəna]
           house-LOC   relax-DM

           I   go   (and   spend   time)   with   my   parents   or   relax   at   home.

KM239   [ir   ha-la   noŋt͡ʃʰən   ponɛ-t͡ʃ'i]

           work   do-CNJ   agricultural village   send-PST.DECL.NEUTR

           They   sent   me   to   do   work   in   an   agricultural   village.

KM2427   [ja   nola   o]

           hey   play.CNJ   come.PRS.IMP.NEUTR

           Hey!   Come   and   play!

While the form of the KM ending appears to be very similar to CSK –l la, its function is quite

different in that it does not connect clauses which refer to a worry about a future event and

the steps to be taken in preparation for it. Rather the functions of these endings in the

transplanted varieties are near identical so they may be considered partial equivalents despite

their divergent forms.

The basic usage of the CSK ending –(u)lye(ko) is similar to that of the endings corresponding to

–(u)le, but the second clause (the clause in which the action being taken to achieve the

intention stated in the first clause) may contain any processive verb rather than only a verb of motion. The VYK ending is almost identical in form and function to the CSK ending:

VYK2544   [ʃint͡ʃʰən   ha-rjəgo   kehøk   ha-go   is'-ɨmnida]

                 application   do-CNJ   plan   do-CNJ   be present-PRS.DECL.POL

                 I   am   currently   planning   making   an   application.

An equivalent ending does not appear in the KM primary data, therefore it is not possible to establish whether or not there are equivalents in the transplanted varieties or what degree of equivalence such endings might be.

5.2.5.3 Adnominalizing Endings

These endings allow predicates or VPs to modify NPs similarly to relative clauses. The adnominalizing endings of the transplanted varieties are near equivalents of each other since they demonstrate considerable overlap of form and function. We examine the forms of the state/result modifier, retrospective modifier, dynamic modifier and future/prospective modifier[24] in turn below and discuss any particular (dis)similarity in their functions.

State/result modifier

In CSK this modifier takes the shape *–n* following verb bases ending with vowels and the form *–un* following verb bases ending with consonants. Depending on whether it is attached to descriptive or processive verbs, it either denotes a present or on-going state or an action which has taken place in the past, respectively (cf. Yeon and Brown 2011 p. 332).

---

[24] In characterising these forms in contrast to other final endings, we use the earlier term 'adnominal(izing ending)' and 'modifier' interchangeably, since the function described by each term is near identical. For differentiating between the modifiers, though, we follow the terminology of Yeon and Brown (2011 pp. 328-361), as presented above strictly, since it characterises the forms of the modifiers more completely than the earlier tense based terminology which referred to these endings as the past, retrospective, present and future adnominals, respectively.

KM1236	[t͡ʃo-ɨn		ir-dɨr		ha-o]
	good-ADN		work-PL		do-PRS.DECL.POL
	I	do	good		work(s).

KM793	[t͡ʃənd͡ʒa		ʃid͡ʒɛŋ-ha-n	nar-e	nas'-o]
	war		start-do-ADN	day-LOC	be born.PST.DECL-POL
	I	was born	the day	that	the war started.

YK2590	[kʰɨ-n		ohɛ	gatʰ-ɨn	gə-n]
	big-ADN		misunderstanding	be similar-ADN	thing-TOP
	Something	like	a	big	misunderstanding

YK256	[tɛəna-n		go-n	jəŋd͡ʒəŋ	i-nde]
	be born-ADN		place-TOP	Yeongjeong	COP-DM
	The place	that I	was born	is Yeongjeong,	you see.

Furthermore, this ending may be used in many similar constructions with bound nouns in both VYK and KM:

CSK Construction	KM	English	VYK	English
-n hwu (ey)	KM928[on hu]	after arriving	VYK1020[t͡ʃoləp han hu e]	after graduating
-n kes	KM176[kɨrən gə]	that sort of thing	VYK159[kɨrən gə]	that sort of thing

Table 80: Shared Constructions Using Adnominalizing Ending -*(u)n* in VYK and KM

These endings are thus considered near equivalents.

Retrospective modifier

This CSK ending is invariant in form and may appear attached to both processive and descriptive verbs. Its marks continuous but completed past actions or states which had some duration in the past no longer continue. Despite the fact that there are few attestations in the primary data, the examples below demonstrate that these functions are performed by formally similar endings in the transplanted varieties:

VYK548   [je͡dʒən-butʰə   is'-ət'ən   ʃiktaŋ   i-go]

                before-from   be present-PST.ADN   restaurant   COP-DM

                There is a restaurant that used to be there from a long time ago.

VYK2844   [t͡sak'u   tɨr-ət'ən   mar   i-mda]

                frequently   hear-PST.ADN   language   COP-PRS.DECL.POL

                That is the sort of language that you used to hear all the time.

KM376   [hak'o-sə   il   ha-dən   saɾam]

                school-LOC   work   do-ADN   Person

                People who used to work in schools

These modifier endings are also near equivalents.

Dynamic Modifier

In CSK this modifier may only appear attached to processive verbs. Where it is taken as an ending it marks on-going or habitual actions.

KM448   [t͡ʃiʃik   i-nɨn   saɾam-do   t͡ʃug-es'-kuma]

                knowledge   be present-ADN   person-ADD   die-CAUS.PST-
                                                                        POL

                People with knowledge (the intelligentsia) were killed, too.

KM616   [jesu-l   mid-nɨn   saram-dɨl   iŋgi   vas'-o]

Jesus-OBJ   believe-ADN   person-PL   here   come.PST.DECL-POL

People   who   believe   in   Jesus   came   here.

KM1363   [hangap'   se-nɨn   ge   kaɾu-t͡ʃimu]

sixtieth birthday   celebrate-ADN   thing   teach-PRS.DECL.NEUTR

I will teach you how to go about the celebration of a sixtieth birthday party.

YK1667   [jəgi   mund͡ʒi   əm-nɨn   go-nɨ   ani-go]

here   problem.SBJ   be absent-ADN   place-TOP   NEG-DM

Here   is   not   a   place   without   (its)   problems

VYK277   [ɾjəŋd͡ʒoŋ-esə   sa-nɨn   gə-e   pi   ha-mjən]

Ryeongjeong   live-ADN   thing-LOC   compare   do-if

You   mean   in   comparison   to   living   in   Ryeongjeong?

VYK1268   [a- nɨn   saram-i-mu   insa   mal   ha-go]

know-ADN   person-COP-   greeting   language   do-DM
                  if

If   it   is   someone   you   know   you   would   greet   them.

The formally and functionally similar endings drawn from the VYK and KM primary data

demonstrate that these modifiers are near equivalents.

Future/prospective modifier

Like the state/result modifier discussed above, this ending has two forms depending on

whether the base of the verb to which it attaches ends with a vowel or consonant: -*l* or –*ul*,

respectively. In its basic usage it may only be attached to processive verbs in situations where it marks events that are yet to take place.

KM280   [ədʒən   t͡ʃib-esə   ha-l   il-i   əps-ɨni]

          house-LOC   do-ADN   work-SBJ   be absent-CNJ

          Anyway,   there   is   no work   to   be   done in   the   house.

VYK294   [ha-l   il   ha-d͡ʒi   an-ɨmjən]

          do-ADN   word   do-CNJ   NEG-if

          If   you   do   not   do   the   work   that   is   to   be   done…

As with the other modifiers its use with specific bound nouns is common to both transplanted varieties to express specific derived meanings:

VYK1480   [mwunhwa-hɛŋmi   is'-ɨl   t'ɛ   t'o

          culture-revolution   be present-ADN time   again

          noŋt͡ʃon-e   gas'-o]

          village-ALL   go.PST.DECL-HON

          At the time of the Cultural Revolution I went to the villages again.

KM844   [irgop-sar   tɛs'-ɨl   t'ɛ

          seven-years of age become.PST-ADN   Time

          hak'o-rɨ   mo   kas'-əjo]

          school-OBJ   NEG   go.PST-DECL.POL

          When   I   turned   seven   I   was   not   able   to   go   to   school.

Also, notably, certain common uses of the prospective modifier are also absent from the primary data, for example a construction corresponding to the periphrastic future of CSK *–l kes ita*.

While there is some variation in the realisation of the final liquid of this modifier in KM, it is considered nearly equivalent to the VYK modifier.

5.2.5.4 Nominalizing Endings

There are two nominalizing endings in CSK, both of which have corresponding forms in Yanbian Korean and *Koryo Mar*. Their forms are near identical to those of CSK, ie. *–(u)m* and *-ki*, and neither ending is a truly productive source of free nouns in either transplanted variety.

Reference	KM/VYK	Translation
KM136	[salgi]	(the act of) living/life
KM1908	[pʰalgi]	(the act of) selling/sale
KM384	[kɨrim kɨrigi]	(the act of) drawing pictures
VYK359	[nolgi]	(the act of) playing
VYK832	[s'ɨgi]	(the act of) writing
VYK1294	[sul mək'i]	(the act of) drinking alcohol

Table 81: Nominalizer *-gi* in KM and VYK

Certain constructions in each transplanted variety require the use of particular nominalizing endings with the KM nominalizer corresponding to CSK *-ki* being used in a wider variety of constructions in *Koryo Mar* than VYK or CSK. This feature is discussed in greater detail in section 5.2.7.3, which is devoted to constructions which may use endings corresponding to CSK *-ki*.

The endings corresponding to CSK *–(um)* in the transplanted varieties are barely productive, but there are nevertheless a large number of predicates which have a final [m]. This may be attributed to either the reduction of a conditional construction (see 5.2.5.2 above) or a

particular polite verbal ending (see 5.2.5.5 below). Despite this, nominalized forms ending

with [m] may be found in both VYK and KM, although they appear largely in lexicalised forms,

as below:

Reference	VYK/KM	Translation
VYK468	[toum]	help
VYK1293	[s'aɨm]	a fight
VYK2133	[əljəum]	difficulties
KM506	[toum]	help
KM2245	[s'aɨm]	a fight
KM1413	[t͡ʃum]	a dance

Table 82: Nominalizer *-gi* in VYK and KM

We conclude by noting, then, that the endings corresponding to CSK *–(u)m* may be considered

near equivalents which appear primarily in fossilised vocabulary in both varieties. The endings

corresponding to CSK *–ki*, on the other hand, are rather only partial equivalents of one another

despite their formal similarity. The functional reasons for this are touched upon in section

below 5.3.3, but now we turn to an examination of the sentence final verb endings of the

transplanted varieties.

This section on non-sentence final endings is brought to a close with a table summarising the

equivalences that it established:

Non-sentence Final Ending	CSK Forms	Near Equivalent	Partial Equivalent	Non-equivalent
Addition/Sequence	*-ko*	✔		
	-myense		✔	
	-ca(maca)	✔		
	-nuntey	✔		

	-(a/e)se	✔		
Cause	*-(u)nikka*		✔	
	-killay	✔		
Contrast	*-ciman*			✔
	-(e)to	✔		
	-kena			
Option	*-tunci*	✔		
	-(u)n/nunci		✔	
Condition	*-(u)myen*	✔		
	-kedun	(✔)		
Causation	*-key*	N/A	N/A	N/A
	-tolok			✔
Intention	*-le*		✔	
	-lyeko	N/A	N/A	N/A
	-(u)n	✔		
	-ten	✔		
Adnominalizers	*-nun*	✔		
	-(u)l	✔		
	-(u)m	✔		
Nominalizers	*-ki*		✔	

Table 83: Summary of Contrasts Between VYK and KM Non-Sentence Final Endings

5.2.5.5   Sentence Final Endings

These endings mark not only the end of a grammatical sentence, but also the mood and level

of politeness (directed toward the interlocutor) of the utterance. The moods distinguished in

the sentence final endings of CSK are declarative, interrogative, imperative, and propositive

moods, each of which is traditionally divided into six distinct levels of formality (Yeon and

Brown p. 171). In practice, though, not all of these are in regular use in any variety of Korean.

Prior research on VYK and KM has identified a variable number of speech levels with a broad consensus of around three: informal, neutral and formal registers. Despite this discrepancy, we continue to turn to CSK for comparative concepts but, informed by earlier work on VYK and KM, select only endings corresponding to three rather than all six levels of politeness as the basis for comparison.

Below we present an inventory of the sentence final endings attested in the KM and VYK primary data over these three speech levels for the declarative, interrogative, imperative and propositive moods. We then go on to present examples drawn from the primary data and compare the endings of KM and VYK directly according to their corresponding CSK endings in turn before summarising the strength of the equivalences between the sentence final endings of each transplanted variety.

Register	CSK	KM	VYK
Informal	*-e/a*	[nda], [nɨnda], [ə], [ta]	[nda], [ə], [ɛ], [ta]
Neutral	*-a/eyo*	[ajo], [əjo], [o], [so]	[ajo], [əjo], [o], [so]
Polite	*-(su)pnita*	[kuma], [sɨk'uma], [mnida], [sɨmnida]	[mda], [sɨmda], [mnida], [sɨmnida], [m], [ɨm], [sɨm]

Table 84: Declarative Sentence Final Endings of KM and VYK

Register	CSK	KM	VYK
Informal	*-ni, -nya*	[nɨŋga], [nja]	[nja], [ni]
Neutral	*-a/eyo*	*-(s)o* (Kwon 2010) *-(s)o, -wu* (Pak 2005)	*-o, -so* (Jeong 2010)
Polite	*-(su)pnikka*	[du], [sɨmdu]	[mk'a], [sɨmk'a], [mnik'a], [sɨmnik'a]

Table 85: Interrogative Sentence Final Endings of KM and VYK

Register	CSK	KM	VYK
Informal	*-a/ela*	[ra]	*-a/ela* (Jeong 2010)
Neutral	*-a/eyo*	*-(s)o* (Kwon 2010)	*-(s)o* (Jeong 2010)
		-(s)o, -wu (Pak 2005)	
Polite	*-sipsio*	[pʃio]	[ʃio]

Table 86: Imperative Sentence Final Endings of KM and VYK

Register	CSK	KM	VYK
Informal	*-ca*	[d͡ʒa]	*-ca* (Jeong 2010)
Neutral	*-a/eyo*	*-gio* (Pak 2005)	*-(s)ukkwey, -kio* (Jeong 2010)
		-geyo (Kwon 2010)	
Polite	*-psida*	*kipso* (Kwon 2010)	[pʃida]
		gipsho (Pak 2005)	

Table 87: Propositive Sentence Final Endings of KM and VYK

Due to the extensive verbal paradigms of Korean, the CSK counterparts of the KM and VYK sentence final endings presented above diverge in surface form a good deal more than the CSK forms presented for other categories of morphemes above. We emphasise at this point, that these CSK examples simply provide illustrations of the underlying comparative concepts which we use to select the forms of KM and VYK to be compared with each other. The establishment of equivalences rests on the formal and functional similarity between the morphemes of the transplanted varieties rather than their relationship with CSK.

It is also necessary to note, before beginning the contrastive analysis proper, that a relatively large number of forms of final endings are not attested in the primary data, most especially forms of the imperative and propositive moods. The forms provided in italics in the tables above are drawn from the existing descriptions of the transplanted varieties. While we do not

present examples of these forms in use, we do discuss their provisional degrees of equivalence in the CA below.

First we turn out attention to the declarative endings, i.e. those which correspond to CSK –*e/a*, -*a/eyo* and –*(su)pnita*. The informal endings of both varieties are similar. In addition to *panmal* forms ending with equivalents of –*e/a*, endings which take the forms corresponding to –*((nu)n)ta* are also found being used to convey a similar level of informality (although with slightly greater frequency in KM, especially in non-present tenses). In the present tense in both transplanted varieties, equivalent distinctions in the conjugation of stative and processive verb when combined with 'plain style' informal endings may be observed (i.e. processive verbs also take present tense marking endings).   Examples of this may be found below.

KM1171/1172	[mo	ka-n-da	…	mo	o-nɨn-da]
	NEG	go-PRS-DECL.INF		NEG	come-PRS-DECL.INF
	I	cannot	go …	I	cannot	come.

KM1465	[kɨ		mam		t͡ʃo-tʰa]
	that		heart		good-PRS.DECL.INF
	That	person	has	a	good	heart/is	kind.

VYK67	[geim-ɨn		an		ha-n-da]
	(computer)game-TOP		NEG		do-PRS-DECL.INF
	I	do	not	play	computer	games.

VYK2041	[uri		kat'a]
	PRN.1.PL		be alike-PRS.DECL.INF
	(They | are)	the	same	as	us.

Below we present further examples of endings corresponding to CSK *panmal* and plain style endings in other tenses in VYK and KM.

VYK766   [koŋd͡ʒaŋ   tɨrəwas'-ə]

            factory   enter.PST.DECL-INF

            I   started   work   at   the   factory.

VYK953   [sərə   irəkʰe   mal   hɛ]

            each other   like this   language   do.PRS.DECL.INF

            They   speak   to   each   other   like   this.

KM159   [t͡ʃʰilʃip'al-jən-do-e   enter.PST.DECL-INF   imi]

            seventy eight-year-degree-   enter.PST.DECL-INF   Already
            LOC

            I   came   (here)   in   (nineteen)   seventy   eight   already.

KM2413   [papa   d͡ʒug- ət'a]

            father   die-PST.DECL.INF

            My   father   died.

KM1177   [ãi   ka-ge-t'a]

            NEG   go-MOD-DECL.INF

            I   will   not   go.

These declarative informal endings of the transplanted varieties are considered near equivalents of one another.

The same degree of equivalence may also be established for the polite endings of the

transplanted varieties. Forms which have realisations very similar to the canonical CSK -a/eyo

appear in the primary data for VYK and KM as do endings with the form [o]/[so]. The latter

appears more frequently than the former in KM data. Examples of each appear below:

YK864 [jəŋgil-poda d͡ʒag-ajo]

Yanji-than small-DECL.PRS.POL

It is smaller than Yanji.

YK889 [ərin ʃid͡ʒər-i kibonsan hɛŋbok-hɛs'-əjo]

childhood era-SBJ fundamentally happiness-do.PST-DECL.POL

My childhood was basically happy.

YK1798 [kɨmjoir-e kɛnt͡ʃʰan-ajo]

Friday-LOC be fine-PRS.DECL.POL

It (my timetable) is fine on Fridays.

YK2482 [song-rjəŋ ira-nɨn pɛu-rɨl

Song Ryeong COP-ADN actor-OBJ

a-ʃi-nɨn-d͡ʒi moru-ges'- əjo]

know-HON-PRS-CNJ not know-MOD-POL

I don't know if you know an actor called Song Ryeong.

KM306 [patʰ-esə ir hɛs'-əjo]

field-LOC work do.PST-DECL.POL

I worked in the fields.

KM1435   [ənu   kolxoz-ɨ   tɛjənas'-əjo]

which   collective farm-OBJ   be born.PST-DECL.POL

I   was   born   on   some   collective   farm   (or   other).

KM1123   [t͡ʃɛmi   ik'e   ponɛ-jo]

enjoyment   be present.ADV   spend time.PRS.DECL-POL

I   spend   (my   time)   enjoyably.

YK2004   [kinjəmpʰum-do   man-so]

souvenir-too   many-PRS.DECL.POL

There   are   a   lot   of   souvenirs,   too.

YK251   [ta   jəŋgil-e   is'-o]

all   Yanji-LOC   be present-PRS.DECL.POL

They   are   all   in   Yanji.

KM699   [kand͡ʒɛ   t͡ʃabɨlo   ha-o]

soy sauce   oneself   do-PRS.DECL.POL

I   make   my   own   soy   sauce.

KM21220   [tul-t͡ʃ'e   orɛbi   samarkand   is'-o]

two-ORD   elder brother   Samarkand   be present-PRS.DECL.POL

My   second   older   brother   is   in   Samarkand.

KM106	[nɛ	sa-ʃim-njən-do-e	nes'-o]
	PRN.1.SNG.SBJ	four-ten-year-extent-LOC	be born.PST.DECL-POL
	I	was born in	1940

KM979	[kɨredo	ir-əps-əs'-o]
	nevertheless	work-be absent-PST.DECL-POL
	Nevertheless,	everything was OK.

A range of endings beginning with the syllable [t͡ʃi] also appeared to be widely used in both transplanted varieties in similar contexts to the other polite endings corresponding to CSK –*a/eyo*. Despite their usage, in earlier works on KM these endings have not been considered a part of the paradigm of sentence final declarative endings. While these are identified as characteristic and innovative KM ending by King (1992 p. 211) they are nevertheless held apart from declarative endings. They are not included in Pak's (2005) inventory of sentence final endings and Kwon classified them as part of a group of endings which he terms *pemyongemi* (widely used endings) in contrast to the 'declarative endings' (*seswulemi*) (Kwon 2010 p. 53). Furthermore, such endings are also absent from both Jeong (2010) and Kim's (2013) descriptions of the inflectional endings of VYK. For this reason we do not consider them here, but we do note that they appear to be with far greater frequency and in a far wider variety of contexts than the near homophonous CSK ending –*ci*. The usage of this ending is considered further in the following quantitative analysis chapter.

Turning to the formal declarative endings (those corresponding to CSK -*(su)pnita*), though, we see a very different picture. These endings may be considered only partial equivalents. The overlapping forms of the transplanted varieties are presented below:

YK141   [t͡ʃe   səŋgjək-e-nɨ   an   ma-sɨmnida]

PRN.1.SG.HON   personality-LOC-TOP   NEG   correct-PRS.DECL.POL

It   does   not   suit   my   personality.

YK581   [əməni-nɨn   jəri-rɨl   t͡ʃal   ha-ʃi-mnida]

mother-TOP   cooking-OBJ   well   do-HON-PRS.DECL.POL

(My)   mother   is   very   good   at   cooking.

YK1579   [jənbjən-esə   ɨmʃik-ɨn   t͡ʃogɨm   tarɨ-mnida]

Yanbian-LOC   food-TOP   a little   different- PRS.DECL.POL

The   food   is   a   bit   different   in   Yanbian.

KM856   [ʃid͡ʒan   t͡ʃom   miã-ha-mnida]

now   a little   sorry-do-PRS.DECL.POL

I   am   a   little   sorry   now.

KM856   [kraslodar   pʲedinstitut   t͡ʃoləp-hɛs'-ɨmnida]

Kraslodar   Teachers' college   graduate-do.PST-DECL.POL

I   graduated   from   Kraslodar   Teacher's   College.

We note that while this form does appear in the KM primary data, it is very rare and appears exclusively in in the speech of consultants who have received some instruction in CSK. In addition, this form appears to have undergone phonological reduction in VYK so that it is most commonly pronounced as in the examples below:

VYK1   [jəŋgil-esə   t͡ɛəna-s'ɨmda]

Yanji-LOC   be born-PST.DECL.POL

I   was   born   in   Yanji.

YK1874   [t͡ʃigɨm   mani   sələŋ-hɛs'-ɨmda]

now   very much   chilly-do.PST-DECL.POL

I   have   been   very   sad   these   days.

YK2248   [nɨl   kəgi-sə   sar-as'-ɨmda]

always   there-LOC   live-PST-DECL.POL

I   have   always   lived   there.

YK1018   [ol-hɛ   t͡ʃorəp   i-mda]

this-year   graduation   COP-PRS.DECL.POL

This   year   is   (my)   graduation.

YK483   [sərə   da   sai   t͡ʃo-kʰe   t͡ʃinɛ-mda]

each other   all   relations   good-ADV   get along-PRS.DECL.POL

We   all   get   along   well   with   each   other.

YK1634   [po-mu   sərə   insa-do

see-if   each other   greeting-even

ha-go   sərə   t͡ʃo-sɨmda]

do-and   each other   good-PRS.DECL.POL

If we saw each other we would greet each other and were polite to each other.

YK2093　　　[ta　　　　　　səlmjəŋ　　　　moɕ'　　　　　ha-gɛs'-ɨmda]

　　　　　　　all　　　　　　explain　　　　NEG　　　　　do-MOD-POL

　　　　　　　I　　　will　　not　　be　　able　　to　　explain　　it.

Less commonly even more extensive phonological reduction of this form may be observed in the YKAP resulting in the forms [sɨm] and [ɨm].　The use of these forms by VYK speakers does not seem to be functionally distinct from other polite declarative endings and they are used interchangeably with them:

YK16　　　　[tɛhag-ɨ　　　　əm-nɨn-səri　　　　isa-was'-ɨm]

　　　　　　　university-TOP　　　not present-AND-CNJ　　move house-come.PST-DECL.POL

　　　　　　　We　　moved　　because　　there　　wasn't　　a　　university　　there.

YK302　　　[t͡ʃə-rɨl　　　　uk'i-nɨn　　　ge　　　　tə　　　t͡ʃo-sɨm]

　　　　　　　PRN.1.SG-　　funny-ADN　　thing.SBJ　　more　　good-PRS.DECL.POL
　　　　　　　OBJ

　　　　　　　I　　　prefer　　funny　　　things.

YK1400　　　[i　　　　　　　　pʰori　　　　　jəgi　　　　apʰa- sɨm]

　　　　　　　this　　　　　　leg　　　　　here　　　　hurt-PRS.DECL.POL

　　　　　　　This　　leg　　here　　hurts.

YK1039　　　[kɨrən　　　　　　ge　　　　man-t͡ʃi　　　　an-sɨm]

　　　　　　　that sort　　　　thing　　many-CNJ　　　NEG-DECL.PRS.POL

　　　　　　　There　　aren't　　many　　of　　those　　sorts　　of　　things

YK1185	[k'ok	pas'-ɨmə̃	t͡ʃo-ges'-ɨm]
	totally	see.PST-if	good-MOD-POL
	If you can see it,	it would be good.	

YK2251	[t͡ʃuŋhak'jo	go-t͡ʃuŋ	ta	kəgi-se	sara-s'ɨm]
	middle school	high-throughout	all	there-LOC	live-DECL.PST.POL
	I lived there all through middle school and high school.				

A further area of divergence is the strength of the influence of the Yukchin variety, traditionally spoken in a small area in the extreme north east of the North eastern dialect zone. These endings have been considered characteristic of both VYK and KM, but they are very infrequently occurring in natural speech and only attested in the KM primary data. This style of formal declarative ending is presented below:

KM521	[t͡ʃɛmi	is'-ɨk'uma	səro-səro]
	enjoyment	be present-PRS.DECL.POL	each other
	You (plainly) enjoy each other's company.		

KM1629	[amu	gə-do	əps-ɨk'uma]
	any	thing-even	be absent-PRS.DECL.POL
	There is not anything there at all.		

We now move on to the interrogative endings of the transplanted varieties. Due to the structure of data gathering and consequent relative status of interviewer and interviewee, very few examples of informal interrogative forms were collected. Such as were attested revealed that the forms of this verb ending in VYK were very similar to those of CSK, as below.

VYK2486	[ət'ən	ijagi-rɨl	t͡ʃad͡ʒu	ha-nja]
	what sort	conversation-OBJ	often	do.PRS-INT.INF

What sort of conversation do you usually have?

VYK 2705	[ni	ai	ga-ge-ni]
	PRN.2.SG.SBJ	NEG	go-MOD-INT.INF

Won't you be going?

The informal interrogatives recorded for KM were similar in form to oblique questions:

KM1420	[dava-j	ədime-sə	sara-nɨnga]
	give-IMP.INF	where-LOC	live-INT.INF

So tell me then, where do you live?

Although not attested in natural speech, one consultant reported a greeting for use in the mornings that appears to contain a fossilised informal interrogative:

KM644	[il	ha-la	kat'ava-nja]
	work	do-in order to	go and return.PRS-INT. INF

Good morning (lit. are you going to work for the day?)

Despite their limited attestation in the primary data, both of these forms are well attested elsewhere for KM (cf. Pak 2005 p. 101), therefore we consider the informal interrogatives of the transplanted varieties to be near equivalents. The problem of balanced attestation where data is gathered from more naturalistic speech data was mentioned at the beginning of this section and for the remaining interrogative and propositive forms this is a particular issue. Our

discussion from this point on relies heavily on reference to prior descriptions of the transplanted varieties and all equivalences are to be regarded as provisional.

Neutral interrogatives are reasonably well attested for KM, but which do not appear at all in the VYK primary data. While we present KM examples below.

KM93   [ədi   gas'o]

            where   go.PST.INT.NEUTR

            Where   did   she   go?

KM436   [ənd͡ʒe   tɛjənas'-o]

            when   be born.PST-INT.NEUTR

            When   were   you   born?

KM640   [pab-u   t͡ʃaps'-o]

            rice-OBJ   Eat.HON.PST-INT.NEUTR

            Hello   (lit.   Have   you   eaten   rice?)

On the basis of the near identical form recorded by Jeong (2010 p. 96), we may provisionally consider these endings near equivalent.

The formal endings, though, may only be considered partial equivalents for the interrogative mood. The VYK formal interrogative endings are somewhat similar in form to CSK –*(su)pnikka*, although as with the declarative endings they seem to have undergone some phonological reduction.

VYK2490　　　[ʃim　　　　　　　njən-i　　　　　　　　nəm-əs'-ɨmnik'a]

　　　　　　　ten　　　　　　　　year-SBJ　　　　　　exceed-PST-INT.POL

　　　　　　　Has　　it　　been　　more　　than　　ten　　years?

VYK324　　　　[kɨ　　　　toŋne-hago　　　　　hak'jo　　　　mər-əs'-ɨmnik'a]

　　　　　　　that　　　neighbourhood-COM　　school　　　distant-PST-INT.POL

　　　　　　　Was　　that　　neighbourhood　　far　　from　　your　　school?

VYK1044　　　[t͡ʃint͡ʃ'a　　　　　　　　　mo　　　　　　　pas'-ɨmk'a]

　　　　　　　really　　　　　　　　　　NEG　　　　　　see.PST-INT.POL

　　　　　　　Did　　you　　really　　not　　see　　it?

VYK2947　　　[i-man　　　　　　　　　ha-mun　　　　　tɛ-mk'a]

　　　　　　　this-only　　　　　　　　do-if　　　　　become.PRS-INT.POL

　　　　　　　Is　　it　　OK　　if　　we　　only　　do　　this　　much　　(recording)?

Unlike the formal declarative ending, though, no similar form appears in the KM primary data. Conversely, we do once more see the influence of the so-called Yukchin endings on the KM data which are absent in the speech of residents of the YKAP:

KM1595　　　　[əmai　　　　　　zat͡ʃʲem　　　　　　　irəna-mdu]

　　　　　　　mother　　　　　why　　　　　　　　　get upPRS-INT.POL

　　　　　　　(I　　asked)　　"mother,　　why　　are　　you　　getting　　up?"

KM453　　　　[mama　　　　　panu-d͡ʒir　　　　　hɛs'-ɨmdu]

　　　　　　　mother　　　　　sewing-work　　　　do.PST-INT.POL

　　　　　　　Did　　your　　mother　　sew?

KM1987   [ad͡ʒubai   a-mdu]

                elder brother   know.PRS-INT.POL

                Do   you   know   (my)   elder   brother?

While imperative endings were very rare in the primary data, in the case of KM several

consultants modelled the informal imperative in direct speech when recalling events. We

include these examples here:

KM684   [maʃi-ra]

                drink.PRS-IMP.INF

                Drink!

KM1015   [nao-ra]

                come out.PRS-IMP.INF

                Come   out!

KM1927   [ha-d͡ʒi   ma-ra]

                do-CNJ   NEG.PRS-IMP.INF

                Don't   do   that!

Forms corresponding to CSK –e/ala did not appear in the VYK primary data, but a form similar

in form and function is attested by Jeong (2010 p. 110), therefore we may once again establish

a provisional relationship of partial equivalence.

KM642   [tɨrɨo]

        enter.PRS.IMP.NEUTR

        Come in!

Imperative forms corresponding to CSK –*sipsio* were attested in the primary data of both VYK

and KM:

VYK1704 [ta   murəbo-ʃio]

        All   ask.PRS-IMP.POL

        Please ask anything.

VYK1347 [jəgi   bo-ʃio]

        Here   look.PRS-IMP.POL

        Please look here.

KM435   [korjə   mal-lo   ha-pʃio]

        Korean   language-INST   do-IMP.PRS.POL

        Please speak in Korean.

Given such a limited number of attestations, we are not able to ascertain whether the initial [p]

of the ending is a distinguishing feature of the forms of the KM and VYK polite imperative

endings or a characteristic of the individual speaker. On the basis of the evidence presented

above, though, we may classify these endings are near equivalents.

The propositive forms suffer most severely from limited attestation in the primary data. Only

one informal KM propositive (a recollection of something that was said in direct speech) and

two polite VYK propositives appear in the primary data:

KM1878	[azubai	ir	ha-la	ka-d͡ʒa]
	elder brother	work	do-CNJ	go.PRS-PROP.POL

	My	elder	brother	said	"let's	go	to	work".

VYK1754	[mal	ha-pʃida	uri]
	language	Do.PRS-PROP.POL	PRN.1.PL

	Let's	talk.

VYK2034	[po-pʃida	irəkʰe]
	lookPRS-PROP.POL	like this

	Let's	look	at	it	like	this.

Secondary sources suggest that the informal propositives of VYK and KM would have similar forms and functions, and are therefore near equivalents. In contrast to the VYK polite propositives presented above, the KM polite propositives differ starkly in form across multiple attestations (Pak 2005 p. 103; Kwon 2010 pp. 85-85), but have near identical functions so would therefore be considered partial equivalents.

The wholly neutral propositives, which were wholly unattested in the primary data, exhibit only partially overlapping forms in KM (Pak 2005 p. 103; Kwon 2010 p. 85) and VYK (Jeong 2010 pp. 104-108). Since the non-shared form appears to be considered primary by Jeong, we provisionally classify these endings are partial equivalents.

5.2.6 Summary

This brings our examination of the sentence final endings of the transplanted varieties of Korean to a close. We present a summary of the equivalences which could be established

below in tabular form. Provisional equivalences based on secondary data are enclosed in brackets.

Mood	CSK Comparative Concept	Near Equivalents	Partial Equivalents	Non-equivalents
	-e/a	✔		
Declarative	*-a/eyo*		✔	
	-(su)pnita		✔	
	-ni, -nya	✔		
Interrogative	*-a/eyo*	(✔)		
	-(su)pnikka		✔	
	-la	(✔)		
Imperative	*-a/eyo*		(✔)	
	-sipsio		(✔)	
	-ca	(✔)		
Propositive	*-a/eyo*		(✔)	
	(u)psita		(✔)	

Table 88: Summary of Contrasts Between VYK and KM Sentence Final Endings

This also concludes our examination of the forms and distributions across VYK and KM of the morphemes which correspond to 98 CSK comparative concepts (31 pertaining to phonology, 26 pertaining to nominal morphology and 41 pertaining to verbal morphology). Of these, 67 were found to be near equivalents of one another, that is, similar in function with only minor differences in form and only five features of (those corresponding to the CSK comparative concepts of the subject honorific pre-final ending, the non-sentence final endings of *–ciman* and *–tolok*, the dative particle *-hanthey* and the comparison particle *-pota*) were found to be non-equivalent. Insufficient data, or only incomparable data, were available for five features

for which equivalences could not be established. The remaining twenty one features were classified as partial equivalents. Thus, we may conclude that while these varieties are globally similar they also exhibit certain notable differences in their phonology and morphology.

In the next sections, we move on to reviewing particular characteristic points of difference in the word order and vocabulary of these varieties.

5.3 Word Order

In contrast with the foregoing, our approach here is more impressionistic. Rather than enumerating an exhaustive catalogue of CSK morpho-syntactic features, instead we focus on particular points of difference between the transplanted varieties.

Aside from the differences examined discussed below, though, KM and VYK have much in common with regards to morpho-synax. The basic word order of SOV word order of CSK is largely shared by both transplanted varieties, which may also be classified 'head final' as they typically manifest all of the concomitant implications of that classification: modifier precedes modified, possessor precedes possessed etc. Nevertheless, close examination of the primary data has revealed certain distinctive features attested which we examine in more depth below.

5.3.1 Verbally Post-posed Arguments

Although SOV is the basic and by far the most frequently occurring word order of Korean and its transplanted varieties, this may vary in order to convey some additional meaning, for example emphasising a particular element of the sentence by moving it to follow rather than precede the verb. While this word order has been argued to be of no great consequence, since it is a device also in common use in CSK (cf. Kwon 2010 p. 126), its use is impressionistically observed to be particularly frequently attested in the KM primary data as compared to that of VYK.

KM441　　　[iro-kʲe　　　　　　　　　　tɨrəva-s'-ɨmdu　　　　　əʃ-i-tɨr-i]

thus-ADV　　　　　　　　　　enter-PST-INT　　　　parent-NOM-PL-NOM

Did　your　parents　arrive　(here　in　Kazakhstan)　like　this?

KM187　　　[t'o　　　　　　　　　　torəva-t͡ʃ'i　　　tashkent]

again　　　　　　　　　　return-PST.DECL　　Tashkent

I　　went　back　to　　Tashkent　　again

KM588　　　[ja　　　əməj-raŋ　　　　kogi-sə　　sa-o

PRO.1.SG　mother-COM　　　　there-LOC　　live.PRS.DECL-POL

uʃtobʲe-ran　　　　　　　te-s]

Ushtobe-COP　　　　　　　place-LOC

I　lived　with　my　mother　there　in　a　place　called　Ushtobe

KM536　　　[pʰur-ɨ　　　　mək'o　　　sara-t͡ʃ'i　　　t͡ʃint͡ʃir]

grass-ACC　　　eat.CNJ　　live-PST.DECL　really

We　　really　ate　grass　to　live

KM1271　　[mo　　　　　　ka-t'a-va　　　　xan-bən-do]

NEG　　　　　go-PST.DECL-come.DECL　one-time-too

I　couldn't　go　back　there　even　once　for　a　visit

KM1071   [teriva-t͡ʃ'i   mama-raŋ]

                bring-PST.DECL.NEUTR   mother-COM

                They   brought   them   with   their   mother

KM1009   [kəgi   and͡ʒ-as'-o   s'irk-esə   uri]

                there   sit-PST.DECL-POL   church-LOC   PRN.1.PL

                We   sat   there   in   the   church.

We may also note that a wide variety of noun phrases appear following the verb (subjects, topics, comitatives, locatives etc.). The primary data stands in contrast to the earlier claim that only subjects, direct objects and adverbs could be placed in post verbal position (Kwon 2010 p. 126).

Post-posed verbal arguments undoubtedly appear in the vernacular Korean of the YKAP, they are found only very rarely in the primary data, as below:

VYK1754   [mal   ha-pʃida   uri]

                language   Do.PRS-PROP.POL   PRN.1.PL

                Let's   talk.

This leads us to the conclusion that such a wide range of post-verbal constituents being used at such a high frequency is characteristic of KM but not VYK.

We further posit that the influence of Russian (perhaps through sentences constructed with Russian as a matrix language) has led to the much higher frequency use of this pattern in Central Asia. The following sentence, which includes Russian code-mixing, provides some support for this theory:

KM1107	[nore-do	t͡ʃal	pur-o		ana	toʒe]
	song-ADD	well	sing-PRES.DECL-POL		PRO.3.SG.F	also
	She	even	sings		well,	too

Russian word order is much freer than that of Korean, with the grammatical relations of constituents being expressed through a wide variety of inflectional and conjugational endings rather than their position in the sentence. The final two Russian words of this code-mixed utterance demonstrate that the verbal post-posing of constituents identified is possible, even natural, for Russian language constituents. We further note that code-mixing or nonce-borrowing from Russian features in many of the examples given above, which further supports the idea that the frequency of use of this word order, if not its existence, is related to the ubiquitous bilingualism of the *Koryo Mar* speech community.

In addition to the influence of Russian bilingualism among the *Koryo Saram* on the level of the sentence, one instance of phrasal word order deviating from the head final tendency of Korean's was recorded. Specifically an attributive modifier [nʲe balʃo-ju], borrowed from Russian, was preceded by the element which it modifies, [t͡ʃibu], as in the below example:

KM993	[nu	t͡ʃib-u	nʲebalʃo-ju		stroi-l	he-t͡ʃʼi]
	so	house-OBJ	small-F.SG.ACC		build-PST.M	do-DECL.PST
	So	they	built	a	little	house

It must be emphasised, though, that the above word order is exceptional and the overwhelming majority of modified noun phrases followed the typologically consistent pattern for an SOV language – modifier preceding modified. Unlike the frequently attested movement of phrases to post-verbal positions, then, this feature is not taken as characteristic of KM, but is simply an instance of interference from the specific consultant's L2.

5.3.2 Negation

Unusual patterns of negation have long been identified with KM (cf. King and Yeon 1992). This is not only in reference to the unusual form of the negative particle corresponding to CSK *an* takes on account of the phonological processes described above, such as /n/ deletion, but also with regard to negation in constructions which feature an auxiliary verb. In this construction negating particles may be inserted between main and auxiliary verbs in verb phrases, unlike in CSK where the negating particle must be positioned before both verbs. While prior research has emphasised this feature as especially distinctive of KM, it was attested in both the VYK and KM primary data.

KM637   [ne   t͡ʃar   ar-a-mo-tɨl-ɨn   kə-n   ar-t͡ʃi]

PRO.1.SG   well   know-CNJ-NEG-hear-ADN   thing-TOP   know-PRS.INT

That's   not   something   that   I   understand   well,   you   know?

VYK2140   [ar-a-mo-tɨl-ɨn   pubun-do   isˈək'o]

know-CNJ-NEG-hear-ADN   portion-too   be present-PST.DM

There   were   also   parts   that   I   could   not   understand.

Furthermore, non-separable patterns of negation in such constructions were attested for KM, as below, but not for VYK:

KM635   [ətʼən   mar   mo   aradɨr-ɨn]

certain   language   NEG   understand-ADN

There   were   certain   words,   incomprehensible   (words)…

A less expected distinguishing feature between the transplanted varieties is the relative

paucity of long form negation after the CKS pattern –*ci anhta* in KM, which is attested only

once for KM, but appears frequently in the VYK primary data.

VYK498   [koŋbu   ha-d͡ʒi   an-ɨl   t'ɛ]

                    study   do-CNJ   NEG-ADN   time

                    (You   mean)   when   I   am   not   studying?

VYK2182   [nəmu   pap'ɨ-d͡ʒi-n   an-ɛsɨmda]

                   too   busy-CNJ-TOP   NEG-PRS.DECL.POL

                   I   am   not   too   busy.

VYK1244   [kɨrə-d͡ʒi   an-ɨm   mo   pɛju-wəjo]

                   to be thus-CNJ   NEG-if   NEG   learn-PRS.DECL.POL

                   You   will   not be   able to learn if you do not do it like this.

VYK44   [t͡sal   kiəg-ɨ   na- d͡ʒi   an-ɨnde]

                   well   memory-TOP   occur-CNJ   NEG-DM

                   I   do   not   remember   all   that   well.

Contrary to expectations, speakers of both KM and VYK appear to have access to all the same

patterns of negation. Rather, it is notable that they are used with very different frequencies

across the transplanted varieties.

5.3.3 Use of Nominalizer {–ki}

The nominalizer /-ki/ is used in a wide range of constructions in all varieties of Korean. It has

been attested (King 1987; King and Yeon 1992) as having uses in KM which are not shared with

CSK, though.

Among these constructions are those using characteristic KM vocabulary pertaining to ease or

difficulty. These constructions are presented below. While they are particular to KM, it must

be noted that they are analogous to CSK *haki swipta/elyepta* and its VYK equivalent.

KM349 [tokir-mar   ha-gi t͡ʃə-nɨ̃   də hor-t͡ʃimɨ]

      Germany-language do-NMZ PRN.1.SG-TOP more easy-DECL.PRS

      Speaking   German is easier for me.

KM583 [sar-gi   pap'-as'-o]

      live-NMZ   busy-DECL.PST-POL

      Life   was   hard.

However, this ending is also used in KM in ways which are more strikingly divergent from CSK:

it replaces CSK *–ko* in the construction corresponding to *hako siphta* and it replaces CSK *–key* in

the construction corresponding to *hakey toyta*. Examples of this may be found below:

KM141 [kundɛsa-r   ga-gi   dɛ-t͡ʃʼi]

      army-OBJ   go-NMZ   become-PST.DECL.NEUTR

      I   wound up enlisting in the army.

KM227   [pʲerem   o-gi   mo   te-nkʼa]

           Perm   come-NMZ   NEG   become-DM

           It was because it turned out that I could not go to Perm.

KM1261   [i   jəd͡ʒa   kɨ   saram-ga

           this   woman   that   person-COM

           tɛŋgi-gi   tɛ-t͡ʃʼi]

           go around-NMZ   become-PST.DECL.NEUTR

           This woman came to hang around with that person.

KM1509   t͡ʃaps-ɨgi   ʃipʰ-ãi]

           eat.HON-NMZ   want-DM

           It   is   because   he   wants   to   eat.

Endings with a closer correspondence to the CSK forms may be found in the VYK primary data:

VYK543   id͡ʒe   paŋag-ɨl   ha-ge   tɛ-go]

           now   holiday-OBJ   do-CNJ   become-DM

           It   turns   out   that   we   are   on   holiday   now.

VYK996   [i   kime   o-ge   dɛsʼ-ɨmda]

           this   in course of   come-CNJ   become.PST-
                                                            DECL.POL

           In   the   course   of   (doing   that)   I   came   to   arrive.

VYK2168	[han	pən	ka	po-go	ʃip-sɨmda]
	one	time	go	see-CNJ	want.PRS-DECL.POL
	I	want	to	try going (there)	one time.

VYK59	[uri	mal-do	pɛu-go	ʃipʰ-ə]
	PRN.1.PL	language-ADD	learn-CNJ	want.PRS-DECL.INF
	I also want to	learn	about our	language.

While it has been proposed that this use of /-ki/ specifically in constructions corresponding to CSK –*hago siphta* is the result of sound change common to the Northeastern dialect zone (King 1992 p. 214), its absence from Yanbian Korean suggests that if ever it was common to that variety it is retained only in Central Asian Korean.

That this is a case of the two connective endings /ko$_{KM}$/ and /key$_{KM}$/ being replaced by the nominalizing ending rather than sound change is may be supported by two arguments. First, the adverbial ending, which is formally identical to /key$_{KM}$/ may be found still in common use in KM.

KM828	[t'al	iɾ-əp-ge	sa-o]
	daughter	work-be absent-ADV	live.PRS-DECL.POL
	My daughter	is living	comfortably.

KM199	[himdɨɾɨ-gey	tɨɾɨwa-t͡ʃ'i]
	difficult-ADV	enter-PST.DECL.NEUTR
	We came here with	great difficulty.

Were this case of regular sound change, we would not expect to find such exceptions.

Second, particularly relevant to the replacement of /ko$_{KM}$/, these changes appear to follow well established patterns of cross-linguistic grammatical replication by which less socially dominant languages imitate the infinitive constructions of more socially dominant languages, but where an infinitive is not present in the less socially dominant language's inventory it is replaced by nominalized verb (cf. Heine and Kuteva 2005).

Nevertheless, since a formally similar ending does not appear in similar constructions in VYK. Therefore, we may conclude that the use of the nominalising ending corresponding to CSK *-ki* is a distinguishing feature of KM and VYK.

We now briefly review certain key differences of vocabulary.

5.4 Lexicon

5.4.1 Kinship Terms

Both transplanted varieties' kinship terms differ from those of CSK. Most strikingly, prevalent in the Central Asian Korean primary data is a term referring collectively to the speakers parents of derived from LMK *e·zi*, realised in the contemporary variety as [əʃi] (cf. King 1992 p. 215). A further unusual development is the native Korean kinship terms for 'father' and 'mother' coming to be used to refer to the grandparent generation. Members of the Central Asian Korean community with otherwise very little knowledge of KM may also be observed to use these forms. Rather than using traditional Korean kinship terms, the parent generation is now more often called by words borrowed from Russian: *mama* and *papa*. The atypical Russian compound word [mamapapa], denoting 'parents' and possibly formed in analogy to KM or colloquial CSK *emeappa* (parents – lit. mother-father), was also attested in the KM primary data. Below we present further examples of the unusual kinship terms of KM in use.

KM797 [kɨ dzən-e uri əs-i-nʲɨ əmaj-ga abədʒi

that before- PRON.1.PL parent-SUBJ- mother- father
LOC TOP COM

puk t͡ʃosən-esə von-dõ-r va-t͡ʃ'i]

North Korea far-East-INST come-DECL.PST.NEUTR

Before that my parents came to the Russian Far East from the North of Korea

KM1846 [ad͡ʒubai-do va-t͡ʃ'i]

older male relative come-DECL.PST.NEUTR

My uncle came, too.

KM2180 [amɛ abai hɛs'-o]

mother father do.PST.DECL-POL

My parents used to do that.

VYK2839 [uɾi abai nulg-a]

PRN.1.PL older male relative be old.PRS.DECL-INF

My uncle is very old.

Although the KM form for the word 'younger sibling' is often considered to be distinct from those of of VYK or CSK, this is simply a result of the combination of the regular phonological processes of umlaut and nasal deletion to produce the form [toŋsɛ̃] from the canonical *tongsayng*. The term [oɾɛbi], which denotes elder male siblings of women, is shared by both transplanted varieties.

KM1075   [t͡ʃogɨm   ha-n   orɛbi   saŋsɛna-t͡ʃ'i]

             a little bit   do-ADN   elder brother   die-PST.DECL.NEUTR

             The   younger   of   my   elder   brothers   passed   away.

VYK2645   [kʰɨ-n   orɛbi]

             big-ADN   elder brother

             (It   was)   my   eldest   brother.

The form [kʰɨn nubi] (cf. LMK *nuiy* and the CSK cognate *nwui* (King and Yeon 1992 p. 217; King

1992 p. 215)) which denotes older female siblings of men was only recorded in KM, though.

While the attested differences between the kinships terms appearing in the KM and VYK in the

primary data were noticeable, we emphasise that the extent of the differences in kinship

terms between KM and all other varieties of Korean is vast. Here, we summarise these

differences by noting that, since many of the KM forms are more similar to LMK than CSK they

may be characterised, along with the subjects of the next sub-section, as archaisms. A fuller

discussion of these kinship terms is outside the scope of this section, but may be found in Pak's

work (2005 pp, 145-158).

5.4.2 Archaisms

In KM certain archaic words or word-forms, which are no longer current in Yanbian Korean or

CSK, are retained. Most notably these include the so-called 'k-stem' nouns, demonstrative

pronouns of place, so-called 'intervocalic stops' and nouns which retain [ɨ] after /s$_{KM}$/, /c$_{KM}$/

and /ch$_{KM}$/. Examples of each of these are presented below.

Only a single relic form of the LMK 'k-stem' nouns was encountered in the KM primary data.

While the CSK word for 'flour' is *kalwu* it is realised as [karg] or [kargi] as below.

KM710 [karg-ɨ … t͡ʃabɨlo t͡ʃib-esə ha-o]

flour-TOP by oneself house-LOC do.PRS.DECL-POL

We make our own flour at home.

KM557 [karg-i se kilo]

flour-SBJ three Kilo

Three kilos of flour

The archaic forms of the demonstrative pronouns of place are quite ubiquitous in KM, as may be seen below:

KM7 [iŋgi-r almata i t͡ʃib- ɨro]

here-ALL Almaty this house-ALL

(I came) here to Almaty, to this house.

KM543 [iŋgi tɨrɨvas'-o]

here enter.PST.DECL-POL

I came here.

KM2271 [iŋgi-sə mjət͡ʃʰ … kilomʲetr i-o]

here-ABL several kilometry COP.PRS.DECL-POL

It is several kilometres from here.

KM1462 [kəŋge- sə drugdrug-mimo aras'-o]

there-LOC each other-by know.PST.DECL-POL

People knew each other there.

298

KM235   [kəŋgi- sə   hɛk'jo   pʰir   ha-go]

there-LOC   school   graduate   do-DM

I   graduated   from   school   there.

KM2193   [kəŋga   saɾas'-o]

there   live.PST.DECL-POL

I   lived   there.

While VYK consultants had knowledge of these forms, in practice demonstrative pronouns of place closer to those of CSK are in more common used in the YKAP.

VYK2871   [kɨre   jəgi   wa-s'o]

so   here   come-DECL.PST.POL

So   I   came   here

VYK402   [ta   jəgi   saɾam   i-d͡ʒi]

all   here   person   COP.PRS.DECL-NEUTR

They are all people from around here.

VYK1720   [kəgi-sə   hak'jo-rɨ   ka-go]

there-LOC   school-OBJ   go-DM

I   went   to   school   there.

VYK2119   [kəgi-e   ka-nɨnde]

there-ALL   go-DM

Well   I   did   there   go   there.

Although we note that there are occasional dialect forms encountered in the VYK primary data, they are differ from the archaic forms retained in KM.

VYK1023   [igi- sə   jənbjən-dɛ-e sə]

                  here-LOC   Yanbian-university-LOC

                  (Do   you   mean)   here   at   Yanbian   University?

VYK2162   [igi-nɨn   kjəɨl   i-mjən   ad͡ʒu   t͡ʃup-go]

                  here-TOP   winter   COP-if   very   cold-DM

                  It   is   very   cold   here   in   winter

In addition, the lenition of many intervocalic stops has not progressed in *Koryo Mar* as it did in LMK and subsequently in CSK and Yanbian Korean. Thus, verbal conjugation may retain stem final bilabial stops or alveolar fricatives in positions or even in archaic forms other parts of speech where they do not appear in other varieties of Korean:

KM827   [kɨrende   ʃibaŋ   hobund͡ʒa   sa- d͡ʒi]

              anyway   now   alone   Live.PRS.DEC-INF

             Anyway,   now   I   live   alone

KM1257   [kob-ɨn   jəd͡ʒa]

              beautiful-ADN   Woman

             (She   was   a)   beautiful   woman.

KM2164   [kɨrəkʰe   kob-at'a]

              like that   beautiful-PST.DECL.INF

              She   was   beautiful   like   that.

Finally, in this sub-section, we turn to words which retain the vowel /u_KM/ following /s_KM/, /c_KM/, and /ch_KM/. While KM retains this string, in VYK (similar to CSK), the reflex of the LMK */u/ in this positions is /i_VYK/. Examples of the KM archaisms are presented below.

KM268   [kɨ   sɨd͡ʒip-d͡ʒib-esə   t͡ʃosən-mar-mã   ha-t͡ʃ'i]

             that   marriage-house-   Korea-language-only   do-DECL.PST
                         LOC

             In   my   marital   home   we   only   spoke   Korean

KM1877   [at͡ʃʰɨm   t͡ʃaps-ɨpdi]

              morning   eat.HON-RETR.POL

              (I   recall   that)   he   ate   breakfast.

The archaic nature of KM, especially KM vocabulary has been argued for elsewhere (e.g. King 1992). The primary data supports these arguments and identifies such archaisms as distinguishing features from VYK.

5.4.3 Semantic Change

There are several vocabulary items attested in VYK and KM which do not share their meanings with the corresponding CSK vocabulary or the vocabulary of the other transplanted variety. We inexhaustively examine several of these widely used words and phrases in this section which are relatively commonly attested in the primary data and which are considered in some way characteristic of the varieties (see, for example, King 1992).

In addition to retaining its meaning of 'to read', the word corresponding to CSK *ilkta* in KM is regularly used with additional meaning of 'to study', especially as part of the phrase /kulul ilkta_{KM}/.

KM322   [kɨ-t͡ʃən-e-nu   uri   abəd͡ʒi

that-before-LOC-TOP   PRO.1.PL   father

vladivostok   univʲersitʲet-sə   kɨr-ɨ   irgə-t͡ʃ'i]

Vladivostok   University-   writing-OBJ   read-PST.DECL.INF
                     LOC

Before   that   my   father   studied   at   Vladivostok   university

KM2125   [kɨr-ɨ   ai   ilgəs'-o]

writing-OBJ   NEG   read.PST.DECL-POL

I   did   not   study/go   to   school

A similar usage was not attested in the VYK primary data where the more common word for study is similar to that of CSK: /kongpwu hata_{VYK}/.

Furthermore, forms shared by the transplanted varieties and CSK may be distinguished from one another in cases where different word senses come to be the primary meanings of cognate forms. For example, the word *papputa* has the primary meaning 'to be busy' in both Yanbian Korean and CSK. This sense is not attested in the KM primar data, rather the word is frequently used in the sense of 'to be difficult'. Dictionaries of CSK only variably list this as a possible wordsense for the entry *papputa* (cf. *cosenmal taysacen k-s* 1991 p. 1096 and *yensey kwuke sacen 1* 1998 p. 777. The meaning given above appears as the third wordsense in the former, but not at all in the latter), however in varieties other than *Koryo Mar* it is used with this meaning only very infrequently.

KM250   [toraga-nik'a   t͡ʃogum   sa-gi   pap'u-gi   dʷe-ne]

die-CNJ   a little   live-NMZ   difficult-NMZ   become-DECL

Because   (my husband)   died,   life   became   a   little   difficult

There are, however, also some cases where the common dialectological heritage of the transplanted varieties if emphasised by the way that words are used with similar word senses which differ from CSK. An example of this would be the words corresponding to CSK *cil* 'work'. This typically a bound noun in peninsula varieties of Korean carries a distinct negative connotation, as evinced by the word *totokcil* – thievery. In KM and VYK, though, this word has lost its negative connotation, so may be freely combined professions in order to describe their activities.

KM305   [kɨ   daɨm-e-nɨn   noŋsa-d͡ʒir   hɛ-s'əjo]

that   after-LOC-TOP   farming-work   do-DECL.PST

After   that   I   did   farm   work

KM821   [kɨre   pir-ago   kĩjaŋ   buxgaltʲer-d͡ʒir   hɛ-t͡ʃ'i]

so   finish-do.CNJ   just   accountancy-work   do-DECL.PST

So   I   graduated   and   just   worked   as   an   accountant

VYK894   [əməni-nɨn   sənsɛnim-d͡ʒir-ɨ]

Mother-TOP   teacher-work-OBJ

My   mother   (did)   work   as   a   teacher.

It would appear, though, that the KM and VYK words may be distinguished in that /cil$_{KM}$/ may even be used as an independent noun.

KM327   [kɨrənde   uri   abod͡ʒi-rɨ   kɨrə-n

        therefore   PRN.1.PL   father-OBJ   thus-AND

        t͡ʃir   ha-rago   pone-s'o]

        work   do-PERP   sent-DECL.PST

        So   my   father   was   sent   to   do   that   kind   of   work

For a feature that distinguishes both VYK and KM from CSK, but is shared between the two, we turn to the phrase corresponding to CSK *il epsta* (lit. there is no work). In both transplanted varieties this has taken on a meaning expressing approval or a positive evaluation.

KM787   [kɨzə   pot^hoŋ   mar-ɨn   ir-əp-g'e   ha-d͡ʒi]

        therefore   normal   language-TOP   work-be absent-ADV   do-PRS.DECL

        So   I   can   speak   *Koryo   Mar*   easily

KM1978 [amɛ   ir-əps- ək'uma]

        mother   work-be absent-PST.DECL.POL

        It   was   easy   for   my   mother.

VYK1470 [ir-əps-o]

        work-be absent-PRS.DECL.POL

        It   is   fine

VYK1969 [məlmi-ga   ir-əpsə   gad͡ʒ-igo]

        travel sickness-SBJ   work-be absent.CNJ   hold-DM

        So   it   is   because   travel   sickness   is   not   a   problem   for   me.

Above we have presented a limited number of specific examples of divergence in the common, inherited lexical stock of VYK and KM. These words may be regarded as emblematic of the varieties, or as 'shibboleths'. While a more the examination of semantic change of the core vocabulary of these varieties would be a worthwhile pursuit, we may conclude that globally, if impressionistically, speaking the meanings of the words held in common by the transplanted varieties do not appeared to have diverged greatly.

5.4.4 Honorific Vocabulary

The predicate endings of both KM and VYK reflect addressee honorification, and VYK further allows for subject honorification. In addition to these means of elevating animate grammatical subjects of sentences and addressees, both varieties may express honorification through the use of specific lexical items.

VYK2749   [tul-t͡ʃʼe   əni-ga   jəgi   ke- ʃi-nɨndɛ   pʰalʃipjadɨlb   i-go]

               two-ORD   elder sister-   here   reside.HON-   eighty eight   COP-DM

                            SUBJ   HON-CNJ

               My   second older   sister   stayed   here   (in   Yanji)   and   is   eighty   eight

VYK321   [ət'ən   ɨmʃik   mani   t͡ʃaps-uʃj-əs'-ɨmnik'a]

               what sort   food   much   eat.HON-HON-PST-INT.POL

               Wha   t sort   of   food   have   you   mostly   eaten?

VYK2366   [kad͡ʒokʼ   soge   hɛ   dɨri-lkʼajo]

               family   introduction   do.CNJ   give.HON-FUT.INT.POL

               Shall   I   introduce   my   family?

KM1889	[ad͡ʒubai	at͡ʃʰim	mai	t͡ʃaps'o]
	father	morning	much	eat.HON.PST.DECL-POL

My father ate a lot in the mornings.

KM228	[əməni-rɨ	dalnʲ-ɨj	vostok-esə	keʃi-n
	mother-OBJ	far-NOM.M	East-ABL	reside.HON-ADN.PST

go		terjo-va-t͡ʃ'i]
Place		bring-come-DECL.PST

They brought my mother here from the place she was living in the Russian Far East

Although these examples overlap to some extent, honorific language appears far more frequently in the VYK data than the KM data and, impressionistically, bears a closer resemblance to CSK honorific speech than KM does. Thus we may say that the transplanted varieties are distinct in their usage of honorific vocabulary.

5.4.5 Loanwords

The speakers of the transplanted varieties of Korean used in China and Central Asia necessarily adopted or coined a large number of words for the purposes of describing the new social and physical surroundings in which they found themselves. Furthermore, since their initial transplantation KM and VYK have looked to different sources from each other and from CSK when it comes to vocabulary pertaining to the technological and social innovations of the twentieth century.

However, without codification, the question of what constitutes a loanword, especially in universally bilingual populations, is a vexed one[25], since at any point a multilingual speaker of either of these transplanted varieties could simply code-switch between their languages and be certain of their interlocutor's comprehension. Here, though, we first examine the use of non-Korean vocabulary in the primary data and then turn to the secondary sources for a fuller account of loanwords in KM and VYK.

As might be expected from the advanced degree of language shift from KM to Russian in Central Asia and the high incidence of language attrition among the remaining speakers, the use of Russian in Central Asia[26], including many instances of apparently uncontrolled code-switching (on both the inter and intra-sentential level), was encountered more frequently than the use of Chinese in the YKAP. Representative examples are provided below:

---

[25] Most problematic is differentiating between true loanwords and so-called "nonce" borrowings (see, for example, Sankoff, Poplack and Vanniarajan 1990) or code-switches (see, for example, Poplack 2000 [1979/1980]). In multilingual communities such as those in Central Asia and the PRC, speakers may draw on the resources of any languages in their shared repertoire in order to facilitate communication so it is not always clear whether an apparent loanword is actually part of the system of a given language or being used, for example, due to a speaker's lack of knowledge of a certain word in one language but not the other. While numerous means for distinguishing true loanwords from code-switches have been proposed (e.g. degree of "nativization" in terms of, phonological adaptation of a word to the receiving language and the possibility of its combination with that language's affixes, or frequency of use by individual speakers or on a community level), it must be acknowledged that there is no categorically reliable test for identifying a true loanword synchronically (Thomason 2001 pp. 131-136). Additionally, given the ultimately imperfect historical attestation of all natural languages, as noted by Haspelmath (2009 p. 38) there is also no way of identifying non-loanwords in an absolute sense.

[26] It must be conceded that this was one of the more extreme observer effects which arose from much of the primary data being gathered by a non-native speaker, non-embedded researcher. One consultant specifically explained that they were simply accustomed to speaking Russian to someone who did not physically resemble *Koryo Saram* and had difficulty answering questions in *Koryo Mar,* even though elicitation was being carried out in Korean which was somewhat accommodated to *Koryo Mar*. While this is certainly an issue to be borne in mind, primary data elicited by a native speaking, embedded researcher is also not without instances of code-switching. Therefore, we may conclude that not all instances of code-switching were brought about due to the presence of an unfamiliar, external researcher and that it constitutes a part of *Koryo Mar* interaction which merits further examination.

KM2367   [kɨ   daɨm-e   zvoni-l   ha-pdi]

                    that   after-LOC   call-PST.M   do-RETR.POL

                    After   that,   I   recall   that   he   phoned   us.

KM1411   [nawa-s ə   t'o   pozdravi   ha-go]

                    come out-CNJ   again   greet   do-DM

                    And   then   they   come   out   to   wish   you   a   happy   birthday   again

KM993   [nu   t͡ʃib-u   nʲe   balʃo-ju   stroi-l   hɛ-t͡ʃ'i]

                  so   house-ACC   NEG   big-F.SG.ACC   build-PST.M   do-DECL.PST

                  So   they   built   a   little   house

KM353   [madadɨr-ɨn   …   avarija   hɛ-s'o]

                  eldest son-TOP   car crash   do-DECL.PST

                  My   eldest   son   …   was   in   a   car   crash

KM838   [nose-mar-tak   ha-ɣo   ʃipʰ-ə   t͡ʃo-ket'a-d͡ʒi]

                  Russia-language-as   do-CNJ   want-CNJ   good-MOD-EMP

                  I   wish   I   could   speak   it   as   well   as   Russian –   that   would   be   great

Despite their many points of typological similarity, the influence of the Turkic languages of Central Asia on *Koryo Mar* has been minimal. Loanwords are restricted largely to nouns denoting common objects of the indigenous material and food cultures of the region.

KM692   [beʃparmak-na   plov-na]

               Beshparmak-or   pilaf-or

               (The sort of Kazakh food I eat) is *beshparmak* or pilaf or something

Turning to the secondary data reveals a similar picture, namely, a large number of words have entered into KM from Russian in a wide range of domains, whereas very little of the vocabulary of the regional Turkic languages has been taken up.

Examining the loanwords of Russian origin into KM it is possible to identify a those words which were adopted to denote new concepts or objects outside of the Korean peninsula of the mid-nineteenth century: KM – *meydurey* Russian – *vedro* (bucket); KM – *shakkay* Russian – *shapka* (fur hat); KM – *ppidzhikkay* Russian – spichki (matches); KM – *pomindori/pomidori* Russian *pomidory* (tomatoes); KM – *nodari* Russian – *lodir'* (lazybones). The phonological forms of these words also provide some diachronic information about the KM that they were borrowed into. For example, while the sound [v] is attested in contemporary KM, even in word initial position, it was rather adapted to early KM phonology as [m] in *meydurey*. Furthermore, we can see the simplification of consonant clusters in the words *shakkay* and *ppidzhikkay*. Most notably the transition from Russian [sp] to KM [p'] is strikingly reminiscent of the simplification of the initial consonant clusters of LMK into the modern reinforced consonants (King 1996 also reports a similar process for the following loanwords: KM – *ppassibai* Russian – *spasibo* (thank you); KM – *ttallobaya* Russian – *stolovaja* (cafeteria); KM – *ttakkai* Russian – *stakan* (glass(drinking vessel)). Another possible borrowing strategy that may be observed, here particularly in the word *ppidzhikkay*, is the use of an epenthetic vowel to break up an otherwise impermissible consonant cluster. This may also be seen, for example in the KM word *kaluman* from the Russian *karman* (pocket).

Also of interest in this set of words is the application of the Head Sound Rule to render a

Russian initial [l] as [n] in KM. While on of the seed dialects of KM, the Yukchin dialect, is well

known for not applying the Head Sound Rule, we may attribute this phenomenon here to the

influence of other dialects which went into the formation of KM in which the Head Sound Rule

does operate.

The former USSR, in which KM has operated for the majority of its existence, contributed a

large number of loans due to its unique culture. These include, but are not limited to the

following:

KM Loan	Russian	English
ssopeythu	sovet	Soviet
pulusuasiya	burzhuazija	bourgeoisie
ppioneylu	pioner	Soviet youth organisation
cupeyno/ucupeyno	zveno	work team
yacheyikka	jachejka	political cell
ppullakhathu	plakat	poster
polseypikki	bol'sheviki	Bolsheviks
alaphakhu	rabfak	Workers' Polytechnic
kkomissiya	kommissija	commission
khulyeymulli	kreml'	Kremlin

Table 89: Soviet Cultural Loanwords into KM (adapted from King 1996 p. 963)

We note that the role of Russian as immediate source language for loans may lead to points of

difference between KM and other varieties of Korean which have adopted words from,

ultimately, the same etymological source. Good examples of this include the measurements of

distance, weight and volume of the international system of units: KM *meyturi* CSK *mithe* (metre); KM *kurami* CSK *kulam* (gram);KM *literi* CSK *lithe* (litre) (Pak 2005 p. 204).

Speakers of Yanbian Korean seem to have more conscious control over their language choice than speakers of *Koryo Mar*. However, given the extensive, long-term influence of Chinese on the vocabulary or peninsula varieties of Korean, the question of what constitutes a loan in Yanbian Korean is also a complex issue. For example, the word *pangcohata* ((帮助 - bāng zhù) to help/abet particularly in legal language or with a pronounced negative connotation) may be found in Korean dictionaries (e.g. *yensey kwuke sacen* p. 815), but its use in Yanbian as the generalised word for helping another is strikingly different from the technical legal term the same form representing in CSK. A further example of Chinese character compounds being borrowed, but then being realised according to the Sino-Korean pronunciation rather than Chinese pronunciation is the case of [hyaŋʧ͡ʼɛ] (香菜 - xiāng cài) which is used in place of CSK *koswu* or the loan *kholianthe* for 'coriander'.

Conversely, there are also cases of traditional borrowing. For example, the verb *chenghata* ((請 - qǐng) to ask/request), which is featured in CSK dictionaries, but rarely used, is not only used with a slightly modified meaning in Yanbian Korean (to invite), but with a pronunciation clearly informed by the contemporary standard Chinese pronunciation of the character which is the etymological root of that verb; that is, Yanbian Korean [ʧ͡ʰiŋkʰɨhada] rather than CSK [ʧ͡ʰəŋhata].

It is safe to conclude that use of loanwords and nonce loans from Chinese and Russian are among the most emblematic distinguishing features of VYK and KM, respectively.

5.4.6 Summary

This chapter has presented a parallel description of the transplanted varieties of Korean spoken in Central Asia and the YKAP using the terms of traditional Korean grammatical

description as a source of comparative concepts. In doing so, we have identified points of similarity and differences in their phonological and morphological forms and tentatively suggested points in which the functions of certain morphemes might diverge. We have also discussed some of the characteristic features of the word order and lexical systems of each other these varieties. While their similarities appear to outweigh their differences, it is clear that there are points of both symbolic and systematic divergence between these varieties.

These detailed and granular qualitative overviews of the transplanted varieties and the differences in their forms and use add weight to the position that, despite their common dialectological heritage, there is linguistic cause for these varieties to be considered separate synchronically.

In the following chapter, we go on to further interrogate the empirical basis for characterising these two varieties as separate synchronically by carrying out a quantitative analysis of the primary data. Having drawn up a feature catalogue of these varieties based on the foregoing descriptions, we shall test to see whether there are any patterns of clustering between speakers of the transplanted varieties which may be identified on the basis of an aggregate description of the features used in each individual consultant's speech. In other words, we employ statistical tests and techniques to determine whether we are able to distinguish between KM and VYK on the basis of the actual usage of the features identified above and, if so, which of those features or combinations of features is distinctive of each variety?

6 Quantitative Analysis

6.1 Introduction

The preceding chapter provided fine-grained insights into the forms and functions of a large number of features of KM and VYK. The majority of features analysed were near equivalents, that is, effectively identical in form and function. This result is suggestive of the strong affinity between these varieties. Conversely, features of only partial and non-equivalence were also identified. The simple categorisation of equivalences between these varieties' features, though, does not answer the question of whether the varieties may be held distinct from one another on linguistic grounds. This can only be done by demonstrating that the variance of linguistic features observed in the CA is closely correlated with the origins of speakers in the YKAP or Central Asia.

A further fact brought to light by the preceding chapter was the large number of shared variants in each transplanted variety for certain common comparative concepts; the forms of the topic particle (CSK –(n)un)or conditional connective verb ending (CSK –(u)myen), for example. The inclusion of these forms in the preceding CA, though, simply indicates that they were used by at least one of the consultants who provided the primary data for this project. In contrast to traditional dialectology, data for this project was gathered from numerous consultants who self-identified as speakers of KM or VYK and in the form of connected, vernacular speech rather than short answers to elicitation questions. The form of this data allows us to compare these varieties not only in terms of the forms that are attested in each variety, but also in terms of the distribution of these forms across a population of self-identified speakers of these varieties. In doing so, we will be able to ascertain whether, despite their many similarities, VYK and KM may be distinguished from one another linguistically on the aggregate basis of their features and, if so, which features contribute most to the variance which distinguishes these varieties.

In this chapter, we employ statistical methods to answer the questions implicit in the above. Explicitly, these questions are:

- Is it possible to distinguish VYK from KM on the basis of their linguistic features?
- Which linguistic features are most distinctive of each of the transplanted varieties?

In order to present answers to these questions this chapter shall be structured as follows. We first provide details of the application of the appropriate statistical tests (cluster analysis and Categorical Principal Components Analysis (CATPCA)) identified as suitable for our purposes in sections 4.6.3 and 4.6.4. We follow this with the creation of a feature catalogue, a pre-requisite of applying these tests, and a brief explanation of the inclusion of each feature in the catalogue. We then present the results of the clustering and CATPCA along with discussion of their significance with regard to the questions posed above before moving on to the concluding chapter of this thesis.

6.2 Statistical Tests

Statistical and quantitative techniques have a wide range of uses in linguistics and have been applied to such problems as authorship attribution, measuring rates of linguistic change, establishing general cross-linguistic patterns of language use and, of course, quantifying linguistic diversity and variation. Here, we provide linguistically oriented overviews of the application of the statistical techniques identified in chapter 4: clustering and CATPCA.

6.2.1 Cluster Analysis

The procedure for this cluster analysis, to be carried out using SPSS (IBM 2013), is as follows. Having established a feature catalogue (see 6.3 below), the presence or absence of each feature in each consultant's speech is to be ascertained and coded using a binary representation in an $N \times p$ (number of variable/linguistic features against number of

cases/consultants) matrix. From this matrix the position of each case in a feature space is generated and the squared Euclidean distance between each of them calculated. Ward's clustering algorithm is applied to these distances and individual cases are agglomeratively clustered until they are all contained in a single cluster.

The primary output of this analysis is a dendrogram which gives a graphical representation of the clustering schedule (i.e. the order in which individual cases and clusters were merged with one another) in addition to the relative similarities between each consultant's language and the groups and sub-groups into which they may be divided. From such a dendrogram we may determine whether the speech of VYK and KM consultants may be distinguished on the basis of their aggregate dissimilarity in terms of the presence or absence of features included in the feature catalogue.

6.2.2 Principal Components Analysis

The same *N x p* data matrix as is used to carry out the cluster analysis described above may be used to carry out CATPCA using IBM SPSS in combination with the supplementary 'Categories Module' (Meulman, Heiser and SPSS 2004).

To carry out CATPCA this matrix is used to optimally quantify of each variable and then calculate the variance between each of these optimally quantified variables. On the basis of this, the variables may be located in a feature space and new principal components – an axis along which the greatest variance in the data may be observed and an axis orthogonal to it – are calculated. The relationship between these principal components and the original variables is expressed numerically by a correlation coefficient termed the component loading. It is these values that provide insight into correlations between the groups of variables/linguistic features and also into the degree of variance for which each variable or group of variables is accountable, i.e. to what extent certain groups of linguistic features may

be considered characteristic of a variety. Where variables have very similar values for their component loadings, it may also be taken to indicate that they are in some way related.

The outputs of this procedure of primary interest to us here are a table of the component loadings for each variable and a plot of these values in relation to a centroid. The further the variable from the centroid (i.e. the larger the difference between the factor loading and zero) the greater the amount of variance it is interpreted as causing in the dataset.

A further important output of CATPCA analysis is a biplot of the component loadings and cases (consultants) onto a normalised two dimensional feature space. While the analysis of the component loading plot focuses on the variables exclusively, by combining this information with the distribution of consultants we are able to ascertain which groups of linguistic features are most strongly indicative of each transplanted variety.

While the plot of component loadings, the biplot and the dendrogram discussed above are clear graphical presentations of the outputs of the statistical techniques of clustering and CATPCA, they do not in themselves present explicit answers to the questions posed at the beginning of this chapter and the ultimate interpretation of their linguistic significance rests on the researcher.

6.3 Feature Catalogue

Feature catalogues have commonly been compiled for earlier inductive, quantitative studies into language variation. Examples include Grieve et al's (2011) catalogue of 40 primarily lexical features of written language appearing in a 26 million word corpus of letters to the editor covering the whole of the USA and Wolk's (2014) examination of British English dialects on the basis of a varied 45 item feature catalogue. The vast majority of feature catalogues used in quantitative linguistic research appear to be constructed on the basis of the background knowledge of the researcher who compiles them. For particularly large scale corpora, it has

316

been suggested that 'bottom up' feature catalogues may be produced in order to further reduce researcher bias in their compilation (cf. Wolk and Szmrecsanyi 2016). Since these appear to require large-scale POS tagged corpora, though, the creation of such a feature catalogue is not possible within the scope of the current project.

Here, rather, we suggest a catalogue of 33 features. Despite other catalogues generally including a slightly larger number of features than the catalogue created for this work, we note that in the context of specifically Korean dialectology 33 features may be considered very extensive[27].

Thus, the catalogue proposed here should offer both a strong linguistic signal of the characteristics of the transplanted varieties and an exceptionally nuanced impression of the aggregate (dis)similarity between them. The features of which the catalogue composed and their encodings (for convenience in interpreting the following plots) are presented in list form below:

Phonological features

1. Alveolar /c/   -   ts
2. Trilled /r/   -   r
3. Word initial /r/ and /n/   -   headsound

---

[27] Shinpei's influential establishment of the conventionally accepted contemporary dialect areas of the Korean Peninsula was made on the basis of only 7 phonological features, 5 vocabulary features and 2 morphological features (c.f. Choi 2001 p. 375) and only very limited features have been used in studies establishing sub-dialects of the ancestral varieties of the transplanted varieties, (c.f. Lee 2003 p. 455). The two most extensive feature catalogues used to identify varieties of the Northeastern dialect zone in dialectological works to date are Lee Ju Haeng's examination of the Yukchin variety (2005), which took into account thirteen features and Kwak Chung Gu's description of traditional Hamgyeong dialect (2001) in which nineteen isoglosses are considered. We further observe that the precise criteria by which linguistic features are selected for inclusion in any of the above feature catalogues of earlier studies are not entirely clear.

4. /n/ dropping before   /i/   -   ndrop
5. Monophthongal /wi/ and /oy/   -   monoph
6. Merger of /ay/ and /ey/   -   Emerger
7. Non-etymological /v/   -   v
8. /k/ to /p/ (e.g. *caybi*)   -   KtoP
9. Synchcronic Palatalisation of /k/   -   Kpal

Nominal morphological features

10. Presence of Subject Particle Form *-ka*   -   ka
11. Presence of Subject Particle Form *-ika*   -   ika
12. Apocope of final /l/ in Object Particle Form   -   OBJru
13. Apocope of final /n/ in Topic Particle Form   -   TOPnu
14. Epenthetic final /u/ in Topic Particle Form   -   TOPunu
15. Dative {ru}   -   DATru

Verbal morphological features

16. Yukchin Endings   -   Yukchin
17. Use of –*a/eyo* Endings   -   eyo
18. Extended use of *-ci*   -   haci
19. Long form negation   -   LongNeg
20. Separable negation   -   SepNeg
21. Use of –*anai*   -   anai
22. Phonological Reduction of –*myen*   -   mwuif
23. Use of –*mnita/-supnita* Endings   -   hamnida
24. Use of *-mda/-sumda* Endings   -   hamda
25. Use of –*killay*   -   killay
26. Use of –*seri/-myenseri*   -   seri
27. Subject Honorification   -   SubjHon

Lexical features

28. Charcteristic kinship terms   -   kinship
29. Archaic demonstrative pronouns of place -   ingikengi
30. /k/-stem nouns   -   Kstems
31. Productive use of *cir*   -   cil

Grammatical features

32. Rightward movement   -   RightMov
33. Extended Use of –*ki*   -   ki

The basis for the inclusion of these features in this feature catalogue is that they have all been

identified, either in prior research or in the foregoing CA, as characteristic of the dialects of the

North eastern dialect zone of the Korean peninsula or specifically VYK or KM. By recording the presence or absence[28] of each these features in vernacular speech of each of the VYK and KM speaking consultants who provided the primary data for this project, we shall be able to determine the extent to which VYK and KM may be distinguished on the basis of their aggregate linguistic (dis)similarity and which of the features contributing to that is most salient in distinguishing these varieties. We briefly review each feature in turn below and then go on to present the results of the clustering and CATPCA.

6.3.1 Alveolar /c/

While the phonemes /c_{YK}/ and /c_{KM}/ were found to be near equivalents, the frequency with which /c_{YK}/ was realised as [t͡s] or [z] appeared to be markedly higher than that of /c_{KM}/. Any realisation of /c_{YK}/ or /c_{KM}/ as [t͡s] or [z], i.e. as an alveolar affricate or fricative rather than a palatal affricate such as [t͡ʃ] or [d͡ʒ] is considered indicative of this feature.

6.3.2 Trilled /r/

The phonemes /r_{YK}/ and /r_{KM}/ were identified as only partially equivalent. Realisation of these phonemes as the trill [r] rather than tap [ɾ] or lateral approximant [l], especially in word final position, is indicative of the presence of this feature.

6.3.3 Word initial /r/ and /n/

The CSK phonological rule *twuumpepchik* restricts the placement of /r/ and /n/ in word initial position. Where these phonemes appear in this word initial position (e.g. in such Sino-Korean words as YK *lihayhata* 'to understand' cf. CSK *ihayhata*) in the transplanted varieties, we

---

[28] Rather than simple omission, here absence connotes the use of an alternative linguistic form in the primary data.

consider this feature to be present. We further note that the appearance of these phonemes in initial position in loanwords is not considered indicative of this feature.

6.3.4 /n/ dropping before /i/

Consultants are considered as having this speech if there is even a single instance the deletion of /n/ prior to /i/ with or without accompanying compensatory nasalisation. The deletion of /ng/ in positions preceding /i/ or /y/ is not taken into account.

6.3.5 Monophthongal /wi/ and /oy/

To establish the presence of this feature, each consultant's attestations of sounds corresponding to CSK /oy/ and /wi/ are examined. Rather than any monopthongal realisations, for example realisations as [i] or [ɛ] may be caused by glide deletion processes acting on dipthongs, only instrumentally verified realisations of either of these phonemes as the high front vowels [y] or [ø] are considered indicative of this feature.

6.3.6 Merger of /ay/ and /ey/

The CA in section 5.1.3.2 above concluded that the phonemes /ay/ and /ey/ were still distinguishable in both KM and VYK on the basis of the average F1 and F2 values taken across a large number of consultants. When both of these mean formant values for a single consultant's realisations of phonemes corresponding to CSK /ay/ and /ey/ are very close (within 100 Hertz) we consider this feature to be present.

6.3.7 Non-etymological /v/

This feature is counted as present when the sounds corresponding to CSK /w/ are realised as [v]. The appearance of this sound in loanwords is not taken into account.

6.3.8 /k/ to /p/

This feature is considered present when sounds corresponding to CSK /k/ are realised as [p] or [b], especially in intervocalic position, e.g. KM [t͡ʃɛbi] cf. CSK *cagi* 'oneself', KM [t'arbi] 'raspberry' cf. CSK *ttalki* 'strawberry', KM [kadorobi] c/f/ CSK *katolok* 'the further one goes'. King (1992 pp. 205-206) also recorded cases of the CSK velar nasal /ng/ corresponding to realisations of [p] and [b] in KM, but this phenomenon was not attested in this project's primary data.

6.3.9 Synchronic Palatalisation of /k/

This feature is signalled by the realisation of /k_YK/ or /k_KM/ as [t͡ʃ] before the high front vowel /i_YK/ or /i_KM/ or the glide /y_YK/ or /y_KM/, for example the realisation of CSK *kilum* 'oil' as [t͡ʃirɨm].

6.3.10 Presence of Subject Particle Form –*ka*

This feature is indicated by the use of a subject particle with the form [ga] appended to nouns with final vowels. It is considered absent where the invariant subject particle is used with the form [i], even after vowel final nouns. Only the active use of this form is taken as characteristic of this feature rather than the exclusive use of the subject particle form –*i* following nouns which end with consonants and the dropping of the subject particle where the allomorph –*ka* would be expected.

6.3.11 Presence of Subject Particle Form –*ika*

This use of a subject particle form [iga], either following nouns with final consonants or those with final vowels is considered characteristic of this feature. While it is yet to be established whether this is a novel form of the subject particle which appears in specific phonologically or semantically conditioned environments or simply a case of subject double marking, this is not

taken into account when deciding on the presence of this feature in a consultant's speech. Only the use of the form [iga] is relevant.

6.3.12 Apocope of final /l/ in Object Particle Form

The use of object particles with any of the forms [ɨ], [rɨ] or [rɨ] is sufficient to indicate the presence of this feature in a consultant's speech.

6.3.13 Apocope of final /n/ in Topic Particle Form

This feature may be marked by the use of any of the forms [ɨ], [nɨ], [ɨ̃] or [nɨ̃] as topic particles.

6.3.14 Epenthetic final /u/ in Topic Particle Form

The realisation of the topic particle as [ɨnɨ] is the marker of this feature.

6.3.15 Dative Use of Particle Form *–ru*

This feature is considered to be present when the forms [rɨ] or [rɨ] are used for such functions as marking indirect objects such as animate beneficiaries (cf. 5.2.1.2 above), the goals of verbs of motion. For this feature to be considered present, it is not required that these forms are used to the exclusion of forms similar to CSK *–ey* or *–eykey*, rather that they appear in positions where the use of an object particles would be considered ungrammatical in CSK.

6.3.16 Yukchin Endings

Consultants were classified as having this feature in their speech if they either used the Yukchin formal declarative or interrogative endings *–(su)kkwuma* or *–(sum)dwu* in their natural speech or if it was modelled in the section of the interview pertaining to meta-linguistic knowledge.

6.3.17 Use of *–a/eyo* Endings

Any use of endings corresponding in form to CSK *–a/eyo* in any tense and regardless of the application of vowel harmony is considered indicative of this feature. We include this feature here since it is a contributing factor in the classification of neutral declarative ending as partially rather than nearly equivalent.

6.3.18 Extended use of *–ci*

Sentence final endings with this form may be found in both transplanted varieties and in CSK. In contrast with CSK, its usage in VYK and KM does not denote a tag question or that "the listener should know or be aware of the content of the remark" (Yeon and Brown 2011 p. 379). Although prior descriptions of the transplanted varieties (and the parallel descriptions presented in section 5.2.5.5) do not classify this ending as a declarative ending, it does appear to be used interchangeably with such neutral declarative endings as *–o/-so* and *–a/eyo* in some contexts. Alternative explanations present it as a narrative or habitual ending (King 1987a p. 261), an innovative ending of unknown function (King 1992 p. 211) or a member of a separate class of endings designated 'commonly used ending' (*pemyongemi*) (Kwon 2010). Where this usage is encountered this feature is considered to be present in a consultant's speech.

6.3.19 Long form negation

Any attestation of the constructions corresponding in form to CSK *–ci antha* or *–ci mos hata* in a consultant's speech is considered characteristic of this feature.

6.3.20 Separable negation

This feature is marked by the separation of verbs in a serial verb construction by a negative particle. For example, if constructions corresponding to CSK *mos alatutta* 'to not comprehend'

or *mos meke poda* 'to not have an opportunity to sample (a food)' were realised instead as *ala mos tutta* or *meke mos poda* this feature would be considered present.

6.3.21 Use of *–anai*

The use of the variant form [anai], which corresponds functionally to the CSK ending *–(u)nikka* is the characteristic of this feature. The use of reduced forms of this ending, for example those which result from the deletion of /n/ before /i/ are not taken into consideration.

6.3.22 Phonological Reduction of *–myen*

This non-sentence final ending in particular appears to have undergone extensive phonological reduction in both transplanted varieties. We consider the reduced form to be present in a consultant's speech minimally when both the glide and final nasal are both. The quality of the final vowel, or indeed its deletion, is not relevant to determining whether this feature is present in a consultant's speech. Forms which delete the final nasal and vowel entirely are also considered to be indicative of this feature.

6.3.23 Use of *–mnita/-supnita* Endings

The inclusion of this feature contributes to assessing the extent to which CSK has influenced the divergence between the transplanted varieties. Use of polite predicate endings which are formally near identical to CSK in any tense is the marker of this feature.

6.3.24 Use of *-mda/-sumda* Endings

Their variation in form contributed to the evaluation of the formal declarative endings of VYK and KM as only partially equivalent (cf. section 5.2.9 above). Phonological reduction to [mda], [sɨmda] is the minimal criterion for the identification of this feature in a consultant's speech. More extensive phonological reduction, for example to [m] or [ɨm] is also considered to mark the presence of this feature.

6.3.25 Use of –*killay*

While this connective ending is formally near identical in CSK and both transplanted varieties, its usage varies. We consider this feature to be present when it has lost its volitional meaning or when it is used to express cause or reason for an action taken not by the speaker, but by a third person.

6.3.26 Use of –seri/-myenseri

The origin of especially the final syllable of these endings is not wholly clear and its appearance in a consultant's individual data may be a distinguishing feature between VYK and KM. Phonologically reduced forms of –*myenseri* (e.g. KM [mesəri]) are also considered indicative of the presence of this feature.

6.3.27 Subject Honorification

This feature is considered to be present in a consultant's speech only if they use a verbal affix to denote subject honorification. The use of specific items of honorific vocabulary, for example words corresponding to CSK *tolakasita* 'to pass away', or the use of an honorific particle corresponding to CSK –*kkeyse* is not taken into account.

6.3.28 Characteristic kinship terms

The kinship terms which have been commonly associated with KM and VYK in prior research are taken into account in assessing the speech of our consultants. They are as follows: *esi, amey, abai, azwubai* or *nwui*. Use of a single one of these is considered sufficient to indicate the presence of this feature.

6.3.29 Archaic demonstrative pronouns of place

The archaic demonstrative pronouns of place are those which retain the LMK velar nasal (cf. King 1992 p. 214. They are most commonly realised as [iŋgi] and [kəŋgi]. Use of either marks

this feature. The velar nasal is a pre-requisite of considering this feature to be present in a consultant's speech, therefore use of the reduced form [igi] is not considered indicative of this feature.

6.3.30 /k/-stem nouns

Examples of the so-called /k/-stem nouns of LMK include *kalgi* 'flour', *nayngi* 'tree' and *kwungi* 'hole' among others (cf. King and Yeon 1992). Where such forms as these are observed in the speech of a consultant, as opposed to forms such as CSK *kalwu*, *namu* or *kwumeng*, this feature is considered to be present.

6.3.31 Productive use of *cir*

The free combination of *cir* 'work' to specify a type of job (e.g. VYK [sənsɛŋnimt͡ʃil] with other nouns or its use as an independent noun is characteristic of this feature.

6.3.32 Verbally Post-Posed Arguments

Post-posed verbal arguments are common in many vernacular varieties of Korean.   Some amount of such word order appears in the vast majority if not all of the consultants' data. It was observed by Kwon (2010) that the post-verbal ordering of subject, object and adverbial arguments is even common in CSK. While calling this phenomenon common might be overstating the case, we nevertheless acknowledge that these arguments may appear following the verb in CSK. Therefore, we only consider the feature to be present if non-subject, non-object and non-adverbial arguments appear in a consultant's speech in that position. Although the frequency with which subject, object and adverbial arguments is post-posed appears to vary greatly between consultants, this is not taken into account in determining the presence of absence of this feature.

6.3.33 Extended Use of –*ki*

There are two particular constructions in which this nominalising ending does not appear in CSK in which it has been attested in KM. Where the ending –*ki* is used in place of –*ko* or –*key* in either of the constructions *hako siphtha* 'to want to do' or *hakey toyta* 'to end up doing', respectively, this feature is classified as being present in the consultant's speech.

6.4 Results and Discussion

In this section we present the results of the statistical tests discussed above and interpret their outputs to answer the questions posed in the introduction to this chapter. We first address the clustering before moving on the CATPCA.

Below may be found a dendrogram generated by applying Ward's Method to a feature matrix recording the presence or absence of each of the 33 features discussed above in primary data supplied by 22 consultants[29] (12 VYK speakers and 10 KM speakers):

---

[29] Of the total number of consultants recorded for the purposes of collecting primary data for this project (see sections 4.3.1 and 4.3.2 above), two KM consultants and two VYK consultants were excluded from this quantitative analysis on the grounds of having provided insufficient data.
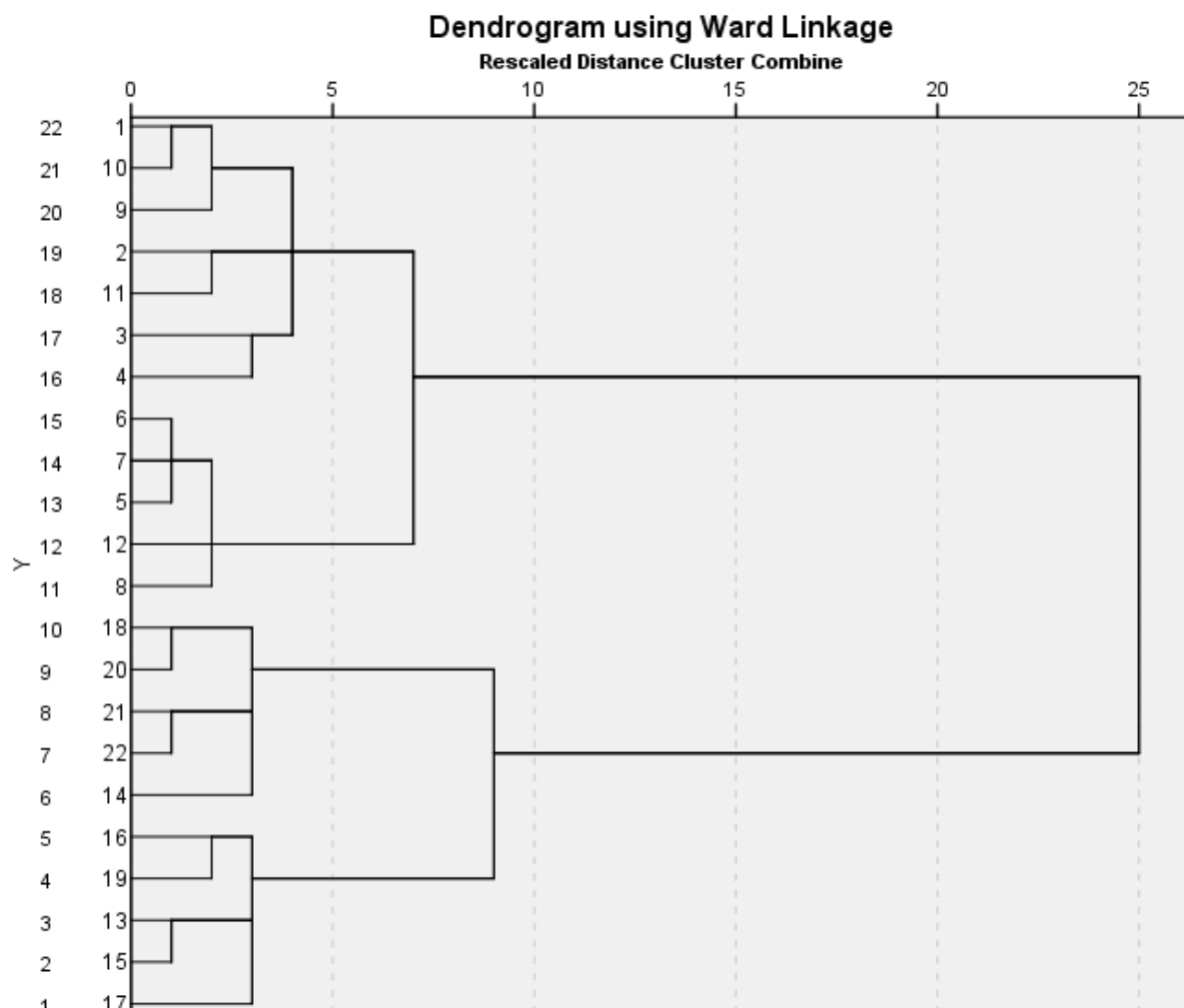
Figure 5: Dendrogram of VYK and KM Consultants

The terminal nodes on the dendrogram represent individual consultants. The nodes

numbered 1 to 12 are VYK speakers while the nodes numbers 13 to 22 are KM speakers. The

most striking conclusion to be drawn from this dendrogram is that VYK speakers and KM

speakers are clearly separated into two groups which exhibit a considerable amount of

aggregate linguistic dissimilarity as shown by the length of the branches separating the two

main clusters. This pronounced division between the VYK group and the KM cluster decisively

answers the question of whether these varieties may be differentiated on the basis of their

linguistic features in the affirmative.

Within the two large clusters, which correspond to VYK (the upper cluster) and KM (the lower cluster), we see that KM is composed of two equally sized, relatively homogeneous groups, which are nevertheless quite distinct from one another. The sub-groups which make up the VYK cluster, however, are a little more variable, although the VYK cluster as a whole may be taken to be more homogenous. The VYK smaller sub-group is exhibits less internal variation than either KM sub-group while the larger is more heterogenous.

While the cases (consultants) in the SPSS output are completely anonymised, we may compare this with the partially anonymised data which allows us to identify certain characteristics of each consultant in order to speculate about the possibility of a relationship between sociological indicators and language use. Although it is far from conclusive, a consideration of the data presented above in that light suggests that age, gender and level of education do not seem to have any correlation with the formation of either VYK or KM sub-groups.

The following CATPCA analysis, in addition to providing insights into the linguistic features behind the grouping of the entire sample in VYK and KM clusters may also shed some light on the linguistic factors behind the formation of sub-groups. Prior research has often made reference to the diversity of KM and VYK in terms the traditional peninsula dialects (e.g. Yukchin dialect, North Hamgyeong Province dialect, Pyeongan Province dialect etc.), therefore it will be worth taking note of whether features which are particularly strongly associated with particular traditional peninsula dialects are implicated in intra-group variation.

We now turn to the CATPCA analysis to determine which of the 33 features were most significant in determining these clusters. Below, we present a table specifying the component loadings of each feature in two dimensions (i.e. along the two principal component) with an accompanying plot to visualise these values and the relationships between them. The values

in the first column are represented along the x-axis as 'dimension one' and the values in the second column represented along the y-axis as 'dimension two'.

Component Loadings

	Dimension	
	1	2
ts	.756	.251
r	-.851	-.403
headsound	.598	.461
ndrop	-.274	.262
monoph	.435	.312
Emerger	.021	.155
v	-.857	-.239
KtoP	-.704	.122
ka	.765	.480
ika	.189	.406
OBJru	-.077	.406
TOPnu	-.280	.450
TOPunu	-.146	.806
Yukchin	-.443	.217
eyo	-.154	-.058
haci	-.625	.327
mwuif	-.357	.800
anai	-.715	.322
hamda	.709	.489
hamnida	.200	-.120
seri	-.226	.709
kinship	-.601	.375
ingikengi	-.631	.094
cil	-.334	.145
killay	-.323	.506
LongNeg	.786	.331
ki	-.505	.323
Kstems	-.556	-.100
Kpal	-.619	.183
SubjHon	.573	.137
DATru	-.695	-.153
RightMov	-.737	.216
SepNeg	-.386	.224

Variable Principal Normalization.
Table 90: Component Loadings of Features Analysed in CATPCA Run 1
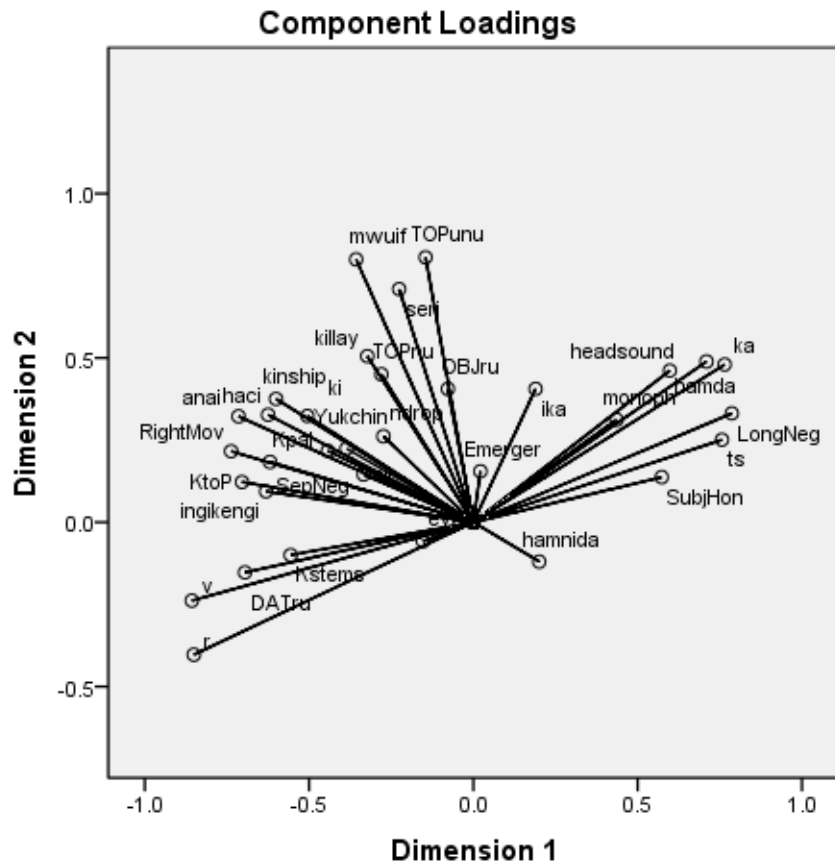
Figure 6: Plot of Component Loadings of Features Analysed in CATPCA Run 1

While it is perhaps possible to impressionistically identify groups of features which together contribute towards the differentiation of the transplanted varieties using the plot and table presented above, this may be clarified by excluding those features which account for minimal variance in the data. The table below, as its title suggests, displays the information which allows us to determine which features to exclude in an additional run of the CATPCA test.

Variance Accounted For

	Centroid Coordinates			Total (Vector Coordinates)		
	Dimension		Mean	Dimension		Total
	1	2		1	2	
ts	.572	.063	.318	.572	.063	.635
r	.724	.162	.443	.724	.162	.886
headsound	.357	.213	.285	.357	.213	.570
ndrop	.075	.068	.072	.075	.068	.143
monoph	.189	.097	.143	.189	.097	.287
Emerger	.000	.024	.012	.000	.024	.024
v	.734	.057	.395	.734	.057	.791
KtoP	.496	.015	.256	.496	.015	.511
ka	.585	.230	.408	.585	.230	.815
ika	.036	.165	.100	.036	.165	.201
OBJru	.006	.165	.085	.006	.165	.171
TOPnu	.078	.203	.141	.078	.203	.281
TOPunu	.021	.649	.335	.021	.649	.671
Yukchin	.197	.047	.122	.197	.047	.244
eyo	.024	.003	.013	.024	.003	.027
haci	.390	.107	.249	.390	.107	.497
mwuif	.127	.639	.383	.127	.639	.767
anai	.511	.104	.307	.511	.104	.615
hamda	.503	.239	.371	.503	.239	.742
hamnida	.040	.014	.027	.040	.014	.055
seri	.051	.503	.277	.051	.503	.554
kinship	.361	.141	.251	.361	.141	.502
ingikengi	.398	.009	.204	.398	.009	.407
cil	.112	.021	.066	.112	.021	.133
killay	.104	.256	.180	.104	.256	.360
LongNeg	.619	.110	.364	.619	.110	.728
ki	.255	.105	.180	.255	.105	.360
Kstems	.309	.010	.160	.309	.010	.319
Kpal	.383	.034	.208	.383	.034	.417
SubjHon	.329	.019	.174	.329	.019	.347
DATru	.483	.023	.253	.483	.023	.507
RightMov	.544	.047	.295	.544	.047	.590
SepNeg	.150	.052	.101	.149	.050	.199
Active Total	9.764	4.594	7.179	9.764	4.593	14.356

Table 91: Variance Accounted for by Features Analysed in CATPCA Run 1

The common criterion for exclusion from the second run of CATPCA is being a variable with a mean value (the sum of the component loading values divided by two) of less than 0.100. This value is displayed in the table above in the fourth column from the right. According to the above table we may exclude the following features: /n/ dropping before /i/, merger of /ay/ and /ey/, apocope of final /l/ in object particle, use of –a/eyo endings, extended use of –ci, and use of –pnita/-supnita endings. It may be observed that the exclusion of these features does not represent an assessment on how typical they are of VYK or KM, but rather reflects their discriminatory power with regard to these transplanted varieties. Thus, while these features may very well be distinctive of these varieties in the broader context of Korean dialectology, they do not count for a great deal of variance between the transplanted dialects. In other words, they are relatively consistently present in the speech of our VYK and KM consultants.

The table and plot displaying the component loadings of the new 27 feature catalogue, which were generated having excluded the above specified six features, are displayed below. Changes in values of each variable in both dimensions – universally increases in value – may be observed:

Component Loadings

	Dimension	
	1	2
ts	.763	.279
r	-.853	-.396
headsound	.621	.386
monoph	.435	.350
v	-.856	-.237
KtoP	-.705	.159
ka	.771	.485
ika	.211	.428
TOPnu	-.246	.453
TOPunu	-.126	.813
Yukchin	-.433	.214
haci	-.603	.265
mwuif	-.344	.812
anai	-.695	.309
hamda	.726	.462
seri	-.219	.769
kinship	-.579	.384
ingikengi	-.637	.141
killay	-.312	.512
LongNeg	.791	.323
ki	-.493	.313
Kstems	-.576	.005
Kpal	-.630	.247
SubjHon	.577	.104
DATru	-.705	-.075
RightMov	-.748	.273
SepNeg	-.391	.218

Variable Principal Normalization.

Table 92: Component Loadings of Features Analysed in CATPCA Run 2
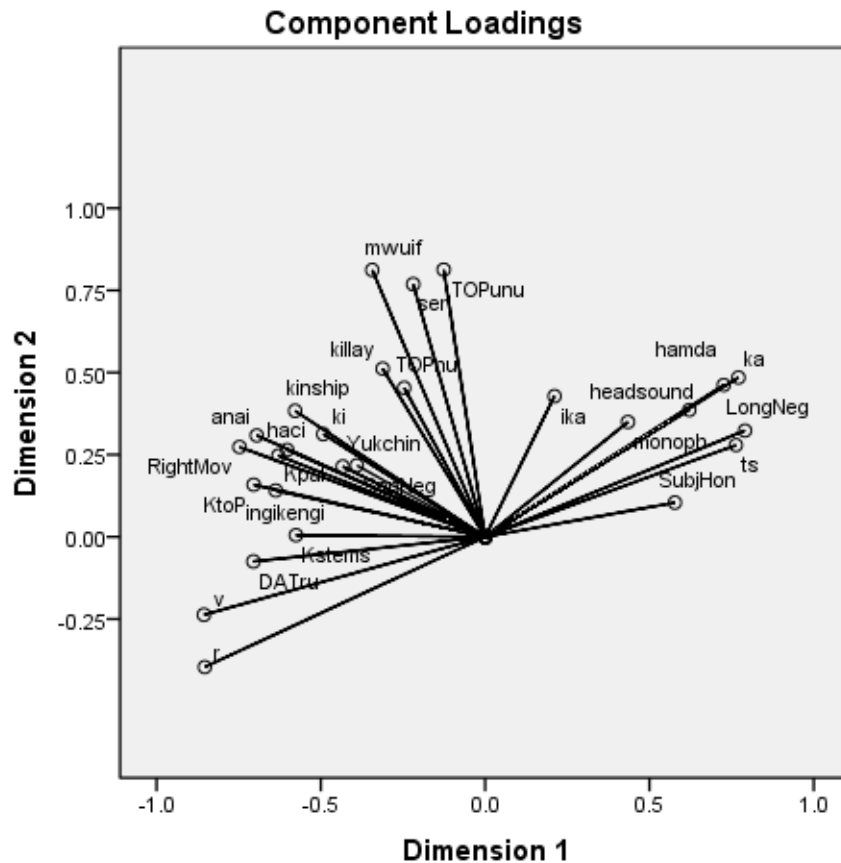
Variable Principal Normalization.

Table 93: Plot of Component Loadings of Features Analysed in CATPCA Run 2

The above graphical representation of this information provides a much clearer picture of the distribution and grouping of the variables/linguistic features in a two dimensional feature space. Along the first dimension, we may observe that the features of alveolar realisation of /c/, word initial /r/ and /n/, subject particle /ka/, used of –mda/sumda endings, long form negation, subject honorification and the monopthongal realisation of /oy/ and /wi/ are grouped together relatively compactly. We see a comparatively large number of features at the other, negative end of the first dimension, but these are not grouped as neatly in that they range widely in the second dimension. Relying on the table rather than the plot for our interpretation, we may identify two further subgroups which are strongly implicated in the variance of the sample: first, eleven features which are at a distance of greater than 0.5 from

the centroid on the first dimension and second a group of four features at a distance of less

than 0.5 from the centroid in the first dimension, but greater than 0.5 from the centroid along

the second dimension. In order to determine the roles that these groups of features may play

in distinguishing VYK from KM, we may examine a biplot which juxtaposes the consultants (still

labelled from 1-22 with the same division into KM and VYK speakers as laid out above) with the

features in the same feature space. The scale is adjusted in this biplot and its units arbitrary,

but, where relevant, for ease of identification of groups of variables, we shall continue to make

reference to their position as defined on the plot showing component loadings above.
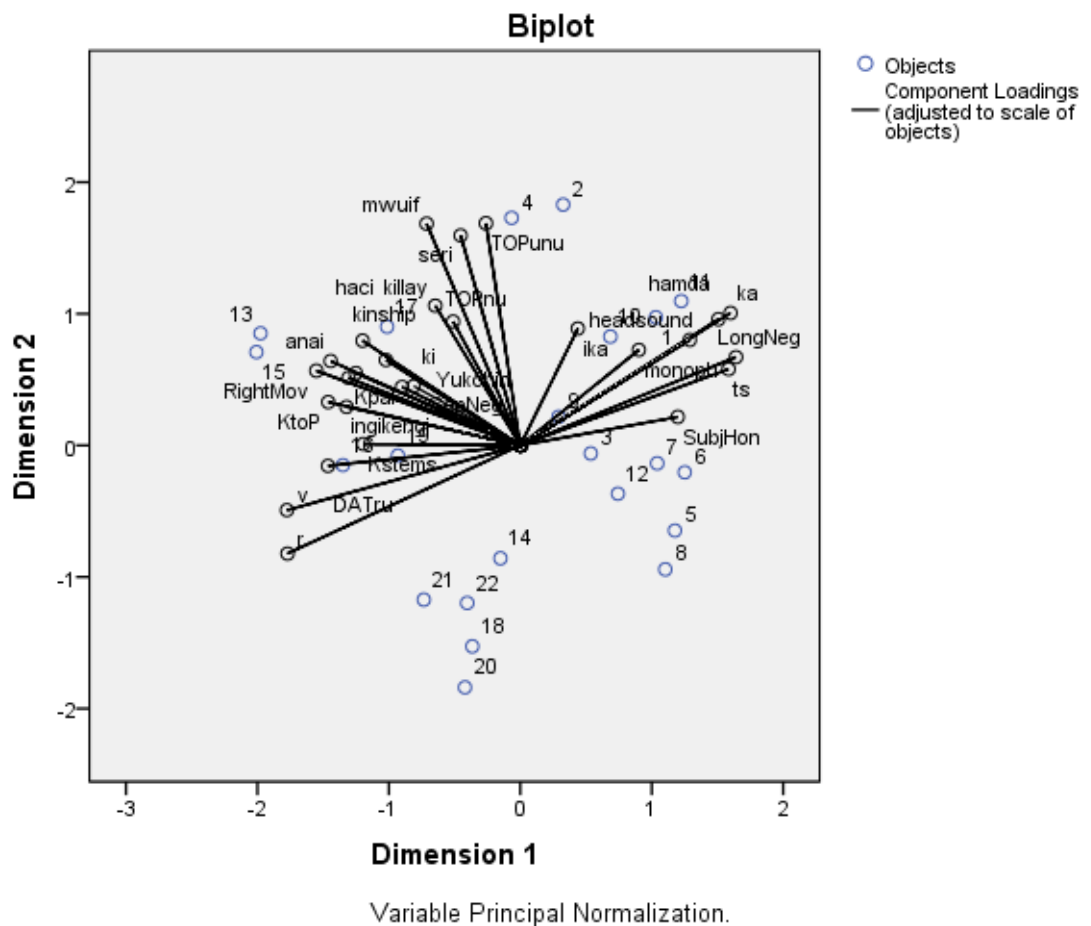
Table 94: Biplot of Component Loadings and Consultants Analysed in CATPCA Run 2

This biplot strongly suggests that the first group features that was characterised above as

being greater than +0.5 in distance from the centroid along the first dimension may be

considered distinctive of VYK speech, and, indeed, any features which are located a positive distance from the centroid along dimension one (i.e. along the first principal component) contribute somewhat towards the linguistic distinctiveness of VYK. We base this conclusion on the fact that cases 1 to 12 (VYK speakers) are all to the right of the centroid (i.e. a positive distance from the centroid along the first dimension) and the majority of them may be characterised as being loosely clustered around the same area as the seven features with the furthest positive distance from the centroid.

KM speakers (cases 13 – 22), on the other hand, appear to be clearly divided into two clusters. One of these is clustered around the large group of eleven features which are at a distance of greater than -0.5 from the centroid along the first dimension. The other group, consisting of consultants 14, 18, 20, 21 and 22, while tightly grouped does not seem to be associated with particular linguistic features in the first dimension. Turning back to the feature matrix, we may account for their position by the absence of certain features from the data which they provided rather than their presence. Specifically, the group of features identified as having a large positive distance from the centroid along the second dimension are present in the other KM consultants' data (and the data of several VYK consultants), but absent from these consultants' data. Furthermore, the speech of all of the members of this cluster does not exhibit two features which are a large negative distance from the centroid along the first dimension (synchronic palatalization of /k/ and rightward movement) and a the majority do not exhibit the labialization of /k/ to /p/, use of archaic /k/ stem nouns or use of characteristic kinship terms. It is the absence of these two sets of features from the speech of consultants 14, 18, 20, 21 and 22 combined which accounts for the position of this cluster as separate from the other KM consultants rather than the presence in their speech of any particular distinctive linguistic feature common to this group.

In more dialectological terms, the foregoing analyses may be summarised as follows. Despite their common dialectological heritage, the transplanted varieties of Korean spoken in Central Asian and the YKAP are synchronically linguistically distinct. The linguistic features which distinguish VYK from KM are alveolar realisation of /c/, word initial /r/ and /n/, subject particle /ka/, used of –*mda/sumda* endings, long form negation, subject honorification and the monopthongal realisation of /oy/ and /wi/. Linguistic or extra-linguistic features underlying the internal variation of VYK cannot be determined. The distinguishing features common to KM speech are trilled /r/, non-etymological /v/, use of –*anai*, use of archaic demonstrative pronouns of place, extended use of the ending -*ci* and use of dative particle –*ru*. A further sub-group of KM speakers may also be identified on the basis of five linguistic features. On the level of their vocabulary by their use of archaic /k/-stem nouns and characteristc kinship terms. In terms of phonology, this group is further distinguished by the synchronic processes of /k/ labialisation to /p/ and /k/ palatalization before high front vowels. Finally, their non-canonical, unusual verbally post-posed arguments also contribute to their distinction form other speakers of KM. There are underlying factors in terms of sociological indicators or dialectological background which clearly correlate with the KM sub-groups.

6.5 Summary

In this chapter we have undertaken a quantitative exploration of the data upon which the preceding parallel descriptions of the transplanted varieties and their Contrastive Analysis were based. The two questions of whether VYK and KM could be differentiated on the basis of their linguistic features and which linguistic features might prove to be distinctive between the two were raised, and statistical tests for answering these questions were identified. A principled feature catalogue of 33 linguistic features found to vary over the transplanted varieties was drawn up and the tests were carried out on data drawn from 22 consultants coded for the presence or absence of these features.

The clear result of the cluster analysis is that, despite the high degree of near equivalence established by the CA, KM and VYK may indeed be distinguished on the basis of their linguistic features. The supplementary CATPCA established that of the 33 features included in our analysis the greater part of the variance in the sample was accounted for by just thirteen features. Of these features, seven in particular may be considered distinctive of VYK (those with relatively large positive factor loadings in the first dimension) and a further six may be considered characteristic of KM (those with a relatively large negative factor loading in the first dimension).

While a certain amount of internal variation in KM and VYK was observed in this data, this analysis does not allow us to identify the linguistic factors underlying the grouping of the consultants included in the VYK cluster. We have, however, identified differences in the linguistic features evident in the two sub-groups of KM speakers.

Turning from the purely linguistic distinctions between the sub-groups observed within the varieties, we acknowledge that we are not able to identify any underlying factors behind this variation in terms of the sociological backgrounds of consultants. Also, tellingly, in contrast with the assertions of prior research, we are not able to conclude that intra-varietal variation may be accounted for in terms of the distinctive features of traditional Korean dialectology (i.e. those features traditionally used to distinguish Hamgyeong dialect from Yukchin dialect).

In the following concluding chapter of this thesis, we present a full summary of our findings and discuss them in the broader context of Korean dialectology and variationist linguistics.

7 Conclusion

At the outset of this thesis we stated that language varies in time and space. We went on to identify a particular aspect of this universal characteristic of language, variation in the geographically diffuse transplanted varieties of Korean which originate in the North east of the Korean peninsula and formulate specific questions concerning their synchronic relationship. We re-state these questions below:

1. What is the form/are the forms of Korean currently in use in China and Central Asia?

2. Do these forms of these varieties differ from one another and, if so, to what extent?

3. If these transplanted varieties can be shown to be distinct from one another, by what linguistic features may we distinguish them?

In addition to providing comprehensive reviews of the socio-historical backgrounds of the communities by which these varieties are used and the research pertaining to them, we provided both general and specific answers to each of these questions, revealed through a qualitative (CA) and quantitative (clustering, CATPCA) analysis of a unique collection of primary data.

We now go on to in this final section to summarise the findings of these analyses, situate them within the broader academic disciplines to which they are relevant and provide suggestions for future directions for research.

7.1 Summary of Findings

The findings of this thesis can be divided into three parts: the conclusions arrived at as a result of conducting a thorough and critical literature review, the results of the contrastive analysis and the results of the supplementary quantitative analysis. We deal with each of these in turn.

The main findings of the literature review concerned the scope and practices of Korean dialectology. The relative paucity of available research on transplanted varieties of Korean and

its near total non-integration with mainstream Korean dialectology (the recent overview by Yeon and Brown (2015) being a notable and welcome exception) highlighted the timeliness and necessity of this research. The exclusive focus of this research on the Korean Peninsula and its surrounding islands is echoed in the methodologies which underlie much of the extant research on the transplanted varieties in that there is a marked tendency for them to be explained mostly with reference to the dialectological heritage of their speakers, rather than their synchronic features. Furthermore, the reliability of certain common sampling and data-gathering procedures was called into question. We examine the contribution made by this thesis towards improving these situations more deeply in the following sub-section.

The results of the CA were summarised in tabular form in sections 5.1.2, 5.1.4, 5.2.3, 5.2.4 and 5.2.6, but in addition to the equivalences established between the large number of comparative concepts, this section produced a high-quality, fine grained parallel description of VYK and KM which provided a nuanced impression of the synchronic features of each variety and extensive examples. Returning to the equivalences, here we re-iterate that of the large number of equivalences established between the VYK and KM instantiations of the CSK comparative concepts, the vast majority were near equivalents. While the similarity of these varieties is implicit in their frequent conflation, systematically establishing their points of similarity in addition to their points of dissimilarity is a unique feature of this CA.

In general, that is, in the sense that the similarities between VYK and KM outnumber their differences, and in many specific points (e.g. extended used of –*ki* in KM, new Sino-Korean coinings in VYK etc.) the findings of the CA reflect the findings of prior research, thereby strengthening the empirical basis of earlier work. In more detail, the CA revealed the following:

- In terms of phonology VYK and KM are very similar. The only categorical differences found were the allophonic realisations of the sounds corresponding to

CSK /l/ and /w/ as [r] and [v] in KM but not VYK. We further impressionistically

identified some differences in frequency of some realisations and phonological

processes, e.g. the more frequent realisation of /c_{YK}/ as $\widehat{[ts]}$ in comparison to /c_{KM}/

and more commonly occurring phonological reduction in KM.

- The case particles and delimiters of the transplanted varieties were generally very

  similar. Most significantly, the subject particle of KM did not exhibit

  phonologically conditioned allophony, honorification was not reflected in the case

  marking of KM, and the particles corresponding to CSK –*hanthey* and –*pota* were

  found to be non-equivalent between KM and VYK.

- The forms of the verb endings of VYK and KM appeared to be considerably more

  divergent than the particles in that cases of partial equivalence (generally as a

  result of considerably divergent forms) were numerous. We must add the caveat,

  though, that many of these contrasts were formulated using secondary data which

  may only represent the most extreme forms of local usage. Nevertheless we may

  summarise crucial points of difference by observing that a subject honorification

  pre-final ending was found to be present in VYK but not KM, forms corresponding

  to CSK –*a/eyo* were found to be in much more widespread use in the YKAP while

  the so-called Yukchin formal verb endings enjoyed much wider use in Central Asia.

- As with the phonological level, KM and VYK were found to be very similar on the

  level of basic word order. Both varieties have a basic head-final SOV word order in

  which modifier precedes modified.   In terms of the features discussed in depth in

  section 5.3.3, only the use of the nominalizer corresponding to CSK –*ki* clearly

  delineates the transplanted varieties in that its use is broader in KM than VYK. The

  question of whether KM has freer word order is still somewhat open. Although

  post-posed verbal arguments appear in all varieties of Korean, it is the appearance

of indirect objects such as goals of verbs of motion and comitative complements in this position which appear to set KM apart from VYK. In addition, we may impressionistically report that the frequency of post-posed verbal arguments appears to be much greater in KM. Lastly, in contrast with prior research, we found that seperable patterns of negation previously only reported in KM were also present in VYK, and that the use of long form negation appeared to be dramatically more widespread in VYK.

- The clearest point of distinction between the transplanted varieties in terms of their vocabulary is the influence that each has received from their respective main contact language. We also observed that the honorific vocabulary of VYK appears to be slightly more elaborated than that of KM. On the other hand, KM appeared to exhibit more archaic forms of vocabulary than VYK.

We now turn to the results of the quantitative analysis. The CA provided us with a detailed picture of the linguistic forms being used in Central Asia and the YKAP, many of which were found to be shared. From this, however, it was not clear which of these variable shared features were in wider use in each transplanted variety and the picture of linguistic dissimilarity between the varieties which it provided was impressionistic only. By examining each consultants' speech for the presence or absence of 33 variable features we were able to determine the distribution of these features across a sample of speakers drawn from each survey site and determine both whether this aggregate distribution of features was distinctive for the transplanted varieties and which among these features contributed most to distinguishing between these varieties. Concretely, our findings may be summarised as follows:

- Clustering reveals that KM and VYK are clearly distinct from one another on the basis of the feature catalogue established in section 6.3. Sub-groups of speakers of each may also be identified but on the basis of the data available to us we are

not able to infer whether there is any regular underlying sociological factor accounting for this.

- The greater part of the variance between VYK and KM may be accounted for by thirteen features.

- There are seven features which appear frequently in the speech of VYK consultants, but not KM consultants: alveolar realisation of /c/, word initial /r/ and /n/, subject particle /ka/, used of *–mda/sumda* endings, long form negation, subject honorification and the monopthongal realisation of /oy/ and /wi/.

- There are six features which appear frequently in the speech of KM consultants but not VYK consultants: trilled /r/, non-etymological /v/, use of *–anai*, use of archaic demonstrative pronouns of place, extended use of the ending *-ci* and use of dative particle *–ru*.

- The linguistic features which distinguish the sub-groups established for VYK could not be identified, but in the case of KM the following features appear more frequently in the speech of one of the sub-groups: use of archaic /k/-stem nouns, use of archaic kinship terms, /k/ labialisation to /p/, /k/ palatalization before high front vowels and unusual verbally post-posed arguments. In contrast with prior studies, these features which distinguish the linguistic sub-groups of KM do not map back onto the linguistic features considered representative of the traditional dialectological divisions between the North Hamgyeong Province dialect and the Yukchin dialect.

We now go on to contextualise these results and the methods by which we arrived at them in the broader disciplines of Korean Studies and linguistics before bringing this thesis to a close by exploring how its central topic and the further questions which were raised in its production may be explored in future.

7.2 Contribution to Korean Studies

A necessary first step in carrying out this research was a radical re-conceptualisation of Korean dialectology to incorporate transplanted varieties in their own right rather than as extensions of the traditional peninsula varieties. While this work is not the first to do so (cf. King and Yeon 1992; Jin 2008; Silva 2010) it nevertheless represents a further challenge the hegemonic narrative of the centrality of peninsula varieties in Korean dialectology and thereby contributes to the re-orientation of the entire field away from its foundational and much disputed questions of how many dialects the language spoken on the Korean peninsula may be divided into and what the distinguishing linguistic features of such dialects might be. Rather it implicitly presents a vision for Korean dialectology which incorporates the contemporary reality of a globalised Korean language and identity.

Concretely, however, the findings of this research have much in common with traditional Korean dialectology. While differences between the transplanted varieties were identified in chapter 5 and the salience of these differences in distinguishing them was demonstrated in chapter 6, we nevertheless acknowledge the existence of a relatively high degree of similarity between these varieties. Due to a lack of available contemporary descriptions or primary data, it is not certain how similar they are with the contemporary varieties spoken by the people now resident in the Northeast of the Korean peninsula, but dialect materials from the early 20th century (Ogura 2009 [1944]) seem to indicate impressionistically that, of the varieties spoken on the Korean peninsula, these are the most similar to the transplanted varieties discussed here. As the region from which the seed communities of both the *Koryo saram* and *Cosencok* originally migrated, and therefore the point of origin for their varieties of Korean, this is perhaps to be expected. Such a conclusion is in-line with the assertions of traditional Korean dialectology.

Two of the techniques employed in gathering this data represent relatively novel, more ecologically valid and naturalistic (but not wholly natural) forms of data gathering in Korean dialectology. The first of these is the staging communicative events in the form of semi-structured interviews as opposed to direct elicitation sessions. Despite the somewhat inconsistent nature of our data with regard to the appearance of certain forms (e.g. informal imperatives and propositives), such techniques mean that we were able to examine non-lexical variation in VYK and KM without relying on the meta-linguistic knowledge of consultants and without the danger of direct elicitation leading to consultants over-correcting and providing more 'authentic' dialect features which are not actually in common contemporary use. An additional advantage of more naturalistic, vernacular language data is that it may be subjected to a wide variety of analytical techniques and thereby contribute to future research in a way which data that is elicited for a very specific purpose cannot.

While this approach to data gathering is rare in Korean dialectology, it must be acknowledged that our study is not unique this regard. Connected speech data from interviews also formed the basis of Kwon's description of KM (2010), for example. Here, though, we must highlight the second contribution of this thesis to the methodology of Korean dialectology, which is the inclusion of a broad sample of self-identified speakers of the varieties under investigation. To again take Kwon's description of KM as an example, this work was created on the basis of interview data gathered from a single consultant yet presented as representative of the entire variety.

The self-identification criterion for including consultants in this study is also radically different from prior research. Previously, consultants providing data for research into transplanted varieties have been selected almost universally on the basis of judgement sampling. This has led to situations in which 'authenticity' is established on the basis of non-linguistic characteristics (e.g. length of settlement in a particular place or geographical and

dialectological origins of distant ancestors on the Korean Peninsula) or to circular situations in

which the presence or absence of certain linguistic features are the criteria which determine

whether consultants should be included in studies investigating the presence or absence of

certain linguistic features. None of these are conducive to the collection of representative or

reliable data. The approach of judgement sampling based on linguistic feature implies that the

researcher has *a priori* knowledge of the distinguishing features of the variety which they

intend to investigate and also has the problematic aspect of putting the researcher rather than

the consultant in the position of determining whether they are a 'real' speaker of their variety.

The self-identification approach, on the other hand, empowers the consultants themselves,

and by extension their community, to determine what constitutes their variety and moreover

allows the researcher to capture an accurate impression of the use of particular varieties in all

their diversity for inductive investigation.

The two contributions made by this thesis to data gathering methodology are wholly

applicable to research into the linguistic variation of Korean in general. The concrete data

gathering procedures which were developed as a result of these underlying methodologies of

broader sampling and speaker self-identification may be replicated not only the globalised

context, but also in the on-going study of the varieties of the Korean peninsula.

Furthermore, the material gathered for this project represents a small step towards the

compilation of a larger, more diverse record of the transplanted varieties. Again, steps to

address the paucity of documentation of these varieties have already been taken by some

pioneering researchers, most notably the documentary materials of of Kwak Chung-gu (2008a;

2008b; 2009; 2010a; 2010b; 2011). Conversely, the urgent need for a greater quantity and

much greater variety of such materials in the fact of advanced (KM) or incipient (VYK) language

shift has been observed (Lee 2005; King 2006b). Therefore, an action to be taken following the

conclusion of this project is to arrange for the archiving of the materials upon the basis of which it was produced.

In summary, then, in addition to the empirical contribution of this thesis to Korean Studies, its key contributions are to the expansion of the theoretical and methodological scope of Korean dialectology. We now turn to the contributions which this thesis has to offer non-area based linguistics.

7.3 Contribution to Linguistics

The finding that varieties of with similar (albeit not identical) dialectological heritage exhibit some synchronic linguistic differences after a period of separation in excess of one hundred and fifty years and settlement in vastly different conditions in which they have been in intense contact with very dissimilar regional languages of wider communication is, when critically assessed, banal to the point of triviality. It reflects the findings of earlier research into transplanted varieties of languages such as Arabic (Kusters 2003) and English (Trudgill 2004a) amongst many others (Lupyan and Dale 2010). While this particular aspect of the findings is not surprising in the context of global linguistics, that it presents a significant challenge to the narrative surrounding these particular transplanted varieties of Korean in Korean scholarship carried out in the framework of traditional dialectology, as detailed above, bears repeating.

Nevertheless, the findings and most especially the methods of this paper make some contribution to the discipline of linguistics outside of the specific sub-field of Korean dialectology. The interrogation of the place of comparison in linguistics and how it is to be carried out is of particular note with regard to the sub-discipline of CA. Rather than focus on the pedagogic possibilities of CA or its prior preoccupation with an 'objective of applicatbility, we followed the definitions of a theoretical CA provided by Fisiak (1981) and Gast (2012) to develop a replicable and principled approach to the comparison of not two languages, but two

related varieties of 'the same language' using traditional grammatical description as a source of cross-varietal comparative concepts.

Linguistics has increasingly come to employ if not rely on statistical techniques for the testing of its hypotheses. Both of the families of statistical techniques used in this thesis, clustering and principal components analysis, have seen wide applications in linguistics, particularly in the sub-field of modern quantitative dialectology. Here we employ these methods somewhat unconventionally in that we use them to ascertain simply patterns of variation between speakers rather than patterns of geographical variation.

We turn first to clustering. Despite its seeming unconventionality, the clustering carried out in this thesis in fact involves the application of precisely the same statistical technique (e.g. clustering using Ward's Method) to a similar kind of data (i.e. a distance measure based on aggregate linguistic (dis)similarity) as is commonly seen in the field of quantitative dialectology. Rather than being a novel method, it represents only a change in the conception of consultants since it conceives of them as individuals rather than linguistically representative avatars of their geographical origins. Such an approach demonstrates that the applicability of this method to inductively establishing or testing hypotheses about the relationships between groups of speakers goes beyond simply the geographical.

Moving on to the PCA, here we see a potentially more significant contribution to variationist linguistics through the employment of the somewhat novel method of categorical rather than classical PCA. While statistical tests from the family of factor analysis tests, including classical PCA, have been applied to continuous linguistic data (e.g. formant frequencies and vowel duration (Clopper and Paolillo 2006) or coding linguistic features on an arbitrary, continuous scale (Shackleton 2010)) our chosen method is suitable for use with categorical data, including our dichotomously coded data which records the presence or absence of linguistic features.

This offers an efficient method of assessing a large amount of categorical data and has a high degree of applicability to other data drawn from other languages and other sources, be they linguistic atlases or field recordings. This method combined with the 'bottom up' construction of feature catalogues from large-scale POS tagged dialect corpora would also represent a particularly interesting approach language variation and perhaps even an objective standard for establishing cross-linguistic patterns of the distribution of significant isoglosses.

Finally, we again note the contribution to documentary and descriptive linguistics, albeit a small one, made by the gathering of data for this project. While the materials gathered for this project do not represent full documentations of either variety, both the primary data themselves and the links forged with the communities will inform future documentation projects and, in so doing, benefit the field by aiding in the recording and, ideally, the maintenance of two specific aspects of the linguistic and cultural wealth of humanity as a whole.

In essence, the main contributions of this thesis to linguistics are twofold. First, and more minor, are its empirical findings, which add further weight to the notion that geographically distant and socially distinct varieties of a language diverge over time. Second, in its attempt to gather high-quality data and analyse it with a high-degree of objectivity a principled and cross-linguistically replicable methods for the comparison of two varieties of a given language have been developed. In the following, final section, we suggest how to follow up this research.

7.4 Future Directions for Research

This thesis makes a great empirical contribution to the understanding and description of the synchronic characteristics of and relationships between the contemporary transplanted varieties of Korean in Central Asia and the YKAP. As noted above, though, distinguishing features of VYK and KM, which were identified by the CATPCA do not find their origins in the

linguistic distinctions between traditional peninsula varieties. In other words, while this thesis

identified differences between the transplanted varieties, it does not in itself explain why

these particular differences between VYK and KM have arisen. We now go on to suggest two

specific possible avenues of enquiry which may contribute towards investigating this issue,

both in general and for particular cases of linguistic divergence.

For investigating the underlying reasons for the linguistic divergence between KM and VYK and

its general characterisation in terms of cross-linguistically observed patterns of language

change, the framework of Sociolinguistic Typology (ST) appear particularly promising. Finding

its roots in explanatory comparative dialectology, which sought to put the geographical

variation described by traditional dialectology in the context of social or sociolinguistic factors

(cf. Trudgill 1982; Trudgill 1997), the ST framework examines the general linguistic

characteristics of a given variety (often in terms of the relationship between linguistic

complexity and linguistic simplicity, as defined by the framework) in light of the sociological

characteristics of the community which uses that variety, to date largely focussing on the

community size, degree of contact with the surrounding languages, social stability and social

cohesiveness. Since the transplanted varieties of Korean examined here, and those spoken

elsewhere around the world, find themselves embedded in social situations which are very

different not only from conditions on the Korean peninsula, but from each other, ST appears to

be an appropriate framework for exploring the divergence of the transplanted varieties of

Korean through the examination of their synchronic linguistic features. The contact between

KM and Russian and VYK and Chinese deserve special attention. While such contact

phenomena are examined on the macro-level in the ST framework, it must be emphasised that

language contact phenomena at all levels, from individual to political, must be examined in

their own right in order to give a full account of dissimilarities between VYK and KM.

Returning to ST specifically, though, we suggest that such an analysis invites the incorporation

of other transplanted and peninsula varieties of Korea and would allow the linguistic diversity of Korean to be placed in cross-linguistic context since the ST framework has already been applied to divergence between Germanic languages (Trudgill 2011), Austronesian languages (Trudgill 2004c) and the dialects of Modern Greek (Trudgill 2004b; 2009) among others.

An example of the particular linguistic features revealed by this thesis which are suitable for further investigation is the extended use of nominalizer *–ki* in KM. In particular, this feature merits a closer examination within the framework of contact induced grammaticalisation (Heine and Kuteva 2002). While this unique feature of KM has been accounted for by appeals both to sound change and analogy with other KM constructions (e.g. King 1992 p. 214) its development closely resembles the grammaticalisation cline and the adoption of the nominalizer of a minority language to replicate the infinitival constructions of a socially dominant contact language in cases where the replicating language does not have this linguistic resource is somewhat widely attested (cf. Heine and Kuteva 2005). This study would complement the ST research suggested above in that it would investigate the specific mechanisms by KM diverged from other varieties of Korean and contribute towards discovering the extent to which particular social structures are correlated with specific pathways of linguistic change.

Outside of studies specific to the transplanted varieties of Korean spoken in Central Asia or the YKAP, similar methods to those employed for this study could be applied to the direct comparison of other related varieties. Indeed, these methods have the possibility of being expanded further yet to the direct synchronic comparison of any two related linguistic systems. If such research were carried out on a sufficiently large scale comparing transplanted varieties, drawn from a broad sample of languages, it so may allow us to discern whether there are any significant cross-linguistic generalisations to be made about patterns of language change or use in diaspora contexts.

For the last words, we return to the topic of VYK and KM. The necessity of the continued documentation of these varieties cannot be over-emphasised. This is no mere exercise in linguistic 'butterfly collecting'. While it may not be possible to prevent the extinction of the critically endangered *Koryo Mar* the creation of a Boasian trilogy of dictionary, descriptive grammar and diverse collection of texts for each transplanted variety of Korean (not just KM and VYK, but Zainichi Korean in Japan and the emergent varieties in Europe and North America) would serve as an enduring record of the linguistic and cultural realities of the Korean diaspora in the early twenty first century for future generations.

In addition to being an important record for subsequent generations of Koreans living in diaspora and humanity at large, such a record could be conceived of as the data which is prerequisite for the research plans formulated here and, if correctly archived and made available, research which has not yet been conceived.

8 Bibliography

Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory 21*(3), pp. 435-483

Armstrong, J.A. (1968). The Ethnic Scene in the Soviet Union: The View of the Dictatorship. In Goldhagen, Erich (Ed.), *Ethnic Minorities in the Soviet Union* (pp. 3-49). New York: Frederick A. Praeger, Publishers.

Back Tae Hyon (2001). The Social Reality Faced by Koreans in Central Asia. *Korean and Korean American Studies Bulletin*, *12* (2/3), 45 – 83

Baayen, R.H. (2008). *Analyzing Linguistic Data: A Practical introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Barbiers, S., Corniers, L. and van der Kleij, S. (2002). *Syntactic Microvariation*. Amsterdam: Meertens Institute.

Bashirova, N. and Solnyshikova, M. (2015). Representation of Ethnic Identity of Tatars through the Ethnonym "Tatar". In Balirano, G. and Nisco, M.C. (Eds). *Languaging Diversity: Identities, Genres, Discourses*, pp. 212-225. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Barnes-Sadler, S. (2015). Romanisation System for Koryo Mar. *Koreevedenie Kazakhstana, 3*, 334 – 349

Bayley, R. (2003). The Quantitative Paradigm. In Chambers, J.K., Trudgill, P. and Schilling-Estes, N. (Eds). *The Handbook of Language Variation and Change*, pp. 85-107. Blackwell Reference Online.

Bell, A. and Holmes, J. (1992). H-droppin': Two Sociolinguistic Variables in New Zealand English. *Australian Journal of Linguistics 12*(2), 223-248

Bellikov, V. (1991). Koreans in the Linguistic Situation of the Russian Far East. *Icwung*

enehakhoyci, 8, 299-305

Berge, A. (2010). Adequacy in Documentation. In Grenoble, L.A. and Furbee, L. (Eds.),

Language Documentation: Practice and Values (51 – 66). Amsterdam; Philadelphia, PA: John

Benjamins Publishing Company

Boas, H. C. (2003). Tracing dialect death: The Texas German dialect project. In *Proceedings of*

the 28th Meeting of the Berkeley Linguistics Society (pp. 387-398). Berkeley: California.

Boersma, P. and Weenink, D (2013). *Praat: doing phonetics by computer Version 5.3.51*.

Brandner, E. (2012). Syntactic Microvariation. *Language and Linguistics Compass 6*(2)L pp. 113-

130

Bossong, G. (1985). *Differentielle Objektmarkierung in den Neuiranischen Sprachen*

[Differential Object Marking in the New Iranian Languages]. Tuebingen: Gunter Narr Verlag

Brugman, H. and Russel, A. (2004). Annotating Multimedia/Multi-modal resources with ELAN,

In *Procceedings of LREC 2004, Fourth International Conference onf Language Resources and*

Evaluation (pp. 2065-2068). Lisbon, Portugal.

Cantos Gomez, P. (2013). *Statistical Methods in Language and Linguistic Research.* Sheffield:

Equinox Publishing

Cedergren, H. J. and Sankoff, D. (1974). Variable Rules: Performance as a Statistical Reflection

of Competence. *Language 50*(2): pp. 333-355

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

Chang Suk-Jin (1996). *Korean*. Amsterdam; Philadelphia, PA: John Benjamins Publishing

Company

Chesterman, A. (1998). *Contrastive Functional Analysis*. Amsterdam/Philadelphia, PA: John

Benjamins Publishing Company.

CIA (2013). *Kazakhstan.* Retrieved April 10, 2013 from

https://www.cia.gov/library/publications/the-world-factbook/geos/kz.html

Chey Youn-Cha S. (1987). Soviet Koreans and Their Culture in the USSR. In Suh Dae-Sook (Ed.),

Koreans in the Soviet Union (60 - 84)*.* Honolulu: University of Hawai'I, Center for Korean

Studies.

China (1953). *Policy Towards Nationalities of the People's Republic of China*. Beijing: Foreign

Language Press.

Choi Myeong Ok (2001). Hankwukpangenkwuhoyk [Korean Dialect Taxonomy] in *Pangenhak*

Sacen [Dictionary of Dialectology] Ed. Pangenyenkwuhoy (pp. 373-384). Seoul: Thayhaksa.

-       (2002). Cwungkwuk yen byen ciyekuy hankwuke yenkwu [Research into the

         Korean Language of China's Yanbian Area], *Hankwuk munhwa, 25*, 17 – 62

Choi Myeong-ok, Kwak Chung-gu and Jeon Hak-seok (2002). *Hampwuk pwukpwuciyeke*

yenkwu [Research on the Language of North Hamgyeong Province's Northern Region]. Seoul:

Thayhaksa.

Choi Yun-kap, Bu Ju-pil and Jeon Hak-seok (1994). *Cwungkwuk cosen hankwuk*

cosenechaiyenkwu [Research on the Differences between China, North Korea and South

Korea's Korean Languages]. Yanji, PRC: Yenbyeninminchwulphansa.

Clopper, C.G. and Paolillo, J.C. (2006). North American English Vowels: A Factor-Analytic

Perspective. *Literary and Linguistic Computing 21*(4): 335-359

Coupland, N. (2010). *The Handbook of Language and Globalisation.* Oxford: Wiley-Blackwell.

Cui, Longhau (1990). The Position and Development of Koreans in China. In Suh Dae-Sook and Schultz, Edward J. (Eds.), *Koreans in China* (pp.78-92). Honolulu: University of Hawai'i, Center for Korean Studies.

Davis, Lawrence M. (1990). *Statistics in Dialectology*. Tuscaloosa, AL: University of Alabama Press.

Dialect Research Society (2001). *Pangenhak sacen* [Dictionary of Dialectology]. Seoul: Thayhaksa.

Dixon, R.M.W. (2010). *Basic Linguistic Theory*. Oxford: Oxford University Press.

Dzharylgasinova, R. (1980). Osnovnye Tendentsii Etnicheskikh Protsessov u Koreitsev Srednei Asii i Kazakhstana [Basic Ethnic Tendencies and Processes of the Koreans of Central Asia and Kazakhstan], *Etnicheskiye Protsessy u Natsional'nykh Grupp Srendnei Azii i Kazakhstana*, 43-73

Eckert, P. (2008). Variation and the Indexical Field. *Journal of Sociolinguistics 12* (4), 453-476

Eddington, D. (2015). *Statistics For Linguists: A Step-by-Step Guide for Novices*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Edwards, J. (2010). *Minority Languages and Group Identity: Cases and Categories*. Philadelphia; Philadelphia, PA: John Benjamins Publishing Company.

Field, A. (2005). *Cluster Analysis*. Retrieved from

http://www.statisticshell.com/docs/cluster.pdf

Filipović, R. (1984). What are the Primary Data for Contrastive Analysis? In Fisiak, J. (Ed.),

Trends in Linguistic: Studies and Monographs: Contrastive Linguistics: Prospects and Problems

(107 – 118). Berlin: Mouton Publishers.

Fischer, J.L. (1958). Social Influences on the Choice of a Linguistic Variant. *Word 14*, 47-56

Fishman, J.A. (1965). Language Maintenance and Language Shift: The American Immigrant

Case within a General Theoretical Perspective. *Sociologicus, 16* (1), 19-39

Fisiak, J. (1981). Editors Introduction. In Fisiak, J. (Ed.). *Theoretical Issues in Contrastive*

Linguistics. Amsterdam: John Benjamin's Publishing Company.

Fox, A. (1995). *Linguistic Reconstruction: An Introduction to Theory and Method*. Oxford:

Oxford University Press.

Galloway, N. and Rose, H. (2015). *Introducing Global Englishes*. New York, NY; Abingdon:

Routledge.

Gast, V. (2012). Contrastive analysis: Theories and methods. In Kortmann, B. (ed.): *Dictionaries*

of Linguistics and Communication Science: Linguistic theory and methodology. Berlin: Mouton

de Gruyter.

-   (2013). Contrastive analysis. *The Routledge Encyclopedia of Language Teaching*

    and Learning (153-158). London: Routledge.

Gelb, M. (1995). An Early Soviet Ethnic Deportation: The Far-Eastern Koreans. *Russian Review,*

54(3), 389 - 412

Goebl, H. (2005). Dialektometrie [Dialectometry]. In Köhler, R., Altmann, G. and Piotrowski, R.G.

(Eds.). *Hanbücher zur Sprach und Kommunikationsvissenschaft: Quantitative Linguistik: Ein*

Internationales Hanbuch II [Handbooks of Linguistics and Communication Science: An

International Handbook of Quantitative Linguistics II]. Berlin; New York, NY: De Gruyter

Mouton.

Gries, S. Th. (2009). *Statistics for Linguistics with R: A Practical Introduction*. The Hague: De

Gruyter Mouton.

Grieve J., Speelman, D. and Geeraerts, D. (2011). A Statistical Method for the Identification and

Aggregation of Regional Linguistic Data. *Language Variation and Change 23*: 1-29

Hale, M. (2007). *Historical Linguistics: Theory and Method*. Oxford: Blackwell Publishing.

Halliday, M.A.K. (1978). *Language as Social Semiotic*. London: Edward Arnold.

Hara, T. (1987). The Korean Movement in the Russian Maritime Province, 1905 – 1922. In Suh

Dae-Sook (Ed.), *Koreans in the Soviet Union* (pp. 1 - 23)*.* Honolulu, HI: University of Hawai'I,

Center for Korean Studies.

Harrell, S. (1993). Linguistics and Hegemony in China. *International Journal of the Sociology of

Language, 103*, 97 – 114

Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. In Haspelmath, M and Tadmor,

U (Eds.), *Loanwords in the world's languages: A comparative handbook* (pp. 35-54). New York,

NY: Mouton de Gruyter.

- (2010). Comparative Concepts and Descriptive Categories in Crosslinguistic

   Studies. *Language, 86* (3), 663 – 687

Hatch, E.M. (1982). *Research Design and Statistics for Applied Linguistics*. Rowley, MA:

Newbury House.

He, J. (1990). China's Policy on Nationalities. In Suh Dae-Sook and Schultz, Edward J. (Eds.), *Koreans in China* (1 - 30)*.* Honolulu: University of Hawai'I, Center for Korean Studies.

Heine, B. and Kuteva, T. (2003). On Contact-Induced Grammaticalization. *Studies in Language 27* (3), 529-572

- (2005). *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.

Hellinger, M and Ammon, U. (1996). Contrastive Sociolinguistics: an introduction. In Hellinger, M. (Ed.), *Contrastive Sociolinguistics* (1-16). Berlin: Walter de Gruyter and Co.

Heo Myeong-Cheol (2013). Yenbyencosencokcachicwuuy senglipkwa cosencok sahoy [The Foundation of the Yanbian Korean Autonomous Prefecture and Chinese-Korean Society]. In Inhatayhakkyo Hankwukhakyenkwuso (Ed.), *Yenbyencosencokuy yeksawa hyensil* [Yanbian Chinese-Koreans' History and Present] (135 - 166), Seoul: Comyengcwhulphan.

Heo Su-dong (2013). Secensoswukkwa cwungkwuk tongpukciyekuy kyuyuk [The *secensoswuk* and Education in China's North Eastern Region]. In Inhatayhakkyo Hankwukhakyenkwuso (Ed.), *Yenbyencosencokuy yeksawa hyensil* [Yanbian Chinese-Koreans' History and Present] (79 - 102), Seoul: Comyengcwhulphan.

Hjulmand, L. and Schwarz, H. (2009). *A Concise Contrastive Grammar of English for Danish Students*. Copenhagen: Sammfundslitteratur

Im Gye-sun (2003). *Wulieykey takaon cosencokun nwukwuinka?* [Who are the Chinese-Koreans?]. Seoul: Hyenamsa.

Inoue Fumio (1999). Subjective Dialect Division in Great Britain. Preston, D.R. and Kretzschmar, W. (Eds.) *Handbook of Perceptual Dialectology: Volume 1* (161 - 176), Philadelphia, PA; Amsterdam: John Benjamins Puiblishing Company.

Ito, C. (2008). Historical Development and Analogical Change in Yanbian Korean

Accent. *Harvard Studies in Korean Linguistics12*, 165-178

Ito C. and Kenstowicz, M. (2009). Mandarin Loanwods in Yanbian Korean II: Tones. *Language*

Research, *45* (1), 85 – 109

James, C. (1980). *Contrastive Analysis*. Harlow: Longman.

Jenkins, J. (2009). *World Englishes: A Resource Book for Students*. New York, NY; London:

Routledge.

Jeon, L. and Cukor-Avila, P. (2015). "One Country, One Lanugage?": Mapping Perceptions of

Dialects in South Korea. *Dialectologia 14*, 17-46

Jeong Hyang Ran (2010). *Yenbyen pangenuykokyongkwa hwalyong* [Inflection and Declension

in Yanbian Dialect]. Paju: Hankwuk hakswul cengpo.

Jeong Seung-cheol (2001). *Hankwuk pangenhaksa* [History of Korean Dialectology]. In

Pangenyenkwukhoy (Ed.). *Pangen Sacen* [Dictionary of Dialectology] (pp.384-391). Seoul:

Thayhaksa.

Jeong Shin-ja (2013). Cwungkwukcosencok yesenguy kwukceykyelhon yenkwu. In

Inhatayhakkyo Hankwukhakyenkwuso (Ed.), *Yenbyencosencokuy yeksawa hyensil* (167 - 223),

Seoul: Comyengcwhulphan.

Ji Dong-eun, Kim Kwang-su, Yun Hui-nam and Kim Gyeol-pyeon (2012). *Cwungkwukeyseuy*

cosene yenkwu loncemoklok [Bibliography of China's Korean Language Research]. Yanbian,

PRC: Yenbyentayhak chwulphansa.

Jin Wenhua (2008). *Sound of Chinese Korean: A Variationist Approach* (Unpublished doctoral thesis). University of Texas at Arlington, Arlington, TX, USA.

- (2012). Variation and Change in Chinese Korean: the case of vowel /y/. *Language Variation and Change, 24*, 79 – 106

Jiphyeonjeon (c.1446). *Hwunmincengumhaylyey* [Examples and Instructions for the Correct Sounds for the Instruction of the People]. Seoul: Jiphyeonjeon.

Johnson, K. (2008). *Quantitative Methods in Linguistics*. Malden, MA; Oxford: Blackwell Publishing.

Jo Mi-Jeong (2017) Social Mobility in a Transnational Context: Marriage Migration of Koryo-Saram from Central Asia to South Korea. AKSE 2017, Prague, 22nd April 2017. Prague: Charles University.

Jones, W. (1796/1967). The Third Anniversary Discourse on the Hindus. In Lehmann, W.P. (Ed.) *A Reader in Nineteenth Century Historical Indo-European Linguistics* (7-20). Bloomington, IN: Indiana University Press.

Kachru, B. (1992). *The Other Tongue: English Across Cultures*. Chicago, IL: University of Illinois Press.

Kang Hui-Suk (2003). Cwungkwuk cosencok haksayngtuluy mokwuke sayongay tayhan kongsicek yenkwu [Research on China's Chinese-Korean Students' Synchronic Mother Tongue Usage]. *Sahoyenehak 11,* (2), 1 – 24

Kho Songmoo (1987). *Koreans in Soviet Central Asia*. Helsinki: Finnish Oriental Society.

Kim, G. (2009). Education and Diasporic Language: The Case of Koreans in Kazakhstan. *Acta slavica iaponica, 27*, 103-123

Kim, G. and King, R. (2001). *Koryo Saram: Koreans in the Former USSR*. New Haven, CT: East Rock Institute.

Kim H-Y and Lee J-S (2008). Heritage Language Learners' Attitudes Motivations and Instructional Needs: The Case of Post-Secondary Korean Language Learners. In Kondo-Brown, Kimi and Brown, James Dean (Eds.), *Teaching Chinese, Japanese and Korean Heritage Language Students: Curriculum needs and Assessment* (pp. 159 - 186). Oxford: Lawrence Earlbaum Associates.

Kim Ki-chang and Jeong Sang-seop (2011). Cosencok sohakkyo kwuke kyoyukkwa hankwukuy chotunghakkyo kwuke kyoyukuy pikyo yenkwu (1) kyoyukkwaceng cwungsimulo[Comparative Research on Chinese-Korean Primary School National Language Education and Korean Primary School National Language Education (1) Focus on Curricula]. *Chenglam emunkyoyuk, 43*, 339 – 370

Kim Kwang-su (2009). *Haypang cen cwungkwukeyse coseneuy pyenhwa palcen yenkwu* [Research on the Change and Development of the Korean Language in China Prior to the Liberation of Korea]. Seoul: Tosechwulphan yeklak.

- (2012a). Yenbyen ciyek wuli mincokuy ene cengcheysengkwa tiasuphola [The Linguistic Sense of Identity and Diaspora of the Korean People in the Yanbian Region]. In Kim Kwang-su (Ed.). *Cosene kochalkwa yenkwu* [Perspectives and Research on the Korean Language] (pp.104 - 117). Yanji, PRC: Yenbyeninminchwulphansa.

- (2012b). Yenbyencosencokcachicu yenkilsi kanphan enecekthukseng [The Linguistic Characteristics of Signs in Yanji City, Yanbian Korean Autonomous Prefecture]. In Kim Kwang-su (Ed.). *Cosene kochalkwa yenkwu* [Perspectives and

Research on the Korean Language] (pp. 151 - 161). Yanji PRC:

Yenbyeninminchwulphansa.

- (2012c). Cwungkwukcosenemalmungchilul liyonghan cwungkwukcosenhakyenkwu

[Using a Corpus of Chinese Korean for Research into the Language of Chinese

Koreans]. In Kim Kwang-su (Ed.). *Cosene kochalkwa yenkwu* [Perspectives and

Research on the Korean Language] (pp.181 -193). Yanji, PRC:

Yenbyeninminchwulphansa.

- (2012d). Puknam mich cwungkwuk cosene munpepyonge syonguy chaiwa

thongiluy pangan [The Difference between North Korean, South Korean and

Chinese-Korean Usage of Grammatical Terms and a Framework for their

Unification]. In Kim Kwang-su (Ed.). *Cosene kochalkwa yenkwu* [Perspectives and

Research on the Korean Language] (pp. 416 - 436). Yanji, PRC:

Yenbyeninminchwulphansa.

- (2012e).Yenbyeneyse pon nampukuy ehwi: cenmunyongelul cwungsimulo [North

Korean and South Korean Vocabulary in Yanbian: Focus on Technical Terms]. In

Kim Kwang-su (Ed.). *Cosene kochalkwa yenkwu* [Perspectives and Research on the

Korean Language] (pp. 437 - 447). Yanji, PRC: Yenbyeninminchwulphansa.

Kim Kwang-su, Kang Mi-hwa and Hwang Hye-yeong (2013). *Hyentay cosene munpeplon*

[Contemporary Korean Grammar]. Yanji, PRC: Yenbyentayhakchwulphansa.

Kim Seon-hui (2013). Yenbyen pangen yenkwu: cosawa cwungkyelemilul cwungsimulo

[Research on Yanbian Dialect: Focus on Particles and Final Verb Endings]. *Hanmincok emunhak*

64, 71 – 98

Kim Thay-gyun (1986). *Hampukpangensacen* [A Dictionary of North Hamgyeong Province

Dialect]. Seoul: Kyenggitay chwulphanpu.

Kim Yeong-hwang (1982). *Cosenpangenhak* [Korean Dialectology]. Pyeongyang: Kimilseong

cwunghaptayhakchwulphansa

Kim-Renaud, Young-Key (Ed.) (1997) *The Korean Alphabet: Its history and Structure*. Honolulu:

University of Hawai'i Press.

-   (2009) *Korean: An Essential Grammar*. New York, NY; Abingdon, UK: Routledge.

Kimura, H. (1987). Korean Minorities in Soviet Central Asia and Kazakhstan. In Suh Dae-Sook

(Ed.), *Koreans in the Soviet Union* (85 - 100). Honolulu: University of Hawai'I, Center for Korean

Studies.

King, R. (1987a). An Introduction to Soviet Korean. *Language Research 23,* (2), 233 – 277

-   (1989). The Korean Dialect Materials in Matveev's 1900 "Reference Book to the
    City of Vladivostok". *Language Research 24,* (2), 281 - 329

-   (1991a). A Soviet Korean Grammar from 1930. *Korean Language Education 3*, 153
    – 178

-   (1991b). Korean Language Studies in the USSR: Past, Present and Future.
    Icwungenehakhoyci 8, 42 – 153

-   (1992). Archaisms and Innovations in Soviet Korean Dialects. *Language Research
    28,* (2), 201 – 223

-   (1994). Dialect Elements in Soviet Korean Publications from the 1920s. In Aronson,
    Howard I. (Ed.). *NSL 7: Linguistic studies in the non-Slavic languages of the
    Commonwealth of Independent States and the Baltic Republics* (151 - 183).
    Chicago: Chicago Linguistic Society.

-   (1996c). Russian Loanwords in Hamkyeng and Soviet Korean Dialects. In Shim Jae-
    Kee et al. (Eds.). *Essays in Honor of Kim-moon Lee* (939 - 966). Seoul: Shin-gu
    Publishing Company.

365

- (1997b). Experimentation with Han'gul in Russia and the USSR. In Kim-Renaud, Young-Key (Ed.). *The Korean Alphabet: History and Structure* (296 - 339). Honolulu: University of Hawaii Press.

- (2001a). Blagoslovennoe: Korean Village on the Amur, 1897 – 1937. *The Review of Korean Studies 4* (2), 133 – 176

- (2006a). Dialectal Variation in Korean. In Sohn Ho-Min (Ed.). *Korean Language in Culture and Society* (264-280). Honolulu, HI: University of Hawai'i Press.

- (2006b). Korean Dialects in the Former USSR: Reflections on the Current State of Research. *Pangenhak 3*, 127 – 153

King, R and Yeon Jaehoon (1992). Cwungang asia hanintuluy ene – kolyemal [Koryo Mar – The Language of Central Asia's Koreans]. *Hangul 217*, 83 – 134

Kolarz, W. (1954). *The Peoples of the Soviet Far East*. New York, New York: Praeger.

König, E. (2012). Contrastive Linguistics and Language Comparison. *Language in Contrast, 12* (1), 3 – 26

Kono Rokuro (1945) *Chosenpangenhaksigo* [Impressions of Korean Dialects]. Tokyo: Tongtosecek.

Kozlov, V.I. (1982). *Natsional'nosti SSR, etnodemograficheskii obzor* [An Ethno-demographic Report on the Nationalities of the USSR]. Moscow: publisher unknown.

Kretzschmar (1996). Quantitative Areal Analysis of Dialect Features. *Language Variation and Change, 8*, 13 – 39

Krzeszowski, T. P. (1990). *Contrasting Languages: the Scope of Contrastive Linguistics*. Berlin/New York: Mouton de Gruyter.

Kusters, W.C. (2003). *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Utrecht: LOT.

Kwak Chung-gu (1987). Nohan hoyhwawa hampuk kyenghung pangen [Russian Korean Speech and Northern Hamgyeong Kyenghung Dialect], *Cintan Hakpo, 62,* 79-124

- (2001a). Tongpuk pangen [North Eastern Dialect]. In Pangenyenkwukhoy (Ed.). *Pangen Sacen* [Dictionary of Dialectology] (pp.97-106). Seoul: Thayhaksa.

- (2001b). Yukchin Pangen [Yukchin Dialect]. In Pangenyenkwukhoy (Ed.). *Pangen Sacen* [Dictionary of Dialectology] (pp. 277-285). Seoul: Thayhaksa.

- (2007). Cwungang asia kolyemaluy calyuwa yenkwu [Resources and Research on Central Asia's Koryo Mar]. *Inmunnonchong 58,* 231 – 272

- (2008a). *Cwungang asia icwu hanmincokuy enewa saynghwal: khacahusuthan almathi* [The Language and Life of the Korean People of Central Asia: Kazakhstan, Almaty]. Paju: Thayhaksa.

- (2008b). *Cwungkwuk icwu hanmincokuy enewa saynghwal: killimsen hoylyongbong* [The Language and Life of the Korean People of China: Jilin Province, Hoelyongbong]. Paju: Tahyhaksa.

- (2009). *Cwungang asia icwu hanmincokuy enewa saynghwal:wucupeykhisuthan thasyukheynthu* [The Language and Life of the Korean People of Central Asia: Uzbekistan, Tashkent]. Paju: Thayhaksa.

- (2010a). *2010nyento kwukoy ciptan icwu hanmincokuy ciyeke cosa 1: cosa ciyek khacahusuthan thaldikhurukan* [The 2010 Dialect Survey of Overseas Korean Communities 1: Site of survey Kazakhstan, Taldikurgan]. Seoul: Kwuklip Kwukewen.

- (2010b). *2010nyento kwukoy ciptan icwu hanmincokuy ciyeke cosa 2: kwukoy 5 cicem ehuy, umun, munpep thonhap calyucip, cosa ciyek, cwunkwuk killimseng*

367

hwunchwunsi, khacahusuthan almathi, khacahusuthan thaltikhurukan, wucupaykhisuthan thasyukeynthu, khilikisusuthan pisyukheyku [The 2010 Dialect Survey of Overseas Korean Communities 2: combined collection of resources from five sites including vocabulary, phonology and grammar. Survey sites PRC, Jilin Province, Hoelyongbong; Kazakhstan, Almaty; Kazakhstan, Talkdikurgan; Uzbekistan, Tashkent; Kyrgyzstan, Bishkek]. Seoul: Kwuklip Kwukewen.

- (2011). *Cwungang asia icwu hanmincokuy enewa saynghwal:khilukisusuthan pisyukheykhu* [The Language and Life of the Korean People of Central Asia: Kyrgyzstan, Bishkek]. Paju: Thayhaksa.

- (2012). *2010nyento kwukoy ciptan icwu hanmincokuy ciyeke censa pokose: censa ciyek, cwungang asia khililkisusuthan pisyukheykhu* [Report on the 2010 Dialect Survey of Overseas Korean Communities: Site of survey Central Asia, Kyrgyzstan, Bishkek]. Seoul: Kwuklip Kwukewen.

Kwon Jae-il (2010). *Cwungangasia kolyemaluy munpep* [A Grammar of Central Asian Koryo Mar]. Seoul: Seoultaehakkyo chwulphanmunhwawen.

- (2013). Cwungangasia koleymaluy kwuewa mune [Central Asian Koryo Mar's Vernacular and Literary Languages]. *Koreevedenie Kazakhstana, 1*, 45-66

Kwon Song-Nim and Zribi-Hertz, A. (2008). Differential Function Marking, Case and Information Structure: Evidence from Korean. *Language 84*(2) 258-299

Labov, W. (1964). Phonological correlates of social stratification. In Gumperz and Hymes (eds.) *Directions in Sociolnguistics* pp. 164-176

- (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

- (2008). *Quantitative Reasoning in Linguistics*. Retrieved 30 March 2015 from http://www.ling.upenn.edu/~wlabov/Papers/QRL.pdf

Lamont, C. (1945). *The Peoples of the Soviet Union*. New york: Harcourt, Brace and Company.

Landau, J.M. and Kellner-Heinkelle, B. (2001). *Politics of Language in the ex-Soviet Muslim States: Azerbayjan, Uzbekistan, Kazakhstan, Kyrgyzstan, Turkmenistan and Tajikistan*. London: Hurst and Company.

- (2012). *Language Politics in Contemporary Central Asia: National and Ethnic Identity and the Soviet Legacy*. London: I.B. Tauris & Co. Ltd.

Ledyard, G.K. (1998). The Korean Language Reform of 1446. Seoul: Sinkwumunhwasa.

Lee Chae-Jin (1986). *China's Korean Minority: the politics of ethnic education.* Boulder, Colo. Westview Press.

Lee Iksop (1987). Kangwonto pangenuy thukcingkwa ku yenkwu [The Characteristics and Research of Gangwon Province Dialect]. *Kwuke Saynghwal* 10, 61-71

Lee Ik-sop, Jeon Kwang-hyeon, Lee Kwang-ho, Lee Byeong-gun and Choi Myeong-ok (2008). *Hankwukenecido* [Linguistic Atlas of Korea]. Paju: Thayhaksa.

Lee Iksop and Ramsey, Robert S. (2000). *The Korean Language*. Albany, NY: State University of New York Press.

Lee Ju-Haeng (2005a). *Hankwuke sahoy pangenkwa ciyek pangenuy ihay* [Understand Korean Social and Regional Dialects]. Seoul: Hankwuk Munhwasa.

- (2005b). Hankwukinkwa cwungkwuk cosencokuy umun silhyen yangsang – 10taywa 20tayuy enelul cwungsimulo [Aspects of the Phonological Phenomena of Koreans and Chinese Ethnic Korean – Focus on Teenagers and Young Adults in their 20s]. *Icwungenehak* 28, 309-331

369

Lee Ki-Moon and Ramsey, S. R. (2011). *A History of the Korean Language*. Cambridge: Cambridge University Press.

Lee Kuk-lo (1932). Cosen mal sathwuli [Korean Dialects]. *Tongkwang 29*, page reference unknown.

Lee Sang-kyu (2003). *Kwukepangenhak* [Korean Dialectology]. Seoul: Hwakyensa.

Lee Sang Oak (2006). *Sewulmal cinkyeng kwue yenkwu* [Research on Real Seoul Vernacular Speech]. Seoul: Tosachwulphan pakiceng.

Lee Sungjae (2004*). Pangen yenkwu: calyoeyse ilonulo* [Dialect Research: From Date to Theory]. Seoul: Thayhaksa.

LeeThae-jin (Ed.) (2011). *Cwungkwuk haninuy yeksa (sang)* [A History of Koreans in China (vol. 1)]. Seoul: Kwuksaphyenchanoywenhoi.

Linting, M., Meulman, J.J., Grownen, P.J.F. and Van der Kooij, A.J. (2007). Nonlinear Principal Components Analysis: Introduction and Application. *Psychological Methods 12*(3), 336-358

Long, D. and Yim Young-Cheol (2002). Regional Differences in the Perception of Korean Dialects. In Preston, D.R. and Long, D. (Eds.) *Handbook of Perceptual Dialectology Volume 2* (247-276). Philadelphia, PA; Amsterdam: John Benjamins Publishing Company.

Lüpke, F. (2010). Research Methods in Language Documentation, *Language Documentation and Description, 2,* 55 – 104

Lupyan, G. and Dale, R. (2010). Language Structure Is Partially Determined by Social Structure. PLoS ONE, *5*, (1): e8559. doi:10.1371/journal.pone.0008559

Ma Yin (Ed.) (1994). *China's Minority Nationalities*. Beijing: Foreign Language Press.

Macaulay, R.K.S. (2009). *Quantitative Methods in Sociolinguistics*. Basingstoke, UK; New York, NY: Palgrave Macmillan.

Martin, S.E. (1992). *A Reference Grammar of Korean: A Complete Guide to the Grammar and History of the Korean Language*. Rutland, VT: Charles E. Tuttle.

Matveev, N.P. (1900). *Spravochnaja Kniga g. Vladivostoka. S Prilozheniem 5 Risunkov, Slovarej: Kitajskogo, Korejskogo I Japanskogo, Plana Teatra i Plana Goroda* [Guidebook to the City of Vladivostok. With Appendix of Five Pictures, Chinese, Korean and Japanese Dictionaries, a Map of the Theatre and a Map of the City]. Vladivostok: Publisher unknown.

Meulman,J.J., Heiser, W.J. and SPSS (2004). *SPSS Categories 13.0*. Chicago: IBM Corp.

Miller, R.A. (1971). *Japanese and the Other Altaic Languages*. Chicago: University of Chicago Press.

Milroy, J and Milroy, L. (1985). Linguistic Change, Social Network and Speaker Innovation. *Journal of Linguistics, 21*, 339 – 384

Ministry of Foreign Affairs and Trade (2011). *Kukkapyel cayoytongphohyenhwang* [The Present Condition of Overseas Koreans by Country]. Retrieved April 10, 2013 from http://www.mofa.go.kr/webmodule/htsboard/template/read/korboardread.jsp?typeID=6&boardid=232&seqno=334627&c=&t=&tableName=TYPE_DATABOARD&px=&dc=&wc=&lu=&vu=&iu=&du

National Language Association (2006). *Ciyeke cosa cilmunci* [Dialect Survey Questions]. Paju: Thayhaksa

- (2016). *Pyocwune saceng wenchik* [Fundamental Principles of the Standard

    Language]. Retrieved 9th December 2016 from

    https://www.korean.go.kr/front/page/pageView.do?page_id=P000085&mn_id=94

Nerbonne, J. and Kretzschmar, W. (2003). Introducing Computational Techniques in

Dialectometry. *Computers and the Humanities 37* (3), 245-255.

Nerbonne, J. and Wieling, M. (forthcoming). Statistics for Aggregate Variationist Analyses. To

appear in Boberg, C., Nerbonne, J. and Watt, D. (eds). *Handbook of Dialectology*. Boston, MA:

Wiley.

Nettle, D. (1999). *Linguistic Diversity*. Oxford: Oxford University Press.

Ogura Shinpei (1940). The Outline of the Korean Dialects. *Memoirs of the Research

Department of the Tokyo Bunko 12*, no page reference.

- (2009 [1940]). *Cosenepangensacen* [Korean Dialect Dictionary]. Seoul:

    Hankwukmunhwasa.

Ornstein, J. (1968). Soviet Language Policy: Continuity and Change. In Goldhagen, Erich (Ed.),

Ethnic Minorities in the Soviet Union (pp. 121-146). New York: Frederick A. Praeger, Publishers.

Pak, N.S. (1991). On Korean Dialects in the USSR. *Icwungenehakhoyci 8*, 617 – 620

- (1997). Besonderheiten der Kasusformen des Yukchin Dialektes des Koreanischen

    in Kazachstan [Characteristics of the Case Particles of the Yukchin Dialect of

    Kazakhstan]. *Kwukehaknoncib, 3,* 234 - 245

- (2005). *Koreyskiy Yazik v Kazakhstane: Problemy I Perspektivy* [The Korean

    Language in Kazakhstan: Problems and Perspectives] Almaty: Kazakh Ministry for

    Science and Education.

Pan, Y. and Kadar, D.A. (2011). Historical vs. Contemporary Chinese Linguistic Politeness. *Journal of Pragmatics 43*, 1525-1539

Park Kyeong-rae (2002). Cwungkwuk cosencoktuluy mokwuke sayong silthay [Mother Tongue Usage among China's Chinese Koreans]. *Sahoy enehak 10*, (1), 113 – 145

Park Youngmae (2003). A Preliminary study of the Language of ethnic Koreans in China: Toward a Sociolinguistic Understanding. *Kyoto University Linguistic Research, 22*, 1-21

Paio, Changyu (1990). The History of Korean in China and the Yanbian Korean Autonomous Prefecture. In Suh Dae-Sook and Schultz, Edward J. (Eds.), *Koreans in China* (pp.44-77). Honolulu: University of Hawai'i, Center for Korean Studies.

Petyt, K.M. (1980). *The Study of Dialect: An Introduction to Dialectology*. London: Andre Deutsch Limited.

Poplack, Shana (2000 [1979/80]). Sometimes I'll Start a Sentence in Spanish y Termino en Español: Toward a Typology of Code-switching. In Wei, L. (Ed.) *The Bilingualism Reader* (pp. 205-240). London: Routledge.

Preston, D.R. (1989). *Perceptual Dialectology: Nonlinguists' Views of Areal Linguistcs*. Providence, RI; Dordrecht: Foris Publications

Preston, D.R. and Long, D. (1999). *Handbook of Perceptual Dialectology*. Amsterdam: John Benjamins Publishing Company

Price, G. (1973). Minority Languages in Western Europe. In Stephens, M. (Ed.) *The Welsh Language Today*. Llandsul: Gomer Press.

Pung Ui-san (2000). Cwungkwuk mincok ene cengchaykkwa yenpyenuy hankwuke sayong silthay [China's Language Policy for Nationalities and the Usage of the Korean Language in Yanbian]. *Cengsinmunhwayenkwu, 23,* (2), 209 – 224

Purnell, T., Salmons, J. and Tepeli, D. (2005). German Substrate Effects in Wisconsin English: Evidence for Final Fortition. *American Speech* 80, 135-164

Putsillo, M. (1874). *Essai et Dictionnaire Russe-Coréen*. [A First Attempt at a Russian-Korean Dictionary]. St. Petersburg: Hogenfelden.

Ramset, R.S. (1978). *Accent and Morphology in Korean Dialects: A Descriptive and Historical Study*. Seoul: Thap.

Ramstedt, G.J. (1949). *Studies in Korean Etymology*. Helsinki: Suomalais-ugrilainen Seura.

Rasinger, S.M. (2008). *Quantitative Research in Linguistics: An Introduction*. London; Ney York, NY: Continuum International Publishing.

Rask, R. (1818/1967). An Investigation Concerning the Source of the Old Northern or Icelandic Language. In Lehmann, W.P. (Ed.) *A Reader in Nineteenth Century Historical Indo-European Linguistics* (29-37). Bloomington, IN: Indiana University Press.

Rein, K. (1983). *Einführung in die Kontrastive Linguistik* [Introduction to Contrastive Linguistics]. Darmstadt: Wissenschaftliche Buchgesellschaft.

Ringe, D.A. (2013). *Historical Linguistics: Toward a Twenty-first Century Reintegration*. Cambridge: Cambridge University Press.

Sahoykwahakchwulphansa. (1992). *Cosenmal Taysacen k-s* [The Greater Dictionary of the Korean Language K-S]. Phyengyancwunghapinswaykongcang: Pyeongyang.

Sankoff, D., Poplack, S. and Vanniarajan, S. (1990). The Case of the Nonce Loan in Tamil. *Language Variation and Change 2*, 71-101

Schilling, N. (2013). *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press.

Seguy, J. (1973). La Dialectométrie dans l'Atlas linguistique de la Gascogne. [Dialectometry in the Linguistic Atlas of Gascony]. *Revue de Linguistique Romane* [Review of Romance Lingusitics] *37*, 1-24.

Shackleton, R.G. (2005). English-American Speech Relationships: A Quantitative Approach. *Journal of English Linguistics 35*: 30-102.

- (2010). *Quantitative Assessment of English-American Speech Relationships*. Unpublished PhD thesis, University of Groningen.

Shinpei, O. (1940). *The Outline of Korean Dialects*. Tokyo: Dongyangmunsa.

Sikma, J. and Gorter, D. (1991). Inventory – A Synthesis Report. In Sikma, J. and Gorter, D. (Eds.) *European Lesser Used Languages in Primary Education*. Ljouwert: Fryske Akademy/Mercator.

Silva, David (2010) Death, Taxes and Language Change: The Inevitable Divergence of Korean Varieties as Spoken Worldwide. In Lee Sang-Oak (Ed.), *Contemporary Korean Linguistics: International Perspectives* (300-319). Paju: Thaehaksa.

Smakman, D. and Heinrich, P. (2015). *Globalising Sociolinguistics: Challenging and Expanding Theory*. New York, NY; London: Routledge.

Social Sciences Research Institute (1992). *Cosenmal taysacen* [Large Dictionary of the Korean Language]. Pyeongyang: Sahoykwahakchwulphansa.

Sohn Ho-Min (1999). *The Korean Language*. Cambridge: Cambridge University Press.

Song Cheol-uy (2001). Tungesenuy tungkup [Grading of Isoglosses]. In Pangenyenkwukhoy (Ed.). *Pangen Sacen* [Dictionary of Dialectology] (pp. 110-112). Seoul: Thayhaksa.

Starostin, S.A, Dybo, A.V. and Mudrak, O.A. (2003). *Etymological Dictionary of the Altaic Languages*. Leiden: Brill.

Tagliamonte, S. (2006) *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

- (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.

Tai, Pingwu (2004). Language Policy and Standardization of Korean in China. In Zhou Minglan (Ed.). *Language policy in the People's Republic of China: theory and practice since 1949*. Boston, MA: Kluwer Academic Publishers.

Thomas, E.R. (2011). *Sociophonetics: An Introduction*. New York, NY: Palgrave Macmillan.

Thomason, S.G. (2001). *Language Contact: An Introduction*. Washington D.C: Georgetown University Press.

Thomason, S.G. and Kaufman, T. (1988). *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

- (1982). The Contribution of Sociolinguistics to Dialectology. *Language Sciences, 4* (2), 237 – 250

- (1997) Typology and Sociolinguistics: Linguistic Structure, Social Strucutre and Explanatory Comparative Dialectology, *Folia Linguistica, 31*(3-4), 349-360

- (2004a). *New-Dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh: Edinburgh University Press.

- (2004b). The Impact of Language Contact and Social Structure on Linguistic Structure: Focus on the Dialects of Modern Greek. In Kortmann, B. (Ed.), *Dialectology Meets Typology: Dialect Grammar from a Cross-linguistic Perspective* (435 – 452). Berlin: Walter de Gruyter and Co.

- (2004c). Linguistic and Social Typology: The Austronesian Migrations and Phoneme Inventories. *Linguistic Typology, 8*, 305 – 320

- Trudgill, P. (2004d). The impact of language contact and social structure on linguistic structure: Focus on the dialects of Modern Greek. *Trends in Linguistics Studies and Monographs*, *153*, 435-452.

- (2006). Dialect Mixture versus Monogenesis in Colonial Varieties: The Inevitability of Canadian English. *The Canadian Journal of Linguistics, 51*, (2/3), 265 – 286

- (2009). Greek Dialect Vowel Systems, Vowel Dispersion Theory, and Sociolinguistic Typology. *Journal of Greek Linguistics, 9*, 165 – 182

- (2011). *Sociolinguistic Typology: The Social Determinants of Linguistic Complexity*. New York, NY: Oxford University Press Inc.

UNESCO (2010). *Unesco Atlas of the World's Languages in Danger*. Retrieved 12 August 2016 from http://www.unesco.org/languages-atlas/en/atlasmap.html

Wada, H (1987). Koreans in the Soviet Far East 1917 – 1937. In Suh Dae-Sook (Ed.), *Koreans in the Soviet Union* (24 - 59)*.* Honolulu: University of Hawai'i, Center for Korean Studies.

Ward Jr. J.H.W. (1963). Hierarchical Grouping to Optimize and Objective Function. *Journal of the American Statistical Association 58*(301), 236-244

Weinreich, U. (1979). *Languages in Contact: Findings and Problems*. Location unknown: Walter de Gruyter.

Weinreich, U., W. Labov & M. Herzog (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (eds.), *Directions for Historical Linguistics*. 97-195. Austin: University of Texas Press.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity.* Cambridge: Cambridge University Press.

Willems, D., Defrancq, B., Colleman, T. and Noël, D. (2003) *Contrastive Analysis in Language: Identifying Linguistic Units of Comparison*. Basingstoke, UK: Palgrave Macmillan.

Wolk, C.B.S. (2014). *Integrating Aggregational and Probabilistic Approaches to Dialectology and Language Variation*. Unpublished Doctoral Dissertation, Albert-Ludwigs-Universitaet, Freiburg.

Wolk, C.B.S and Szmrecsanyi, B. (2016). Top-down and Bottom-up Advances in Corpus Based Dialectometry. In Côté, M-H., Knooihuizen, R. and Nerbonne, J. (eds.) *The Future of Dialects* (pp. 225-244). Berlin: Language Science Press.

Yeon Jaehoon and Brown, L. (2011). *Korean: A Comprehensive Grammar*. Abindgon, UK: Routledge.

-   (2015). Varieties of Contemporary Korean. In Brown, L and Yeon Jaehoon. (Eds.). *The Handbook of Korean Linguistics*. Oxford: Wiley Blackwell.

Yeon Jaehoon (2003). *Korean Grammatical Constructions: Their Form and Meaning*. London: Saffron Books.

- (2012). Korean Dialects: A General Survey. In Tranter, N. (Ed.). *The Languages of Japan and Korea*. Abingdon, UK: Routledge.

Yi Kikap, Kim Cwuwen, Choy Tongcwu, Yen Kyutong and Yi Hencong (2000). Cwungang asia hanintuluy hankwuke yenkwu [Research into the Korean Language of Central Asia's Koreans]. *Hangul 247*, 5 – 72

Yonsei Tayhakkyo Ene Cengpo Kaypalyenkwuwen. (1998). *Yonsei Hankwuke Sacen* 1 [The Yonsei Korean Dictionary 1]. Tusan Tonga: Seoul.

Young, R. and Bayley, R. (1996). VARBRUL Analysis for Second Language Acquisition Research. In Bayley, R. and Preston, D.R. (Eds.) (253-306). *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins.

Yun In-jin (2004). *Kholiantiasuphola: cayoyhaninuyicu, cekung, cengcheyseng* [Korean Diaspora: The Movement, Adaptation and Identity of Overseas Koreans]. Seoul: Korye Tayhakkyo Chwulphanpwu.

- (2012). *Caycwung tongpho ene silthay cosa* [Survey of the Linguistic Reality of Chinese Koreans]. Kwulipkwuekwen: Seoul.

Zedeck, S. (2014). *APA Dictionary of Statistics and Research Methods*. Washington, DC: American Psychological Association.

Zhang J. and Li Y (2007). A Comparative Study on Bilingual Language Policy in Quebec and in Yanbian Korean Autonomous Prefecture. *Khaynata nonchong, 13* (1), 47 – 68

Zheng Xi Shu (2010). Cosencok munhwapyentongkwa munhwacengceyseng [Chinese-Korean Cultural Change and Cultural Identity], *Yeksamunhwayenkwu, 35*, 555-580

Zhou, M. (2003). *Multilingualism in China: The Politics of Writing Reforms for Minority Languages 1949 – 2002*. Berlin: Mouton de Gruyter.

- (2012) Historical Review of the PRC's Minority/Indigenous Language Policy and Practice: Nation-state Building and Identity Construction. In Beckett, Gulbahar H. and Postiglione, Gerard A. (Eds.). *China's Assimilationist Language Policy: The Impact on Indigenous/Minority Literacy and Social Harmony* (18 – 30). Abingdon, UK: Routledge.

Zhou, Q. (2004). Population Bahaviors and Language Maintenance – A Case of the Korean Speaking Community in China. *Collegium Antropologicum, 28* (1), 83-96

Zipf, George. (1935). *The Psychobiology of Language: An Introduction to Dynamic Philology.* Cambridge, MA: M.I.T. Press.