ISBN 978-94-92910-26-4

Printed and bound by Ipskamp Printing
Cover photo: Sign on Taketomi Island reading “This way”, fieldwork 2017.

Variation in form and meaning across the Japonic language family with a focus on the Ryukyuan languages

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 30 maart 2021
om 10.30 uur precies

door

Johan Leonardus Antonius Huisman
geboren op 1 november 1987
te Kerkrade

Promotoren
Prof. dr. A. Majid
Prof. dr. R. W. N. M. van Hout
Prof. dr. A. P. J. van den Bosch

Manuscriptcommissie
Prof. dr. H. de Hoop
Prof. dr. A.M. Backus (Tilburg University)
Dr. P. Heinrich (Università Ca' Foscari, Italië)
Dr. A. Kopecka (Université Lumière Lyon 2, Frankrijk)
Prof. dr. P.C. Muysken

The research reported in this was supported by the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (grant number: 322-70-009).

To my parents, José and Peter

Contents

7 Summary and conclusion

Acknowledgments

I’ve crossed paths with many people over the course of this project and I would like to thank everyone who has been part of the journey. Your guidance, kindness and support have helped me reach where I am today.

First, I am deeply grateful to all the Ryukyuan and Japanese consultants for their contribution. Without your assistance and endless patience, this project would have been impossible. Thank you all for your warm welcome, your generosity in answering my many questions about your languages, and teaching me about your local culture and history:

本研究の方言調査の参加者の皆様にご協力いただいたことに改めて感謝致します。遠く離れたオランダから来たにもかかわらず、私の数々ご質問に対してわざわざ分かりやすく説明をいただいたりと、皆様のご厚意に感激致しました。お教えいただいた有益な情報を活用させていただき、博士論文を仕上げました。重ねてお礼申し上げます。青森県の皆様、ありがんどごす。奄美群島の皆様、ありがっさまりょーた。沖縄諸島の皆様、にふぇーでーびる。宮古諸島の皆様、たんでぃがーたんでぃ、やぐみすでぃがふー。八重山諸島の皆様、しかいとぅにーふぁいゆー。

I would also like to offer my sincere gratitude to those who made my fieldwork life a lot easier, whether by introducing me to speakers, driving me around to teach me about everything local, giving me that at-home feeling when staying over, or just having fun over food and drinks. Thank you Barry and Hiroko, Tony and everyone at the English Cafe, Taishi, Yoko and the rest of the Amami FM team, Satoshi, Mr. Nakazato, Marino and family, Aya and Yukito, Hiro and the Hiraraya Allstars, the Miyako Municipal Cultural Association, the Nishihara Community Centre, fellow fieldworkers Chris, Gijs, Martha, Nana, Natsuko, and Sōichiro, and thanks to Arisa for the translation help. I’d also like to give my special thanks to Inoue-sensei for our engaging converations, and to the senior researchers who were kind enough to free up time in their own busy fieldwork schedules: thank you Iwasaki-sensei, Maeda-sensei, and Takubo-sensei.

I am extremely grateful to my supervisors, whose unwavering belief in me helped bring this thesis to a good end. Asifa, thank you for your continued support, from my time as a Master's student to this thesis and beyond. Your guidance in the various projects we have worked on together, as well as in the academic world as a whole, has helped me grow into a better scholar and will benefit me for the rest of my career. Roeland, thank you for always being available for feedback and thoughtful advice, and never failing to point me to new challenges. You have helped me acquire skills and knowledge I never thought I could acquire, and opened my eyes to many exciting opportunities. Antal, thank you for your support for this project, providing me with an additional audience to discuss my work with, and giving me the freedom to structure the project as I envisioned it.

I would also like to extend my sincere thanks to the members of the thesis committee and examination board, Helen de Hoop, Ad Backus, Anetta Kopecka, Pieter Muysken, Patrick Heinrich, Michael Dunn, Mutsumi Imai, for finding the time to read and evaluate this thesis.

Also very important were the many engaging and insightful conversations with everyone in the Meaning, Culture & Cognition ~~lab~~ family. Thank you Ilja, Laura, Ewelina, Patricia, Josje, Artin, Lila, Carolyn, Alex, Saskia, Ezgi, Hasan, Lilia, Mariann, Claudia, Simge, Afrooz, Simeon, Maggie, Parla, Marisa, Julia, Feifei, Doris, and Hilário. Our interactions, both serious and fun, over lunch, in our meetings, and outside of uni have helped me grow both as a scholar and as a person.

I would like to thank the International Max Planck Research School for Language Sciences and the Graduate School for the Humanities at Radboud University for providing an excellent environment to grow as an academic, and thank you Kevin, Peter and Rob for listening to me go on about my work and keeping me on track through our various progress meetings.

I would also like to thank the *Nederlandse organisatie voor Wetenschappelijk Onderzoek* (PhDs in the Humanities project number 322-70-009). Their financial support enabled me to carry out this research project.

I am also grateful to all the language warriors of CoLang 2016, for inspiring me to see beyond the academic value of linguistic research ad think about its value for the community. I would also like to thank everyone at the Miyako workshop for a wonderful first practical experience with Ryukyuan. I would like to extend my particular thanks to the organisers, Toshihide Nakayama and Yoshi Ono, and of course to our consultant Nakama-sensei, who has also helped me tremendously during my fieldwork.

I would like to thank my office mates and paranymphs Ilona and Nina, whose support will surely help me get through the defense succesfully. Thanks also to fellow office mates Judith, Hans, and of course the best of all: Koos. While I might have been away for long times during fieldwork, coming back to 4.23 was like coming back home.

My gratitude also goes to my family and friends, for their support during this project and for providing me with the necessary off-time over games, music and beer. Special thanks go to my parents, José and Peter, for all their love support throughout everything I have done up until now. You provided me with the perfect environment to finish this thesis in lockdown, and I dedicate this thesis to you.

Finally, thank you Bonnie for your endless love and encouragement—not to mention your fabulous food, which kept me going through the long days of thesis writing. I could not have done this without you, and I hope to provide you with the same support whenever you need it. Looking forward to many more years of fieldwork ‘n’ fun.

List of abbreviations

1	first person
2	second person
3	third person
ACC	accusative
ADN	adnomial
CAUS	causative
COND	conditional
CONT	continuous
DAT	dative
FOC	focus
GEN	genitive
IMP	imperative
INF	infinitive
INS	instrumental
LOC	locative
NEG	negative
NOM	nominative
NPST	non-past
PST	past
Q	question marker
QUOT	quotative
SG	singular
TOP	topic

CHAPTER

1

1 General introduction

1.1 Background

For centuries, scholars have discussed the variability of the many languages spoken across the world. In fact, linguistic diversity is so common and persistent that "everyone knows that language is variable" (Sapir 1921, p. 147). A look at the language databases *Ethnologue* (Eberhard, Simons & Fennig 2020) and *Glottolog* (Hammarström et al. 2020) tell us that there are several thousands of languages at present, each with their own sounds, words and grammar that "embod[y] the intellectual wealth of the people who use it" (Hale 1992, p. 35). Studying this diversity helps uncover systematicity in the variation and increases our understanding of the nature and limits of language itself.

However, linguistic diversity in the world today is not the same as it was thousands of years ago, nor will it be the same in the future. As has been pointed out in earlier work, this diversity "is not something whose future can be taken for granted" (Hale 1992, p. 35). A recent review of the status of the world's languages estimated that more than half of them are threatened or endangered to at least some degree (Seifart et al. 2018). This is an important point to take into account, given that the death of a language equals the "irretrievable loss of diverse and interesting intellectual wealth" (Hale 1992, p. 36). However, Seifart et al. (2018) also concluded that steady progress has been made in research of endangered languages since the call to action by Hale et al. (1992), which itself followed earlier warnings on the vulnerability of endangered languages by e.g., Dorian (1977), and Hill (1977).

At the same time, the review by Seifart et al. (2018) pointed out several points for improvement related to the documentation and description of endangered languages. The first one listed is a lack of material collected through controlled methods that allows for coordinated comparison (Seifart et al. 2018, p. e335). The authors argue that conducting data collection with systematic cross-linguistic comparison in mind can enhance the value of the individual language data points significantly, as the work becomes more than a description on its own. Specifically, they mention the use of semantic field stimuli as one of the pathways to cross-linguistically comparable data (Hellwig 2006), which is important given that "the semantic side of language documentation has been relatively neglected" (p. e335; see also Evans & Sasse 2007).

As such, this thesis aims to contribute to cross-linguistic comparison, and with an emphasis on the study of semantics, while also contributing to the documentation and description of endangered languages. To do so, I have chosen the Japonic language family, spoken across the Japanese archipelago, specifically focusing on the endangered Ryukyuan languages spoken in the south (see Chapter 2 for an overview of the language family). Linguistic typology, and semantic typology in particular, generally focuses on samples of diverse languages (see Dryer 1989; Perkins 1989; Rijkhoff & Bakker 1998), which has left semantic variation across related languages understudied (for some exceptions, see e.g., (Majid et al. 2007, 2015; Wright 1990). The study of the Ryukyuan languages thus serves multiple purposes. In addition to describing the semantic categories in the languages themselves, their study contributes to our understanding of semantic variation across related languages, and puts previous work on Japanese into cross-linguistic perspective. Japanese has been included in many cross-linguistic studies—e.g., on colour in Berlin and Kay (1969); space in Levinson and Wilkins (2006); separation events in Majid, Boster and Bowerman (2008); human locomotion in Malt et al. (2014), and body parts in Majid and van Staden (2015). However, the Japonic languages have no confirmed affiliation to other languages (see also Chapter 2) and so, data from the Ryukyuan languages can provide valuable perspective on the position of Japanese in previous studies as such data will show whether previously described features of Japanese are typical of the Japonic languages as a whole, or specific to Japanese only.

The thesis investigates patterns of linguistic variation in the Japonic language family. Section 2 of this chapter introduces the processes and drivers of linguistic diversification, with specific attention to the role of geography, as the Japonic languages are spoken across an archipelago.

Section 3 introduces two major approaches to the study of linguistic variation (dialectology and typology), and how they have been influenced by the dialect vs. language distinction. Also in this section, quantitative approaches to studying linguistic variation are discussed as these are central to the studies and data analyses presented in the core chapters of this thesis. After that, Section 4 provides an overview of the study of meaning and introduces the three semantic domains that form the empirical basis of this thesis: colour, body parts, and cutting and breaking events. Section 1.5 presents the methodology used in Chapters 3 to 6, after which the final sections summarise the aim, scope, and structure of the thesis.

1.2 Linguistic diversification

1.2.1 Processes of linguistic diversification

Linguistic diversity in the present is the result of processes of language change in the past. Individual speakers differ in their specific knowledge and usage of a language (see e.g., Labov 1966), but because language is used in human interaction, they adapt their speech patterns to accommodate to their conversational partners to ensure efficient communication. This strive for mutual understanding ensures a high degree of homogeneity across a group of speakers that frequently interact—their speech community (Bloomfield 1933). In essence, each speech community is its own self-contained linguistic microcosm with an inevitable, but generally limited, amount of variability among its individual members. Members may introduce new linguistic features, which can then spread across the community along patterns of frequent communication (Labov 2010).

Zooming out beyond a single speech community, contact—or isolation, as the other side of the same coin—is an important factor in linguistic diversification. As with interaction between individual speakers, when there is contact between communities, accommodation to each other's speech patterns will cause the languages of these communities to resemble each other (Chambers & Trudgill 1998; Heeringa & Nerbonne 2001). Through these patterns of contact, linguistic features can spread across language areas through a process called diffusion—see e.g., Gerritsen and van Hout (2006); Trudgill (1974); and Wolfram and Schilling-Estes (2017). These features will first diffuse between communities that share dense interaction, but as the diffusion process takes time, peripheral communities often only adopt features later—see e.g., Kawaguchi and Inoue (2002); Onishi (2010); and Yanagita (1930), for Japan.

Conversely, members of isolated communities only adapt to each other, without accommodating to the speech patterns of other communities. The lack of contact prevents the diffusion of linguistic features into the isolated community. Instead, speakers may develop and adopt linguistic traits that are unique to their specific community, and over time their language becomes different from other communities around them. Initially, these differences will be small—as the process of diversification takes time—but eventually, they may become so large that the language of the isolated community is no longer recognisable as belonging to those around it. Since contact drives accommodation (and isolation drives diversification), we can study the factors that influence patterns of contact in order to understand patterns of linguistic diversity.

1.2.2 Drivers of linguistic diversification

In the linguistics literature, several factors that influence diversification have been discussed, ranging from geographic (e.g., Gavin & Sibanda 2012; Lee & Hasegawa 2014; Séguy 1971) to sociocultural and political (e.g., Britain 2010; Honkola et al. 2018; Labov 2001). Considering the geography and history of the Japonic language family, three potential drivers of diversification will be discussed: (1) geographic distance, (2) the features and configuration of the landscape, and (3) expansion and settlement in new territories.

For logistic reasons, interaction between communities will be more frequent and intense when they are close to each other. As a result of interaction and accommodation, the languages of neighbouring communities generally differ only slightly (e.g., Chambers & Trudgill 1998).

These differences accumulate over increased distance as members of communities that are further apart will interact less, accommodate less, and therefore resemble each other less. This gradual increase of linguistic differences over geographic distance results in a language area known as a dialect continuum or dialect chain, where the languages of neighbouring communities are similar enough to allow their speakers to understand each other without problems, but where speakers from communities at opposite ends of the area might have trouble understanding each other (Chambers & Trudgill 1998).

In addition to distance itself, the features and configuration of the landscape also influence patterns of interaction as they determine the routes available for travel. People are more likely to travel a distance over flat land than they are to travel the same distance over a mountain pass. Mountainous areas and the isolation they bring about have received considerable attention in linguistics—perhaps, in part at least, inspired by Gauchat's (1905) study on the Charmey community in the Swiss Alps. Languages spoken in mountainous areas often show remarkable levels of variation (e.g., Axelsen & Manrubia 2014; Stepp, Castaneda & Cervone 2005), which often puts them in contrast with neighbouring languages spoken at lower altitudes (e.g., Merriam 1907; Post 2013)—see Urban (2020) for an overview. Linguistic diversity across archipelagos is less explored. Work on population genetics has argued that archipelagos are an ideal setting to investigate the role of geography in diversity (Clegg & Phillimore 2010), but linguistic studies rarely have this specific factor in mind and it is an open question whether there is such a thing as "island languages" (Nash et al. 2020). Nevertheless, previous work has shown that the presence of an oceanic barrier leads to more differences between languages (Lee & Hasegawa 2014), and the diffusion of linguistic features happens differently through sea travel as opposed to land travel (Gerritsen & van Hout 2006). The Japanese archipelago thus presents an excellent opportunity to study the effects of geography on linguistic diversification across islands.

Finally, the expansion of a language into new territory can itself bring about language change. The linguistic characteristics of the new community are determined by the language(s) of its original settlers—the founder principle (Mufwene 1996, 2001). As individual speakers differ in their knowledge and use of their language, this influences the overall variation of the new community. Previous work has shown that linguistic diversification coincides with human settlement, which appears to happen in punctual bursts (see, e.g., Atkinson et al. 2008, for Austronesian). In addition, the use of language as part of social identity can cause new communities to either exert conservatism to maintain their heritage, or to innovate deliberately to differentiate itself from the original population (see e.g., (Chambers 1995; Labov 1994). Alternatively, when the new population comprises speakers of several different, but related language varieties, this can also result in the merging and mixture of dialect features as a result of accommodation (see e.g., Kerswill 1996, 2002). The smaller island clusters of the Ryukyus contrast with the larger and connected islands of the Japanese mainland. They were also settled later than the Japanese mainland (Asato et al. 2004), making it possible to study how geographic factors interact with population history, which will be the focus of Chapter 3.

1.3 The study of linguistic variation

The factors that contribute to diversification discussed in the previous section emerged through the study of linguistic variation which has been approached from two distinct—and mostly independent—perspectives: dialectology and typology. While both fields aim to explain patterns of variation, dialectology targets structural variation *within* languages, whereas typology investigates structural variation *between* languages. The focus and methodology of the two fields also differ traditionally. Where dialectology largely focuses on the geographical and social diffusion of linguistic features, typology studies ask whether there are universal patterns or constraints in the range of attested variation across languages (Kortmann 2004, for a comparison). Dialectological data is generally directly collected through questionnaires conducted with informants, which is then published as atlases and/or dictionaries (e.g., Boberg, Nerbonne & Watt 2017; Chambers & Trudgill 1998; for overviews). In contrast, many typological studies gather their data from descriptive grammars and other descriptive texts—although elicitation tasks have become more widespread as well in recent decades (for overviews, see e.g., (Aikhenvald & Dixon 2017; Song 2010).

In analysing patterns of language variation, linguists in both fields often focus on separate linguistic phenomena. However, several methods have been developed to establish an *overall* measure of linguistic similarity or dissimilarity, which will be introduced in Section 3.2. Before that, I will discuss the language vs. dialect problem, how it has influenced the approaches taken by dialectology and typology, and how this will be treated in this thesis, given that it is inextricably linked to Japanese linguistics.

1.3.1 The language vs. dialect problem

The different perspectives in dialectology and typology find their origins in the age-old question of what constitutes a "language" and what makes something a "dialect". The most used linguistic criterion for distinguishing dialects from languages is mutual intelligibility (see, e.g., Hammarström 2008). If two speakers are (despite some linguistic differences) able to understand each other, they are said to speak dialects of the same language; but if they are unable to do so, they speak different languages. While the concept of mutual intelligibility is fairly simple to describe, studies measuring functional intelligibility are rare (although see, e.g., Gooskens 2006; Gooskens et al. 2018; Gooskens & Schneider 2016; Tang & Van Heuven 2009; and Takubo 2018, for Japan). Instead, it is more common to find anecdotal evidence of varieties that are hard to understand. Moreover, in reality language status is more often than not based on non-linguistic criteria, such as social prestige and political power (Weinreich 1945). For example, speakers of Norwegian, Danish and Swedish are able to understand each other without too much trouble, but they are commonly referred to as speaking different "languages". In contrast, speakers from the Amsterdam area in the Netherlands have considerable trouble to understand the highly divergent "dialects" spoken in the peripheries of the country (e.g., van Bezooijen & van den Berg 1999). In a similar fashion, all varieties spoken across Japan were considered "dialects of Japanese" for the longest time, (see Heinrich 2005, for an overview), but several distinct "languages" are now recognised—see Chapter 2.

The problematic distinction between dialect and language has led to several proposals for more neutral terms. For example, Cysouw and Good (2013) suggest the terms *doculect* to refer to a "named linguistic variety as attested in a specific resource", and *languoid* to represent "a collection of doculects or other languoids, which are claimed to form a group" (p. 356). However, as not all Japonic varieties have been attested in written sources, I will instead use the term "(linguistic) variety" throughout this thesis to cover "any system of linguistic acts which can be distinguished in some way from other related systems"—following Musgrave (2016, p. 391). Moreover, use of the term *variety* better enables description through a geographic label and allows for the discussion of different levels within the hierarchy of a language family. For example, the variety spoken in a particular village (e.g., *Koniya*) can also be classified as a variety of the macro area (in this case, *Amami*), which in turn makes it one of the Ryukyuan varieties, and a variety of Japonic. In Chapter 5, I also empirically test whether individual speakers cluster together in meaningfully distinct linguistic varieties to better understand the attested variation in Japonic.

1.3.2 Quantitative measures of linguistic variation

The earliest examples of mathematically computed similarities between languages were explored by early pioneers such Dumont d'Urville (1834) and Broca (1862). Later, correlational methods were imported from anthropology (Czekanowski 1928a; Kroeber & Chrétien 1937, 1939). An approach that is perhaps more widely known is the work of Swadesh (1950, 1952, 1954), who further developed the field by basing his comparisons on a small set of basic vocabulary, and focusing on estimating the time depth of language families. While Swadesh' original method has fallen out of use due to weaknesses in the methodology—e.g., the assumption of a constant rate of change (see Lees 1953)—endeavours in recent decades have built upon this approach, expanding the toolkit by adopting phylogenetic approaches as used in biology (e.g., Chang et al. 2015; Dunn et al. 2005 2011; Gray, Drummond & Greenhill 2009; Gray & Atkinson 2003; Greenhill, Heggarty & Gray 2020), which allow for variable rates of changes and specific calibration of certain points in time to produce more robust analyses.

With Czekanowski (1928b) as an early exception, quantitative measures of similarity and dissimilarity did not become popular within dialectology for a long time. A more concerted effort to quantitatively study variation between dialects was started by Séguy (1971, 1973), who named this new subfield *dialectometry*. Where work derived from Swadesh's glottochronology focused mainly on the time scale of linguistic differences, Séguy instead focused on the patterns of variation over geographic distance (Séguy 1971, 1973), an approach that was expanded upon by Goebl (1981, 1983, 1984). Where Séguy and Goebl focused on individual linguistic features, Kessler (1995) worked with lexical items and quantified differences between dialects by calculating Levenshtein distances (Levenshtein 1966) over pairwise string alignments. As the Levenshtein algorithm is able to pick up small differences *within* items, the strength of this method is the potential to compare language varieties that share many cognates. Most earlier studies used dialect atlases whose entries often focus on linguistic features specific to the area, and the use of basic vocabulary is relatively new for dialects (Bakker & van Hout 2012). The string edit distance approach has been met with criticism for large-scale comparisons (e.g., Greenhill 2011, as its effectiveness decreases rapidly when used on larger language families like Indo-European or Austronesian (e.g., Serva & Petroni 2008, as the decrease in shared cognates means that character correspondences are more likely to be coincidence.

While quantitative methods have been applied to linguistic variation in Japan (e.g., Inoue 1996a, 1996b, 2019), this thesis introduces several innovative uses of such methods to Japonic data, such as the Levensthein algorithm to quantify differences in basic vocabulary (see Chapter 3). In addition, quantitative methods are used to analyse the three semantic domains explored in-depth (colour in Chapter 4, body parts in Chapter 5, and cutting and breaking events in Chapter 6). First, however, I will briefly introduce the study of meaning in general, the approach this thesis will use, and the choice of the domains.

1.4 The study of meaning

Creating and expressing meaning is a fundamental aspect of language. Humans categorise and organise their experiences into units and systems based on (perceived) common features so that they can be expressed through linguistic labels. Highly similar experiences are grouped into a single conceptual category and consequently described by the same label. The diversity of human experiences across cultures has led to a wide variety of conceptualisations, expressed through thousands of languages spoken across the globe. Meaning can be found not only in individual words, but also in the way that different parts of sentences combine and how the context of a particular situation can influence our interpretation of what someone is saying. While meaning is at the core of language, its complexity also has the consequence that other variable features of language, such as grammatical or phonological phenomena, have received more attention in linguistic typology.

Nevertheless, the study of meaning has a long tradition, with one tradition identifying and classifying the types of change in meaning (see e.g., Blank 1999; Bréal, 1897; Stern 1931; Ullmann 1951). Relying primarily on meanings that can be extracted from texts, this line of work has taught us the many general processes of how the meaning of individual words can change. Examples are how a specific meaning can become more general (a process called broadening, e.g., Old English *docga* ‘powerful dog’ changed to *dog* ‘generic dog’ in contemporary English), or the opposite (called narrowing, found in the change of Old English *hund* ‘generic dog’ into *hound* ‘hunting dog’ in contemporary English). Words can also acquire positive or negative connotations (melioration or pejoration, respectively—e.g., German *Knecht* ‘servant’ and English *knight* ‘nobleman’ both derive from the same term that once just meant ‘boy’). Other ways used to depict new meanings are through contiguity (metonymy, e.g., German *Bein* ‘leg’ from Old High German *bein* ‘bone’), and similarity (metaphor, e.g., German *Bein* used for ‘leg of a table’ in *Tischbein* from generic *Bein* ‘human leg’).

In recent decades, other methods developed, aimed at the systematic comparison of lexical meaning across languages. These can be categorised into three main approaches: (1) Natural Semantic Metalanguage (NSM), which uses a limited set of basic conceptual elements (semantic primes) that are proposed to be shared by all languages, as the basis for cross-linguistic comparisons (Goddard 2001; Goddard & Wierzbicka 1994; Wierzbicka 1999); (2) distributional approaches, which study the meaning of words through their combinatorial properties in large-scale corpora, arguing that the way words can co-occur with each other is motivated by their semantic properties (Landauer & Dumais 1997; Lund & Burgess 1996); and (3) denotational approaches, which use non-linguistic stimuli to study how words are used to categorise different types of entities that belong to a single semantic domain—a coherent set of meanings—e.g., space (Bowerman 1996; Levinson & Wilkins 2006), events (Evans et al. 2011; Kopecka & Narasimhan 2012), and perception (Majid et al. 2018; Majid & Levinson 2011).

Each method can provide insights into variation and change in word meaning, and each comes with its own strengths and challenges. While the suggested universality of semantic primes in NSM allows for a systematic investigation of semantic patterns across languages, there has been debate about their expression in actual languages (Evans 2010;

Riemer 2006). An obvious limitation of distributional approaches is the need for sufficiently large linguistic corpora, which unfortunately are not available for many most languages—especially endangered ones, although work on this has picked up in recent years. Finally, critics of the denotational approach have pointed to the decontextualised nature of the stimuli (Levinson 2000; Wierzbicka 2005), or the difficulty of using audio-visual stimuli to straightforwardly represent subjective experiences such as emotions or pain, or abstract notions such as possession (Koptjevskaja-Tamm, Rakhilina & Vanhove 2016).

I have chosen to study semantic variation in the Japonic languages using the denotational approach. A standard set of stimuli makes it easy to collect and compare data from different speakers and different language varieties, making the method particularly suited for underdescribed languages (Koptjevskaja-Tamm, Rakhilina & Vanhove 2016), such as the Ryukyuan languages. The three semantic domains studied—colour (Chapter 4), body parts (Chapter 5), and cutting and breaking events (Chapter 6)—are each described in more detail below. These domains were chosen for several reasons. First, they represent three different types of concepts (properties, objects/parts, and events, respectively), which in Japonic can be expressed through property concept roots, nominals, and verbs, respectively (see Chapter 2 for further details). Secondly, all three domains have received sustained cross-linguistic attention, and best practices for stimulus sets are therefore already established, which enhances comparability. Finally, and importantly, while (Standard) Japanese has been included in cross-linguistic studies on all three domains, work on the Ryukyuan languages is limited to a small number of case studies (e.g., Kusakabe 1964; Nakama 1978, 1984, 1985). These case studies have showed considerable differences between Japanese and Ryukyuan, which calls for further systematic investigation.

1.4.1 Colour

Berlin and Kay (1969) pioneered the denotational approach by adopting a methodology in which non-linguistic stimuli—colour chips, in their case—

served as a prompt for linguistic elicitation. This method allows for the systematic collection of comparable data as it provides the researcher with an objective referential grid against which similarities, differences, and change in extension (the referential ranges, or "boundaries" of words) can be systematically quantified and compared (Majid 2011), creating a neutral basis for language comparison that minimises the researcher's influence on the results. The original methodology by Berlin and Kay (1969) was further developed in the World Color Survey (Kay et al. 2009), a follow-up study that collected data from 110 unwritten languages, spoken by small-scale, non-industrialised communities. Speakers of each language were shown a set of colour chips, one by one, and were asked to name the colour. The study revealed important differences between languages: some languages have three colour terms, while others may have eleven, or even fifteen. Nevertheless, across diverse languages the boundaries for these colours are claimed to be orderly, with a limited number of evolutionary trajectories a colour vocabulary can take (Berlin & Kay 1969; Kay 2015). Kay and colleagues measured similarity in meaning across languages, and found that the semantics of colour terms are not random, but instead reflect cognitive and communicative principles (e.g., Conway et al. 2020; Kay & Regier 2003; Regier, Kay & Khetarpal 2007; Zaslavsky et al. 2018). Additional work using the World Color Survey database has tried to account for the apparent variation by appealing to other factors such as the physical environment (e.g., Lindsey & Brown 2004) or cultural practices (e.g., Majid et al. 2018).

Despite the long study of colour, many cross-linguistic quantitative studies are based on data that was collected decades ago—both the World Color Survey and the Mesoamerican Color Survey were conducted in the 1970s. As such, in Chapter 4, I study the impact of modern society on colour language and semantics. The Ryukyu Islands make an ideal testbed for this given societal changes (resulting in their endangered status) and language contact (with the standard language, but also English through globalisation) in recent decades.

1.4.2 Parts of the body

Early cross-linguistic work on body parts proposed that 'the body' as a whole was the starting point for a hierarchically structured lexicon, in which each subsequent level consisted of 'parts of' the previous level, with a maximum depth of up to six levels (Andersen 1978; Brown 1976). However, more recent studies in several unrelated languages found that a hierarchical organisation of the body part lexicon is not as universal as previously thought (see contributions in Majid, Enfield & van Staden 2006). In addition, languages have been shown to differ in the granularity of distinctions made for body parts (compare Terrill 2006, on Lavukaleve and Burenhult 2006, on Jahai). In another cross-linguistic study, speakers were asked to colour-in body parts (Majid & van Staden 2015), which showed that languages can differ in e.g., the extensional range of limb terms, and that there need not be parallelism between the upper and lower limbs as previously suggested (e.g., Andersen 1978; Brown 1976). Moreover, the colouring-in study found that some parts in the face (e.g., 'eye', 'nose', 'forehead') showed little variation across languages (Majid & van Staden 2015) which raises the question of why some body parts show high variability in meaning, and others low variability.

These findings raise broader questions about the organisation of the body part domain. In fact, it has been argued there might not be a single organisational principle for the body part lexicon as a whole. Instead, if principles exist "they are more likely to be limited to distinct sub-systems such as the face, internal organs, or limbs" (Majid & Enfield 2017). Further evidence for such a proposal comes from Wilkins (1981, 1996), who studied semantic shifts of cognates across several major language families. He formulated five natural tendencies of semantic change in the body part lexicon, one of which—terms for parts shifting to mean the whole, e.g., 'navel' → 'belly'—was suggested to be purely unidirectional. As part of these tendencies, he found four distinct chains of semantic shift ending in the head, body, arm, and leg, but no evidence of shifts between chains, suggesting there might indeed be distinct sub-systems. In Chapter 5, I study the body part domain across the Japonic languages using a multi-method approach to investigate whether there are in fact distinct structuring principles for different parts of the body, and how this relates to levels of lexical and semantic variability.

1.4.3 Separation ("cutting and breaking") events

Separation events, including cutting and breaking events, encompass the "separation in the material integrity" of an object (Hale & Keyser 1987). Verb argument structure alternations show a distinction between "break" semantics (I broke the vase; the vase broke) and "cut" semantics (I cut the bread; *the bread cut)—see Bohnemeyer (2007). A comparative study found that semantic variation across languages was captured in a small number of dimensions: a continuous dimension that represents the predictability of the point of separation, a second dimension distinguishing "tearing" events, and one more dimension that distinguishes "snapping" from "smashing" events [1] (Majid, Boster & Bowerman 2008). At the same time, in-depth investigations of individual languages have revealed some interesting differences in the number of specific categorisations, as well their structure. The number of categories has been shown to range from only a handful (e.g., Yélî Dnye, Levinson 2007) to dozens (e.g., Tzeltal, Brown 2007). In addition, the categories can be organised hierarchically—such the English hyponyms *to slice* and *to chop* compare to the semantically boarder *to cut*—but also include obligatory subdivision—e.g., there is no verb in Dutch that can describe both *snijden* 'to cut with a single-bladed instrument' and *knippen* 'to cut with a double-bladed instrument' (Majid et al. 2007).

In Chapter 6, I revisit the cutting and breaking domain for Japanese, which has previously been considered a cross-linguistic outlier in its semantic categorization. I collected new data to re-examine the semantic structure of the cutting and breaking domain in Japanese. In addition, I collected data in several Ryukyuan varieties to assess their position in

[1] Majid, Boster and Bowerman (2008), p. 242) describe 'snapping' as the breaking of one-dimensional rigid objects into two pieces by application of pressure to both ends, and 'smashing' as the breaking of rigid objects into many pieces by a blow.

cross-linguistic perspective, and to establish the amount of semantic variation for this domain across Japonic. Finally, this new data was used to investigate how lineage-specific developments influence variation through a direct comparison with data from another language family—Germanic.

1.5 Methodology

The thesis adopts a multi-method approach to analysing variation in the Japonic languages, using both quantitative and qualitative methods. Chapter 3 studies general patterns of variation in basic vocabulary using methodology from dialectometry to compute linguistic differences between varieties, and approaches from population genetics to explain these patterns. Chapters 4 through 6 on the three semantic domains pay attention to both variation in form and variation in meaning, at the level of the individual speaker as well as on broader levels such as linguistic varieties and language areas. Triangulating different techniques allows for a more holistic analysis of semantic variation.

1.5.1 Fieldwork

While Standard Japanese has been included in several cross-linguistic studies, the Ryukyuan languages have not, and there is no comparative data available for these languages. As such, I collected new primary data through a series of interviews involving stimulus-based elicitation tasks, which were conducted over several fieldtrips. During each fieldtrip, I collected data from locations ranging from north of the main island Honshū to the southern parts of the Ryukyu Islands—locations that are separated by almost 2,500km. Six language areas were chosen as field sites as they were thought to represent the breadth of variation across the Japonic languages based on previous work (see also Chapter 2): the Tohoku region, the Tokyo area, the Amami Islands, Okinawa and its surrounding islands, the Miyako Islands, and the Yaeyama Islands. The fieldwork locations are shown in Figure 1.1.

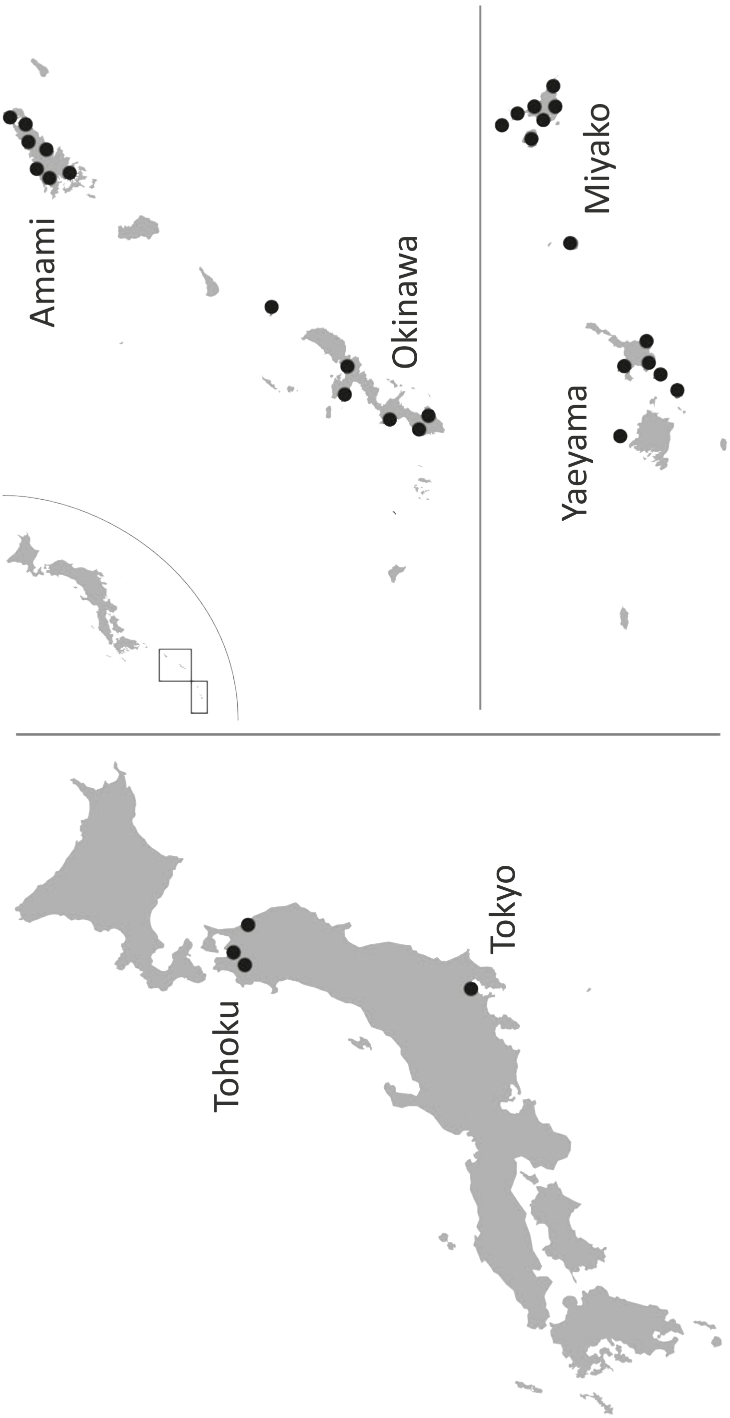


Figure 1.1. Map of fieldwork locations on the Japanese mainland (left panel) and the Ryukyu Islands (right panel).

I travelled to Japan on four occasions and ended up spending a total of around nine months there. My main goal of my first fieldwork trip in January-March 2017 was to become more familiar with the six language areas and introduce myself to speakers and other scholars working on Ryukyuan. After a first successful trip to Japan—in which I was able to collect data for all three semantic domains—a shorter fieldtrip to the Miyako and Yeayama islands was made alongside an international conference in Tokyo in August 2017. This summertime fieldtrip made it clear that a fieldtrip in late winter/early spring was preferable as the humid heat of the Ryukyuan summer was tiring—not only for the (often elderly) speakers, but also for me. As such, I planned my remaining fieldwork in the period January-April in both 2018 and 2019 and the main challenge became packing both for meters of snow in the Tohoku area and the 20°C–25°C across the Ryukyu Islands.

For each area, I collected data from multiple localities, i.e. from multiple linguistic varieties, in order to capture a range of variability in each language area. Native speakers were found by visiting community centres, attending cultural events, contacting local radio stations, and through introduction by other scholars and speakers. Speakers were asked to name and describe audio-visual stimuli: Munsell colour chips, a line drawing of the human body, and video clips showing various cutting and breaking events. In addition to the structured nature of the elicitation tasks, further inquiry into specific features of items that came up during the interviews was conducted to better understand the semantic range of words and descriptions that were elicited. Usually this happened when speakers completed the tasks, but as there was not always time to do so, I also conducted several informal, less structured interviews to help put the elicited items into broader perspective and contextualise their denotational meanings separately. The interviews were conducted in a combination of Standard Japanese and the local Ryukyuan variety and the sessions were audio—and sometimes video—recorded for later transcription with the speakers' consent. Data availability is discussed in each respective chapter.

1.5.2 Additional data

In addition to finding speakers and conducting the interviews, I used my fieldwork time to engage with existing Japanese academic sources—which are often difficult to access outside of Japan—to incorporate knowledge from this scholarship into my analyses. I consolidated data from existing sources such as word lists and dictionaries—to study overall patterns of variation in basic vocabulary, see Chapter 3—as well as corpora, for further insights into terms elicited for the three semantic domains. The main dialect sources used were the *Gendai Nihongo Hougen Daijiten* [Dictionary of Contemporary Japanese Dialects] (Hirayama 1992) and the *Miyara Tōsō Zenshū* [Complete Works of Toso Miyara] (Miyara 1980). The main corpus used was the *Balanced Corpus of Contemporary Written Japanese* (BCCWJ; Maekawa et al. 2014), although the usage of some rarer lexical items was also explored in the Japanese web corpus part of the *TenTen Corpus Family* (Jakubíček et al. 2013). Relevant discussion of historical forms is based on the *Jidaibetsu Kokugo Daijiten* [Periodised Dictionary of the Japanese Language] (Jodaigo Jiten Henshu Iinkai 1967) for Old Japanese, and the *Okinawa Kogo Daijiten* [Dictionary of Old Okinawan] (Okinawa Kogo Daijiten Henshu Iinkai 1995) for older forms of Ryukyuan. These data sources were invaluable to further contextualise the data elicited in the stimulus-based elicitation tasks, and to enrich its subsequent interpretation.

1.6 Aim and scope of the thesis

As mentioned in the introduction, the aim of this thesis is to address several theoretical issues in linguistics while also contributing to the documentation and description to the endangered Ryukyuan languages. The value of the Ryukyuan data is strengthened through its incorporation in the study of broader questions. Given the geography of the Japanese archipelago, the main issue that Chapter 3 tackles is:

Are there differences in the patterns of linguistic variation between connected land and islands?

This issue will recur throughout the discussion of the patterns of variation in the semantic domains in later chapters as well. However, as the Ryukyuan languages are underdescribed, particularly for semantics (Shigeno et al. 2015), the first overarching goal across the three chapters (Chapters 4-6) that focus on semantics is:

What are the semantic categories in the Ryukyuan languages for the domains of colour, body parts, and cutting and breaking?

A comparison between the semantic categories in the Ryukyuan languages and Japanese will be made to assess the amount of semantic variation across the language family. In addition, the Ryukyuan data for each semantic domain will be used to address more specific issues:

To what extent are semantic categories in endangered languages affected by contact with majority languages?

Are some subparts of a semantic domain more variable than others?

How do language-specific developments interact with cross-linguistic constraints?

Question (1) is particularly important for endangered languages, especially if they are indigenous minority languages, as their use is both in decline and under influence from a standard language. This issue will be addressed using the domain of colour in Chapter 4, as this is the only semantic domain for which historical and comparable data is available in both Japanese and Ryukyuan. Question (2) will be addressed using the body parts domain (see Chapter 5), as recent cross-linguistic work has found no clear organising principle for the overall domain that is applied cross-linguistically, but the organisation and variation within subparts within this domain have shown more consistency across languages. Finally, the cutting and breaking domain will be used to address question (3) in Chapter 6, as semantic variation in this domain is constrained by a small number of dimensions cross-linguistically, and for which Japanese was shown to be an outlier.

1.7 Structure of the thesis

In the remainder of the thesis, Chapter 2 will briefly introduce the Japonic language family and present a short overview of comparative work on the three semantic domains that include Japanese or Ryukyuan. Chapter 3 uses basic vocabulary to investigate general patterns of linguistic diversity and how the geographic configuration of a language area influences these patterns. Chapters 4, 5 and 6 investigate variation in three semantic domains. Chapter 4 looks at semantic variation and change in the colour vocabulary of modern Ryukyuan speakers, focusing on the influence of changes in the linguistic landscape—i.e. increased exposure to, and use of, Standard Japanese and English. Chapter 5 explores the semantics of body part terminology, assessing whether there are different structuring principles across subsets within the domain. Chapter 6 revisits expressions for cutting and breaking events in Japanese—which was previously found to be unique in its semantic organisation of this domain—and expands on this by also testing the related Ryukyuan languages to study the effect of lineage-specific developments. Finally, Chapter 7 puts the findings of the thesis into a general overview, discussing their methodological and theoretical contributions, as well as offering potential avenues for future research.

CHAPTER

2

2 The Japonic language family

2.1 Introduction

This chapter briefly describes the Japonic language family and gives an overview of its location, history and internal subdivisions, before providing a short typological overview. The goal is to provide a general background in sections 2.2 to 2.6 for readers who are less familiar with the languages studied in this thesis. It does not claim to be a comprehensive and complete overview of the Japonic language family; it simply provides the basics for understanding the core empirical chapters and providing relevant contextualisation. Section 2.7 will provide an overview of previous research on linguistic diversity across the Japonic languages, as well as work on the three semantic domains covered in Chapters 4, 5 and 6 of this thesis.

2.2 Geographic location and population history

2.2.1 Geographic location

The Japonic languages are spoken across the Japanese archipelago[1], a chain of 6,582 islands in the Pacific Ocean—over 400 of which are inhabited. The four biggest islands—Hokkaido, Honshu, Shikoku and Kyushu, from north to south—make up 96% of the total surface area of Japan and are home to almost 99% of a total population of around 126 million. Major metropolitan areas include the Greater Tokyo Area (approx. 40 million people) and the Kyoto-Osaka-Kobe area (approx. 20 million people). Stretching from the south of Kyushu to Taiwan are the Ryukyu Islands, an arc of almost 200 smaller islands with a contemporary population of approximately 1.5 million on some 70 islands. Extending from approximately 24° to 46° north latitude and from 122° to 146° east longitude,

[1] Japanese is also spoken by Japanese diaspora across the world, whose members number close to 4 million, the largest groups of which are found in Brazil and the United States (mainly Hawai'i)—see the website of the Association of Nikkei & Japanese Abroad (http://www.jadesas.or.jp/).

the Japanese archipelago stretches over more than 3000km on multiple tectonic plates, whose movements are the cause of volcanic activity, frequent earthquakes, and tsunami. Most of Japan's terrain is mountainous with more than 60% of the country's surface covered in forest. The wide latitudinal range results in several climate zones in which the four seasons are often very distinct. The northern regions experience warm summers with average August highs around 25°C, but also around six meters of snow during winter. The central part of the country has a humid subtropical climate, with hot summers (30~35°C in July and August) and mild winters. The southernmost islands lie in the tropical rainforest climate zone with considerable rainfall throughout the year, where highs below 20°C are uncommon even in winter, and where summer temperatures do not change much between night (25~27°) and day (30~32°). Several typhoons pass over Japan each year—which sometimes affect even northern parts of the country. Figure 2.1 is a map of the Japanese archipelago with a zoomed inset map of the Ryukyu Islands.

2.2.2 Population history

The earliest evidence of human activity on the Japanese archipelago has been dated to around 35,000 BP, with the earliest human fossils found in the Ryukyu Islands dating back to around 30,000 BP and stretching until around 18,000 BP (Etler 1996; Nakagawa et al. 2010; Shinoda & Adachi 2017). There is no archaeological evidence that suggests this population survived beyond the Palaeolithic, so it is unlikely that these populations are related to the modern inhabitants of Japan. Instead, from around 15,000 BP the main islands were inhabited by the Jōmon, a Neolithic hunter-gatherer culture with considerable complexity and regional variation (Crawford 2008; Kuzmin 2006), named after the cord-markings that characterise their pottery.

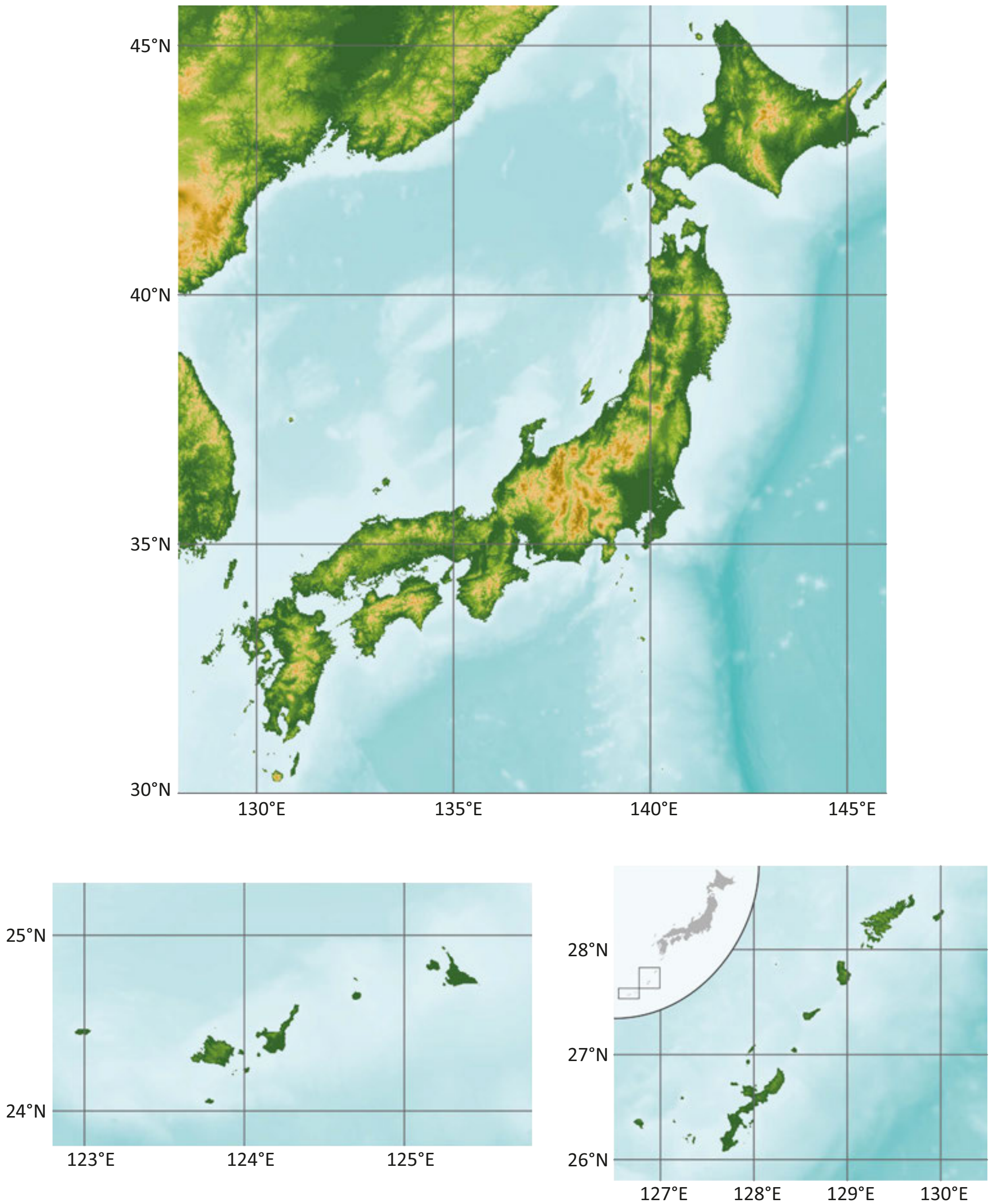


Figure 2.1. Map of the Japanese mainland (top panel), the Amami and Okinawa islands (bottom right), and the Miyako and Yaeyama islands (bottom left).

There is evidence that around 6,500 BP the Northern Ryukyus were home to a culture similar to, but distinct from, the Jōmon culture of the mainland (Pearson 2013; Takamiya et al. 2016). This so-called Shellmound (midden; shell heap) culture were also hunter-gatherers with pottery. There is evidence of contact and trade with Kyushu (with an exchange of pottery and seashells). In the Southern Ryukyus, a distinct pottery culture—*Shimotabaru*; likely of Austronesian origin (Summerhayes & Anderson 2009)—was present from around 4,500 BP, but then disappear in the early 4th millennium BP.

On the mainland, the Jōmon period lasted up to the first millennium BCE, at which time a wave of Bronze Age migrants, the Yayoi, arrived from the Korean peninsula, bringing with them rice agriculture and iron tools (Hudson 2002). The general consensus is that during this period, the Yayoi immigrants interbred with the original Jōmon population, giving rise to what would become the contemporary population of Japan (this is called the "dual structure model", see Hammer et al. 2006; Hanihara 1991; Hudson, Nakagome & Whitman 2020). There is evidence of Yayoi contact with the population in the Northern Ryukyus, but not in the Southern Ryukyus—where a new distinct group without pottery and agriculture appeared. Archaeological evidence suggests that it was not until the 10th century that agriculture was adopted in the Ryukyus (Crawford 2011; Takamiya 2001, 2005), and studies from population genetics suggests that it was a mixed Yayoi-Jōmon population that spread across all of the Ryukyus, replacing the original populations there.

2.3 Affiliation to other language families

Japanese was long considered an isolate with many dialects (e.g., Shibatani 1990; Tōjō 1927, 1954), and while the recognition of Japonic[2] as a "language family" (Matsumori 1995) means it is now connected to its sister languages in Japan, its relations to other language families is still unclear. The mixed genesis of the current inhabitants of Japan described above provide three possible scenarios for the origins of the Japonic languages: (1) the Japonic

[2] The term "Japonic" was coined by Serafim (2003).

languages developed from the language(s) of the original *Jōmon* people; (2) the Japonic languages developed from the language(s) of the immigrant *Yayoi* people; or (3) the Japonic languages developed from a mixed language comprising both *Jōmon* and *Yayoi* elements.

The overall degree over linguistic variation found in the Japonic language family is comparable to e.g., the Germanic languages (Kindaichi 1978) or the Romance languages (Heinrich, Miyara & Shimoji 2015), which suggests a shallower time-depth than macro-families such as Austronesian or Indo-European. This makes it most likely that the Japonic languages have developed from the language(s) of the Yayoi people. Who exactly they were and how they relate to other contemporary populations remains, however, an open question—even after over a century of continuous scholarship (see Elmer 2019, for an overview).

Most theories that have tried to identify a genetic relationship with other languages are based on comparisons with Standard Japanese, and typically suggest a shared descent with geographically close languages. Earlier theories examined a potential relationship with the Ainu languages, traditionally spoken in the northern parts of Japan, Sakhalin, and the Kuril Islands[3] (Shibatani 1990), but current scholarship does not generally consider a genetic relationship between the two anymore[4] (Satō 2010; Vovin 2016). The bulk of contemporary work focuses on Korean, based on correspondences in the both the lexicon and morphology (e.g., Martin 1966, 1991; Robbeets 2005; Unger 2001, 2009). However, as there is still no irrefutable evidence for a connection with Korean, an alternative view suggests that the contemporary inhabitants of Korea descend from a population different from the original population from which the Yayoi originated (Vovin 2013; Whitman 2011). Further theories that have been put forward suggested a connection with Austronesian (Kawamoto 1977; Ōno 1970) or other languages spoken in parts of Asia (e.g., Benedict 1990), but

[3] Sakhalin Ainu and Kuril Ainu are now extinct, and Hokkaido Ainu is moribund.
[4] While the origins of the Ainu languages also still unclear, it has been theorised that Ainu might have developed from a language spoken by the Jōmon population (Hong 2005).

these theories have generally not received widespread support (see e.g., Blust 2014).

For now, the origins of the Japonic languages remain an open question, which might eventually be solved by combining insights from linguistics, archaeology, and population genetics.

2.4 General subdivisions

Despite claims made about the unintelligibility of some Japonic varieties (e.g., Kindaichi 1978), there has been little research that functionally tests mutual intelligibility. Takubo (2018) is a recent exception, who showed that varieties of Amami and Miyako are unintelligible to each other and to speakers of Standard Japanese. Earlier work showed that Okinawa and mainland Japanese varieties share around 65 to 70 percent of the basic vocabulary (Hattori 1954, 1961). Recent linguistic work has therefore pushed for the recognition of multiple distinct languages within Japan. It is now generally accepted that there are at least several distinct Ryukyuan languages, an effort that has been supported through the inclusion of several entries in the UNESCO *Atlas of the World's Languages in Danger* (Moseley 2010). However, this sentiment is not yet common for varieties of the Japanese mainland, despite the unintelligibility of Tohoku varieties (spoken in the north of the main island Honshū) to speakers of Standard Japanese, for example (Takubo 2018).

Exactly how many "languages" there are will always be up for debate, but it has been argued that they might even number into the dozens if mutual intelligibility were used as a criterion (Takubo 2018). Leaving aside the discussion about dialects versus languages (see Chapter 1, section 3.1), several distinct subgroups can be recognised within the Japonic language family—shown in Figure 2.2.

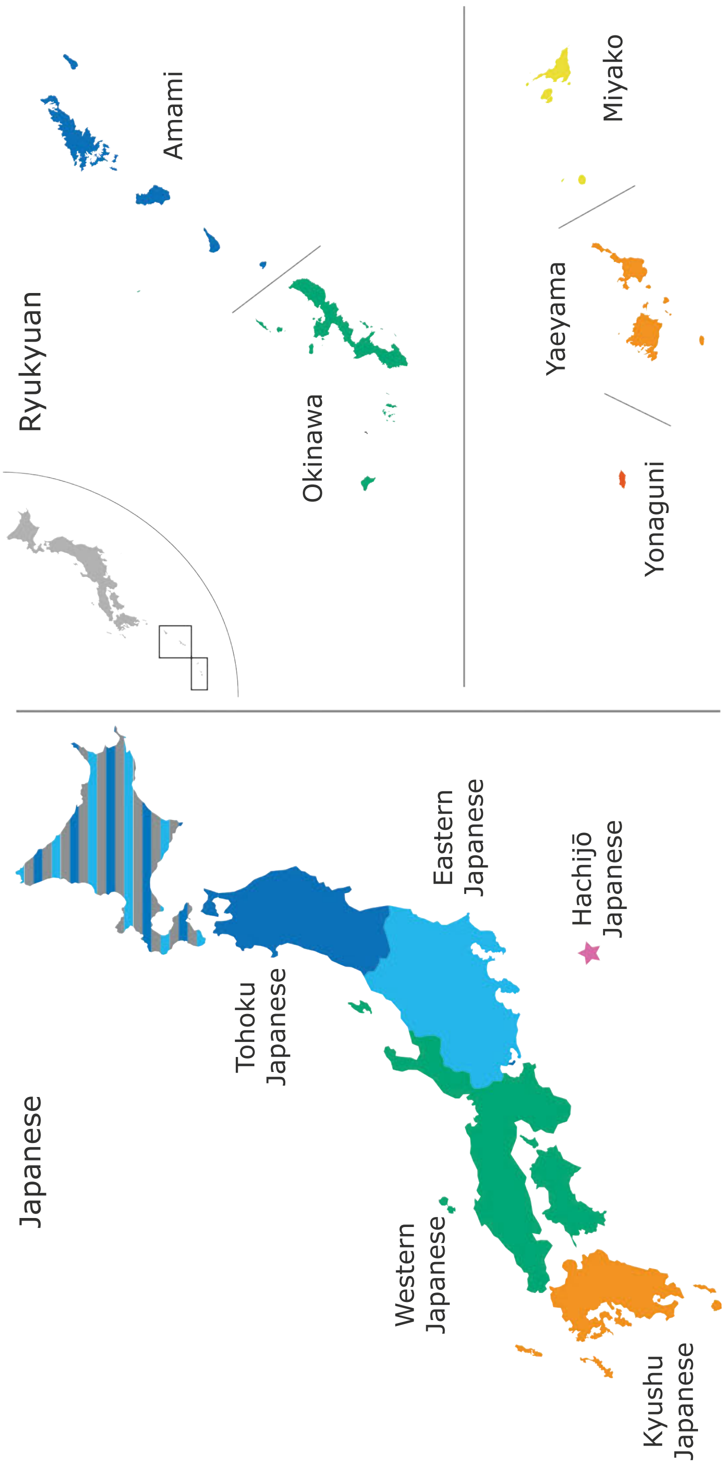


Figure 2.2. Map of the Japonic language family (left panel: Japanese, right panel: Ryukyuan), with major subdivisions

The extant Japonic language family consists of two main branches[5]: (1) Japanese, spoken on the Japanese mainland—i.e., Hokkaido, Honshu, Shikoku and Kyushu—and its surrounding islands, and (2) Ryukyuan, spoken across the majority of the Ryukyuan Islands arc in the south of Japan—with Amami-Ōshima and Kikai as the north-eastern boundary, and Hateruma and Yonaguni as the south-western boundary.

2.4.1 The Japanese branch

Traditionally the primary division within the mainland Japanese varieties is between Western Japanese and Eastern Japanese (e.g., Kindaichi 1955; Tōjō 1927), which is largely based on several grammatical isoglosses (Tokugawa 1981). The varieties spoken on the southernmost main island Kyushu, and the varieties spoken in the Tohoku region in the north of Honshu form two cohesive and distinct subgroups (see Shibatani 1990), but there are different views on their position within the language family. Some see them as distinct subbranches of Japanese, whereas others include them as part of Western Japanese (for Kyushu) and Eastern Japanese (for Tohoku). The varieties spoken on Hachijō-jima and Aogashima[6], which are located approximately 300 kilometres south of Tokyo, are also highly divergent, but their exact position within the Japanese branch is uncertain. They are often referred to as the only living descendants of Eastern Old Japanese and have therefore been argued to constitute a main subbranch within Japanese (e.g., Hattori 1976). This assessment is based on several grammatical features that were unique to Eastern Old Japanese that are preserved in the Hachijō varieties, e.g., the attributive forms of verbs and adjectives[7] (Hōjō 1948, 1966; see Iannucci 2019, for a recent discussion).

[5] A third branch, Peninsular Japonic, is a term that has been used to represent the varieties of the now extinct Japonic languages formerly spoken on the Korean peninsula (e.g., Vovin 2013).

[6] Related, quite distinct, varieties were also spoken on the now abandoned Hachijō-kojima (Hirayama 1965).

[7] Other modern Eastern Japanese varieties also preserve several features unique to Eastern Old Japanese, such as the imperative suffix *-ro* for vowel-stem verbs (all Eastern Japanese varieties), or the negative auxiliary *-noo* (thought to derive from Eastern Old Japanese *nafu*; used in the Ikawa and Narada dialects, see Iitoyo et al. 1983).

2.4.2 The Ryukyuan branch

For the second main branch of the Japonic language family, Ryukyuan, the main division generally recognised is between Northern Ryukyuan and Southern Ryukyuan (Pellard 2015; Shibatani 1990), which are separated by the approximately 250 kilometre wide Miyako Straight—the largest gap between any two neighbouring islands across Japan. Further subdivisions of the Ryukyuan languages are often made along geographical lines. As such, Northern Ryukyuan is generally divided into an Amami subgroup and an Okinawa subgroup (Pellard 2015; Shibatani 1990). Some add Kunigami, comprising the southernmost Amami and northernmost Okinawa varieties as a distinct subgroup (e.g., Karimata 1999; Uemura 1997). Southern Ryukyuan can be considered to comprise three distinct subgroups: Miyako, Yaeyama and Yonaguni, of which the latter two are sometimes grouped together into a Macro-Yaeyama group (Pellard 2015). Others, however, consider Yonaguni to be distinct enough to warrant classification as distinct third Ryukyuan subbranch alongside Northern Ryukyuan and Southern Ryukyuan (e.g., Hirayama 1966; Thorpe 1983; see also Tranter 2012).

While the main focus of Chapter 3 lies on investigating the influence of geography on linguistic diversity, the nature of the data and methodology also enables a classification of varieties across the Japonic language family based on dialectometry—an approach that has not been taken before for lexical data. This brief discussion serves as a backdrop for the results from Chapter 3 where the dialectometric classification will be compared against the traditional view presented above. To briefly anticipate the results, the novel dialectometric analysis confirms the traditional classification, with some interesting caveats.

2.5 Language history

2.5.1 Japanese branch

Generally, the history of Japanese is divided into five stages that roughly correspond to political periods (Table 2.1). Frellesvig (2010) summarises the linguistic changes between these periods as follows: (1) the shift from Old Japanese to Early Middle Japanese brought about many changes related to syllable structure and segmental phonology, (2) there was increased Sinification of the lexicon during Early Middle Japanese, (3) the shift from Early Middle Japanese to Late Middle Japanese led to many grammatical changes related to morphology and syntax, and (4) Modern Japanese saw increased Westernisation of the lexicon.

Table 2.1

Overview of stages of the Japanese language.

Century	Language stage	Political period
8^{th}	Old Japanese	Nara
9^{th}–12^{th}	Early Middle Japanese	Heian
12^{th}–16^{th}	Late Middle Japanese	Kamakura / Muromachi
17^{th}–19^{th}	Early Modern Japanese	Edo
mid-19^{th} onwards	Modern Japanese	Meiji onwards

Although some inscriptions on artefacts go back as far as the 5^{th} or 6^{th} century CE, comprehensive written attestations of Japanese date back to the 8^{th} century CE. The *Kojiki* (712 CE), the *Nihon Shoki* (720 CE) and the *Man'yōshū* (759 CE) are the most famous earliest sources[8]. These sources, as well as sources representing the subsequent stages of Japanese, largely reflect the language spoken in (the area around) the ancient capitals of Nara and Kyoto. Nevertheless, even the oldest sources contain poems written in dialects spoken in what roughly corresponds to the

[8] Dates listed are compilation dates; their contents are generally believed to predate compilation.

modern-day Kanto region. These are commonly grouped together as Eastern Old Japanese and some of their features have been preserved in contemporary geographic patterns of dialectal variation.

Modern Standard Japanese (*hyōjungo* or *kyōtsūgo*) is primarily based on the language variety spoken by the upper class of Tokyo. However, the establishment of Edo (the old name for Tokyo) as the capital at the beginning of the 17th century also caused an influx of migrants from the region around Kyoto and Osaka (Iwasaki). As a result, the modern standard language also contains elements from this region. For example, while the prosodic system (pitch accent), negative auxiliary *-nai* in colloquial speech, and imperative suffix *-ro* for vowel-stem verbs find their origin the Tokyo region, the negative auxiliary *-sen* in polite speech is an influence from the Kyoto-Osaka region.

2.5.2 Ryukyuan branch

In contrast to the long history of written Japanese, there was and is no standard orthography for any of the Ryukyuan language varieties (Heinrich, Miyara & Shimoji 2015). As a result, comprehensive attestations of Ryukyuan only date back to the 16th century. The most extensive work is the *Omoro Sōshi*, a collection of songs and poetry from the Amami and Okinawa islands, which was compiled in the 16th century—but parts of which are thought to go back to the 12th century. Compared to the study of older Japanese sources, however, these older Ryukyuan sources have received considerably less scholarship, particularly in the international literature. Older language forms can also be found in local songs and poetry, traditionally transmitted orally without a long established written tradition. Chapter 4 of this thesis provides a small foray into incorporating historical data into a contemporary study of colour semantics in Ryukyuan.

2.6 Typological overview

As a coherent set of related languages, the varieties of Japonic share many features. Here, I present a number of features that apply to the vast majority of Japonic varieties, although it should be noted that given the variability present in the language family as a whole, there can be exceptions to most of these features—some of which are highlighted in the following paragraphs. The examples presented here are not taken from my own fieldwork, but come from existing descriptive work on the Japonic languages. For an international audience, I have focused on English work, and I refer to the sources of the examples for further in-depth discussion of the specific language varieties cited below.

The number of vowel quality features in Japonic varieties ranges from three in Yonaguni (/i,u,a/, see e.g., Yamada, Pellard & Shimoji 2015) to eight in some varieties spoken in Central Honshu (e.g., /i,y,u,e,ø,o,æ,a/ in Nagoya dialect, see e.g., Hirayama 1992). Some vowels may appear only as long vowels, such as close-mid vowels in many Ryukyuan varieties (e.g., Nakamoto 1981). The Ōgami dialect (Miyako; Ryukyuan) is the only Japonic variety that has no voicing contrast for obstruents and so only has 9 consonants (Pellard 2009), but most Japonic varieties have around 14 to 18 consonants.

Most Japonic varieties prefer an open (CV) syllable structure, although a limited number of consonants can take the coda slot—mainland varieties generally only allow an (assimilating) moraic nasal /N/, whereas varieties of Amami and Miyako have a wider range of consonants that can do so (see Example 1a-c).

(1) Examples of consonantal codas across Japonic

a. Standard Japanese
/siNbuN/ [ɕimbuɴ] 'newspaper', /siNkaNseN=de/ [ɕiŋkaɰ̃sen=de] 'bullet train=INS'

b. Koniya, Amami (Nakamoto 1981)
hap 'snake', *ˀiɴ* 'dog', *mik* 'right', *mït* 'water', *nam* 'wave', *wunak* 'woman'

c. Ōgami, Miyako (Pellard 2009)
im 'sea', *iŋ* 'dog', *kf:* 'to make', *psks* 'to pull'

Standard Japanese is well known for its mora-based speech rhythm (e.g., Bloch 1942; Han 1962; Port, Dalby & O'Dell 1987), in which a mora is a phonological unit with a constant duration. Many other Japonic varieties are mora-based and exceptions are mostly found in the Tohoku and Kyushu areas (e.g., Hirayama 1992). Vowels carry moraic weight, with long vowels analysed as a sequence of two vowel, thus constituting two morae. In addition, consonants in the coda position (see examples above), as well as the first part of geminate consonants (transcribed as /Q)/ carry moraic weight (see Example 2).

(2) Mora versus syllable in Japonic

a. Standard Japanese

	Morae	Syllables
/siNbuN/ 'newspaper'	*ɕi.m.bu.ɴ*	*ɕim.buɴ*
/to:kyo:/ 'Tokyo'	*to.o.kjo.o*	*to:.kjo:*
/su:Qto/ 'quickly'	*su.u.ʔ.to*	*su:ʔ.to*

b. Tarama, Miyako (Aoi 2015, p. 407–408)

	Morae	Syllables
/Qsu/ 'white'	*s.su*	*ssu*
/mim/ 'ear'	*mi.m*	*mim*
/tur/ 'bird'	*tu.ɭ*	*tuɭ*

Pitch accent, where the meaning of a word can be determined by the position or absence of a drop or rise in pitch, is a distinguishing feature in most varieties of Japonic. In Standard Japanese, for example, *hashi* can have three meanings: 'chopsticks', 'bridge', and 'edge'. In isolation, *háshi* 'chopsticks' and *hashi* 'bridge/edge' can be distinguished by their pitch accent, with the former having high-low accent, and the latter having flat accent. The meanings 'bridge' and 'edge' can be distinguished when the subject marker *ga* is added, where *hashí=ga* with accent on the second mora means 'bridge', and where flat accented *hashi=gá* means 'edge'. Some varieties in the Tohoku and Kyushu areas do not use pitch accent (see e.g., Hirayama 1960).

The basic constituent order in Japonic languages is SOV (see Examples 3–5), but besides being strictly predicate-final, constituent order can vary according to pragmatic factors, and there can be extensive ellipsis. Questions are formed by adding a question particle (which can differ between open and polar questions—compare Examples 6a and 6b) without changing the order of constituents. Most varieties have nominative-accusative case alignment, but this is neutralised in the Hateruma variety of Yaeyama Ryukyuan (see Aso 2010, 2015). Particles are used for case marking. The Japonic languages employ a modifier-head order, which extends to subordinate clauses, which are dependent-marking (see Example 7).

(3) Mutsu, Tōhoku (Kibe & Aoi 2020, p. 113—glossing adapted)

inu=ŋa	*sika=ba*	*midea*
dog=NOM	deer=ACC	see:CONT

‘The dog sees a deer.’

(4) Nakanogō, Hachijō (Kibe 2013, p. 159—glossing adapted)

ar=a	*kiː=wa*	*isogaɕikja*
1SG=TOP	today=TOP	busy

‘I am busy today.’

(5) Yoron, Amami (adapted from Kibe 2016, p. 121)

ɸunu	*ʔwa:gi=ja*	*ɸune:da*	*ukina:=nunti*	*ho:taɴ*
this	coat=TOP	recently	Okinawa=LOC	buy:PST

‘I bought this coat in Okinawa the other day.’

(6) Shuri, Okinawa (adapted from Miyara 2015, p. 396)

a.

taa=ga	*ʔiʧ-u-ga*
who=TOP	go-NPST-Q

‘Who is going?’

b.

ʔare=e	*ʔiʧ-u-mi?*
3SG=TOP	go-NPST-Q

‘Is he going?’

(7) Okinoerabu, Amami (van der Lubbe & Tokunaga 2015, p. 359)

ʔmaa=ni	*ʔu-nu*	*ʔmaa*
there=LOC	exist-ADN	horse

‘The horse that is there.’

Information structure includes sensitivity to the syntactic notion of subject as well as the more pragmatically oriented notions of topic (Examples 8a and 8b)—which occurs in both Japanese and Ryukyuan—and focus (Examples 9a and 9b)—which now only occurs in Ryukyuan, but was present in older stages of Japanese (Example 10).

(8) Standard Japanese

a.

sakana=ga	*oyogu*
fish=NOM	swim

‘The/a fish is swimming.’

b.

sakana=wa	*oyogu*
fish=TOP	swim

‘Fish swim.’

(9) Shuri, Okinawa (Miyara 2015, p. 393–340—glossing adapted)

a.

Kamadee=ga	*maŋgoo*	*ʧuku-ju-mi*
Kamadee=NOM	mango	grow-NPST-Q

‘Will Kamadee grow mangoes?’

b.

Kamadee=ga=ga	*maŋgoo*	*ʧuku-ju-ra-jaa*
Kamadee=NOM=FOC	mango	grow-NPST-Q-I.wonder

‘Is it Kamadee who will grow mangoes?’

(10) *Ise Monogatari* (10th c.; Frellesvig 2010, p. 255—glossing adapted)

wotoko=pa	*kono*	*wonna=wo=koso*	*eme*	*to*	*omopu*
man=TOP	this	woman=ACC=FOC	get	QUOT	think

‘The man thought that it was this woman (and her alone) that he wanted.’

Nominals and verbs are unambiguously distinguished in all varieties. Nominals are defined syntactically through lack of inflectional morphology, whereas verbs are defined by their inflection. Japonic varieties are agglutinative, with single morphemes rarely encoding more than one meaning. A rich array of verbal suffixes is used to inflect for tense, polarity, aspect, mood, and honorific value (Examples 11–12).

(11) Yuwan, Amami (Niinaga 2015, p. 339—glossing adapted)

a.	Past tense		b.	Negation	
	waŋ=ga	*koo-ta*		*wan=na*	*jum-aɴ*
	1SG=NOM	buy-PST		1SG=TOP	read-NEG
	‘I bought (it).’			‘I don’t read.’	

(12) Ura, Amami (Shigeno 2010, p. 25—glossing adapted)

a.	Passive	b.	Causative
	jum-ar-i		*jum-as-jur-i*
	read-PASS-NPST		read-CAUS-IMP-NPST
	‘be read’		‘make (somebody) read’

Japonic varieties possess an adjective class that can be identified as property concept roots with a variety of conjugational endings (see Examples 13–16). Especially in the Ryukyuan languages, these endings are often shared with verbs, and some classify adjectives as a subclass of verbs. However, as a result of their semantic properties, they are still often referred to as “adjectives” in the literature.

(13) Standard Japanese

a.	*taka-i*	b.	*taka-katta*	c.	*taka-kunai*
	high-NPST		high-PST		high-NEG:NPST
	‘(It) is high’		‘(It) was high’		‘(It) is not high’

(14) Sonai, Yonaguni (adapted from Yamada et al. 2015, p. 463)

a.	*thaga-n*	b.	*thaga-tan*	c.	*thaga-minun*
	high-NPST		high-PST		high-NEG:NPST
	‘(It) is high’		‘(It) was high’		‘(It) is not high’

(15) Ogami, Miyako (Pellard 2010, p. 141)

takaa-taka=nu	*kii*
high-high=NOM	tree
‘A tall tree’	

(16) Hateruma, Yaeyama (Aso 2015, p. 427—glossing adapted)

kuma-kuma	*sïs-i=ba*
small-small	cut-IMP=COND

'Cut (these) into small pieces.'

In addition, the Japonic languages are known for possessing a lexical class of ideophones (Example 14; also called mimetics in the Japanese tradition, e.g., Akita 2009; Hamano 1998; Iwasaki, Sells & Akita 2017), with some variation in the types of meanings that this class can express between varieties (McLean 2020).

(17) Standard Japanese (Akita & Tsujimura 2016, p. 142)

Taroo=ga	*eda=o*	*bokiboki=to*	*otta*
Taro=NOM	branch=ACC	MIM=QUOT	break:PST

'Taro broke branches with a forceful snap.'

2.7 Previous work on Japonic relevant to this thesis

As introduced in Chapter 1, this thesis investigates the influence of geography on linguistic diversity (Chapter 3), and examines diversity across three semantic domains: colour, body parts, and cutting and breaking events (Chapters 4–6). The following paragraphs give an overview of previous work on these topics, particularly work featuring different Japonic varieties. This brief section is meant as a general introduction and overview; for a more detailed discussion please refer to the respective chapters. As we will see, most studies of semantics—especially in cross-linguistic comparisons—are limited to Standard Japanese. This highlights the needs for comparative work on other Japonic varieties, which is why this thesis will systematically study semantics across the entire Japonic language family. In addition, many dedicated studies are written in Japanese only, so this thesis aims to introduce these studies to the international audience.

2.7.1 Linguistic diversity

Dialectology has a long tradition in Japan—e.g., Tōjō (1938); see Grootaers (1967, 1982) for overviews of early work—and it has produced at large collection of descriptive work, often published in Japanese, such as the *Linguistic Atlas of Japan* (published in 6 volumes, National Insitute for Japanese Language and Linguistics 1966-1974), the *Grammar Atlas of Japanese Dialects* (published in 6 volumes, National Insitute for Japanese Language and Linguistics 1989-2006), and the more recent *Field Research Project to Analyze the Formation Process of Japanese Dialects* which ran from 2010 to 2015, which is aimed to produce the *New Linguistic Atlas of Japan* by the National Institute for Japanese Language and Linguistics. Other works include the *Zenkoku Akusento Jiten* [Countrywide Pitch Accent Dictionary] (Hirayama 1960) and the *Gendai Nihongo Hōgen Daijiten* [Dictionary of Contemporary Japanese Dialects] (Hirayama 1992), as well as a large collection of conversational data in the *Hōgen Danwa Siryo* [Texts of tape-recorded conversations in Japanese dialects] (in 10 volumes, National Insitute for Japanese Language and Linguistics 1978-1987) and the *Nihon no Furusato Kotoba Shūsei* [Collection of Japanese Hometown Language] (in 20 volumes, National Insitute for Japanese Language and Linguistics 2001-2008). While the Ryukyuan languages are often represented—albeit sparsely—in these works, more specific descriptive work has also been compiled, such as Miyara's (1930) *Yaeyama Goi* [Yaeyama vocabulary], the *Okinawago Jiten* [Dictionary of the Okinawan Language] by the National Insitute for Japanese Language and Linguistics (1963), and the atlas *Zusetsu Ryūkūgo Jiten* [Illustrated Dictionary of the Ryukyuan Languages] (Nakamoto 1981).

While the focus of Japanese dialectology is traditionally descriptive, several studies have used the databases produced by this tradition in quantitative studies. Inoue (1996a, 1996b) provides an early overview of computational dialectology, with several examples from work in Japan. In many cases, non-standard varieties are studied in comparison to Standard Japanese (e.g., Inoue & Kasai 1989). An approach perhaps unique to Japan is the use of railway distance as a predictor of distinctness from the standard languages (e.g., Inoue 2004a), which is possible as railroads in Japan approximate the historical road network (Inoue 2019). Studies that compare all varieties to each other rather than to Standard Japanese alone include Yarimizu (2009), who conducted pairwise comparisons using the Levenshtein algorithm based on grammatical data, and Lee and Hasegawa (2011), who used phylogenetic methods. The influence of geographic factors on linguistic diversity in Japan has been studied by e.g., Lee and Hasegawa (2014), who showed that the presence of an oceanic barrier increases linguistic variation, and Jeszenszky et al. (2019), who used geographic distance and travel time estimates to predict patterns of lexical variation. Recent work has also put forward a mathematical model that predicts the distribution of contemporary variation based on the historical process in which linguistic innovation spread from a centre into the periphery (Takahashi & Ihara 2020).

In Chapter 3, I combine insights from these studies to investigate the influence of geography on linguistic diversity, by applying dialectometric methods to a newly consolidated lexical database and incorporating information on both geographic distances and oceanic barriers.

2.7.2 Colour

Chapter 4 examines how colour is lexicalised in Japonic varieties. The colour lexicon of modern Standard Japanese has been studied extensively. Japanese was included in the classic study by Berlin and Kay (1969) where it was shown to have 11 basic colour terms. Later work by Uchikawa and Boynton (1987) and Kuriki and colleagues (2017) provide us with almost five decades of colour naming studies on Japanese. Through this, we can see how the Japanese colour lexicon has changed (see Kuriki et al. 2017; Uchikawa & Boynton 1987). The traditional native terms *daidai(iro)* 'orange' and *momo(iro)* for 'pink' have slowly been replaced by their English loanword counterparts *orenji* and *pinku*, respectively. The term *nezumi(iro)* 'grey'—from *nezumi* 'mouse/rat'—has fallen out of use in favour of *hai(iro)*—from *hai* 'ash'. Finally, 'light blue'—mainly described with *mizu(iro)*, from *mizu* 'water'—seems to have established itself as a twelfth basic colour.

In addition, the long history of writing in Japan has made it possible to study historical developments in the colour lexicon. The oldest writings (from around the 8th century) already contain eight colour terms. Stanlaw (2007, 2010) has proposed a four-term system for the proto-language, in which the terms were (also) used to distinguish brightness vs. darkness, clarity vs. opacity and warmness vs. coolness. Gradually, the terms focused on hue, after which 'yellow' as a fifth colour concept was added as a result of influence from the Chinese philosophy of the five elements. The Japanese expanded on this by adding another virtue with 'purple' as the corresponding colour. This early appearance of a purple category is cross-linguistically uncommon—although not unheard of (see MacLaury 2001). Other work on the historical colour lexicon in Japanese includes discussion of traditional colour names based on the plants that were used as the sources of dye (McNeill 1972). In modern times, most Japanese speakers use a set of colour terms borrowed from English that supplement, rather than replace, the native colour vocabulary (see e.g., Haarmann 1989; Hinds 1974; Stanlaw 1987, 1997).

For the Ryukyuan languages, Kusakabe (1964) is the only in-depth study into the colour lexicon based on the systematic collection of colour naming data—based on 15 chromatic colour cards from the colour system by the Japanese Colour Institute. The study showed that most Ryukyuan speakers use two, three or four chromatic colour terms, which is considerably different from the nine that contemporary Standard Japanese speakers were using (Berlin & Kay 1969). Systematic colour naming data has not been collected for varieties spoken on the Japanese mainland, but data from dialect dictionaries suggests that not all varieties use eleven basic colour terms as in Standard Japanese (see e.g., Hirayama 1992). Particularly, the distinction between 'green' and 'blue' is not always made across non-standard varieties (e.g., Hirayama 1992—see also Conlan 2005, for an in-depth study on the encoding of green and blue shades in Standard Japanese).

For Chapter 4, I collected new colour naming data from Standard Japanese and three Ryukyuan language areas (Amami, Miyako, and Yayeama) to study how societal changes in the Ryukyus in recent decades (e.g., endangered language status, and contact with the standard language and English through globalisation) have influenced colour language and semantics.

2.7.3 Parts of the body

Chapter 5 investigates the semantics of body part terminology across Japonic. Compared to colour, few studies have systematically investigated the extensional range of body part terms in Japanese. Majid and van Staden (2015) compared the extensional range of several body terms in Standard Japanese, Dutch and Indonesian and found several differences between these three languages. No equivalent study has been conducted for non-standard varieties. Instead, work on non-standard varieties remains limited to vocabulary lists and dialect dictionaries—although it must be noted that this work is very extensive in its descriptive value, and one can often find comments on the semantics of specific terms. Parts of the body are a common topic in vocabulary lists for dialect research and an entire volume of the Linguistic Atlas of Japan is dedicated to “the body” (National Insitute for Japanese Language and Linguistics 1968), which shows extensive lexical variation across the language family.

In contrast, non-literal uses of body part terms have received considerably more attention. Several studies have looked at metaphorical extensions of body parts to describe emotion and cognition in Standard Japanese. These studies include both descriptions of the Japanese system on its own (e.g., Hasada 2002; McVeigh 1996), but also comparisons between Japanese and other languages, e.g., Thai and English (Berendt & Tanita 2011). In addition, there have been multiple studies on the use of body part terms in idiomatic expressions and proverbs, again looking at the Japanese system as such, as well as in comparison to e.g., Chinese and Korean (Haegyoung 2009; Yoshida & Zhi 1999).

Chapter 5 explores lexical and semantic variation in the body part lexicon using both a body part naming task and a body colouring task. Combining these two approaches with a novel application of statistical analyses presents a holistic view of body part semantics in Japanese and Ryukyuan, highlighting key similarities and differences between the language areas.

2.7.4 Separation ("cutting and breaking") events

The final semantic domain that is studied in this thesis comprises cutting and breaking events (Chapter 6). Japanese cutting and breaking verbs have been studied from several perspectives, including descriptive work on Standard Japanese—both modern and historical—work on non-standard varieties, contrastive and cross-linguistic comparison, and studies in the context of second language learning of Japanese.

In the earliest work on Standard Japanese, Kunihiro (1970) and Kaetsu (1979) compared several 'destruction verbs' based on a number of features, such as how they encode the method and range of the applied force, and characteristics of the object and the resulting fragments. Several experimental studies were carried out by Hojo. The first study (Hojo 1983) had participants make similarity judgements between verbs in a triad task. In another study (Hojo 1991), participants were first asked to divide a square into pieces based on whether the square had been *waru*-ed, *kudaku*-ed, or *kaku*-ed[9], three verbs than can be used to talk about 'breaking'. The results showed that the verbs differed in the number of fragments they produce, as well as the size of those fragments. In the second part of the study, participants were asked to choose which of the three verbs best described a generated set of fragments, which was found to match the results of the drawing task. Finally, Hojo (1993) had participants do a pile-sorting task and found that participants grouped separation verbs based on a distinction between 'cutting' versus 'dividing', and the number of fragments that were produced. For historical work, Hashimoto (2007) and

[9] Following the author, the Japanese verbs are used as if they were English for descriptive purposes.

Woo (2010) examined the meaning and use of several cutting and breaking verbs in Early Middle Japanese.

Studies on non-standard varieties are limited to case-studies, but often draw upon comparisons with the standard language. For Ryukyuan, Nakama (1978) describes three cutting verbs in Nishihara (Miyako) dialect, which showed distinctions that are not made in Standard Japanese—such as the manner of cutting. A two-part follow-up study discussed additional verbs in more detail (Nakama 1984, 1985). For the mainland, Hashio (1992) describes the meanings of five breaking verbs in Kobe (Western Japanese) dialect.

Standard Japanese was included in a cross-linguistic comparison of 28 typologically distinct languages using a set of video clips showing various cutting and breaking events. Japanese participants did not appear to use the same core set of dimensions that speakers of other languages used to categorise cutting and breaking events (Majid, Boster & Bowerman 2008). Based on this finding, Fujii, Radetzky and Sweetser (2013) provide additional explanation for why Japanese might be deviant, by discussing the different ways Japanese categorises these events. Further contrastive work has compared Japanese to e.g., English (Ogawa 1984), French (Itou 2006), and Korean (e.g., Kwon 2013; Li 2018). Seol (2016) used the original Cut and Break Clips to examine differences in the use of Japanese cutting and breaking verbs between native speakers and Korean learners of Japanese.

Chapter 7 revisits the semantic structure of cutting and breaking events in Japanese to investigate whether the organisation of this domain in Japanese does indeed fall outside the suggested cross-linguistic constraints. In addition, data from Ryukyuan is collected for the first time using the same standardised stimuli to assess how much overall variation there is within this semantic domain for Japonic, and to compare the Japonic data to previously collected data from the Germanic languages to examine the interaction between lineage-specific developments and cross-linguistic constraints.

2.8 Summary

The Japonic language family and its varieties have received considerable scholarship, but many questions remain—particularly regarding the endangered, yet underdescribed, Ryukyuan varieties. This thesis contributes to our understanding of semantic variation in the Japonic language family as a whole by studying data from Ryukyuan, as well as to our understanding of semantic variation across related language in general through direct comparison between Japonic and another language family (Germanic).

CHAPTER

3

3 The geographical configuration of a language area influences linguistic diversity [1]

Abstract

Like the transfer of genetic variation through gene flow, language changes constantly as a result of its use in human interaction. Contact between speakers is most likely to happen when they are close in space, time, and social setting. Here, we investigated the role of geographical configuration in this process by studying linguistic diversity in Japan, which comprises a large, connected mainland (less isolation, more potential contact) and smaller island clusters of the Ryukyuan archipelago (more isolation, less potential contact). We quantified linguistic diversity using dialectometric methods and performed regression analyses to assess the extent to which distance in space and time predict contemporary linguistic diversity. We found that language diversity in general increases as geographic distance increases and as time passes—as with biodiversity. Moreover, we found that (I) for mainland languages, linguistic diversity is most strongly related to geographic distance—a so-called isolation-by-distance pattern, and that (II) for island languages, linguistic diversity reflects the time since varieties separated and diverged—an isolation-by-colonisation pattern. Together, these results confirm previous findings that (linguistic) diversity is shaped by distance, but also goes beyond this by demonstrating the critical role of geographic configuration.

Keywords: linguistic diversity, geographic isolation, isolation-by-distance, isolation-by-colonisation, Japanese, Ryukyuan

[1] This chapter is based on Huisman, J. L. A., Majid, A. & van Hout, R. (2019). The geographical configuration of a language area influences linguistic diversity. PLOS ONE, 14(6), e0217363. I certify that I performed data collection, analysis and writing of the manuscript, with feedback from the co-authors.

3.1 Introduction

The diversity found across the world's languages today is not the same as it was a hundred or 10,000 years ago, nor will it stay the same in the future. As the processes of diversification need time to run their course, we often find more diversity in areas where a language has been used for longer—compare, for example, English in the United Kingdom with English in Australia (Blair & Collins 2001). On top of this temporal dimension, we also see that linguistic diversity increases over geographical distance. Several patterns of linguistic diversity have been shown to exist, ranging from gradually accumulating differences (Heeringa & Nerbonne 2001), to more burst-like diversification (Atkinson et al. 2008). The specific role that the geographical configuration of a language area plays in this process is less explored. The current study aims to investigate to what extent a cultural process such as language diversification follows the same patterns as a biological diversification. To do this, we investigate patterns of linguistic diversity in the context of an island setting by applying insights from population genetics.

There are two notions from population genetics that we investigate in detail here. First we consider dispersal, which is defined as any movement that has the potential to affect gene flow, i.e. the transfer of genes between populations (Ronce 2007). If dispersal can occur without restriction, genes are transferred across all populations and we find evenly-spread genetic variation and high levels of homogeneity (Hutchison & Templeton 1999). However, the physical characteristics of the individual put a limit on its dispersal range and this reduces gene flow between distant populations. With this reduced gene flow, genetic differentiation between populations will increase and the end result is increased diversification over geographic distance; a pattern that has been dubbed isolation-by-distance (Wright 1943).

The same idea can be applied to language. Speakers adapt their speech patterns to accommodate to their most common conversational partners, their speech community (Bloomfield 1933). The use of language in human interaction can be thought of as linguistic gene flow. This interaction will, for logistical reasons, be more intense between people that are close to each other: linguistic features first spread across communities that share dense interaction, and then expand into the rest of a language area—a process called diffusion (Gerritsen & van Hout 2006, for an overview). As a result, the language of neighbouring communities will differ only slightly (Chambers & Trudgill 1998). However, contact between geographically distant communities will be less frequent and accommodation will occur to a lesser degree. This limited linguistic gene flow over increasing geographic distance means that speech communities will resemble each other less and less the farther apart they are (Heeringa & Nerbonne 2001)—the isolation-by-distance pattern described above. Linguists often call this a dialect continuum and it has been shown to hold over several language areas. Nerbonne (2010) investigated language varieties in six areas (Bantu, Bulgaria, Germany, US East Coast, the Netherlands, and Norway), and found linguistic diversity increased over geographic distance.

Although compelling in some ways, the areas investigated to date have focused on land-connected language areas (cf., Gavin & Sibanda 2012). It is unclear whether the same generalizations hold for island languages as other factors play a role there. Linguistic dispersal, i.e. contact, requires travel and travel across connected land can, in principle, be done on foot. This lowers the threshold for contact between neighbouring communities, making it easier to maintain connections over longer periods of time. In contrast, travel across islands requires seafaring technology and this limits the amount of contact between island communities.

As such, a second issue to consider is colonisation history (Orsini et al. 2013). From population genetics, we know that when a new population is started by a small subgroup of a larger one, it will only represent part of the overall diversity found in the original population—known as the founder effect (Mayr 1942). In isolation, the new population undergoes local genetic adaptation and in time, this leads to a significant divergence from the original population. This divergence reduces the chances of successful colonisation by later waves of migrants from the original population (De Meester et al. 2002). As such, the diversity we find reflects the time that has passed since the two populations separated and diverged, a pattern that is called isolation-by-colonisation (Orsini et al. 2013).

Similarly, for language, when subgroups of speakers expand into new territory, isolation caused by large distances between island communities has been shown to increase language diversification after settlement (Pawley & Green 1973). We find that languages diverge in pulses that coincide with each wave of colonisation (Nettle 1998). While islands have been argued to require wider resource networks due to a greater ecological risk (Nettle 1998)—which would increase contact and in turn decrease linguistic diversity—Lee and Hasegawa (2014) show that the presence of a body of water acts as a barrier that promotes diversification. Sustained contact between communities will depend on the distance between islands (Marck 1986).

The two factors involved in diversification discussed above (dispersal and colonisation history) result in predictable patterns of genetic diversity (isolation-by-distance and isolation-by-colonisation; Orsini et al. 2013). Moreover, these factors have been shown to play different roles in specific geographic configurations (Spurgin et al. 2014). Fragmented landscapes, such as archipelagos, have been considered a good setting to investigate how genetic diversity is influenced by geography (Clegg & Phillimore 2010). Therefore, if the same processes apply to language, as has been argued above, we should be able to make predictions about patterns of linguistic diversity too. To test this, we investigated linguistic diversity in Japan.

The Japanese archipelago is an arc of islands stretching over 2,500 kilometres and comprising over 400 contemporary inhabited islands. Approximately 70% of the land area consists of forested mountains. Ecological risk seems to be low across islands (cf., Nettle 1998). Their climate provides self-sufficiency through abundant food sources (Koyama & Thomas 1984), which is further evidenced by the relatively late arrival of agriculture to the archipelago, despite it being inhabited for a long time

(Asato et al. 2004). The switch to agriculture happened even later in the southern islands, showing that the survival of its first settlers was supported by the resources available and did not require broader social networks beyond the scope of the island on which they lived.

Spoken across the archipelago is the Japonic language family. Japonic has not been convincingly linked to other languages or language families, but a distant connection to Koreanic seems plausible (Whitman 2011; Unger 2014). The language family consists of two main branches: (I) Japanese, which can be subdivided into Eastern, Western, Kyūshū and Hachijō Japanese; and (II) Ryukyuan, which can be subdivided into Amami and Okinawa (Northern Ryukyuan) on one hand, and Miyako, Yaeyama and Yonaguni (Southern Ryukyuan) on the other (Shibatani 1990; Pellard 2015). Both traditional dialectology and computational approaches have shown a clear split between Japanese and Ryukyuan based on the shared presence of Standard Japanese forms (Inoue & Kasai 1989), the shared presence of linguistic innovations (Pellard 2009), and phylogenetic analyses based on shared cognacy of basic vocabulary (Lee & Hasegawa 2011). The split is corroborated by politico-cultural history (Asato et al. 2004), and population structure studies (Sato et al. 2014; Takeuchi et al. 2017). Importantly, Japanese is spoken on the large islands that are close to each other, whereas Ryukyuan is spoken across a number of small island clusters that have relatively large distances between them. We investigated whether these specific geographic configurations influence patterns of linguistic diversity. In addition to Japonic, varieties of Ainu have traditionally been spoken by a distinct indigenous non-Japonic group in the northern parts of Japan. Ainu is critically endangered with few speakers remaining. However, we do not consider Ainu in the current investigation.

While dispersal and colonisation history are both expected to influence language diversification in Japanese and Ryukyuan, we predict that they do so to different degrees. Owing to the relative ease of travel across connected land, dispersal—contact between speakers—is less restricted by natural barriers across the Japanese language area and therefore, gene flow—accommodation between speakers—can occur more freely. As such, we predict that linguistic diversity in Japanese will mostly be a reflection of the distance that speakers can travel: an isolation-by-distance pattern. In contrast, the technological requirements of sea travel limit contact and accommodation across the Ryukyuan language area and local diversification will occur to a larger degree. Therefore, we predict that linguistic diversity in Ryukyuan will mostly reflect the time since language varieties diverged: an isolation-by-colonisation pattern.

3.2 Methods

3.2.1 Linguistic data

3.2.1.1 Vocabulary database

We created a new comparative dataset based on the 100-item Swadesh List (Swadesh 1955; see also Table 3.1)—a list of what are considered to be basic concepts, such as body parts and everyday actions. The Swadesh List is well-established in both large-scale and small-scale comparative studies (Gray & Atkinson 2003; Bakker & van Hout 2012). In light of recent findings that the lexicon may be more stable over time than grammatical features (Greenhill et al. 2017), we take this list of basic concepts to be a good starting point for comparison. We built on the database collated by Lee and Hasegawa (2011), like them using the six-volume *Dictionary of Contemporary Japanese Dialects* (Hirayama 1992), but additionally coding the data to preserve all distinctions present in the original material (unlike Lee and Hasegawa, see their Data Supplement 2). Furthermore, we include an additional 11 (mostly island) varieties over the original Lee and Hasegawa database. In addition, we collated data from Volumes 1–3 and 7 of *The Complete Works of Tōsō Miyara* (Miyara 1980), to add another 22 Ryukyuan varieties. Miyara was a Ryukyu-born phonetician, and speaker of one of the local varieties, whose works have been used as a reliable source of contemporary variation, e.g., for the reconstruction of Proto-Ryukyuan (Bentley 2008). In total, 58 Japanese and 32 Ryukyuan varieties are represented in the data set—see Figure 3.1a and 3.1b for maps with location names).

Japanese locations

1	Aomori	30	Shiga
2	Hirosaki	31	Mie
3	Hachinohe	32	Nara
4	Iwate	33	Kyoto
5	Akita	34	Osaka
6	Miyagi	35	Wakayama
7	Yamagata	36	Hyogo
8	Fukushima	37	Okayama
9	Aizutakada	38	Hiroshima
10	Niigata	39	Yamaguchi
11	Sado	40	Tottori
12	Ibaraki	41	Shimane
13	Tochigi	42	Tokushima
14	Chiba	43	Kagawa
15	Tokyo	44	Kochi
16	Saitama	45	Ehime
17	Gunma	46	Oita
18	Kanagawa	47	Fukuoka
19	Yamanashi	48	Saga
20	Shizuoka	49	Nagasaki
21	Akiyama	50	Nakadori
22	Nagano	51	Fukue
23	Hachijo	52	Miyazaki
24	Toyama	53	Kumamoto
25	Nanao	54	Kagoshima
26	Ishikawa	55	Kamikoshiki
27	Fukui	56	Shimokoshiki
28	Gifu	57	Tanegashima
29	Aichi	58	Yakushima

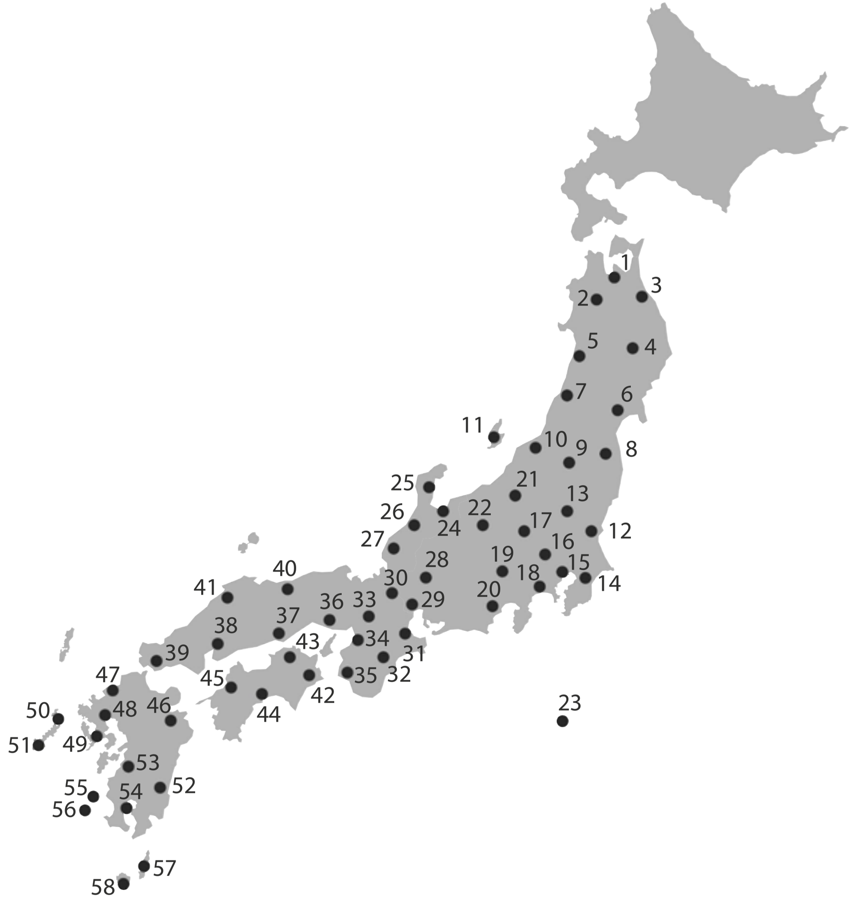
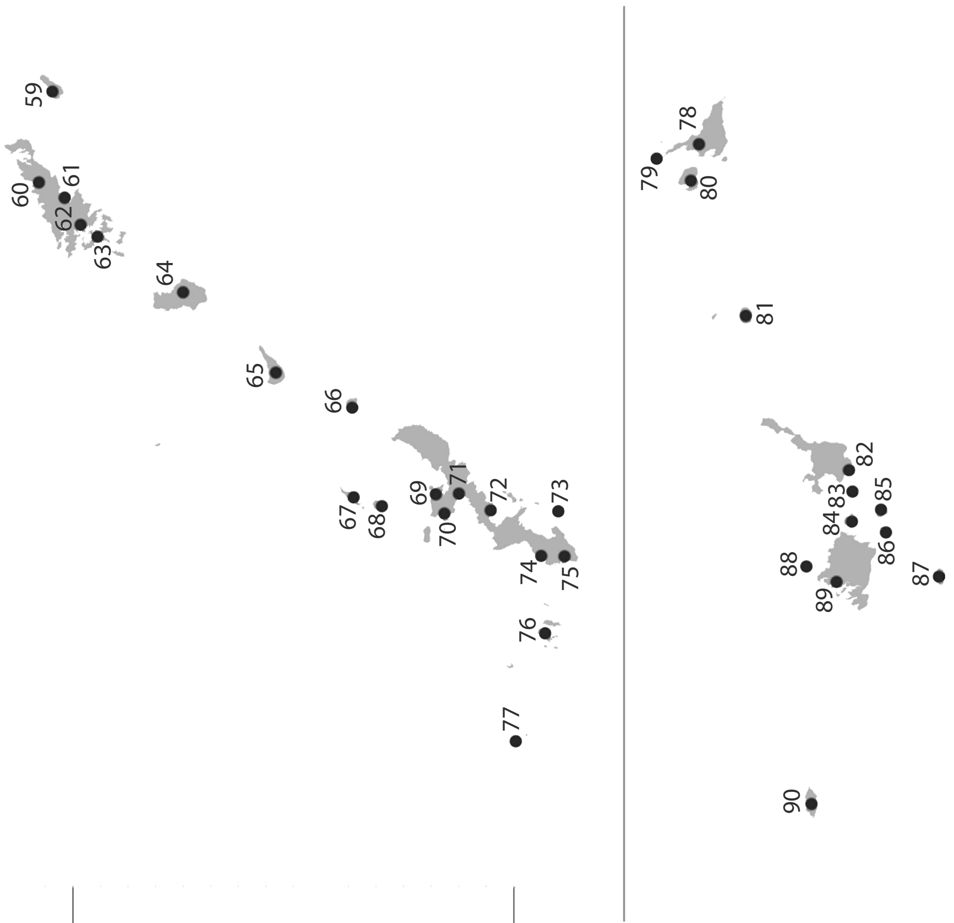


Figure 3.1a. Map of Japanese varieties included in the analyses.



Ryukyuan locations

59	Kikai	75	Itoman
60	Naze	76	Zamami
61	Sumiyo	77	Kume
62	Koniya	78	Hirara
63	Kakeroma	79	Ikema
64	Tokunoshima	80	Irabu
65	Okinoerabu	81	Tarama
66	Yoron	82	Ishigaki
67	Iheya	83	Taketomi
68	Izena	84	Kohama
69	Nakijin	85	Kuroshima
70	Motobu	86	Aragusuku
71	Nago	87	Hateruma
72	Kin	88	Hatoma
73	Kudaka	89	Iriomote
74	Shuri	90	Yonaguni

Figure 3.1b. Map of Ryukyuan varieties included in the analyses.

Due to incomplete source material, the eventual dataset contained data for 98 out of the 100 Swadesh List items (Table 3.1).

Table 3.1

Items of the 100-item Swadesh List.

all	full	new	to die
ash	to give	night	to drink
bark	good	nose	to eat
belly	green	not	to kill
big	hair	one	to know
bird	hand	path, road	to lie down
black	head	person	to say
blood	to hear	rain	to see
bone	heart	red	to sit
breasts	horn	root	to sleep
claw	hot	round	to stand
cloud	I	sand	to swim
cold	knee	seed	to walk
dog	leaf	skin	tongue
dry	liver	small	tooth
ear	long	smoke	tree
earth, soil	louse	star	two
egg	man	stone	water
eye	many	sun	we
fat, grease	meat, flesh	tail	what?*
feather	moon	that	white
fire	mountain	this	who?
fish	mouth	to bite	woman
to fly	name	to burn*	yellow
foot	neck	to come	you

Note. Items marked with an asterisk were omitted from this study due to a lack of data.

3.2.1.2 *Measuring linguistic diversity*

Various methods of quantifying linguistic distance have been used in previous research. One approach has been to compare varieties to one "standard", and calculate distances accordingly (Inoue & Kasai 1989). However, comparing to one standard variety does not reveal how different non-standard varieties are from each other, which is important as these non-standard varieties can differ in both the linguistic features they retain, as well as the innovations they pick up. Another approach is to focus on a number of language-specific innovations, e.g., examining vowel shifts or voicing patterns characteristic of one language area (Pellard 2009). However, this requires both an in-depth knowledge of the language varieties that are being studied, and it limits the number of features that can be compared in a single analysis. Finally, phylogenetic approaches applied to language data require cognate-coding (Lee & Hasegawa 2011, 2014), which in turn require broad linguistic judgements, and critically reduce the amount of data as non-cognate forms are excluded.

Instead, we adopted a measure of linguistic distance commonly used in dialectometry, based on edit distance—specifically Levenshtein distance (Levenshtein 1966). The Levenshtein distance between two strings (e.g., dialect word forms) is calculated as the minimum number of single-character edits needed to turn one into the other. Edits can entail any combination of character additions, deletions, or substitutions. This method was first used in the study of Irish dialects (Kessler 1995) and is a novel approach to analysing linguistic diversity in the Japonic language family. We used Gabmap (Nerbonne et al. 2011), a free online tool for dialect analysis, to perform the calculations. Gabmap normalises edit distance based on the length of the word forms to take into account the differential impact edits have on short versus long items. Linguistic distance between two locations is then calculated by aggregating Levenshtein distance over a large number of items, an approach that finds its roots in the works of Séguy (1971; 1973) and Goebl (1984). Gabmap also allows for multiple entries per item.

We opted to use the software's algorithm that assigns linguistically informed costs to the edits involved. In this approach, to preserve syllable, structure substituting a vowel with a consonant, or vice versa, receives double weight. Furthermore, diacritic marks—used to indicate smaller degrees of modification like devoicing or aspiration—are counted as half an edit as they are seen as a smaller deviation from the character they modify than a completely different character would entail. Vowel-consonant substitutes are rare in Japonic varieties given their rigid CV mora structure. While syllabic (moraic) fricatives do occur in Miyako Ryukyuan, e.g., in the Ōgami dialect (Pellard 2009), the source material used for the varieties in this study's dataset did not include such cases. However, diacritic changes are not uncommon. For example, the underlying phonological contrast of front versus back high vowels is maintained across both Tokyo Japanese and the Tohoku dialects, but the phonetic realisation of these vowel in Tohoku is more central, so this is represented as a change in diacritics rather than as a change in characters, coded as /i/ vs. /ï/ and /ɯ/ versus /ɯ̈/. Another example is devoicing of the vowel in the first mora, which is common in some Yaeyama varieties, as found in e.g., *pḁna* 'nose' in Hateruma. This is a small, non-phonemic, modification when considering *pana* in Yoron (Amami). However, in comparison to *hana* 'nose' in Tokyo Japanese there is a change of the initial consonant that is phonemic, which is represented by a character change.

Calculating aggregate distances over all items for all locations within a dataset creates a location-by-location linguistic distance matrix. The method has a number of advantages over previous approaches. It can: (I) make direct comparisons between all varieties of interest, (II) compare all segments in all words, increasing the number of data-points and expanding the comparison beyond specific predetermined items of interest (Nerbonne 2013), and (III) analyse linguistic data based on surface forms without the need for additional linguistic coding and judgements that potentially decrease the amount of data considered. Finally (IV), it has the additional advantage of examining diversity within a language, rather than merely counting the number of separate languages (cf., Gavin & Sibanda 2011).

3.2.2 Non-linguistic data

3.2.2.1 Colonisation history

The time-depth and phylogeny of a language family reflects its colonisation history (Gray, Drummond & Greenhill 2009) and as such, we used that as a basis to code a *time since divergence* variable. Lee and Hasegawa (2011) estimated the time-depth of the Japonic language family in years before present (YBP) using Bayesian phylogenetic analyses. For this, they collated basic vocabulary data for a number of contemporary varieties, and for two older forms of the language (Old Japanese and Middle Japanese). They calibrated the age ranges of Old Japanese (1216–1300 YBP) and Middle Japanese (437–674 YBP), as well as the divergence of the Kyoto and Tokyo varieties (the historical and current capitals, respectively; dated 142–549 YBP), and then constructed a phylogeny of the Japonic language family based on a model incorporating varying rates of linguistic evolution. They found a median age for the split between Japanese and Ryukyuan of 2182 years before present. Using Lee and Hasegawa's maximum clade credibility tree, we determined the approximate age of the most recent common ancestor (MRCA) for each pair of language varieties, but we generalised time since divergence over all varieties within major subgroups that diverged before 250 YBP. This date was chosen because at this time point, all generally accepted subdivisions in both Japanese (Eastern, Western, Kyushu) and Ryukyuan (Amami, Okinawa, Miyako, Yaeyama, Yonaguni) are represented in the tree.

Within these subgroups, pairwise time since divergence was defined as 50 years younger than the age of the subgroup to which language varieties belonged. This allowed us to include the additional language varieties missing in Lee and Hasegawa's tree with minimal additional assumptions—particularly in the Ryukyuan language area. For example, the MRCA for Amami and Okinawa in Lee and Hasegawa's tree was dated at approximately 400 YBP, but since their data set only included one variety of each, we dated the MRCA for the Okinawa varieties in our dataset at 400 – 50 = 350 YBP. We did not adopt a more fine-grained coding as more recent, relatively small divergences were not expected to have a substantial impact on the outcome since the older divergence between major groups occurred much longer ago—see also the last paragraph in *Analysis* section (3.2.4) below. Importantly, this coding scheme takes the time-depth of larger subgroupings within the two language areas into account, which can be important as language diversity in general increases over time (Nettle 1999). Time since divergence was coded in a location-by-location matrix.

3.2.2.2 Geographic distance

All locations included in the linguistic data were marked in a KML map file using Google Earth. The geospatial data from their coordinates was used to calculate straight-line geographic distances, which were entered into a location-by-location distance matrix. As language distance decay has been shown to be sublinear (Nerbonne 2013), we created a second distance matrix by performing a natural logarithmic transformation on straight-line geographic distance.

3.2.2.3 Separation by water

As the presence of an oceanic barrier has been shown to influence language diversification, we coded a *separation by water* variable for each pair of locations, with value “1” if a body of water separates the two, and with value “0” if not, following Lee and Hasegawa (2014). As our dataset includes a range of both water and land distances, we included this variable to be able to look at the effect of separation by water individually, and along with geographic distance. The binary values were coded as a location-by-location matrix.

3.2.3 Analysis

We began by verifying the commonly accepted subgroupings of Japanese and Ryukyuan within our data. For this, we analysed the linguistic distance matrix of the Swadesh List data using a hierarchical clustering algorithm based on Ward's method (Ward 1963), in R (`hclust` function; R Core Team 2019).

Next, we tested to what extent the factors discussed above (geographic distance, time since divergence, and separation by water) are related to linguistic distance. Because we expected the effect of geographic distance to differ between island versus mainland languages (Marck 1986), we also included an interaction between geographic distance and separation by water in our analyses. Using Mantel tests (`mantel` function, *ecodist* package; Goslee & Urban 2007), we correlated the four factors with each other to test their relatedness, and then correlated linguistic distance with those same four factors, using partial Mantel tests to control for their mutual influence. All Mantel tests were carried out using 10,000 permutations and 1,000 bootstrap iterations on 95% confidence intervals. To further model linguistic diversification, we performed multiple regression over distances matrices (`MRM` function, *ecodist* package; Goslee & Urban 2007), using the four factors as independent variables and linguistic distance as the dependent variable.

However, MRM analysis has limitations in that it cannot include random effects. We therefore performed an additional linear mixed model analysis on the full distance matrices (`lmer` function, *lme4* package; Bates et al. 2015) to predict linguistic diversification using the same four variables as before, while adding random intercepts for language varieties to account for their inherent uniqueness. For all mixed models, we will report standardised coefficients (`beta` function, *reghelper* package; Hughes 2018), and include estimates of *p*-values (*lmerTest* package; Kuznetsova, Brockhoff & Christensen 2017), as well as pseudo-R^2 values (*piecewiseSEM* package; Lefcheck 2016).

A preliminary analysis of the Japonic language family as a whole showed that time since divergence was the most important factor across all Mantel and regression analyses. However, the correlation between time since divergence and a binary coded Japanese-vs-Ryukyuan—in which a comparison between one Japanese and one Ryukyuan variety was coded as "1", and a comparison between two Japanese varieties or two Ryukyuan varieties was coded "0"—was r = .980, indicating that the time since divergence variable for all of Japonic primarily represents the split between Japanese and Ryukyuan.

3.3 Results

3.3.1 Japanese and Ryukyuan form distinct subgroups

The results of the cluster analysis (Figure 3.2) are in line with both traditional classification in Japanese dialectology (Shibatani 1990), and with Lee and Hasegawa's (2011) phylogenetic tree. Critically, the cluster analyses confirmed that Japanese and Ryukyuan are distinct, showing a clear split between all Japanese and all Ryukyuan varieties, replicating previous findings (Pellard 2015). Discussing all the specific subgroups is unfortunately beyond the scope of this paper. However, for Japanese (Figure 3.2, left panel) it is noteworthy that while the cluster analysis confirmed the accepted main division between Eastern and Western Japanese varieties, both the peripheral varieties in the north (Tohoku Japanese) and those in the south (Kyushu Japanese) formed distinct subgroups. For Ryukyuan (Figure 3.2, right panel) the cluster analysis confirmed a main division between Northern Ryukyuan (Amami and Okinawa), and Southern Ryukyuan (Miyako and Yaeyama).

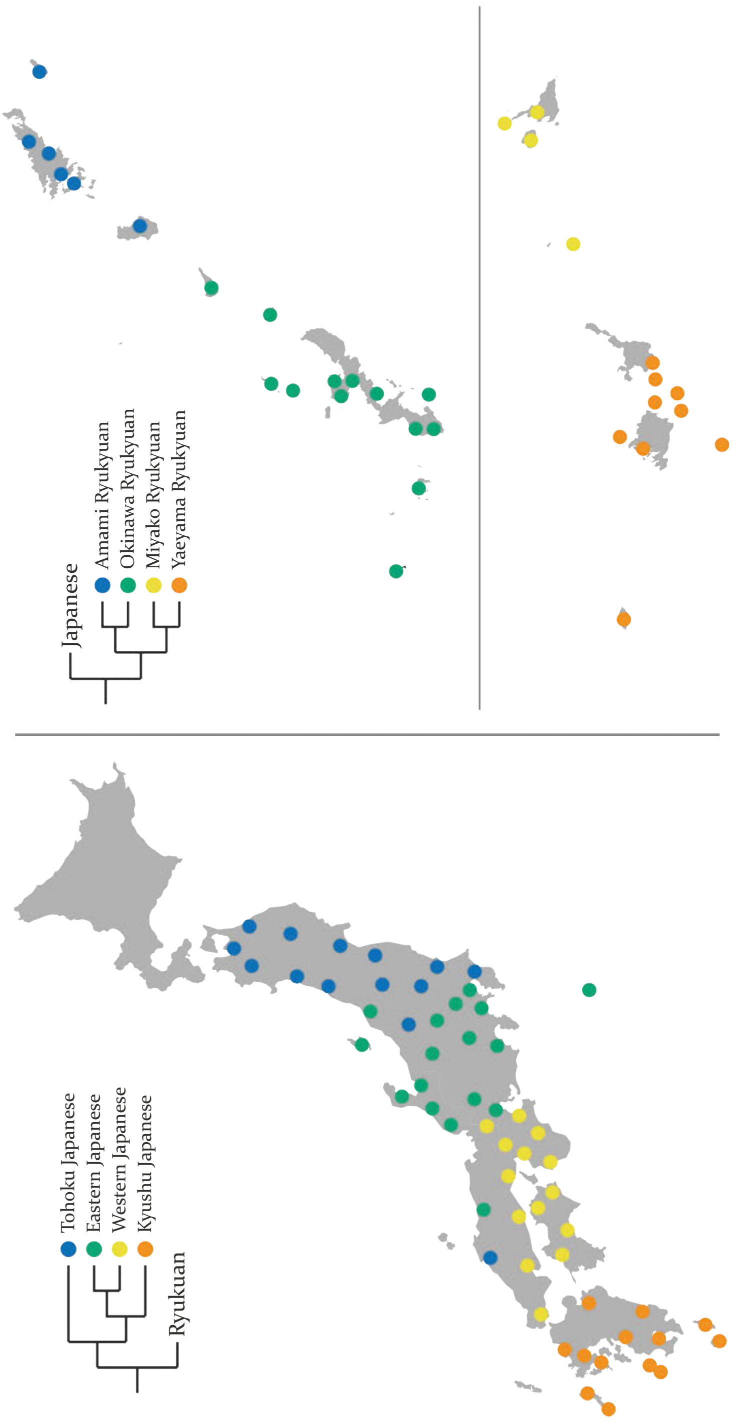


Figure 3.2. Cluster analysis results for Japanese and Ryukyuan.

Figure 3.3 shows the distribution of linguistic distances within and between Japanese and Ryukyuan. The Japanese distances (blue) show a bimodal distribution, where the second peak corresponds to the large differences between the two peripheral subgroups, Kyushu Japanese and Tohoku Japanese. For Ryukyuan (orange), we see a quadrimodal distribution that corresponds to the four subgroups (Amami, Okinawa, Miyako, and Yaeyama). The four separate modes show that linguistic distances between the subgroups is large, i.e. these subgroups are pronounced in their distinctiveness. Average linguistic distance within Ryukyuan (M_{Ryu} = 0.256, SD = 0.068) was significantly larger than the distance within Japanese (M_{Jap} = 0.205, SD = 0.061), $t(751.1) = 14.88$, $p < .001$, Cohen's d = 0.78. Linguistic distances between the Japanese language area and the Ryukyuan language area (grey) were larger overall and showed a normal-like distribution, indicating that there are no Japanese-Ryukyuan subgroups between which linguistic distances were small.

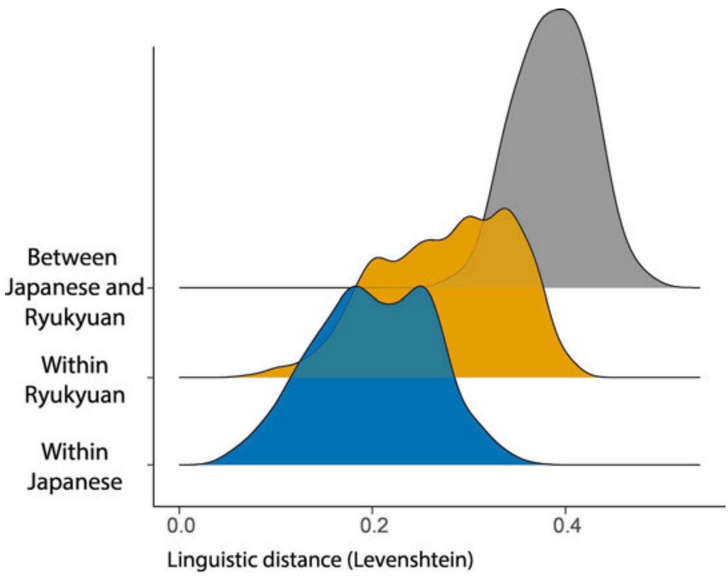


Figure 3.3. Linguistic distances within Japanese (blue), within Ryukyuan (orange) and between the two language areas (grey).

Figure 3.4 shows the distribution of linguistic distances along geographic distance for Japanese (blue) and Ryukyuan (orange), together with a Loess smoothing curve. As described above, linguistic distances in Ryukyuan are larger than in Japanese—despite occurring over smaller geographic distance. In addition, Ryukyuan shows a sharp increase that tapers off quickly, while Japanese showed a moderate increase that continues linearly. This points to Japanese as being more continuum-like where linguistic differences slowly accumulate over geographic distance, which is evidence for an isolation-by-distance pattern. The initial increase in linguistic distance for Ryukyuan shows that this language area also shows continuum-like characteristics on the small scale, but the fact that this levels off fairly quickly shows that beyond a certain point—i.e., beyond the island cluster—linguistic differences are large in genera without a clear connection to geographic distance, evidence for an isolation-by-colonisation pattern.

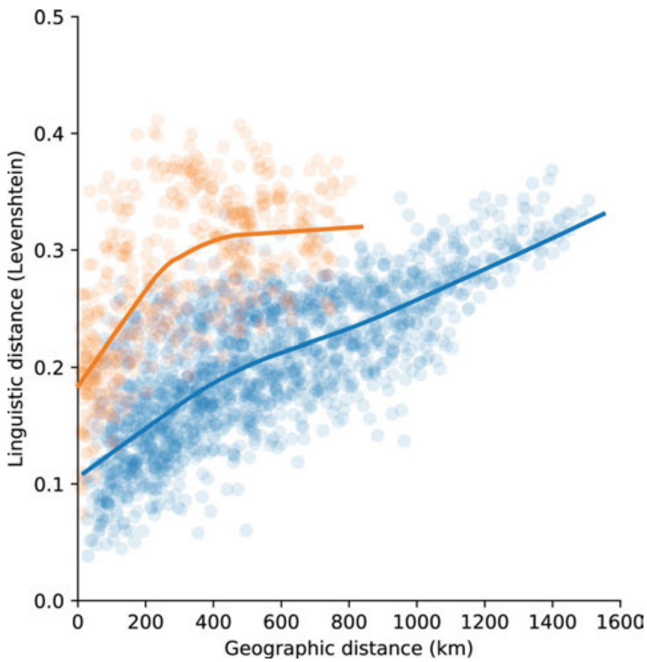


Figure 3.4. Linguistic distance over geographic distance in Japanese (blue) and Ryukyuan (orange) with Loess smoothing.

3.3.2 Geography and linguistic diversity across the Japanese mainland

Mantel tests confirmed that geographic distance, time since divergence, and separation by water are related to each other across the Japanese mainland (Table 3.2). Partial Mantel tests then showed that geographic distance was strongly correlated with linguistic diversity (Table 3.3): linguistic distance between language varieties increased with increased geographic distance. Contrary to what has been previously reported (Nerbonne 2013), there was no significant difference between linear geographic distance and logarithmic geographic distance in the strength of their association with linguistic distance, z = 0.53, p = .596. In fact, the correlation with linear distance was numerically higher (r = .545 versus r = .532). There was no significant correlation between linguistic diversity and time since divergence, nor between linguistic diversity and separation by water for the Japanese varieties. The interaction between geographic distance and separation by water was significant, however, and its negative value indicates that the effect of geographic distance was smaller for varieties separated by a body of water. These findings were supported by the MRM analysis, which confirmed that geographic distance was a significant predictor of linguistic distance, as was the interaction between geographic distance and separation by water, in a model that accounted for 58% of the variation (Table 3.4).

Table 3.2

Simple Mantel correlations between time since divergence, geographic distance, and separation by water for Japanese.

	Time since divergence	Separation by water
Geographic distance	.501	.452
Separation by water	.060	

Table 3.3

Partial Mantel correlations between linguistic distance, time since divergence, geographic distance and separation by water for Japanese.

	Linguistic distance			
	r	95% CI		p
Time since divergence	-.097	-.160	-.040	.129
Geographic distance	.549	.504	.598	<.001
Separation by water	-.001	-.049	.054	.999
Distance * Water	-.097	-.158	-.058	.041

Table 3.4

Results for predicting linguistic distance in Japanese using multiple regression over distances matrices.

	Estimate	*p*
Intercept	0.146	
Time since divergence	$-8.52 \cdot 10^{-5}$	.119
Geographic distance	$1.76 \cdot 10^{-4}$	<.001
Separation by water	$-1.54 \cdot 10^{-5}$	.999
Distance * Water	$-2.93 \cdot 10^{-5}$	.037

R^2 = .579.

Coefficients produced by the mixed model analysis (Table 3.5) were largely in line with results from the Mantel tests, except that all predictors turned out significant in the analysis after including random effects for language varieties. VIF values for the main effects were all < 2.0. The model confirmed the strongest predictor of linguistic distance across the Japanese mainland to be geographic distance—once again, linear geographic distance (AIC = -12648.7) provided a better model than logarithmic distance (AIC = -12443.6). This geographic distance effect was weaker for varieties separated by a body of water. In line with Nettle's proposal that the increased ecological risk of islands calls for wider social networks (Nettle 1999)—more contact and accommodation, and thus less diversity—varieties separated by water exhibited smaller linguistic distance. The effect of time since divergence, while significant, is much weaker than that of geographic distance. In fact, the negative coefficient indicates that varieties that diverged longer ago are *more* similar to each other, which is a sign that sustained contact (through geographic proximity) can negate the effects of previous isolation. Taken together, these findings show a strong effect of geographic distance on linguistic distance, which confirms our hypothesis that the patterns of linguistic diversity on the mainland should largely reflect contact between speech communities, as predicted by isolation-by-distance.

Table 3.5

Results for predicting linguistic distance in Japanese using linear mixed effect modeling.

	β	*SE*	*t*	*p*
(Intercept)	.046	.045	1.02	
Time since divergence	-.040	.013	3.14	<.001
Geographic distance	.809	.016	51.54	<.001
Separation by water	-.111	.014	7.88	<.001
Distance * Water	-.101	.013	7.87	<.001

Conditional R^2 = .667, Marginal R^2 = .551.

3.3.3 Geography and linguistic diversity across the Ryukyu Islands

Mantel tests confirmed that the predicting factors of linguistic distance are correlated for Ryukyuan as well (Table 3.6). The partial Mantel tests showed that only the correlation between linguistic distance and time since divergence was significant (Table 3.7). The longer ago two varieties diverged from each other, the more linguistic distance there was between them. Geographic distance failed to reach significance, and logarithmic geographic distance showed no difference in its correlation with linguistic distance when compared with linear distance, z = 0.05, p = .960. Moreover, there was little numerical difference between the two; r = .067 versus r = .064. In contrast with the findings by Lee and Hasegawa (2014), separation by a body water did not lead to increased linguistic distance, which can be attributed to the fact that Ryukyuan is spoken on island clusters and the presence of a body of water is not a defining characteristic. Finally, the interaction effect indicated that the influence of geographic distance decreased when language varieties are separated by water, but it was not of significant strength. These results were supported by the MRM analysis (Table 3.8), in which time since divergence was the only significant predictor of linguistic distance. The model accounted for 60% of the variation in linguistic diversity across the Ryukyu Islands.

Table 3.6

Simple Mantel correlations between time since divergence, geographic distance, and separation by water for Ryukyuan.

	Time since divergence	Separation by water
Geographic distance	.824	.365
Separation by water	.210	

Table 3.7

Partial Mantel correlations between linguistic distance, time since divergence, geographic distance, and separation by water for Ryukyuan.

	Linguistic distance			
	r	95% CI		p
Time since divergence	.438	.359	.515	<.001
Geographic distance	.067	.033	.092	.094
Separation by water	.051	.023	.089	.269
Distance * Water	-.025	-.056	.001	.559

Table 3.8

Results for predicting linguistic distance in Ryukyuan using multiple regression over distances matrices.

	Estimate	p
Intercept	0.046	
Time since divergence	$1.12 \cdot 10^{-4}$	<.001
Geographic distance	$2.15 \cdot 10^{-2}$	.092
Separation by water	$5.78 \cdot 10^{-2}$	.270
Distance * Water	$-8.08 \cdot 10^{-3}$	.563

$R^2 = .603$.

The linear mixed model produced results confirming the findings from the Mantel tests (Table 3.9). Time since divergence and geographic distance were significant predictors of linguistic distance, indicating that the longer ago varieties diverged and the further apart they are, the larger the linguistic distance between them was. The strength of the effect of time since divergence was slightly stronger than the effect of geographic distance. The inclusion of logarithmic geographic distance provided a better model (AIC = -3594.4) than when linear distance was included (AIC = -3525.3). VIF values for the main effects were all < 3.0. As already shown by the Mantel tests above, and reflecting their status as island languages, there was no effect of separation by a body of water for Ryukyuan. Taken together, the effects that time since divergence and geographic distance have on linguistic diversity in Ryukyuan suggest that the patterns of diversity are a reflection of the time since the language varieties diverged—diversity *between* island clusters—but also a reflection of contact between speech communities *within* the island clusters. This is in line with what we predicted for the isolation-by-colonisation situation expected across isolated island clusters that require technology for travel.

Table 3.9

Results for predicting linguistic distance in Ryukyuan using linear mixed effect modeling.

	β	*SE*	*t*	*p*
(Intercept)	.010	.063	0.15	
Time since divergence	.472	.034	13.75	<.001
Geographic distance	.282	.035	8.02	<.001
Separation by water	.018	.053	0.33	.739
Distance * Water	-.027	.027	0.97	.333

Conditional R^2 = .694, Marginal R^2 = .575.

3.4 Discussion

It is clear that geography influences linguistic diversity, just as it influences biological diversity. However, the exact nature of this relationship in the context of languages is still poorly understood. Here we discovered that the geographical configuration of a language area affects the role of two known diversification processes: dispersal and colonisation history. After a cluster analysis based on linguistic distance measures confirmed the legitimacy of Ryukyuan—spoken across isolated island clusters—as a language group distinct from Japanese—spoken across a connected land—we examined the relationship between geographical distance and linguistic diversity, as well as time since divergence and linguistic diversity in these two language areas. As expected, linguistic diversity in both language areas increased with larger geographic distances, and with increased time since speech communities separated for the Ryukyuan area. Importantly, we found that the effect of geographic distance was stronger for Japanese, while the effect of time since divergence was stronger for Ryukyuan—a result of two different processes that have shaped linguistic diversity.

The separation of Japanese varieties has slowly been negated by sustained contact between communities that are geographically close: contact leads to accommodation, which causes varieties to resemble each other more and more as time passes. As a result, we found negative coefficients for time since divergence in our analyses. This effect appears to be strongly driven by the Tokyo variety. The time calibration by Lee and Hasegawa (2011) puts it among the oldest clade, but its status as mixed variety (of Eastern and Western Japanese characteristics) that has become the de facto standard has caused it to resemble varieties from both subgroups over time. Interestingly, the relationship between geographic and linguistic distance was linear throughout the entire area, which goes against the general sublinear trend found in other language areas (Bantu, Bulgaria, Germany, US East Coast, the Netherlands, and Norway; see Nerbonne 2010). This indicates that Japanese is a true dialect continuum without any gaps, whereas the sublinear trend found in previously studied language areas could point to the presence of clearly defined, i.e., more isolated, subgroups. It appears that the isolation of subgroups disrupts linguistic continuity in a language area. To demonstrate this, we took the characteristics of the prototypical isolation-by-distance and isolation-by-colonisation patterns (see Orsini et al. 2013), and conducted a simulation of linguistic distances between 20 locations across four subgroups. While in this case, isolation-by-adaptation—a scenario in which diversity arises through local adaptation to the natural landscapes (Orsini et al. 2013); local adaption to a socio-political environment for language—would actually be a better comparison, the contrast with isolation-by-distance remains the same as diversity is not directly related to geographic distance. The simulations indeed showed that increasing the isolation of just one subgroup creates the sublinear trend reported by Nerbonne (see the Appendix). In this light, it would be worthwhile to revisit these previously studied language areas to establish whether they differ in the heterogeneity of their linguistic landscapes, which could be an explanation for why linguistic distance appears to reach ceiling at different distances, and moreover, why they vary at all.

In Ryukyuan, the separation of varieties happened a long time ago and has remained largely intact within Ryukyuan due to the difficulties in maintaining contact across isolated islands. Nevertheless, we do find an effect of geographical distance for Ryukyuan—albeit a small effect. This shows that continuum-like characteristics do arise as a result of contact within islands clusters at least for short distances, in line with results from studies that focused on small-scale language areas (Nerbonne & Heeringa 2007; Stanford 2012). However, geographic isolation decreases contact beyond the island cluster, which prevents the formation of a continuum across the island chain as a whole. An interesting further step would be to study linguistic diversity in different types of island configurations. The size of islands, as well as the distances between them, affects the potential and frequency of contact between populations, which in turn affects the patterns of overall linguistic diversity, as well as linguistic continuity within a dialect chain.

We also found that overall linguistic diversity was more abundant within Ryukyuan. This goes against what usually happens in population genetics, where a loss of genetic variation usually occurs in a new population as a result of the limited diversity present in its founders (Hundertmark & Van Daele 2010). There has been some discussion about whether overall diversity is also reduced in new linguistic communities: suggestions of a decrease in size of the phoneme inventory have been made (Atkinson 2011), but this idea is not uncontroversial (Cysouw, Dediu & Moran 2012; Wang et al. 2012). It is hard to put the specific linguistic distances reported here into broader perspective, as there has been little comparative work across different language/dialect areas. While Nerbonne (2010) summarises the general patterns from six language areas, each study utilised different units of measurement, providing little opportunity for direct comparison. However, it is not inconceivable that the Ryukyuan language area shows greater overall variation than the ones summarised by Nerbonne, so further work in other island languages is needed to confirm the pattern. Since most fine-grained dialectometric analyses have been applied to land-connected dialect areas, investigating island languages with this approach is an important addition to our knowledge of linguistic diversity. Gavin and Sibanda (2012) showed that the *number* of languages per island across the Pacific decreased with each subsequent expansion, but they did not examine dialectal variation *within* each language. The methodology applied here creates an opportunity to look at linguistic diversity in a more detailed manner that goes beyond merely counting languages (cf. Gavin et al. 2013; but see Honkola et al. 2018).

Finally, the current study used straight-line geographic distances as in population genetics studies, as well as several dialectology studies. An alternative approach would be to measure actual travel time—as has been done for Norway, which is topographically similar to Japan, i.e. mountainous. While travel time between islands will strongly depend on straight line distances over sea, travel across a larger mainland can be hindered by mountain ranges. Modern train distances as a proxy for travel time have been linked to the amount of Standard Japanese vocabulary in dialects across the mainland (Inoue 2004b), but the focus lies on two capital locations (Tokyo and Kyoto) as a starting point rather than a location-by-location comparison. Moreover, as land and sea travel have been shown to affect the diffusion of linguistic features differently (Gerritsen & van Hout 2006), further exploration of historical travel and trade practices—and how they have changed over time—can provide additional insights into the patterns of linguistic diversity we find today.

3.5 Conclusion

To conclude, we have shown that cultural processes—language diversification—are influenced by geography in ways similar to biological processes—species diversification. We examined the role of geographic configuration in diversification and showed that: (I) mainland languages display a typical isolation-by-distance pattern, with gradually increasing diversity over geographic distance, as a result of the higher potential for sustained contact, while (II) island languages display a typical isolation-by-colonisation pattern, where diversity is a reflection of time since divergence, as a result of limited contact due to the geographic isolation of islands. Language variation and change is, of course, influenced by other (historical and socio-political) factors as well, and a more global and multi-dimensional concept of distance—comprising spatial, temporal, and social factors—is needed to help us understand patterns of language diversification. Our results show that the geographical configuration of a language area is one important component of a more comprehensive distance concept to explain language variation and change.

CHAPTER

4

4 Stability and change in the colour lexicon of the Japonic languages[1]

Abstract

Previous work on colour lexicons focussed on universal patterns in their structure and evolution. We collected new colour naming data in Japanese and three Ryukyuan languages (Amami, Miyako and Yayeama) to investigate semantic variation and change in the colour lexicon of related languages in a modern context. We found several new colours terms (e.g., *midori* and *guriin* for 'green') in the lexicon of Ryukyuan speakers, apparently resulting from contact with Standard Japanese and English. A comparison of our data with earlier collected data suggests that modern Ryukyuan colour systems are closer to modern Japanese than they are to their historic pasts. However, we also found that modern-day Ryukyuan languages are more similar to each other than they are to Japanese. These findings show the scope of semantic changes that can occur through outside influence and highlight the need for fresh empirical data in the study of semantics in related languages.

Keywords: colour, semantics, language variation and change, Japonic, Ryukyuan

[1] This chapter is based on Huisman, J. L. A., van Hout, R. & Majid, A. Stability and change in the colour lexicon of the Japonic languages. Submitted to *Studies in Language*; currently under review. Formatting follows the journal's guidelines. I certify that I performed data collection, analysis and writing of the manuscript, with feedback from the co-authors.

4.1 General introduction

As the world around us changes, our language changes with it. Technological and societal developments can increase the cultural salience of a concept and with that boost the need to talk about it. For example, as societies gain greater technological control of colour, the need to communicate about colour becomes more important and more terms are added to partition the colour spectrum (Berlin & Kay 1969; Levinson 2000). For the colour lexicon, an immense body of work—particularly influenced by the classic study of Berlin and Kay (1969)—has shown that its structure and evolution are cross-linguistically constrained. Several large-scale projects, such as the World Color Survey (Kay, Berlin & Merrifield 1991) and the Mesoamerican Color Survey (MacLaury 1986), were launched to further test Berlin and Kay's hypotheses across a wide range of languages, leading to refinements and additions to the original framework (see e.g., Kay et al. 1997; Kay et al. 2009; MacLaury et al. 1992; MacLaury 2001; Jäger 2012), as well as eliciting notable challenges (Lucy 1997; Levinson 2000; Wierzbicka 2008).

This work primarily focused on addressing the universality of constraints on the colour lexicon, so data collection was aimed towards lesser-described, unrelated languages to allow broad cross-linguistic comparison. Since typological studies in general tend to favour diverse language samples, the scope of semantic variation across related languages is often not well explored, and variation in colour terms and categories across related languages has received considerably less attention (although see, e.g., Kristol 1980; Majid, Jordan & Dunn 2015; Haynie & Bowern 2016; Lillo et al. 2018). In addition, it is notable that even modern explorations of colour language (e.g., Regier, Kay & Khetarpal 2007, 2009; Jäger 2012) are based on data that was collected several decades ago—surveys for both the World Color Survey and the Mesoamerican Color Survey were conducted in the 1970s.

This paper departs from this approach by collecting new data from speakers of several related unwritten languages from traditionally low technology cultures. These are interesting to look at today because these communities have undergone many societal changes in recent times as globalisation and interconnectedness has led to increased worldwide exposure and the influence of Western culture and the English language. Contact with other languages and cultures can lead to the introduction of new colour categories as speakers are exposed to colour system that differ from their own. For example, several languages in the World Color Survey (Kay et al. 2009) and the Mesoamerican Color Survey (MacLaury 1986) have forms of *verde* and *azul* for green and blue in their colour lexicon as a result of influence from Spanish.

The current study focuses on Japanese and the Ryukyuan languages spoken across the Japanese mainland and the Ryukyu Islands. The colour lexicon has been a topic of study in these languages since the 1960s (Kusakabe 1964; Berlin & Kay 1969), but more recent work has mainly concerned Japanese (e.g., Uchikawa & Boynton 1987; Kuriki et al. 2017). These later studies revealed several developments in the Japanese colour lexicon, but it is unclear to what extent the Ryukyuan languages have also changed. Given the societal changes and language contact in recent decades, the Ryukyu Islands make an ideal testbed to investigate the impact of modern society on these languages. Furthermore, the region provides an opportunity to investigate variation in the colour lexicon of a single language family—an understudied arena of the colour literature. Section 4.1.1 shortly introduces the language family itself and discusses some sociocultural differences between the Japanese mainland and the Ryukyu Islands, after which Section 4.1.2 summarises previous work on colour vocabulary in the Japonic languages, and Section 4.1.3 introduces the current study.

4.1.1 The Japonic language family

The Japonic language family consists of two major branches. The first is the Japanese branch, comprising the varieties spoken on the main islands, which are generally subdivided into Eastern, Western, Kyushu, and Hachijo Japanese (Shibatani 1990). Modern Standard Japanese is a predominantly Eastern Japanese variety with some influences from

Western Japanese (Iwasaki 2013). The second branch of the Japonic language family is Ryukyuan, which comprises the varieties spoken across the smaller islands in the south. Ryukyuan is generally subdivided into Amami and Okinawa (Northern Ryukyuan) on one hand, and Miyako, Yaeyama and Yonaguni (Southern Ryukyuan) on the other (Pellard 2015; Huisman, Majid & van Hout 2019).

Since the turn of the 20th century, when the Ryukyu Islands were incorporated into Japan as a country, more and more people have been exposed to and eventually become bilingual in Standard Japanese in addition to their local Ryukyuan variety. Various language policies over the course of the 20th century then diminished the overall use of Ryukyuan, resulting in Japanese becoming the first language for many speakers, and leading to the eventual endangered status for Ryukyuan in the 21st century, with no monolingual speakers remaining (Heinrich 2005).

While both Japanese and Ryukyuan have been intensively exposed to Western culture and the English language in the immediate post-war period, the Ryukyu Islands (Okinawa in particular) have their own unique linguistic situation as a result of prolonged American occupation, which led to the incorporation of English loanwords not found in Standard Japanese (Tsuhako 1992). In addition, the increased use of Standard Japanese across the Ryukyus means that speakers of Ryukyuan are now also exposed to English loanwords that have become integrated into the standard language.

In addition to the linguistic situation, there are also some cultural factors to take into consideration. First, as more elaborate material culture is often associated with a larger colour lexicon (Naroll 1970; Ember 1978), it is relevant to note that there is little difference between the general levels of traditional material culture in Japan and the Ryukyus. Both regions have extensive dyeing practices with elaborate colour designs since the Middle Ages, seen in the Japanese *kimono* and Okinawa's counterpart *bingata*, for example. Next, it is also important to consider that the Ryukyu Islands were politically and culturally independent from mainland Japan for several centuries, including when Japan closed itself off from the outside world during the *sakoku* policy. The former Ryukyu Kingdom traded freely with countries across the region, allowing greater exposure to other cultures and languages with different technologies, concepts, and vocabulary.

This combination of cultural interest in colour through material culture and the prolonged contact with both Japanese and English would predict that the Ryukyuan languages have large colour lexicons comparable to the larger languages documented in the colour literature (e.g., Berlin & Kay 1969), but this is not the case as will become clear in Section 4.1.2.

4.1.2 Studies on colour in the Japonic language family

There have been several studies on the colour lexicon of Standard Japanese. Its inclusion in the original study by Berlin and Kay (1969), combined with a now classic naming study by Uchikawa and Boynton (1987), followed by a further replication by Kuriki and colleagues (2017), provide us with around five decades of data on the Japanese colour lexicon. These studies reveal that the Japanese colour lexicon has seen changes within this period—see Table 4.1 for a summary.

Standard Japanese has (at least) the 11 basic colour terms set out by Berlin and Kay (1969). The term *mizu(iro)* for ‘light blue’ has been suggested as a twelfth term based on both its frequent and consistent use across speakers (Uchikawa & Boynton 1987; Kuriki et al. 2017), and on linguistic grounds (Stanlaw 1987; Conlan 2005). Secondly, the findings in both Uchikawa and Boynton (1987) and Kuriki and colleagues (2017) seem to indicate that the term *nezumi(iro)* for grey has fallen out of use since the original study by Berlin and Kay (1969). Finally, the effects of internationalisation are clearly noticeable, as the native words *momo(iro)* ‘pink’ and *daidai(iro)* ‘orange’ have largely been replaced by their English loanword counterparts *pinku* and *orenji* (Uchikawa & Boynton 1987; Kuriki et al. 2017). Uchikawa and Boynton (1987) reported a similar tendency for the term *hai(iro)* ‘grey’, however Kuriki et al. (2017) showed that this replacement did not occur[2]. It is important to note here that the large number of basic colour categories already reported in Japanese by Berlin and Kay (1969) mean that the changes reported in later studies mainly relate to the colour terms used to describe those categories. While there has been no direct comparison between the colour system of Japanese at different points in time, the introduction of 'light blue' as basic category appears to be the most significant semantic change, and mirrors the distinction in other languages, such as Turkish (Özgen & Davies 1998), Russian (Paramei 2005) and Greek (Athanasopoulos 2009).

Table 4.1

Japanese colour terms across three major colour chip naming studies.

	B&K	U&B	K et al.
WHITE	*shiro*	*shiro*	*shiro*
GREY	*haiiro / nezumiiro*	*hai / guree*	*hai*
BLACK	*kuro*	*kuro*	*kuro*
BROWN	*cha(iro)*	*cha*	*cha*
RED	*aka(iro)*	*aka*	*aka*
ORANGE	*daidai(iro)*	*daidai / orenji*	*orenji*
YELLOW	*ki(iro)*	*ki*	*ki*
GREEN	*midori(iro)*	*midori*	*midori*
LIGHT BLUE	-	*mizu (?)*	*mizu*
BLUE	*ao*	*ao*	*ao*
PURPLE	*murasaki(iro)*	*murasaki*	*murasaki*
PINK	*momoiro*	*momo / pinku*	*pinku*

Note. B&K = Berlin & Kay (1969), U&B = Uchikawa & Boynton (1987), K et al. = Kuriki et al. (2017).

[2] The fact that 'orange' as a colour category was not established until the Meiji Era (1868-1912)—a period with significant Western influence—combined with the slightly earlier appearance of *orenjiiro* (1897~1897, cf. *daidaiiro* 1905; Shogakukan 2002) could raise some questions about whether the native term *daidai(iro)* ever, in fact, achieved basic colour term status.

While the contemporary colour system of Japanese conforms to the Berlin and Kay framework, its developmental trajectory shows a few anomalies. The long history of writing in Japan has made it possible to track changes in the colour lexicon (Stanlaw 2007, 2010). Stanlaw theorises about earlier stages of colour terminology based on linguistic and cultural data—see Table 4.2 for an overview.

Table 4.2

Colour terms added in first four steps of the evolution of the Japanese colour lexicon, with approximate dates for each stage (Stanlaw, 2007).

I 400 CE	II 500 CE	III 650 CE	IV 750 CE
aka ‘red’ *ao* ‘grue’ *kuro* ‘black’ *shiro* ‘white’	*ki* ‘yellow’	*murasaki* ‘purple’	*midori* ‘green’ *momo* ‘pink’

Note. Stage IV represents that of the first writing, i.e. there were eight terms.

At the time of the earliest writings in the 8th century, there were eight colour terms in use. Strikingly, ‘purple’ appears early in this data (see also Wnuk, Levinson & Majid submitted), which is cross-linguistically uncommon and not predicted by the original Berlin and Kay framework. However, Stanlaw (2007) argues there is strong evidence for purple being a culturally salient colour. The limited availability of purple dye, and its resulting exclusivity, has made the colour purple important in many cultures, often linked to high status and power (Hendrick-Wong 2013). While the exact timing of the divergence between Japanese and Ryukyuan is still under debate, most theories place it well before the 8th century (Pellard 2015, for an overview), meaning that the Japanese and Ryukyuan colour lexicons developed independently from an ancestral language with six or fewer colour terms.

For some initial insights into the colour lexicon of the Ryukyuan languages, we can turn to dialect dictionaries such as the *Dictionary of Contemporary Japanese Dialects* (Hirayama 1992), which includes data for several Ryukyuan varieties and has entries for seven colour terms: *aka* 'red; a warm color', *kiiro* 'yellow color', *midori* 'green; verdure', *ao* 'blue; sky blue; azure; indigo blue; green', *murasaki* 'purple, amethyst', *kuro* 'black color', and *shiro* 'white' (translations as found in the dictionary). Table 4.3 is a summary of entries for the five chromatic colour terms in Standard Japanese and the Ryukyuan varieties.

Table 4.3

Entries for the five chromatic colours in the *Dictionary of Contemporary Japanese Dialects* (Hirayama 1992) for Standard Japanese and four Ryukyuan languages.

	red	yellow	green	blue	purple
Standard	*aka*	*ki:ro*	*midori*	*ao*	*mɯrasaki*
Ryukyuan					
Amami	*haː*	*kˀiʔiru*	*midori*	*ʔoʔiro*	*murasaki*
Okinawa	*ʔaka:*	*ki:ru:*	*ʔo:ru:*	*ʔo:ru:*	*ʔoːruː*
Miyako	*aka*	*kˢɨ:ru~tsɨ:ru**	*o: ~ au*	*o: ~ au*	*murasaki**
Yaeyama	*ʔaga*	*ki:ru*	*ʔau*	*ʔau*	*murasaki**

Note. Asterisked entries are specifically mentioned to be relatively new expressions resulting from standard language influence.

The table shows that a distinction between 'green' and 'blue' is not generally made in the Ryukyuan languages. Furthermore, it seems that Okinawa extends the use of *ʔo:ru:* to include 'purple', whereas the other languages distinguish it as a separate category. However, the Miyako and Yaeyama entries specifically mention that the term *murasaki* for purple is a relatively new expression that has been introduced through contact with the standard language, which is also suggested by the resemblance to the Standard Japanese form. Finally, some entries for the 'yellow' term in Miyako also mention standard language influence. This colour system with fewer distinctions is expected if we assume that Ryukyuan indeed split from Japanese before the 8th century.

Aside from information in such dictionaries, there is one large-scale study that uses colour stimuli in a naming task. Kusakabe (1964) collected naming data for 15 chromatic colour cards from 89 speakers across 68 locations in Okinawa Prefecture[3]. The study showed that Ryukyuan speakers used between two and four chromatic colour terms, largely cognate with the oldest colour terms in Japanese—*aka* 'red', *ao* 'grue', *ki* 'yellow' and *murasaki* 'purple' (Stanlaw 2007, 2010). The Miyako and Yaeyama 'purple' terms in Kusakabe (1964) were listed as having the form *mura(t)sɨkɨ*[4], which differs from the form *murasaki* in Table 4.3. The note in Hirayama (1992) that *murasaki* is relatively new thus likely comments on the specific form, rather than the colour category as such. Unfortunately, there has been no systematic study of Ryukyuan colour terminology since, leaving it unclear what changes have occurred more recently.

To summarise, while the Standard Japanese colour lexicon has been studied since the 1960s, the most extensive inquiry into the colour lexicon of the related Ryukyuan languages is more than 50 years old. Given the developments in the Japanese colour lexicon during that time, as well the changing linguistic landscape across the Ryukyu Islands described above, the question arises what impact this has had on the colour lexicon of contemporary Ryukyuan speakers.

4.1.3 The current study

The current study aims to gain insight into the modern colour lexicon of Ryukyuan and assess the influence from Standard Japanese and English. In the first part, we describe how colour naming data was collected from modern Ryukyuan speakers, and outline the colour categories in the languages under consideration. We then explore semantic variation in a

[3] The study includes data from four Ryukyuan languages (Okinawa, Miyako, Yaeyama and Yonaguni), but excludes Amami Ryukyuan, which is spoken in Kagoshima Prefecture.
[4] Miyara (1930) also lists the similar form *muraçiki* for Taketomi dialect (Yaeyama).

core set of colour terms—*aka* 'red', *ki* 'yellow', *ao* 'grue', and *murasaki* 'purple'—which were chosen because of their longevity (Stanlaw 2007, 2010) and their use in both Japanese (Berlin & Kay 1969; Uchikawa & Boynton 1987; Kuriki et al. 2017) and Ryukyuan (Kusakabe 1964; see also Hirayama 1992). We compare the semantic distances between these colour categories using a statistical procedure applied by Jäger (2012). Finally, we investigate the semantic changes that have occurred in Ryukyuan over time by comparing the newly collected naming data to earlier data from Kusakabe (1964).

4.2 Methods

4.2.1 Languages and speakers

Data was collected from speakers in three Ryukyuan language areas as well as from Standard Japanese for comparison—see Figure 4.1—during four fieldtrips conducted between 2017 and 2019. An attempt was also made for data collection with Okinawan with two speakers in a session; but this was stopped at the request of speakers after they were unable to name several colour chips. As there is no standardised variety of any of the Ryukyuan languages (Heinrich, Miyara & Shimoji 2015), data was collected from multiple localities, i.e., in multiple dialects. We will use the term "language area" when describing Ryukyuan for the remainder of this paper, e.g., the Amami language area, so we can refer to each (mutually unintelligible) Ryukyuan language without having one specific variety as its standard. As the Ryukyuan languages are endangered, data was collected from elderly native speakers; we also collected data from older speakers in Standard Japanese to provide a comparable sample. The data presented here also constitutes the first systematic investigation of colour naming in the Amami language area, since it was not featured in the Kusakabe (1964) study.

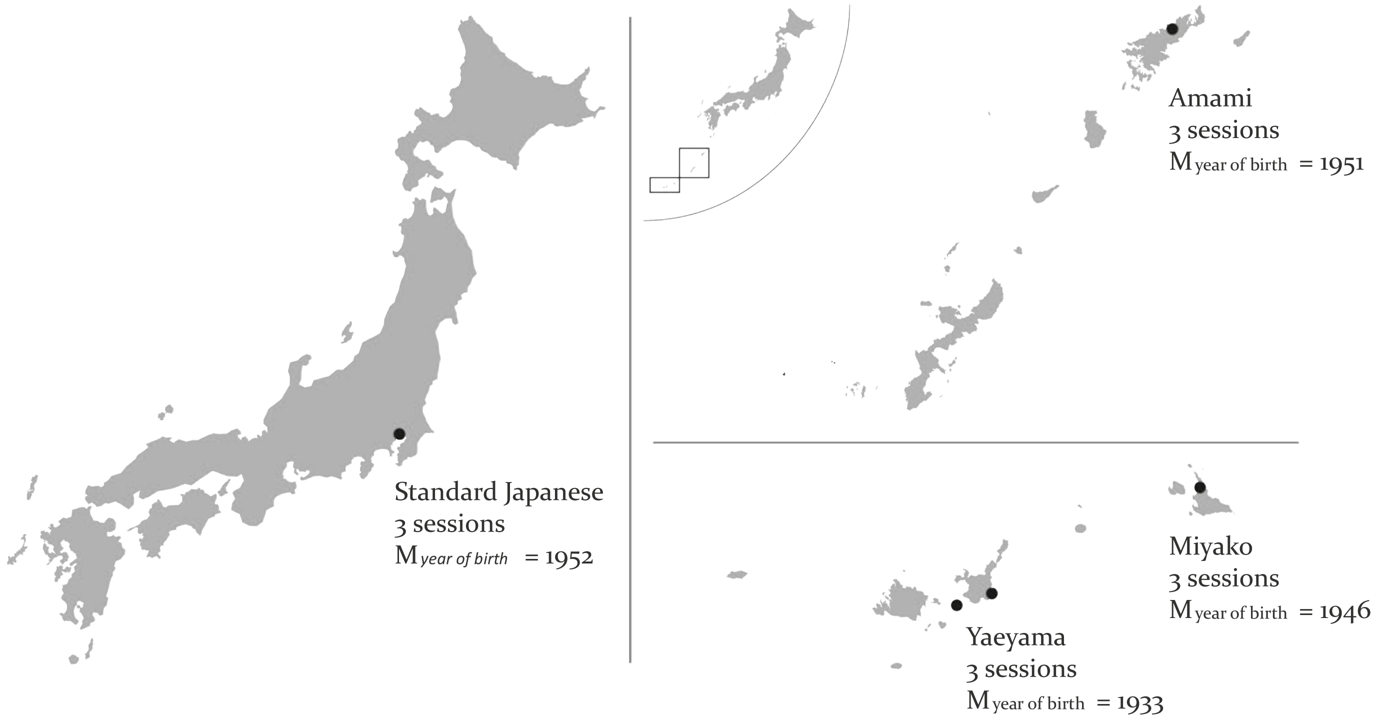


Figure 4.1. Map of Japan (left) and the Ryukyu Islands (right) with fieldwork locations marked, as well as session and participant information.

4.2.2 Materials and procedure

Some of the elderly speakers tested had little experience in performing abstract, reflective language tasks, and therefore, some naming sessions were conducted with multiple speakers simultaneously. As a result, all analyses were performed on sessions rather than speakers—see Section 4.2.3 *Coding*. Speakers were tested indoors in natural lighting conditions. Speakers were pre-tested for their colour vision using an Ishihara test, and we only conducted further testing with speakers that passed it. For the colour naming task, we used a set of 84 Munsell colour chips (Majid, Jordan & Dunn 2011)—see Figure 4.2. The colour chips were presented one by one, in two fixed orders—one being the reverse of the other—after which speakers were asked to freely name the colour of the chip in either Standard Japanese or their local Ryukyuan variety, as appropriate in each site. Speakers were allowed to give multiple responses of any length. Sessions were audio (and sometimes video) recorded for later transcription. Speakers gave informed consent before participating, and all data was collected under the Ethics Assessment Committee of the Centre for Language Studies at Radboud University.

4.2.3 Coding

All full responses were transcribed. Colour terms can appear in several forms in Japanese and Ryukyuan varieties. Moreover, as speakers were told to freely describe the colour chips, they used a range of different naming strategies—see Examples 1 through 4.

Standard Japanese			Coded response:
<u>*ao*</u>=*ni*	*chikai*	→	<u>*ao*</u>
<u>blue</u>=DAT	near:NPST		
'(it is) close to <u>blue</u>.'			

Amami Ryukyuan (Tatsugo) → Coded response:
o:san → *o:*
blue:NPST
'(it is) blue.'

Miyako Ryukyuan (Nishihara) → Coded response:
au-munu → *au*
blue-thing
'blue.'

Yaeyama Ryukyuan (Shiraho) → Coded response:
o:-o:-sero → *o:*
blue-blue-do:CONT
'(it is) blue.'

In addition to the full responses we extracted, per session, the main response(s) for each colour chip. We defined the main response as either the root of the basic colour term, or the source term that is the core descriptor for each colour chip. All nominal-, adjectival- and verbal-suffixing, as well as reduplications and modifier phrases were excluded—see *Coded response* on right-hand side of Examples 1 through 4.

4.3 Results

We first provide a general overview of the colour terms used by contemporary speakers of the Ryukyuan languages and assess the influence from Standard Japanese and English in Section 4.3.1. We then focus on the colour categories distinguished by contemporary speakers of the Ryukyuan languages in Section 4.3.2. Next, we present an analysis of the semantic variation across a core set of colour terms: *aka* 'red', *ao* 'grue', *ki* 'yellow' and *murasaki* 'purple' in Section 4.3.3. Finally, Section 4.3.4 compares our newly collected data to the findings previously reported in Kusakabe (1964).

4.3.1 Colour terms in contemporary Ryukyuan

We used the frequencies of coded responses to get a general overview of the colour terms elicited in the naming task. As the languages under investigation in this study are related, we were interested in establishing the extent to which each term is used across the language family. In addition, such an overview provides some initial clues about the extent of outside influences such as Standard Japanese terms and English loanwords. Table 4.4 shows, per language area, the responses that occurred at least three times and that were used in more than one session. Each row in the table contains one cognate set, with rows arranged by an approximate colour gradient. We grouped together four types of responses: (1) abstract colour terms, (2) terms that have been designated as basic colour terms in Standard Japanese, but that are/were originally source-based, (3) source-based terms that are generally not considered basic colour terms, and (4) English loanwords.

The Standard Japanese data largely recapitulates previous naming studies (Berlin & Kay 1969; Uchikawa & Boynton 1987; Kuriki et al. 2017; see Table 4.1). The native word *momo(iro)* 'pink' (N = 19) outnumbered the English loanword *pinku* (N = 6) and all speakers used *daidai(iro)* instead of *orenji*. The relatively large number of native colour terms used for 'orange' and 'pink', as opposed to their English loanword counterparts, resembles the findings reported by Uchikawa and Boynton (1987) more than Kuriki and colleagues (2017). This is most likely because sample of speakers, who were born in the 1940s-1950s, are demographically most similar to Uchikawa and Boynton (1987). Nevertheless, the term *mizu(iro)* was also frequently used, indicating the emerging salience of 'light blue' as a category as reported in Kuriki et al. (2017).

Overall, the total number of terms elicited in each of the four language areas was comparable, although the number of colour terms produced per individual session ranged between 13 and 22, with the exception of one Miyako speaker who used only 4 terms, specifically mentioning that "there are only four colours in Miyako". Several other Ryukyuan speakers commented that their language variety "does not have many words for colours", also stating more specifically that "there are no separate words for blue and green", although they did appeal to Standard Japanese colour terms, English loanwords or source-based descriptions to make finer distinctions on occasion. Overall, most speakers demonstrated at least some meta-linguistic awareness of their bilingualism, as well as differences between the colour lexicons of both languages.

Most of terms used across the Ryukyuan languages have cognates in Standard Japanese. The six oldest colour terms *aka*, *ao*, *kuro*, *shiro*, *ki* and *murasaki* (Stanlaw 2007, 2010) were used in all sessions. Of these, the four oldest (*aka*, *ao*, *kuro* and *shiro*) were virtually always used in their local pronunciations—see Table 4.4. The 'yellow' term *ki* had a mixture of local and standard-like pronunciations, but the 'purple' term *murasaki* was most often in Standard Japanese pronunciation, in contrast to the reported *mura(t)sïkï* in Kusakabe (1964). Several other terms were produced in Standard Japanese forms, such as *momo(iro)* 'pink; lit. peach' where *mumu* was expected, and *tsuchi(iro)* 'lit. soil (colour)' *sora(iro)* 'lit. sky (colour)' by Miyako speakers even though the local word for 'soil' is *mta~nta* and 'sky' is *tin*. For the Yaeyama language area, we find *tin nu iru* '[sky GEN colour]'. The strongest influence of Standard Japanese on Ryukyuan is found in the high frequency of *midori* 'green' across all Ryukyuan sessions, which as we discuss in Section 4.3.4 was not reported in the Ryukyuan data of Kusakabe (1964). In addition, the form of this colour term was predominantly *midori* as in Standard Japanese. If the term had been integrated in the Ryukyuan languages, the various sound changes across the languages would have produced *miduri*, *midui*, *midurï* or *midu̹*, but these were not attested in the naming data.

Table 4.4

Responses produced at least three times in more than one session in the naming task.

Abstract colour terms

Japanese	Amami	Miyako	Yaeyama	Gloss
shiro	*shiro*	*ssu*	*ssu*	'white'
kuro	*kuro*	*ffu*	*ffu*	'black'
aka	*ha:*	*aka*	*aga*	'red'
ki	*ki:*	*ki:*	*kiN*	'yellow'
ao	*o:*	*au*	*o:*	'blue'
		kon		'dark blue'

Originally source-based terms designated as basic colour terms in Standard Japanese

Japanese	Amami	Miyako	Yaeyama	Gloss	
				Colour	Source
hai	*hai*		*hai*	'grey'	'ash'
		nezumi		'grey'	'mouse'
cha	*cha*		*cha*	'brown'	'tea'
daidai	*daidai*	*daidai*	*daidai*	'orange'	'*C. ×daidai* citrus fruit'
midori	*midori*	*midori*	*midori*	'green'	'verdure'
mizu	*mizu*			'light blue'	'water'
murasaki	*murasaki*	*murasaki*	*murasaki*	'purple'	'*L. erythrorhizon* plant'
momo	*momo*			'pink'	'peach'

Source-based terms

Japanese	Amami	Miyako	Yaeyama	Gloss
		azuki		‘adzuki bean (*V. angularis*)’
hada				‘skin'
			ka:ra	‘roof tile'
tsuchi	*tsuchi*	*tsuchi*		‘soil'
oudo	*oudo*			‘yellow soil'
			fumbutu	‘fukugi tree (*G. subelliptica*)’
kusa			*ssa*	‘grass'
			pa:	‘leaf'
sora	*sora*	*sora*		‘sky'
			tin	‘sky'
			tu:	‘sea'
ai				‘indigo plant (*P. tinctorial*)’

English loanwords

Japanese	Amami	Miyako	Yaeyama	Gloss
		orenji	*orenji*	'orange'
	guriin	*guriin*	*guriin*	'green'
		buruu		'blue'
pinku	*pinku*	*pinku*	*pinku*	'pink'

Note. As the naming data was collected from several dialects, the responses listed in the table list the most frequent form in each language area.

Ryukyuan speakers used several English loanwords, of which *pinku* was the most frequent. Even though *orenji* has been shown to have all but replaced the native term *daidai(iro)* in Standard Japanese across several studies (e.g., Kuriki et al. 2017), Ryukyuan speakers did not use *orenji* to a comparable level. The strongest influence from English is shown in the frequent use of *guriin* to distinguish green parts of the spectrum even though no Standard Japanese speakers used an English loanword to do so, indicating a potential Ryukyuan-specific development.

Speakers from all language areas used source-based terms and the Ryukyuan sessions produced a few that were specific to the local area such as *ka:ra* referring to the orange/brown roof tiles found on Taketomi Island, or *fumbutu* referring to the *G. subelliptica* tree species commonly found around the Ryukyu Islands used for yellow dye. Yaeyama speakers in particular used relatively more source-based terms and interestingly, these were again used for green parts of the spectrum as with the Standard Japanese terms and English loanwords. Green shades were described with the source-based terms *ssa* 'grass' (< Standard Japanese *kusa*), *pa:* 'leaf/leaves' (< Standard Japanese *ha*), or a combination of the two (*ssa nu pa:* '[grass GEN leaves]').

4.3.2 Colour categories in contemporary Ryukyuan

We used the modal responses per colour chip to uncover the most salient colour categories in each language area. Figure 4.2 (top panel) shows the 84 colour chips arranged by hue (coded 0 through 20) and brightness (coded A through D). The panels below show the modal response for each colour chip in each of the four language areas. We determined, for each colour chip, the most frequent response per language area. Where speakers specifically commented that terms were synonyms, e.g., *midori*, *guriin* and *ssa* for 'green' in Yaeyama, we counted these together for the purpose of these figures, but elsewhere keep them distinct. We then coloured all chips with the same modal response the same colour. Blank chips are those for which there was no modal response, i.e. there were different responses across sessions.

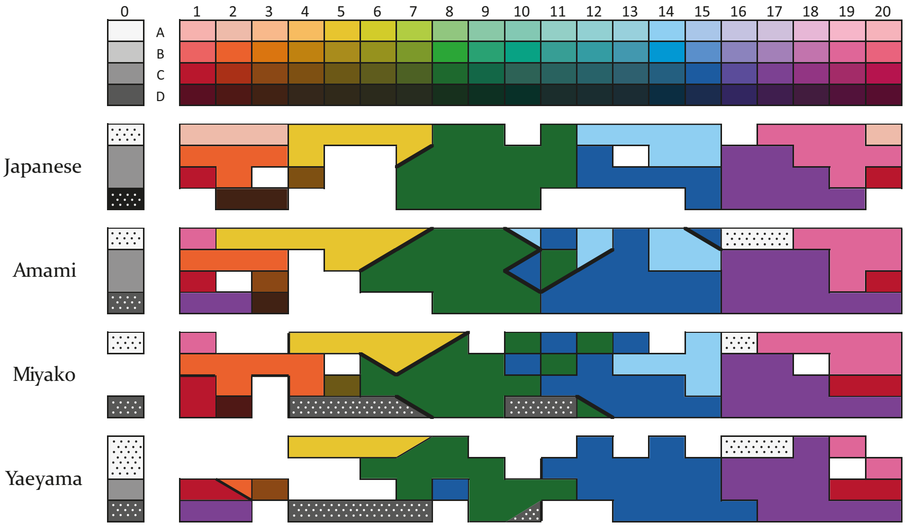


Figure 4.2. Modal responses per language for the 84 colour chips. Top grid presents an approximation of the colour chips used for elicitation.

The Japanese data is consistent with what has been reported previously: the eleven basic colours 'white', 'grey', 'black', 'brown', 'red', 'orange', 'yellow', 'green', 'blue', 'purple', and 'pink' from Berlin and Kay (1969) were clearly distinguished. In addition, a 'light blue' category emerged as previously suggested (Uchikawa & Boynton 1987; Kuriki et al. 2017), as well as 'peach' and 'light brown', consistent with Kuriki et al. (2017).

Fewer categories emerged in the Ryukyuan language areas, with 13 colours in Amami, 12 in Miyako and 11 in Yaeyama—compared to the 14 in Standard Japanese. Overall, the Ryukyuan languages areas and Standard Japanese showed evidence of the eleven basic colours suggested by Berlin and Kay (1969). The 'light blue' category emerged in Japanese, Amami and Miyako language areas but not in Yaeyama; and 'light brown' was also evident in Amami, but not in the other two language areas. There was no evidence of a 'peach' category in Ryukyuan.

The figure also shows a number of differences between Japanese and Ryukyuan in the 'grue' area. Japanese neatly divides 'green' and 'blue', but this division is not established across the Ryukyuan languages where the two categories seem to overlap and intersect in the boundary area. In addition, while 'light blue' was present in Amami and Miyako, it is not as clearly defined as in Standard Japanese, Also, whereas *murasaki* is only used for the 'purple' part of the spectrum in Japanese, the Ryukyuan languages seem to extend its range to include darker shades of 'red'. Finally, while the use of Standard Japanese *shiro* and *kuro* is limited to achromatic colours ('white' and 'black', respectively), the three Ryukyan language areas extend the use of these terms to include lighter and darker shades in general—see chips in rows A and D, respectively, suggesting they are still being used as 'dark' and 'light'. This usage resembles that of the oldest stages of the Japanese colour system (Stanlaw 2007, 2010).

Figure 4.2 also hints at some differences between the three Ryukyuan language areas. In contrast to Amami and Miyako speakers, Yaeyama speakers did not seem to distinguish shades of blue. Furthermore, while Amami extends the use of its 'white' term to include lighter shades in general, it does not extend the use of its 'black' term to include general dark shades as Miyako and Yaeyama do. Finally, 'orange' and 'pink' did not emerge as prominently in Yaeyama as they did in the other language areas.

To summarise, while the modal responses from the naming task indicate that Ryukyuan varieties have fewer colour categories, several categories not previously reported emerged, e.g., 'green', 'orange'—cf. Kusakabe (1964). However, the boundaries of these new categories appear to be less clearly defined than in Japanese.

4.3.3 Semantic variation in a core set of colour terms

We next zoom in on the semantic variation across a core set of colour vocabulary. We looked specifically at the meanings of the cognate sets *aka*, *ao*, *ki* and *murasaki*, which were chosen because of their longevity and their widespread use in the language family. To statistically compare the cognate sets across the language areas, we performed principal component analysis (PCA) to uncover the units that formed the basis for these categories, using the approach set out by Jäger (2012), in base R (R Core Team 2019, `prcomp` function).

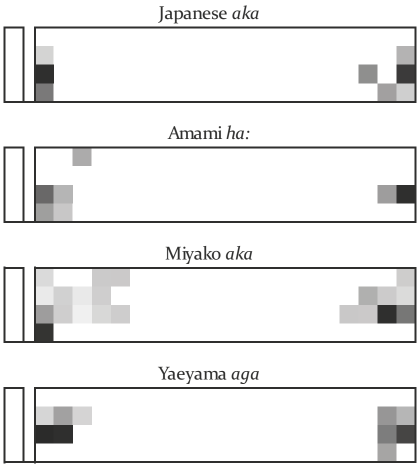
To minimise the effect of idiosyncratic responses, we only included colour terms that were used in more than one session and were used more than three times—i.e., terms in Table 4.4. Following Jäger (2012), we created a colour term * colour chip matrix for each session, in which we coded the frequencies of the terms used for each chip, and then summed the matrices for all sessions per language area. Next, each row was divided by the number of sessions in which the term was used, after which each row was copied as many times as the response occurred to give more weight to more frequent responses. Finally, we stacked the four language area matrices, resulting in a 190x84 contingency table that was used for the principal component analysis.

We used the scree test (Cattell 1966) to determine the number of principal components and continued with a solution of 12 components accounting for 75.5% of the variance.[5] Ten of the eleven basic colour categories proposed by Berlin and Kay (1969) were represented as principal components, with only *grey* missing. Following Jäger (2012), we rotated the PCA solution using Varimax rotation to maximise the components (`varimax` function; R Core Team 2019). As principal component analysis separates statistically important variation from noise, we used the 12 extracted components to project the four chromatic cognate sets (*aka*, *ao*, *ki*, *murasaki*) back onto the colour chip array (see Jäger 2012, p. 526-531). The PCA solution was used to calculate cosine similarities between each set of colour terms as a measure of similarity between languages, which were then used in separate cluster analyses for each cognate set, as well as an aggregate cluster analysis (`hclust` function; R Core Team 2019).

[5] Jäger (2012) used the Kaiser criterion (all factors with an eigenvalue > 1; Kaiser 1960) to select the number of principal components, which in our cases would mean selecting 17 principal components, together accounting for 82.8% of the variance in the data. However, inspection of this solution showed several components that were hard to interpret—especially in light of the naming data presented in Sections 4.3.1 and 4.3.2.

4.3.3.1 The aka *'(macro) red' cognate set*

Figure 4.3a shows the range of the *aka* cognate set across the four languages, Figure 4.3b shows the cosine similarities, and Figure 4.3c is a plot of the cluster analysis based on these cosine similarities. The cluster analysis revealed Miyako to be the most distinct, and that Amami and Yaeyama were more similar to each other than they were to Japanese. As the figure shows, speakers of Ryukyuan varieties tend to extend the use of *aka* further into the orange part of the spectrum, whereas Japanese speakers restrict its use to a small set of red chips. Miyako is the only variety where *aka* is extended to include yellow hues as well.

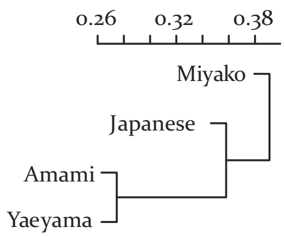


(a)

Cosine similarities for the *aka* cognate set

	Amami	Miyako	Yaeyama
Japanese	.736	.806	.627
Amami		.917	.860
Miyako			.850

(b)

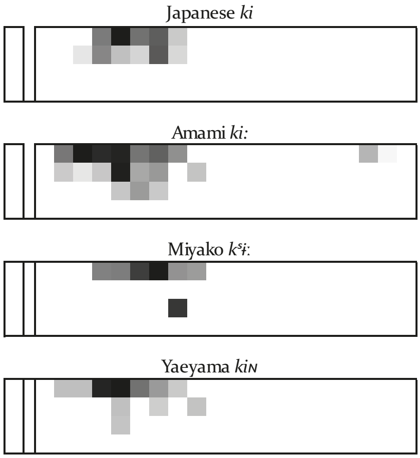


(c)

Figure 4.3. The range of the *aka* cognate across the four language areas (4.3a), cosine similarities calculated based on principal component analysis (4.3b), and cluster analysis results based on these cosine similarities (4.3c).

4.3.3.2 The ki *'yellow' cognate set*

As with *aka*, the cluster analysis revealed Miyako to be the most distinct, with Amami and Yaeyama more similar to each other than to Japanese. Figure 4.4a shows that speakers of Ryukuan tend to extend the use of *ki* further into orange and green, as well as for darker shades than Japanese speakers. Figure 4.4b shows the cosine similarities across language areas, and Figure 4.4c plots these cosine similarities according to a cluster analysis.

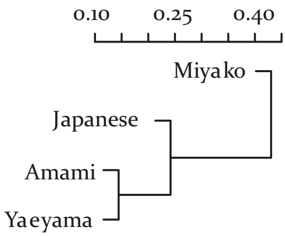


(a)

Cosine similarities for the *ki* cognate set

	Amami	Miyako	Yaeyama
Japanese	.750	.666	.812
Amami		.549	.856
Miyako			.691

(b)

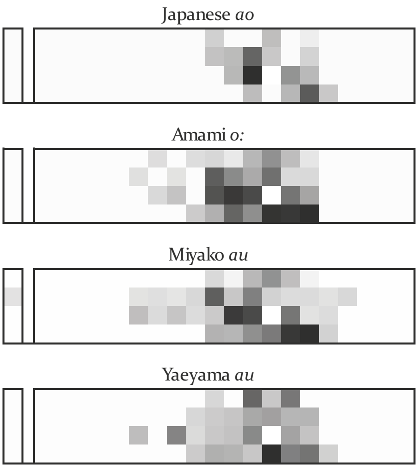


(c)

Figure 4.4. The range of the *ki* cognate across the four language areas (4.4a), cosine similarities calculated based on principal component analysis (4.4b), and cluster analysis results based on these cosine similarities (4.4c).

4.3.3.3 The ao *'grue/blue' cognate set*

The cluster analysis in Figure 4.5c reveals the three Ryukyuan language areas were more similar to each other than they were to Japanese. Speakers of the Ryukyuan varieties tended to extend the use of *ao* further into green, whereas Japanese speakers restricted its use to blue. Within Ryukyuan, Amami and Miyako speakers tended to have larger green categories than Yaeyama speakers, as seen in Figure 4.5a which depicts the range of the *ao* cognate set across the four languages.

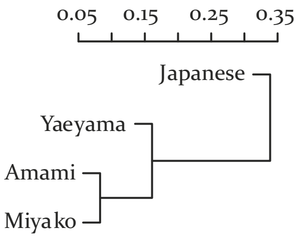


(a)

Cosine similarities for the *ao* cognate set

	Amami	Miyako	Yaeyama
Japanese	.717	.596	.613
Amami		.701	.725
Miyako			.606

(b)

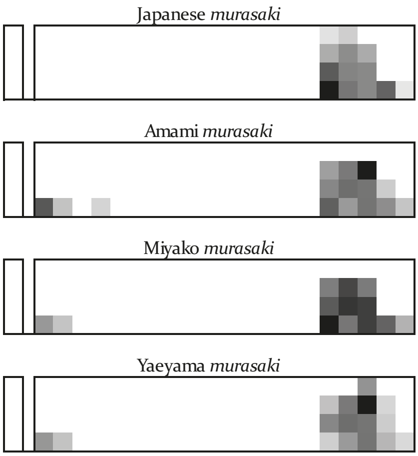


(c)

Figures 4.5. The range of the *ao* cognate across the four language areas (4.5a), cosine similarities calculated based on principal component analysis (4.5b), and cluster analysis results based on these cosine similarities (4.5c).

4.3.3.4 The murasaki *'purple' cognate set*

As with *ao*, the cluster analysis revealed the three Ryukyuan varieties were more similar to each other than they were to Japanese. Figure 4.6a shows that speakers of Ryukyuan tended to use *murasaki* to include dark shades of red, whereas Japanese speakers tended to limit the use to chips between blue and red. As another point of contrast, Ryukyuan speakers limited the use of *murasaki* to darker colour chips, whereas Japanese *murasaki* also included lighter chips. Figure 4.6b shows the cosine similarities across language areas, and Figure 4.6c gives a visualisation of the similarities.

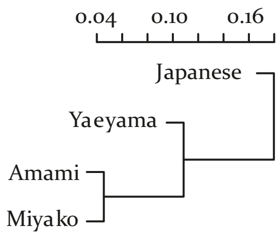


(a)

Cosine similarities for the *murasaki* cognate set

	Amami	Miyako	Yaeyama
Japanese	.858	.941	.781
Amami		.954	.926
Miyako			.885

(b)



(c)

Figure 4.6. The range of the *murasaki* cognate across the four language areas (4.6a), cosine similarities calculated based on principal component analysis (4.6b), and cluster analysis results based on these cosine similarities (4.6c).

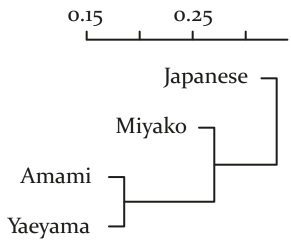
4.3.3.5 Overall cluster analysis for the cognate sets

Finally, we computed an aggregate similarity measure between the four language areas by calculating the mean of the cosine similarities for the four colour terms, and performed a cluster analysis using this aggregate similarity measure—see Figure 4.7. This overall cluster analysis confirmed that the Ryukyuan languages are distinct from Japanese. Within the Ryukyuan languages, Miyako seems to be the most distinct, likely due to the fact that one speaker only used four colour terms in the naming task.

Cosine similarities for the four cognate sets combined

	Amami	Miyako	Yaeyama
Japanese	.776	.671	.698
Amami		.730	.815
Miyako			.732

(a)



(b)

Figure 4.7. Cluster analysis results based on the mean cosine similarity for four cognate sets (*aka*, *ki*, *ao* and *murasaki*).

To summarise, a comparison of four colour cognate sets (*aka*, *ki*, *ao* and *murasaki*) using principal component analysis showed that the Ryukyuan languages are always distinct from Japanese in their use. The analysis also revealed that the languages group together differently for each cognate set, which suggests the meanings of colour terms develop along individual trajectories in each language.

4.3.4 Semantic change in the Ryukyuan colour lexicon

As mentioned in the introduction, the only systematic study of colour naming across the Ryukyus is Kusakabe (1964). We compared our data to this older Ryukyuan data in order to put our new naming data into further historical context and study semantic changes in the Ryukyuan colour lexicon. Since the two studies used different methods, we reanalysed the Kusakabe data for comparability. Kusakabe (1964, p. 62) used 15 colour cards from the Japan Colour Research Institute, which he labelled *a* through *O*[6]. We determined the closest approximations of these colour cards in our sample of 84 colour chips, shown in Figure 4.8. Kusakabe's colour cards *g*, *M*, *N* and *O* fall halfway between two of our colour chips. Colour cards *a* and *E* are not represented because these are distinctions not captured in our sample—in the red and yellow parts of the spectrum, respectively.

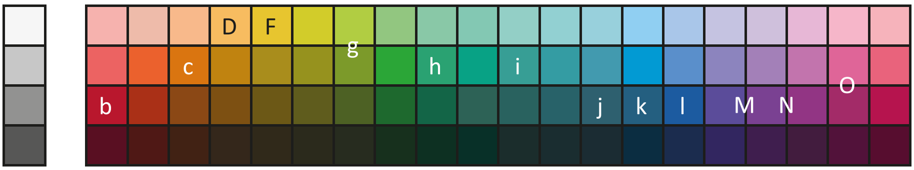


Figure 4.8. Overview of the 84 Munsell colour chips used in the current study, with closest approximations in the Kusakabe (1964) study marked.

Kusakabe (1964) focused on four chromatic colour term cognates: *aka* 'red', *ao* 'grue/blue', *ki* 'yellow', and *murasaki* 'purple'. As a result, his coding scheme only codes these four terms and all other responses were coded as 'other'[7]. For a direct comparison, we therefore recoded responses

[6] The author labelled the colour cards using both lowercase and uppercase (abcDEFghijlklMNO) as part of his coding scheme representing their (idealised) naming in Standard Japanese. We follow these original labels.
[7] In Kusakabe (1964), 24 speakers (27%) used terms designated as 'other'.

from our naming sessions following Kusakabe's conventions, i.e. only counting (cognate forms of) *aka*, *ki*, *ao* and *murasaki*, and coding all remaining responses as *other*. We created five new matrices in which we coded how many times each response (*aka*, *ki*, *ao*, *murasaki*, *other*) was used for each of the 13 colour chips shared between the current study and Kusakabe (1964). Two matrices contained Miyako data (1964; current), two contained Yaeyama data (1964; current), and one contained Standard Japanese data[8]. We do not consider the Okinawan data from Kusakabe (1964), as we were not able to elicit Okinawan data from speakers. As mentioned earlier, data collection was attempted with this variety, but was not possible.

The chip-by-response matrices were used to calculated distances between the five language areas, past and present. We chose chi-square distance to preserve the original frequency data and to compensate for the different number of data points per language area in the current study and Kusakabe (1964). The language-by-language distances were used as input for multidimensional scaling (MDS, using the `cmdscale` function; R Core Team 2019). Figure 4.9 shows the MDS solution in two dimensions—Stress-I = .085, indicating a fair to good fit (Kruskal 1964). The closer two points are in the figure, the more similar to each other they are.

Four things can be learned from Figure 4.9. First, the old colour systems of Miyako and Yaeyama were more similar to each other than they were to Standard Japanese. Second, the new colour systems of Miyako and Yaeyama are more similar to Standard Japanese than they are to their respective historic pasts. Third, while the new colour systems of Miyako and Yaeyama resemble each other more than they used to, they are still distinct. Finally, whereas the older colour systems of Miyako and Yaeyama were equally distinct from Standard Japanese, Miyako now resembles Japanese more than Yaeyama does. These last two points indicate that the

[8] While Kusakabe (1964) does not contain actual Japanese naming data, the Japanese data presented in Berlin and Kay (1969) and subsequent studies (Uchikawa & Boynton 1987; Kuriki et al. 2017) suggest negligible differences in the core set of responses for these colour chips. As such, we consider our newly collected Japanese data to closely resemble the Japanese colour system of the 1960s.

two language areas have had their own trajectories towards their current forms. Nevertheless, the first two points seem to indicate that the changes have happened fairly rapidly.

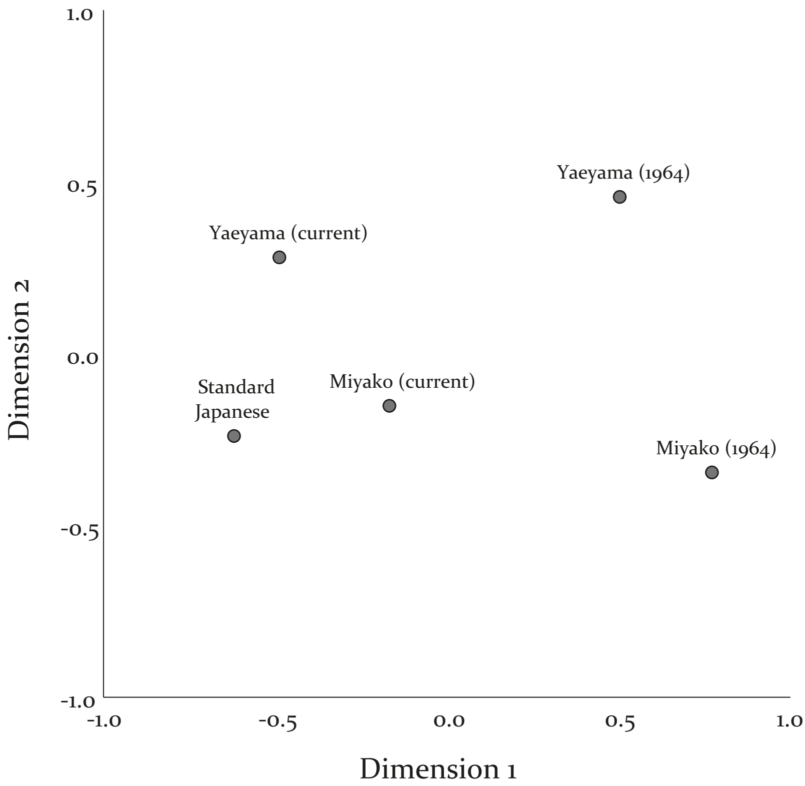


Figure 4.9. Multidimensional scaling solution for colour naming of 13 colour chips common between Kusakabe (1964) and the current study.

4.4 Discussion

We set out to investigate the colour lexicon of contemporary Ryukyuan speakers, in the context of the many changes in the cultural and linguistic landscape in recent decades. While the colour lexicons of Ryukyuan and Japanese were systematically examined in the 1960s (Berlin & Kay 1969; Kusakabe 1964), subsequent research on colour terminology in the Japonic language family mainly concerned Standard Japanese (e.g., Kuriki et al. 2017; Uchikawa & Boynton 1987). Statistical analysis of a core set of colour terms chosen for their longevity and widespread use across the language family showed that the Ryukyuan languages are distinct from Standard Japanese, hinting at the apparent stability of the colour system. This was most noticeable in the term *ao* for 'grue', which was used for a wider range encompassing green and blue across the Ryukyuan languages, whereas the term was only used for blue in contemporary Standard Japanese. Further evidence comes from the cognates of *shiro* and *kuro* which are used for achromatic 'white' and 'black' in Standard Japanese, but extend to 'light' and 'dark' in Ryukyuan, as has been hypothesised to be the case in pre-Old Japanese (Stanlaw 2007, 2010), again a sign of the longevity of the colour system.

At the same time, our naming data also exhibit considerable differences with previous data. A direct comparison between Kusakabe (1964) and the current study indicates that the modern Ryukyuan languages are closer to Standard Japanese than they are to their historic pasts. The responses elicited in the naming task of the current study revealed that several new terms, often standard(-like) phonological forms or loanwords, have entered Ryukyuan speakers' vocabulary through influence from both Standard Japanese (e.g., *midori*) and English (e.g., *guriin*). Most of these newly introduced terms were used for categories not previously reported for Ryukyuan, such as 'green' and 'orange' (Kusakabe 1964; but see also Hirayama 1992). Moreover, 'light blue'—not part of the original Berlin and Kay framework—has also emerged in two of the three Ryukyuan languages included in this study, mirroring recent developments in Standard Japanese (Uchikawa & Boynton 1987; Kuriki et al. 2017).

However, the modal responses from the naming task revealed fewer salient colour categories in the Ryukyuan varieties overall, both when compared to our Standard Japanese data and previously reported Japanese data (e.g., Kuriki et al. 2017). Moreover, the boundaries of newer categories—i.e., those not reported in Kusakabe (1964), such as 'green' and 'orange'—were less pronounced, with lower consensus between speakers shown by the absence of clear modal responses for chips in these parts of the colour spectrum. These findings likely represent variation between individual speakers resulting from ongoing changes. Individual members of a society can be at different stages of the evolutionary sequence of colour lexicons (MacLaury 1991). Ryukyuan speakers can have different levels of exposure to Japanese and English colour terms, which in turn influences their use, resulting in the variation we found in our data.

While it can be difficult to pinpoint the exact sources of specific new colour terms, there are several factors that could help explain how they arise. First of all, if we assume that all speakers end up refining their division of the colour spectrum as technology and other cultural factors encourage this over time, then all languages—including Ryukyuan—will eventually do so by themselves. Given the extensive dyeing practices in the Ryukyu Islands, one could have expected the languages spoken there to have more extensive colour lexicons earlier, so perhaps languages are only now catching up to the needs of the material culture. The fact that Ryukyuan speakers used several source-based terms specific to their local environment—and that sometimes referred to materials used in dyeing, such as *fumbutu* 'tree'—indicates that the languages were indeed already expanding their colour lexicon, doing so independently to at least some degree. An interesting side note is the fact that Ryukyuan language areas did not group together in the same way for each of the colour cognate sets, showing that language-specific developments may affect older colour terms in different ways, which result in separate trajectories of semantic change for colour terms across a language family.

Secondly, the fact that there are no monolingual speakers of Ryukyuan today also needs to be considered. Bilingualism in itself has been shown to affect the stability of both the boundaries and foci of colour categories, with bilinguals showing more variation (Caskey-Sirmons & Hickerson 1977; Jameson & Alvarado 2007). In our case, speakers are bilingual in a standardised language with institutionalised education and an endangered minority language that was for the longest time seen as a dialect of Japanese (Heinrich 2005), and it is currently not well understood how bilingualism, language loss, and semantic change interact in such situations. Moreover, there might be effects that arise as a result of contact through specific language learning. Colour words are commonly part of primary school instruction and the words *aka* 'red', *ao* 'blue', *shiro* 'white', *kuro* 'black', *kiiro* 'yellow', *midori* 'green' and *chairo* 'brown', as well as their kanji characters, are learned by the third year of primary school. Such early exposure might be one of the reasons why many speakers use the Japanese form *midori* in the naming task.

Third, there could be a role for contact with foreign material culture—rather than language itself—that changes colour naming systems, as the cultural environment has been shown to alter colour naming patterns. A comparison of Japanese speakers in Germany, Tokyo and Yamagata Prefecture found that those living in Germany showed the strongest influence of Western culture. More importantly, speakers living in Tokyo—where there is relatively more exposure to Western culture—showed more influence than those living in Yamagata—where there is less exposure to Western culture (e.g., Iijima, Wenning & Zollinger; Zollinger 1988). These results mainly pertained to the categorisation of colour, but contact with Western culture has also led to introduction of new terms. While this was not the case in our naming task data, several languages in the World Color Survey use source-based colour terms that are based on items introduced after Western influence, such as *coffee* and *chocolate*.

Finally, another dimension that applies to endangered languages such as Ryukyuan is the social value of retaining tradition. As described above, one Miyako speaker insisted on the use of only four colour terms, even though they knew about the additional terms that other speakers were using. Several other speakers commented on the traditionally smaller colour lexicon of Ryukyuan compared to Japanese. Similar attitudes towards the conservation of language and culture have been shown to lead to the maintenance of "archaic" colour categories in e.g., Navenchauc Tzotzil (see MacLaury 1991). Such attitudes can favour an older colour system and increase stability, which is an important factor to consider in the broader context of the notion of a constantly evolving and expanding colour lexicon (cf. Berlin & Kay 1969).

4.5 Conclusion

To conclude, the interaction between standard and minority languages, and the modern backdrop of globalisation, has led to the use of new forms for new meanings by contemporary speakers of Ryukyuan (e.g., *midori* or *guriin* for 'green'). However, these new colour categories are not yet fully entrenched, indicating ongoing changes and suggests the establishment of form precedes the establishment of meaning. A direct comparison with data from several decades ago shows that the modern Ryukyuan colour systems are closer to Standard Japanese than they are to their respective historic pasts, pointing to the speed at which semantic changes can occur. At the same time, analysis of four colour cognate sets reveals that recent Ryukyuan colour systems resemble each other more than they resemble Standard Japanese, hinting at the apparent stability of language-specific developments in semantics. Together, these findings highlight the need for additional, contemporary colour naming data to compare to historical data in related languages. Only then can we better understand the evolution of colour terms and their meaning more broadly.

CHAPTER

5

5 Patterns of semantic variation differ across body parts: Evidence from the Japonic languages[1]

Abstract

Human conceptual structure is grounded in the body, so it can be surprising to find that parts of the body singled out for naming vary across languages. Previous research suggests that although diverse languages differ in their body part lexicon, closely related languages show less variability. However, this conclusion may be premature as it is only based on a single study of the Germanic languages. This paper investigates the body part lexicon across the Japonic languages through both a body part naming task (Study I) and a body colouring-in task (Study II). We found that lexical similarity for body part terminology is notably differentiated in Japonic, which is reflected in semantics too. Novel application of cluster analysis on naming data revealed a relatively flat hierarchical structure for parts of the face, whereas parts of the body were organised with deeper hierarchical structure. The colouring data revealed that, rather than clear differences between the face and the body, bounded parts show more stability than unbounded parts across both. This study demonstrates that there is not a single universal conceptualisation of the body as is often assumed, and that in-depth, multi-method explorations of under-studied languages are urgently required.

Keywords: body parts, semantics, semantic variation, Japonic, Ryukyuan

[1] This chapter is based on Huisman, J. L. A., van Hout, R. & Majid, A. Patterns of semantic variation differ across body parts: Evidence from the Japonic languages. Submitted to *Cognitive Linguistics*. Formatting follows the journal's guidelines. I certify that I performed data collection, analysis and writing of the manuscript, with feedback from the co-authors.

5.1 Introduction

According to embodied theories of meaning, humans use their bodies to conceptualise the world around them, making the body central to human cognition. For example, Lakoff and Johnson (1999, p.19) state that "What is important is that the peculiar nature of our bodies shapes our very possibilities for conceptualization and categorization". In many languages, body part terminology is the source domain for the grammar of space (Svorou 1993; Heine 1997), emotions (Enfield & Wierzbicka 2002; Kövecses 2003), as well as knowledge, reasoning, social interactions, and values (Kraska-Szlenk 2014). Given this, studying how the body itself is conceptualised across different languages is important to understand its broader use.

It goes without saying that all languages have terms for parts of the body. Some parts are considered so universal that they are included in basic vocabulary lists intended for translation (e.g., Swadesh 1952; Greenhill, Blust & Gray 2008; Tadmor 2009). These approaches, however, assume that the parts are the same across languages. For example, the inclusion of 'nose' in most major basic vocabulary lists assume it refers to the exact same part of face across languages, but this is not always the case (Tarascan; Andersen 1978). Translation using putative "basic concepts" can provide a first view into a language's lexicon, but it can be misleading for establishing the exact referential meaning of a term (Majid 2019).

Nevertheless, early cross-linguistic studies on the semantics of body part terminology proposed several concepts as universals. Of these, 'the body' as a whole was considered the starting point for a hierarchically structured lexicon, in which each subsequent level consisted of 'parts of' the previous level (Brown 1976; Andersen 1978; see also Wierzbicka 2007). Other proposed universals include 'head' and 'hand' (Brown 1976; Andersen 1978; Wierzbicka 2007), as well as several parts of the face—'eyes', 'nose' and 'mouth' (Andersen 1978; Wierzbicka 2007)—which has its own dedicated neural circuitry (Kanwisher, McDermott & Chun 1997).

However, recent work in semantic typology provides reason to question the equivalence of body part terms across languages. A series of studies of unrelated languages found the body part lexicon is not universally organised in a hierarchical fashion (see contributions in Majid, Enfield & van Staden 2006) contrary to previous claims (Brown 1976; Andersen 1978). In addition, the granularity of distinctions made for body parts varied across languages, with some languages have a general term encompassing the arms and legs (Lavukaleve; Terrill 2006), and others lack a generic term but distinguish upper arm, lower arm, upper leg, and lower leg (Jahai; Burenhult 2006).

In another cross-linguistic study of body part categories, Majid and van Staden (2015) asked speakers of Japanese, Dutch, and Indonesian to colour-in body parts. Although Japanese and Dutch both have terms that would be translation equivalents to 'arm', their extensional meaning was not equivalent—Japanese speakers never included the 'hand' when colouring-in *ude* 'arm', but Dutch participants did colour-in hand when prompted with *arm*. Interestingly, the same was not the case for 'leg', showing that parallelism between upper and lower limb is not a given (contra Brown 1976; Andersen 1978). At the same time, other parts showed far less variation across languages (e.g., 'eye', 'nose', 'forehead'), suggesting that some body parts may indeed be more universal than others.

The cross-linguistic work to date samples a diverse array of languages, leaving variation across related languages understudied. However, a recent study of semantic variation of body parts within the Germanic language family used a body part naming task, and found body parts show considerable semantic (meaning) similarity in closely related languages (Majid, Jordan & Dunn 2015), although differences could be found for specific body parts (see, e.g., Levisen 2015). The current study aims to contribute to our understanding of semantic variation of body part vocabulary in related languages by studying the Japonic language family, which was chosen because of the considerable lexical (form) variation found for body parts in The Linguistic Atlas of Japan (National Institute for Japanese Language and Linguistics 1968).

The Japonic language family is spoken across the Japanese archipelago and consists of two major branches. The first branch, Japanese, comprises the varieties spoken on the main islands. The Japanese branch is generally subdivided into Eastern, Western, Kyushu, and Hachijo Japanese (Shibatani 1990). The second major branch, Ryukyuan, includes the varieties spoken across the smaller islands in the south. Ryukyuan is generally subdivided into Northern Ryukyuan (Amami and Okinawa varieties) and Southern Ryukyuan (Miyako, Yaeyama and Yonaguni) (Pellard 2015; Huisman, Majid & van Hout 2019). Previous work on the semantics of body part terms in Japanese have looked at diachronic change (e.g., in terms for 'head'—Miyaji 1973), the role of body parts in conceptualising emotion (Hasada 2002), and the extensional range of body part terms (Majid & van Staden 2015).

Rather than assume semantic equivalence, as in dictionary-based approaches, we collected new primary data from native speakers through two standardised tasks. In Study I, we collected body part naming data from speakers of six languages: Standard Japanese and Tohoku Japanese (Eastern Japanese), which are among the most divergent mainland dialects (Huisman, Majid & van Hout 2019), plus four Ryukyuan languages, Amami, Okinawa, Miyako, and Yaeyama. Importantly, while Standard Japanese body part vocabulary has been the subject of experimental study previously (Majid & van Staden 2015), the Ryukyuan languages—which are lesser-described and endangered (Moseley 2010)—have not. This paper provides the first in-depth quantitative study of body parts within the Japonic language family.

Following Majid, Jordan and Dunn's (2015) study of Germanic languages, we collected body part naming data through a stimulus-based elicitation task in which speakers were asked to name various parts of the body. A standardised set of non-linguistic stimuli provides a frame of reference against which similarities and differences across languages and their varieties can be systematically compared. In Study II, we collected body part colouring-in data from five languages (Tohoku, Amami, Okinawa, Miyako, and Yaeyama), in which speakers are asked to colour in the range of various body part terms on a line drawing of the human body (see van Staden & Majid 2006). In combining both methods, we look for converging evidence as to the semantics of body part terms.

We first aimed to establish how similar body part vocabulary is within the Japonic language family. We predicted that body part vocabulary would be more similar for languages that are more closely related, i.e., that variation in body part vocabulary reflects the overall geographical differences between the Japonic languages, and that speakers from the same language are more similar to each other. To address this hypothesis, we analysed the body part naming data from Study I, examining both variation between languages, as well as variation across speakers.

Next, we investigated the semantics of body parts by applying cluster analysis for the first time to body part naming data. We separately examined parts of the face and parts of the body, following the observation above that there may be less variation for the face than body. Since the parts of the face are generally well delimited, in contrast to the diverse principles along which the body can be divided (see Majid, Enfield & van Staden 2006), we expected less variation in the extension of terms for parts of the face than body. To test this, we considered both semantic distinctions and lexical (form) variation (Study I).

In addition, we also investigated the semantic extension of face and body part terms in Study II. While a naming task provides us with some information about the extension of the terms elicited, competition between terms at different levels of granularity can obscure the complete picture. So, we collected additional data by asking speakers to colour-in face and body parts in a drawing of the human body, providing us with a common frame through which we can directly compare the extensions of specific terms.

5.2 Study I: Body part naming task

5.2.1 Methods

5.2.1.1 Languages and speakers

Data for the body part naming task was collected from 66 speakers in six language areas (two Japanese and four Ryukyuan) during four fieldtrips conducted between 2017 and 2019. For all areas, data was collected from multiple localities, i.e. in multiple dialects. Apart from Tokyo Japanese, which serves as the de facto national standard, there is no standardised variety of Tohoku Japanese or the Ryukyuan languages (Heinrich, Miyara & Shimoji 2015). As such, we will refer to Tokyo Japanese as "Standard Japanese", and use the term "language area" for the other five varieties for the remainder of this paper, e.g., the Amami language area. Given the endangered status of Ryukyuan, the data was collected from elderly native speakers, some of whom had little experience in performing standardised linguistic elicitation tasks, so some interview sessions were conducted with multiple speakers simultaneously. To minimise potential confounds, all analyses were conducted on sessions rather than speakers—see also Section 5.2.4 *Coding*. Table 5.1 shows information about the number of speakers and sessions per language area.

Table 5.1

Speaker and session information for body part naming task.

Japanese	
Tokyo	8 speakers (6 female) in 8 sessions
Tohoku	8 speakers (3 female) in 8 sessions
Ryukyuan	
Amami	17 speakers (7 female) in 12 sessions
Okinawa	7 speakers (3 female) in 4 sessions
Miyako	18 speakers (11 female) in 16 sessions
Yaeyama	8 speakers (5 female) in 8 sessions

5.2.1.2 Materials and procedure

Data was collected using 52 line drawings of the human body: 39 showed the full human body from the front and back, and 13 depicted the head and face (see Figure 5.1). In each drawing, a red dot was placed somewhere on the body, and participants had to say where the dot was located. The drawings were presented in three blocks: a first block of 39 drawings of the full body, a second block of 10 drawings of the head with the mouth closed, and a third block of 3 drawings of the head with the mouth opened. The presentation order of the blocks was kept constants across speakers, but stimuli were present in two fixed orders—one the reverse of the other—within each block, which was randomised across speakers. This stimulus set was newly created for this project, but was inspired by earlier stimuli (Majid, Jordan & Dunn 2011). Figure 5.1 consolidates the separate drawings into a single image, with stimulus order indicated.

Speakers saw the drawings one by one on a tablet, i.e., they saw one red dot at a time, and were asked to name the body part marked by the red dot, by answering the question "What is the place of the red dot called?". Speakers could give responses of any length. All sessions were minimally audio-recorded, and sometimes video-recorded for later transcription. Speakers gave informed consent before participating. Data collection was approved by the Ethics Assessment Committee of the Centre for Language Studies at Radboud University.

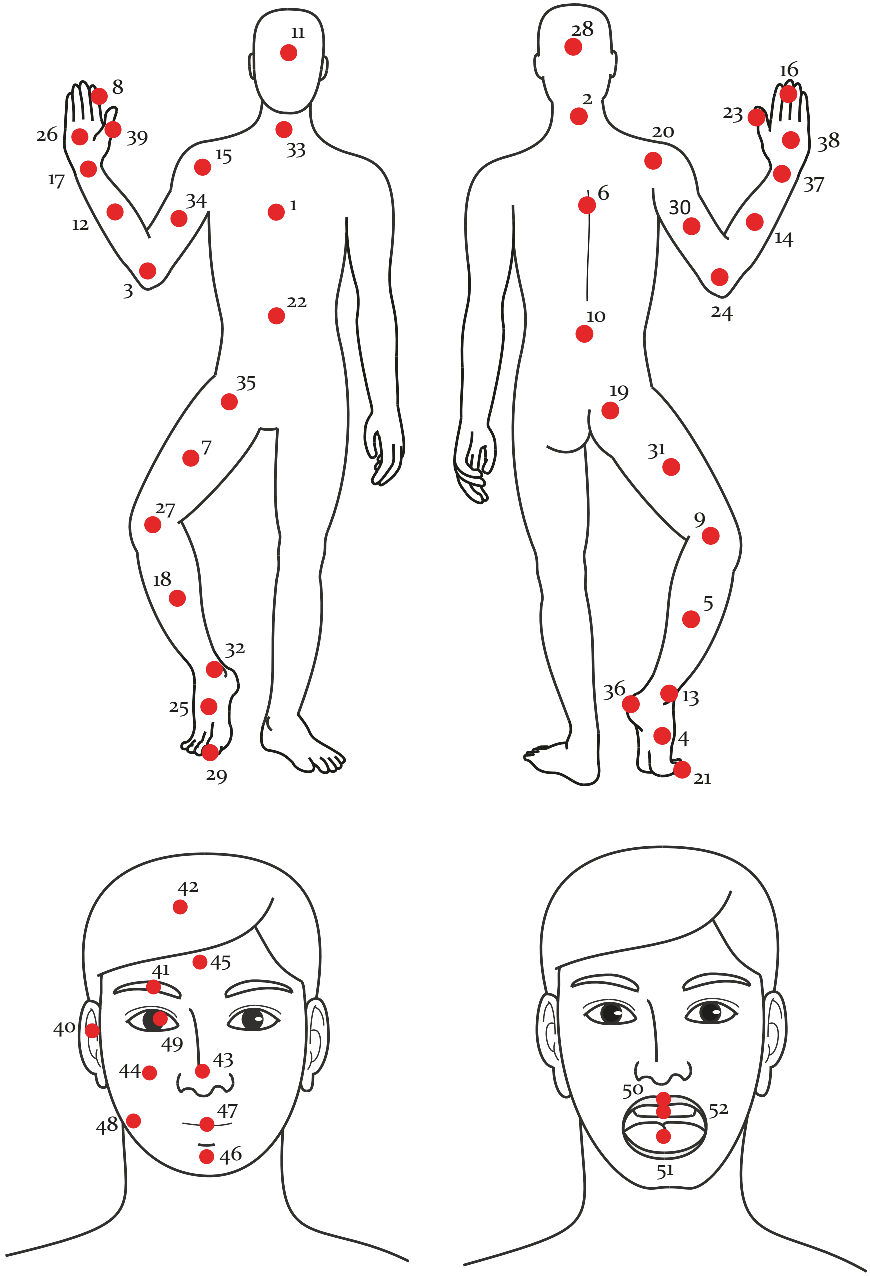


Figure 5.1. Stimuli used in the naming task. In the elicitation task, only one dot appeared in each figure, and participants had to identify where the dot was located.

5.2.1.3 *Coding*

We extracted, per session, the full response(s) for each stimulus, which could include multiple responses. In sessions with multiple speakers, we coded all unique responses, i.e., if two speakers named the stimulus differently, we coded both responses. Next, we coded main response(s) using the following coding scheme: main responses were monolexemic responses (e.g., English *arm*) and polylexemic responses that are conventionalised and untransparent (e.g., English *forearm*). Locatives such as left/right, front/back, upper/lower were excluded when they were compositional and transparent (e.g., the meaning of English *upper arm* can be derived from the elements *upper* and *arm*, so only *arm* was coded). If the meaning of combined elements referred to a different body part which could not be reduced to a single element (e.g., English *between the eyes*), the full response was coded. There were not many such cases in the Japonic data: 2 in Standard Japanese and 1 in the Amami language area, and none in the other four language areas.

5.2.2 Results

5.2.2.1 Regional and individual variation of body part vocabulary in Japonic

The first aim of the current study was to establish the variation in body part lexicons of related languages. To assess this, we ask whether body part vocabulary is more similar for languages that are more closely related. We investigated variation by comparing individual sessions to each other, hypothesising that sessions from the same language area would be more similar than sessions from another language area.

We created a session-by-stimulus matrix in which we coded the main responses for cognacy—i.e. whether the main responses have a common etymological origin. For example, stimulus 22 (‘belly’), received the monolexemic responses *o-naka* (Standard Japanese), *hara* (Tohoku) and *wata~bata* (Ryukyuan). These were coded as *belly-A*, *belly-B* and *belly-C*, respectively because none of these are cognate. For polylexemic responses, we coded the cognacy of individual elements. As the focus of the current study is body part vocabulary, only content elements were coded for cognacy, and so the case marking particles and the copula were excluded in the coding. So, for example, the elements in the Standard Japanese response *aɕi=no oja.jubi* [leg=GEN parent.digit] to stimulus 29 ('big toe, front') were coded as *leg-A*, *parent-A* and *digit-A*, whereas the elements in the Yaeyama response *paɴ=nu bu:.jubi* [leg=GEN large.digit] were coded as *leg-B*, *large-A* and *digit-A*. In cases where there were multiple, (partially) overlapping responses, each element was only counted once.

Next, we used the session-by-stimulus matrix in a series of pairwise comparisons to assess the similarity between sessions. The analysis was conducted in GABMAP (Nerbonne et al. 2011), an online tool for dialectometry. For each pair of sessions, cognacy overlap was calculated using the Jaccard Index (*J*; Jaccard 1901, 1912) for each stimulus separately. Thereafter, the mean cognacy overlap over all stimuli was taken as a general measure of similarity between sessions. Doing this for all sessions creates a session-by-session similarity matrix that we analysed with multidimensional scaling, as well as cluster analysis using Ward's method (Ward 1963), both in base R (`cmdscale` and `hclust` functions, R Core Team 2019).

Figure 5.2 shows the results of the multidimensional scaling analysis, and reveals Tohoku and Standard Japanese sessions are much closer to each other than the Ryukyuan sessions—average similarity between the mainland Japanese sessions was $M_J = .65$ ($SD = .09$), whereas average similarity between the Ryukyuan sessions was $M_J = .48$ (SD = .10). The Amami sessions show a wide spread in the figure, suggesting more variation in that language area in particular. Some Amami sessions show similarities to both the Japanese mainland and other Ryukyuan sessions. Several Yaeyama sessions also fall in between the Northern and Southern Ryukyuan sessions. Miyako sessions appear to be the most divergent.

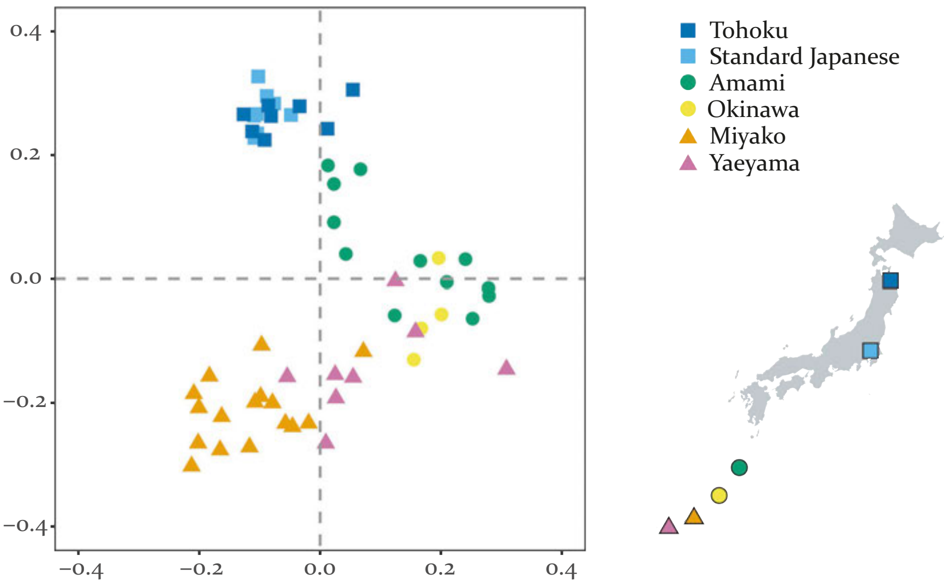


Figure 5.2. Multidimensional scaling analysis of body part naming data, left. Each symbol represents a session. Plotting the second dimension on the x-axis and the first dimension on the y-axis roughly recreates the geographic layout of the language area, shown bottom right.

A cluster analysis (Figure 5.3) of the same data recapitulated the results of the multidimensional scaling for the most part. Mainland sessions (Japanese and Tohoku) were grouped together, and most Ryukyuan sessions (Amami, Okinawa, Miyako and Yaeyama) grouped together, roughly corresponding to how the results of the multidimensional scaling analysis (Figure 5.2). In addition, the cluster analysis shed further light on distinctions within subgroups. Within the mainland cluster, all Standard Japanese and Tohoku sessions were grouped separately, and within the Tohoku cluster, two subgroups were distinguished that correspond largely to the traditional division between Nambu dialects (Hachinohe) and Tsugaru dialects (Aomori and Hirosaki). Consistent with the multidimensional scaling analysis, four Amami sessions were grouped together with the mainland clusters.

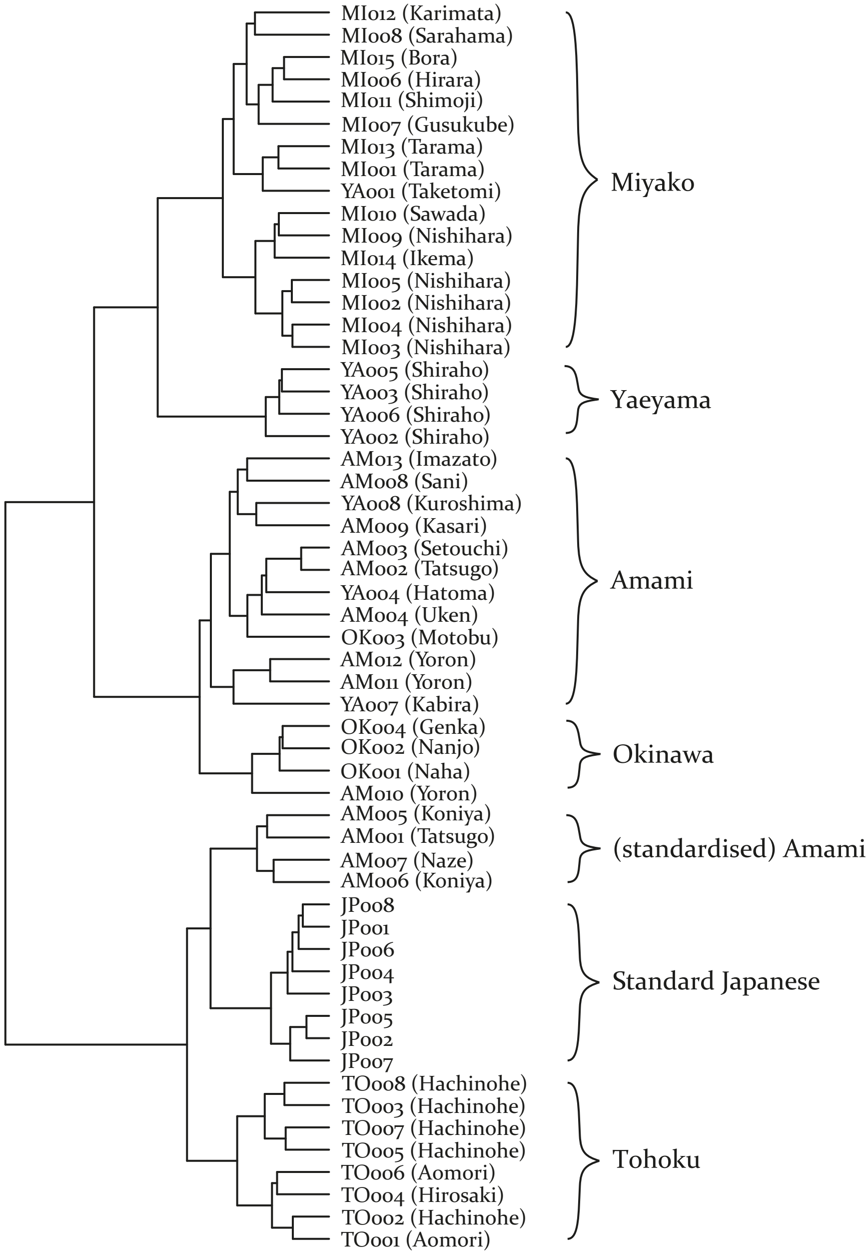


Figure 5.3. Cluster analysis of body part naming data by session recapitulates language areas.

Closer inspection revealed these speakers to be some of the youngest in the sample, and they used several Standard Japanese terms not generally used in the Ryukyuan languages (e.g., *çiza* for 'knee' rather than *tsubusɨ*, and *kakato* for 'heel' rather than *ado*). Within Ryukyuan, there was a large cluster of sessions from the Northern Ryukyuan area, split into an Amami subgroup and a distinct Okinawa subgroup. Finally, there was a large cluster of only Southern Ryukyuan sessions, split into a subgroup of all Miyako sessions plus a single Yaeyama session, and another subgroup comprising the remaining Yaeyama sessions—all of the Shiraho variety.

Overall, the multidimensional scaling and cluster analyses show body part naming data largely reflect the geographical differences between the Japonic languages (see, e.g., Shibatani 1990, Pellard 2015), with individual sessions mirroring larger patterns of variation. The Japanese mainland varieties resemble each other more than the Ryukyuan varieties (cf. Huisman, Majid & van Hout 2019)—we return to the high variability of Ryukyuan in the General Discussion.

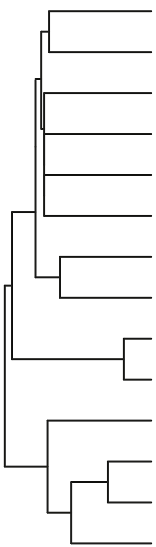
5.2.2.2 Semantic distinctions in parts of the face and parts of the body

We then examined the organization of the lexicon for parts of the face and parts of the body for all Japonic languages as a whole. To do this, for each language area, we create a stimulus-by-cognate frequency matrix where, for each stimulus, we coded how many sessions a stimulus was described by each cognate term in the naming task. To uncover the body part categories, and the overall structure of the lexicon, we used these stimulus-by-cognate matrices to determine similarities between stimuli. We calculated, for each pair of stimuli, the cosine similarity based on naming responses—in R; `cosine` function in the *lsa* package (Wild 2015). Doing this for all stimulus pairs creates a stimulus-by-stimulus similarity matrix that encodes, for each language area, the body part categories and the relationship between them.

To generate an overall frame of reference that covers the shared structure of the body part lexicon across the Japonic language family as a whole, we averaged the six stimulus-by-stimulus similarity matrices to create an overall matrix, on which we performed cluster analysis using Ward's method in base R (`hclust` function). We conducted separate cluster analyses for the face and the body. These analyses showed that while the face has a relatively flat hierarchy (Figure 5.4), the body parts exhibit deeper hierarchy, with clusters embedded within higher-order clusters (Figure 5.5).

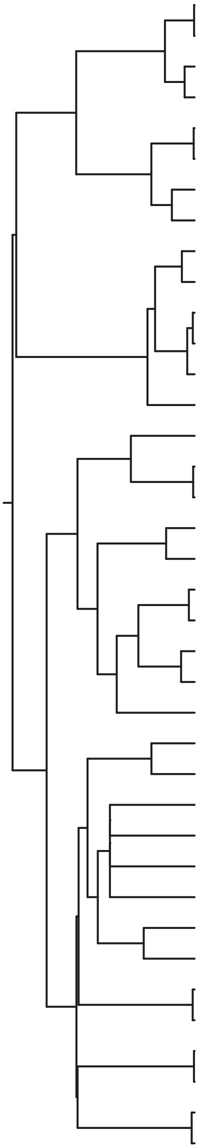
Speakers of all languages tended to distinctly name 'hair', 'forehead', 'tongue', 'tooth', 'nose', 'ear', 'eye', 'eyebrow and 'mouth'. There were a few exceptions to this. First, 'eye' and 'eyebrow' were grouped together because some Amami and Okinawa varieties use the lexeme for 'eye' in a polylexemic term for 'eyebrow'—as in English. Second, the terms 'mouth' and 'lips' show a closer relationship as many varieties use the lexeme for 'mouth' in a polylexemic term for 'lips'. Next, a subgroup comprising 'face', 'cheek, 'jaw' and 'chin' emerged, as some Ryukyuan varieties do not distinguish between 'face' and 'cheek', and Standard Japanese does not distinguish 'chin' from 'jaw'. In addition, while most non-standard varieties have a separate term for 'chin', not all speakers use it. Finally, the 'cheek' and 'jaw' stimuli received similar responses in some sessions, which could be due to the placement of the red dot, which led to varied interpretations of these stimuli across sessions.

In contrast to the high level of distinctness found for parts of the face, Figure 5.5 revealed a deeper hierarchical relationship between parts of the body. There were four major subgroups: the upper limb (arm and hand), the digits (fingers and toes), the lower limb (leg and foot), the torso (its various parts), as well as some joints as separate subgroups (shoulder, elbow, knee).



	Lexical similarity	Standard Japanese	Tohoku (Hachinohe)	Amami (Naze)	Okinawa (Naha)	Miyako (Nishihara)	Yaeyama (Shiraho)
hair	M_{cos} = .210	*kami*	*kami*	*kamatsɨ*	*karadʑi*	*aka*	*amadʑi*
forehead	M_{cos} = .254	*çitai*	*nandzugi*	*mɨttɕu*	*çitɕe:*	*futai*	*futeɴ*
tongue	M_{cos} = .689	*ɕita*	*bero*	*ɕitɕa*	*ɕiba*	*sɨta*	*ɕita*
tooth	M_{cos} = .845	*ha*	*ha*	*ha*	*ha:*	*pa:*	*futsɨ nu paɴ*
nose	M_{cos} = .996	*hana*	*hana*	*hana*	*hana*	*pana*	*pana*
ear	M_{cos} = .960	*mimi*	*mimi*	*miɴ*	*mimi*	*mim*	*misɨkurumiɴ*
eye	M_{cos} = .833	*me*	*managu*	*mɨ*	*mi:*	*mi:*	*miɴ*
eyebrow	M_{cos} = .708	*majuge*	*konoŋe*	*maju*	*mi:maju*	*maju*	*maju*
mouth	M_{cos} = .986	*kutɕi*	*kudzi*	*kutɕi*	*kutɕi*	*futsɨ*	*futsɨ*
lip	M_{cos} = .701	*kutɕibiru*	*kudzibiru*	*kutɕibiru*	*kutɕibiru*	*sɨba*	*supa*
chin	M_{cos} = .669	*ago*	*odoŋɛ*	*agu*	*utuge:*	*utugai*	*agu*
jaw	M_{cos} = .279	*ago*	*aŋo*	*agu*	*kakudʑi*	*kamagita*	*mutɕi*
cheek	M_{cos} = .361	*ho:*	*hoppeda*	*fu:*	*fu:dʑira*	*kamatsɨ*	*mutɕi*
face	M_{cos} = .317	*kao*	*tsura*	*tsɨra*	*tɕira*	*mipana*	*mutɕi*

Figure 5.4. Cluster analysis of the face stimuli, with average lexical similarity (cosine similarity), and the most commonly elicited cognate for each language variety.



	Lexical similarity	Standard Japanese	Tohoku
upper arm (front)	M_{cos} = .581	*ni no ude*	*u^{n}de*
upper arm (back)	M_{cos} = .615	*ni no ude*	*u^{n}de*
lower arm (front)	M_{cos} = .728	*ude*	*u^{n}de*
lower arm (back)	M_{cos} = .712	*ude*	*u^{n}de*
wrist (front)	M_{cos} = .968	*tekubi*	*teku^{m}bi*
wrist (back)	M_{cos} = .960	*tekubi*	*teku^{m}bi*
hand (palm)	M_{cos} = .785	*te no çira*	*te no sura*
hand (back)	M_{cos} = .774	*te no ko:*	*te no ko*
middle finger (back)	M_{cos} = .878	*jubi*	*ju^{m}bi*
index finger (front)	M_{cos} = .803	*çitosaɕijubi*	*ju^{m}bi*
thumb (front)	M_{cos} = .838	*ojajubi*	*ojaju^{m}bi*
thumb (back)	M_{cos} = .827	*ojajubi*	*ojaju^{m}bi*
big toe (front)	M_{cos} = .828	*aɕi no ojajubi*	*asu no ojaju^{m}bi*
little toe (back)	M_{cos} = .602	*tsumasaki*	*asu no ju^{m}bi*
hip joint	M_{cos} = .320	*aɕi no tsukene*	*asu no tsukene*
upper leg (front)	M_{cos} = .764	*futomomo*	*momota*
upper leg (back)	M_{cos} = .772	*futomomo*	*momota*
lower leg (front)	M_{cos} = .555	*sune*	*sune*
lower leg (back)	M_{cos} = .489	*fukurahagi*	*fukurahaŋi*
ankle (inner)	M_{cos} = .402	*aɕikubi*	*kuru^{m}busu*
ankle (outer)	M_{cos} = .388	*aɕikubi*	*asuku^{m}bi*
foot (instep)	M_{cos} = .356	*aɕi no ko:*	*asu no ko*
foot (sole)	M_{cos} = .268	*aɕi no ura*	*asu*
heel	M_{cos} = .348	*kakato*	*agudo*
throat	M_{cos} = .643	*kubi / nodo*	*no^{n}do*
neck (back)	M_{cos} = .528	*kubi*	*bonnogo*
head (back)	M_{cos} = .090	*ko:to:bu*	*usuroko^{n}do*
chest	M_{cos} = .899	*mune*	*mune*
belly ; navel	M_{cos} = .476	*onaka ; heso*	*hara ; hettɕo*
buttocks	M_{cos} = .175	*oɕiri*	*kettsu*
lower back	M_{cos} = .803	*koɕi*	*kosu*
upper back	M_{cos} = .450	*senaka*	*senaga*
shoulder (front)	M_{cos} = .961	*kata*	*kada*
shoulder (back)	M_{cos} = .935	*kata*	*kada*
elbow (front)	M_{cos} = .924	*çi^{d}ʑi*	*su^{n}dzu*
elbow (back)	M_{cos} = .952	*çi^{d}ʑi*	*su^{n}dzu*
knee (front)	M_{cos} = .462	*çi^{d}za*	*su^{n}dzaka^{m}bu*
knee (back)	M_{cos} = .435	*çi^{d}za*	*su^{n}dzaka^{m}bu*

Figure 5.5. Cluster analysis of the body stimuli, with average lexical similarity (cosine similarity), and the most commonly elicited cognate for each language variety.

Amami	Okinawa	Miyako	Yaeyama
ʔudɨ	*ke:na*	*udi*	*(ti: nu) udi*
ʔudɨ	*ke:na*	*udi*	*(ti: nu) udi*
tɨ	*ti:*	*udi*	*ti: / udi*
tɨ	*ti:*	*udi*	*ti: / udi*
tɨɴkubi	*ti: nu kubi*	*ti:fugᶻɨ*	*ti: nu kubi*
tɨɴkubi	*ti: nu kubi*	*ti:fugᶻɨ*	*ti: nu kubi*
tɨɴçira	*ti: nu wata*	*tibɨdza*	*ti: nu pisa*
tɨɴko:	*ti: nu ʔura*	*ti:*	*ti: nu pisa*
ʔibɨ	*ʔi:bi*	*uibi*	*ti: nu:bi*
ʔibɨ	*saɕiʔi:bi*	*uibi*	*ti: nu:bi*
ʔujaʔibɨ	*ʔufuʔi:bi*	*upuuibi*	*bu:jubi*
ʔujaʔibɨ	*ʔufuʔi:bi*	*upuuibi*	*ti: nu bu:jubi*
ʔujaʔibɨ	*çisa nu ʔi:bi*	*pagᶻɨ nu upuuibi*	*paɴ nu bu:jubi*
hagiɴʔibɨ	*çisa nu ʔi:bi*	*pagᶻɨ nu uibi*	*paɴ nu:bi*
hagiɴtsukene	*mata*	*mumuni*	*paɴ nu ni:*
mumu	*mumu*	*mumuni*	*mumu*
mumu	*mumu*	*mumuni*	*mumu*
sunɨ	*kuɴda*	*karasɨni*	*paɴ*
hagi	*kuɴda*	*kuvva*	*paɴ nu taru*
haginkubi	*çisakubi*	*amambuni*	*paɴ (nu katu)*
kurubuɕi	*çisakubi*	*amambuni*	*paɴ (nu *)*
hagiɴko:	*çisa*	*ssabidʑa*	*paɴ nu pisa*
hagi	*çisawata*	*pˢɨsa*	*paɴ nu pisa*
ʔado	*ʔadu*	*adu*	*paɴ nu adu*
nudɨ	*nu:di:*	*nudu*	*nudu*
kubi	*kubi*	*nubui*	*nubuɕiɴ*
kamatɕi	*ʔuɕi nu kubu:*	*ussɨ*	*ussoɴ*
munɨ	*ɴni*	*mmivtsɨ*	*ɴni*
wata ; fuɕu	*wata ; fusu*	*bata ; m:bu*	*bata ; putsu*
mari	*tɕibi*	*tɕibi*	*ɕipi*
kusɨ	*kuɕi*	*kusɨ*	*buɕa*
kusɨ	*nagani*	*kusammi*	*futɕa*
kata	*kata*	*katamusɨ*	*kata*
kata	*kata*	*katamusɨ*	*kata*
çidzɨ	*çidʑi*	*pidzɨ*	*p̥itɕi*
çidzɨ	*çidʑi*	*pidzɨ*	*p̥itɕi*
tsɨbusɨ	*tɕiɴɕi*	*tsɨgusɨ*	*sɨpuɕiɴ*
tsɨbusɨ	*tɕiɴɕi*	*tsɨgusɨ*	*sɨpuɕiɴ*

Within the upper limb subgroup, the upper and lower arm were distinct from the hand and wrist. For the hand, speakers from all varieties except those from the Yaeyama language area distinguished the palm and the back of the hand—albeit through polylexemic responses for both. Distribution of the modal responses suggest the upper limb parts are lexicalised differently across language areas (see Figure 5.6). Responses belong to the *te* cognate set responses were elicited for all parts of the upper limb across the Ryukyuan varieties, whereas they were restricted to the hand in the mainland varieties. In contrast, responses belong to the *ude* cognate set were elicited for the upper and lower arm (lower arm only for Okinawa), but never the hand. Yaeyama speakers (of the Shiraho variety in particular) were the only ones to use the genitive construction *ti:=nu udi* when describing the upper or lower arm. Finally, response belonging to the *kaina* cognate set (elicited in Tohoku, Amami, Okinawa and Miyako) also showed variation: they only appeared for the upper arm in Amami and Okinawa, but for both the upper and lower arm in Tohoku and Miyako where the term was used interchangeably with *ude*-type responses. For Tohoku, this appears to be individual speaker variation, but for Miyako, it is the case that some varieties only use *udi*, whereas others only use *kaina*.

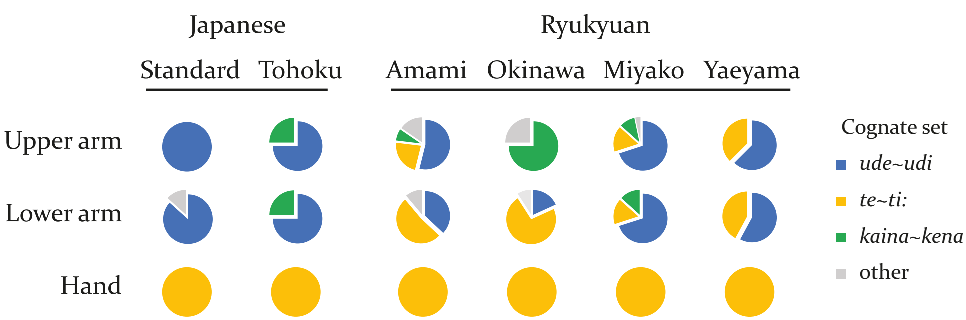


Figure 5.6. Main responses for the upper limb across the Japonic languages.

Moving to the digits, all varieties used cognate forms for fingers and toes, with the toes more likely to be named with the genitive construction *limb=GEN digit*. In addition, speakers of the mainland varieties also used variants of the compound *tsuma.saki* to describe toes, but never fingers. Finally, there was some variation between speakers in whether they named specific digits distinctly, with the thumbs and big toe more likely to be named with a dedicated polylexemic term.

In general, parts of the lower limb were named with specific terms, suggesting finer-grained naming for the lower limb than upper limb. The core set of cognates was fairly similar across the language areas (Figure 5.5), and often occurred with language-specific affixes, e.g., *momo.ta* 'upper leg' in Tohoku (cf. *momo* 'upper leg' in Standard Japanese) or *kara.suni* 'front of the lower leg' in Miyako (cf. *sune* 'front of the lower leg' in Standard Japanese). Three cognate sets meaning 'lower limb' (*aɕi~asɯ̈*; *hagi~pag*^{z}*ɨ~paɴ*; and *çisa*) were elicited either monolexemically or as part of a compound or genitive construction (see Figure 5.7).

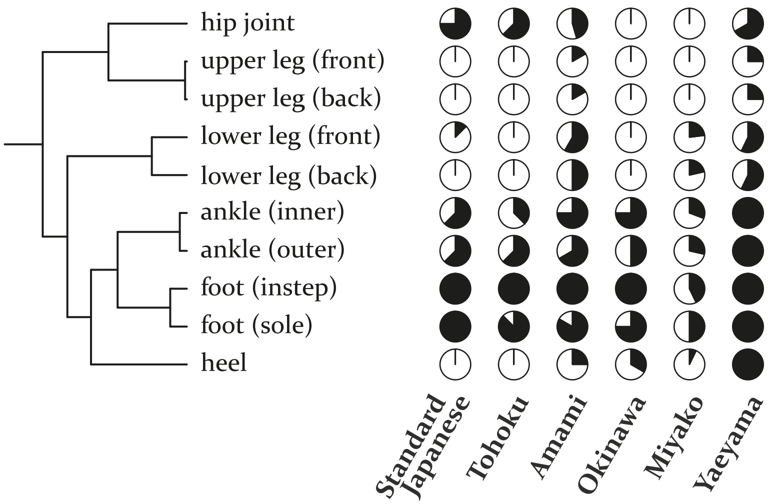


Figure 5.7. Proportion of sessions in which a term that refers to the lower limb as a whole was used in responses describing its subparts (in black).

Ryukyuan speakers were more likely to use a term referring to the entire lower limb term when naming its different subparts—either in a monolexemic responses or as *lower limb=GEN [part]*. Yaeyama speakers in particular tended to use this genitive construction, mirroring naming for the upper limbs. The use of a 'leg' term for thigh and hip suggests that speakers extend its meaning across the entire lower limb. Only Miyako and Yaeyama speakers seem to distinguish between 'leg' and 'foot', both using *pisa* for the foot.

Within the torso subgroup, 'head', 'chest', 'belly', 'navel' and 'buttocks' were all distinctly named across the six language areas. For some parts there was considerable lexical variation (e.g., Standard Japanese *atama*; Tohoku *adama*; Amami *kamatsɨ*; Okinawa *tɕiburu*; Miyako *kanamai*; and Yaeyama *amasi̥kuru* for 'head') whereas for others there was less (e.g., for Standard Japanese *heso*; Tohoku *hettɕo*; Amami *fuɕu*; Okinawa *fusu*; Miyako *m:bu*; and Yaeyama *putsõ:* for 'navel'). The front of neck ('throat') was often named separately, but some speakers named it using the same term they also used for the back of the neck. The upper and lower back were distinguished in most varieties (except most Amami varieties), which suggests a clean subdivision between the two, rather than a part-whole relationship.

Finally, three joints turned up as separate branches in the cluster analysis, two of which showed high form similarity across the entire Japonic language family. Only one cognate is used to describe the shoulder across all sessions (e.g., Standard Japanese *kata*; Tohoku *kada*; Amami *kata*; Okinawa *kata*; Miyako *katamusɨ*; Yaeyama *kḁta*). The elbow is also named with a single cognate set across all sessions (e.g., Standard Japanese *çi*^{d}*ʑi*; Tohoku *çï*^{n}*dzï*; Amami *çidzɨ*; Okinawa *çidʑi*; Miyako *pidzɨ*; Yaeyama *pi̥tɕi*). There were two cognate sets for the knee, distributed between the mainland varieties (Standard Japanese *çi*^{d}*za*; Tohoku *çï*^{n}*dʑaka*^{m}*bü*), and the Ryukyuan varieties (Amami *tibuɕi̥*; Okinawa *tɕiɴɕi*; Miyako *tsɨgusɨ*; Yaeyama *si̥puɕiɴ*).

Overall, our results suggest that parts of the face showed a relatively flat hierarchy with little variation between languages, whereas parts of the body were organised in a more hierarchical structure with relatively more variation between languages.

5.2.2.3 Lexical similarity in terms for face and body parts

The previous section showed differences in the hierarchical structure of face and body parts in the Japonic language. In addition, while some parts are described with cognate terms, others show more variation in forms. To investigate whether face parts show more lexical similarity than body parts, we measured lexical similarity—i.e., degree of overlap between languages in terms used for each part—by calculating the cosine similarities between all language pairs for each stimulus, using the `cosine` function in the *lsa* package (Wild 2015). If two languages use the same cognates to name a stimulus, the cosine similarity is 1, and if they use different cognates, 0 (see Table 5.2).

There was no significant difference between the lexical similarity for parts the face (M_{cos} = .63, *SD* = .29) and parts of the body (M_{cos} = .64, *SD* = .24), $t(50) = 0.12$, $p = .9$, contrary to the prediction that face parts may be more similar.

Table 5.2

Lexical similarity (the degree of overlap between languages in terms used; M_{cos}) for face and body parts, calculated as the mean cosine similarity between all language pairs for each stimulus.

Face		**Body**			
	M_{cos}		M_{cos}		M_{cos}
nose	.996	wrist (front)	.968	neck (front)	.643
mouth	.986	shoulder (front)	.961	upper arm (back)	.615
ear	.960	wrist (back)	.960	little toe	.602
tooth	.845	elbow (back)	.952	upper arm (front)	.581
eye	.833	shoulder (back)	.935	lower leg (front)	.555
eyebrow	.708	elbow (front)	.924	neck (back)	.528
lip	.701	chest	.899	lower leg (back)	.489
tongue	.689	middle finger	.878	belly	.476
chin	.669	thumb (front)	.838	knee (front)	.462
cheek	.361	big toe	.828	upper back	.450
face	.317	thumb (back)	.827	knee (back)	.435
jaw	.279	lower back	.803	ankle (inner)	.402
forehead	.254	index finger	.803	ankle (outer)	.388
hair	.210	hand (palm)	.785	foot (instep)	.356
		hand (back)	.774	heel	.348
		thigh (back)	.772	hip joint	.320
		thigh (front)	.764	foot (sole)	.268
		lower arm (front)	.728	buttocks	.175
		lower arm (back)	.712	head (back)	.090

5.2.3 Summary

We found body part vocabulary between languages was more similar the more closely related and physically close the languages. The multidimensional scaling analysis largely captures the geographic layout of the Japonic language family, except for Miyako. The distinctness of Miyako within the body part domain is likely the result of a combination of lexical innovations (e.g., *kanamai* for 'head', *mipana* for 'face'), and semantic innovations (e.g., a distinct 'foot' category).

We also compared parts of the face with parts of the body, addressing both the semantic organisation of the lexicon and lexical similarity between forms. We found that parts of the face had a relatively flat hierarchy, whereas parts of the body were organised in a more hierarchical structure. The modal responses also show that face parts were generally named using monolexemic terms, whereas polylexemic terms were more common for body parts. The use of polylexemic terms partially contributed to the hierarchical structure of body parts, especially in cases where genitive constructions (e.g., limb=GEN digit) were used. In addition, some speakers chose to be more specific in naming body parts (e.g., 'thigh' instead of 'leg'), perhaps because of pragmatic pre-emption.

We uncovered a (covert) hierarchical structure for body parts by applying cluster analysis to naming data, rather than asking people to make explicit linguistic judgements (e.g., "Is the hand part of the arm?") which has been how linguists have previously endeavoured to establish hierarchies. To our knowledge, this is the first time cluster analysis has been applied in this way (cf. Crowe & Prescott 2003), and we believe this opens interesting possibilities for future cross-linguistic work on body part categorisation. We return to this in the General Discussion.

Finally, lexical similarity did not differ between parts of the face and body, despite the earlier suggestion that face parts may be more stable. Within parts of the face, we see highest lexical similarity for bounded parts (i.e., nose, mouth, ear and eye), whereas for parts of the body, lexical similarity was generally highest for parts of the upper limb.

5.3 Study II: Colouring-in of body parts

Study I provided information about how a part is named. However, the same part can be named at different levels of granularity and the nature of the naming task results in a choice between choosing specific terms (e.g., 'thigh') or generic terms (e.g., 'leg'), limiting the ability to establish the exact extension of body part terms. Therefore, Study II asked speakers to colour in body parts on a drawing of the body, to provide further information about the extension of terms.

5.3.1 Methods

5.3.1.1 Speakers

Data for the colouring-in task was collected from 37 speakers in five language areas (one Japanese and four Ryukyuan) during a single fieldtrip in 2019. As with the naming task, data was collected from multiple localities, i.e. in multiple dialects. Where possible, speakers had not participated in the naming study, although this was not always possible (given the endangered status of Ryukyuan). As with Study I, some interview sessions were conducted with multiple speakers, as some elderly native speakers had little experience in performing standardised linguistic elicitation tasks, so analyses were conducted on sessions (see Table 5.3). In some sessions, younger family members helped with use of the tablet (see Section 5.3.1.2), but these are not included in the counts in Table 5.3.

Table 5.3

Speaker and session information for body part colouring in task.

Japanese	
Tohoku	8 speakers (4 female) in 7 sessions
Ryukyuan	
Amami	7 speakers (3 female) in 4 sessions
Okinawa	10 speakers (6 female) in 6 sessions
Miyako	7 speakers (3 female) in 6 sessions
Yaeyama	5 speakers (2 female) in 4 sessions

5.3.1.2 Materials and procedure

Data was collected using the blank line drawings of the human body used in Study I (Figure 5.1). Two drawings were used: a unitary image of the front and back views of the whole body, and a close-up image of the head/face with the mouth closed. We selected several terms of interest from each subgroup of the cluster analysis from Study I (see Table 5.4).

Table 5.4

List of body part terms used as prompts in the colouring-in task for each language area.

	Tohoku
Face	*aŋo, hana, hoppeda, kudzi, managu, mimi, na^{n}dzugi, odoŋe*
Torso	*adama, hara, ketsu, koɕi, ku^{m}bi, mune, senaga*
Upper limb	*kena, te, u^{n}de*
Lower limb	*aɕi, ko^{m}bura/fugurahaŋi, momota/yorota, sunegara*
	Amami
Face	*agu, fuː, hana, kutɕi, mɨ, mɨttɕu/maki, miɴ, utugə*
Torso	*kamatʃi, kubi, kuʃi, mari, munɨ, wata*
Upper limb	*kəːnja, tɨ, udɨ*
Lower limb	*hagi, kubura, mumu, sunɨ*
	Okinawa
Face	*çitɕeː/mukoː, fuːdʑira, hana, kakudʑi, kutɕi, miː, mimi, utugeː*
Torso	*kubi, kuɕi, nagani, nni, tɕibi, tɕiburu, wata*
Upper limb	*ke:na, ti:, udi*
Lower limb	*çisa, kuɴda, mumu, ɕini*
	Miyako
Face	*agu, futai, futsɨ, kamatsɨ, miː, mim, pana, utugai*
Torso	*bata, kanamai, kusammi, kusɨ, mnifutsɨ, nubui, tɕibi*
Upper limb	*kaina, ti:, udi*
Lower limb	*karasumi/sukara, kuvva, mumuni, pag^{z}ɨ, pisa*
	Yaeyama
Face	*agu, futai, futsɨ, miː, miɴ, pana*
Torso	*bata, buʃa, kusɨ, nubi, nni, tɕibi, tsɨburu/amasɨkuru*
Upper limb	*ti:, udi*
Lower limb	*dabura, mumu, paɴ, pisa, sɨni*

Note. Forms listed in the table are for illustrative purposes; specific forms of the body part terms differ per local dialect. Body part terms separated by a slash represent variants.

We did not restrict our choice to terms that were modal responses across all language areas, but instead chose cognate terms and non-modal terms if these were of relevance to the current study (e.g., terms that are cognate with Okinawa *ke:na* were used in other language areas as well, but not frequently). We excluded joints and digits for further elicitation.

The task was performed on an iPad Pro tablet using an Apple Pen stylus using the Adobe Photoshop Sketch app. The blank line drawings of the human body were used as the background layer of an image file. All colouring was done in a single file with multiple layers, with a separate layer used for each prompt. After completing the colouring for a given prompt, the layer containing that colouring was hidden and a new layer was created, so that speakers started with a 'clean sheet' for every prompt. Participants were presented with each body part term one at a time and were asked to colour in the part it referred to. There was only a single image for the face, so speakers simply coloured in the part of the face referred to by the prompt term. For the body, participants were instructed to colour in the left, right, front, and back, as applicable.

5.3.1.3 Image processing

We superimposed all colourings that were elicited with the same term, creating a multi-layered image in which each layer represented one session, thus creating a single image showing the extensional range of each body part term.

5.3.2 Results

The body colouring data is presented according to the subgroups in Table 5.4 (i.e., face, torso, upper limb, lower limb).

5.3.2.1 *The face*

As is clear from Figure 5.8, there is considerable similarity between the language areas in the extension of terms for parts of the face. Bounded areas (e.g., eyes, nose, ears and mouth) are virtually the same across languages, whereas more variation is evident in unbounded parts (e.g., forehead, cheeks)—see also Majid and van Staden (2015). Most parts of the face were named with distinct terms across languages in Study I, except the 'cheek' in Yaeyama, which all but one speaker named using a term that refers to the face as a whole, so this was not included as a prompt[2]. The 'cheek' also showed the most variation, especially in Okinawa where its range was more circumscribed than in the other Ryukyuan languages.

We found there little to no overlap in the extension between different parts, even though participants coloured in each part independently (without seeing what they had coloured in previously), suggesting that the parts of the face are highly distinct and do not form a clear hierarchy. The only part-whole relationship that might be present is between the jaw and the tip of the jaw (chin): when prompted with forms of the *otogai* cognate set, speakers only coloured in the tip of the lower jaw, whereas the whole jaw was coloured for the corresponding terms (TO: *aŋo*; AM/MI/YA: *agu*, OK: *kakudʑi*). However, not all speakers and varieties distinguish between 'chin' and 'jaw', and speakers often coloured in a wider area below the mouth for *otogai*, whereas the *aŋo~agu* and *kakudʑi* generally only prompted colouring of the bony structure.

[2] Speaker 2 (Shiraho variety) in Study I used the term *kamutɕi* for 'cheek'—see also *kamutsɨ* in the related Hateruma variety (Miyara 1980, p. 1250)—but the speakers that performed the colouring task for the current study were not familiar with this term.

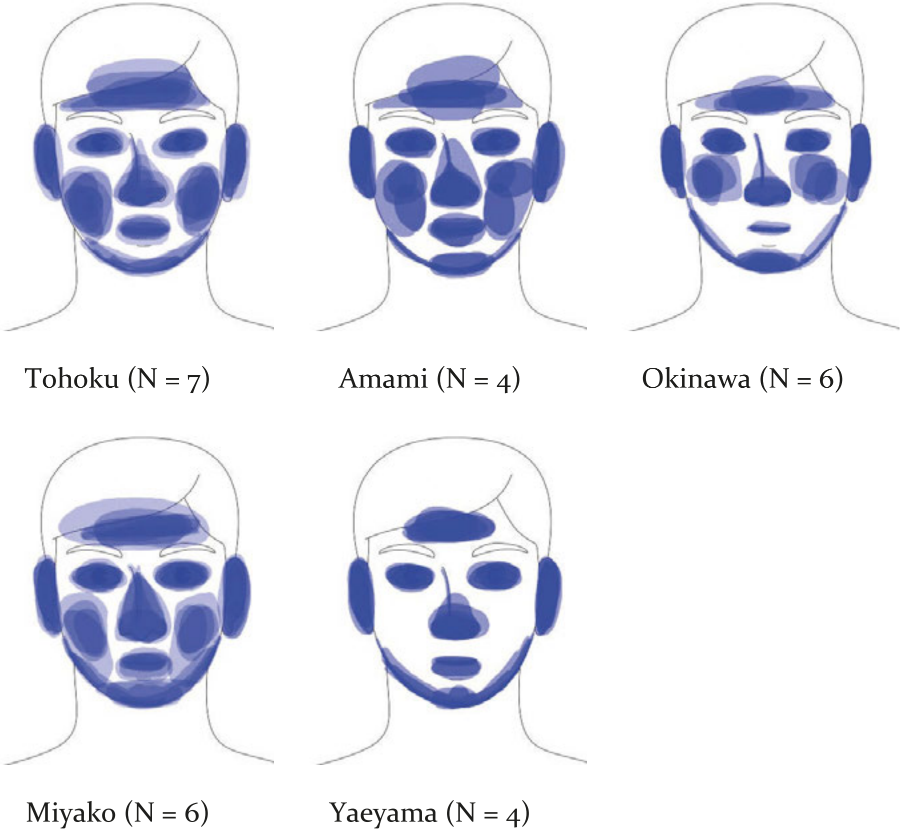


Figure 5.8. Extension of terms for parts of the face. Figures are composite images of the colouring-in data for participants in each of the five language areas.

5.3.2.2 *The torso*

The colouring data for the neck, chest, belly, and buttocks were highly differentiated with no overlap between them (Figure 5.9). In Tohoku, most speakers coloured in the entire back for *senaga*, and only the lower back for *kosu*, indicating a part-whole relationship between the two. Amami speakers generally recognised only one 'back' term, but the exact range differed between speakers. Okinawa speakers showed a pattern similar to Tohoku, with one term referring to the entire back (*nagani*) and another to the lower back (*kuɕi*). Finally, speakers of the Miyako and Yaeyama language areas used distinct terms for 'upper back' (*kusammi* and *buɕa*, respectively) and 'lower back' (*kusɨ~kutɕa*) that did not overlap in extensional range. Some Miyako speakers only coloured in the spine for the *kusammi*, and one speaker of the Tarama dialect coloured in a larger area for *kusɨ*, stating *kusammi* is not used in that variety.

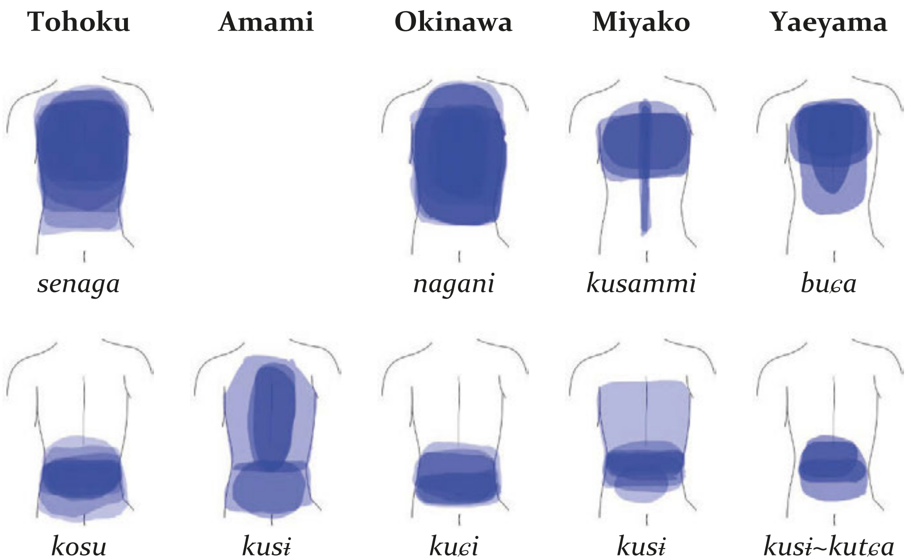


Figure 5.9. Extension of terms for the back of the torso. Figures are composite images of the colouring-in data for participants in each of the five language areas.

Study I showed large lexical variation across the Japonic varieties for 'head', and the colouring in data likewise found considerable variation in the extension of 'head' terms (Figure 5.10). Tohoku speakers' data for *adama* was similar to what has previously been reported for Standard Japanese *atama* (Majid & van Staden 2015), in that speakers generally excluded the face. Similarly, Amami and Okinawa speakers also do not include the face, and not all speakers coloured the back of the head either, focusing on the top of the head instead (similar to Jahai, Burenhult 2011, p.169). In contrast, Miyako and Yaeyama speakers coloured the front and back parts of the head as well.

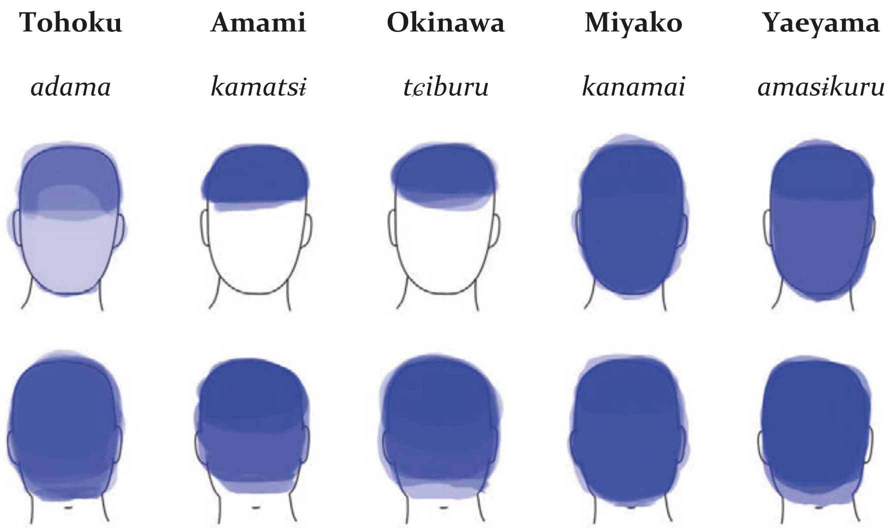


Figure 5.10. Extension of terms for the head (front and back). Figures are composite images of the colouring-in data for participants in each of the five language areas.

5.3.2.3 *The upper limbs*

The Tohoku varieties generally distinguish a separate hand category *te*, and an arm category *u*^{n}*de* that covered the area between the shoulder and wrist (Figure 5.11). Most speakers also recognised a third term *kɛna* that refers primarily to the upper arm, but some speakers in the colouring in task also included the lower arm. Ryukyuan varieties did not have a specific hand term. Instead, the data reveal an upper limb category *tɨ~ti:* encompassing the arm and hand, and *udɨ~udi~udẓi* (alternatively *kaina* in some Miyako varieties, e.g., Karimata) which covers the area between the shoulder and the wrist. In Okinawa, *udi* does not refer to the upper and lower arm, but instead seems to be restricted to the lower arm, with *ke:na* being used for upper arm. However, as in Tohoku, some speakers coloured in both upper and lower arm.

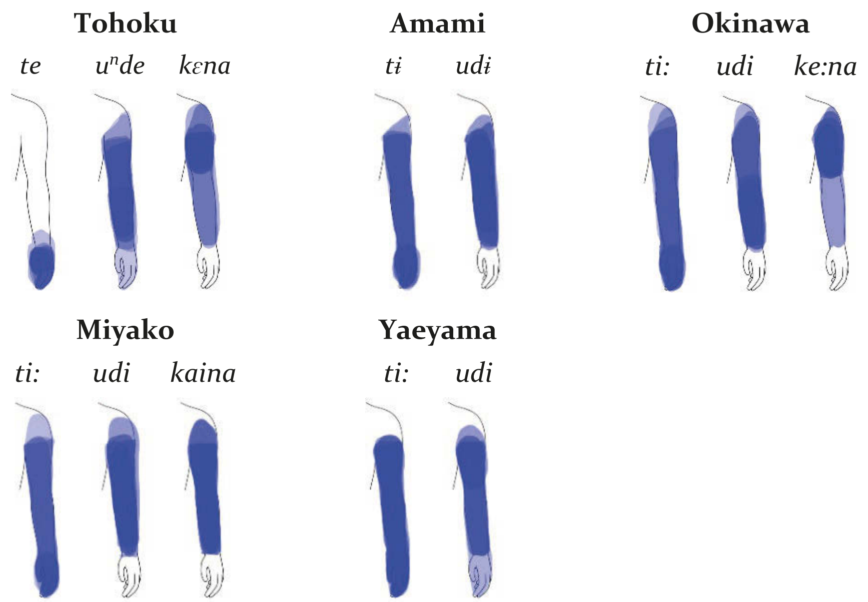


Figure 5.11. Extension of terms for parts of the upper limb (front). Figures are composite images of the colouring-in data for participants in each of the five language areas.

5.3.2.4 *The lower limbs*

There was substantial similarity in the extension of the lower limb parts across languages (Figure 5.12), in contrast to the variation found for upper limbs. Speakers of all varieties tended to colour in the entire lower limb (including foot) for *asu, hagi~pag^{z}ɨ~paɴ*, and *çisa*, respectively—although some speakers of Tohoku only coloured in the foot. For *sune* and its equivalents, some Tohoku and Amami speakers coloured in the entire lower leg, but mostly speakers only coloured in the shin. The two Ryukyuan varieties Miyako and Yaeyama distinguished a separate ‘foot’ category, whereas the other languages did not. ‘Foot’ appears to be a subpart of ‘leg’ based on Figure 5.12.

Tohoku	**Amami**	**Okinawa**	**Miyako**	**Yaeyama**
1. *momota*	1. *mumu*	1. *mumu*	1. *mumuni*	1. *mumu*
2. *sune*	2. *sɨnɨ*	2. *ɕini*	2. *karasɨni*	2. *ɕini*
3. *asu*	3. *hagi*	3. *çisa*	3. *p^{s}ɨsa*	3. *pisa*
			4. *pag^{z}ɨ*	4. *paɴ*

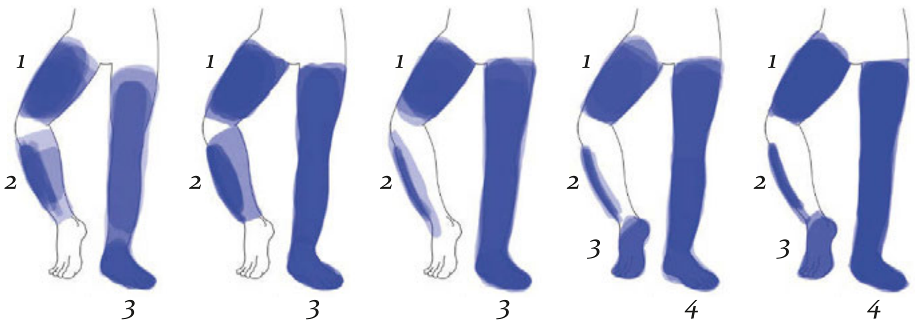


Figure 5.12. Extension of terms for parts of the lower limb (front). Figures are composite images of the colouring-in data for participants in each of the five language areas.

5.3.3 Summary

The body colouring task supported the broad patterns that emerged from the naming task in Study I. Parts of the face were for the most part highly differentiated with little to no overlap. Within bounded parts of the face (eyes, nose, ear, mouth), there was substantial similarity between languages. In contrast, unbounded parts varied more, with most variation found in the cheek area—Okinawan participants indicated a smaller referential area for 'cheek' than other languages, whereas Yaeyama participants included in this study did not recognise a specific term for 'cheek' at all.

Moving to parts of the body, the torso data revealed distinct categories with little overlap, with the exception of throat/neck and upper back/lower back. This pattern was similar across basically all languages. For the limbs, there was clear evidence of part-whole relationships between subparts, and evidence of clear cross-linguistic variation. Tohoku participants displayed distinct 'hand' and 'arm' categories, which is likely a feature of the mainland Japanese varieties (see Majid & van Staden 2015 for Standard Japanese). In contrast, Ryukyuan participants did not have distinct 'hand' category counter to claims of 'hand' being a universal category (e.g., Wierzbicka 2007). If there is a distinction here, it is to have a separate 'upper+lower arm' or even a distinct 'upper arm' category (e.g., Tohoku and Okinawa[3]). Most of the languages had a single 'leg' category that included the 'foot', with only Miyako and Yaeyama have a distinct category for foot as well. All languages in addition had further distinctions between 'upper leg', 'shin' and 'calf', all of which were sub-parts of the leg.

[3] Also in Amami, where it might be restricted to some varieties; only one of the speakers included in the sample that performed the colouring-in task distinguished it through the term *gotə*.

5.4 General discussion

In the current study, we investigated the body part lexicon by examining both lexical and semantic variation. We found that when there is little lexical (form) variation, there is also relatively little semantic variation. In contrast, parts with high lexical variability are associated with semantic variation. As predicted, we found more variation in the body part lexicon across the Ryukyuan varieties than the Japanese mainland. While this is in line with a previous study showing there is more overall linguistic variation in the Ryukyuan branch of Japonic (Huisman, Majid & van Hout 2019), the amount of variation appears to be amplified for body parts, especially at the individual level. This could be because the current study included several varieties of each Ryukyuan language, whereas Standard Japanese has only one "variety". Ryukyuan languages lack standardisation leaving room for individual variation, in contrast to Japanese where standardisation has resulted in conventionalisation and conformed responses. If language standardisation does lead to more conformity in body part naming, this may explain the high similarity found between Germanic languages previously (Majid, Jordan & Dunn 2015). Future studies could systematically study whether body part semantics is more diverse in non-standardised languages.

Another possibility is that the Ryukyuan languages differ more from one another because of language loss associated with their endangered status. Several speakers reported knowledge of terms for specific body parts, but were not always able to produce these in the moment during the naming task in Study I. Speakers would then respond with a generic term (e.g., 'leg') or would not respond at all. Lexical retrieval is more difficult for elderly people, further challenging data collection (Wulff et al. 2019). Despite this, there were some striking similarities across the Ryukyuan languages with respect to missing responses—e.g., there were missing responses for 'hip joint' and 'shin' in all languages suggesting that there might be genuine lexical gaps rather than idiosyncratic failures of retrieval in individual participants. This demonstrates an additional, perhaps unexpected, value of stimulus-based elicitation. As well as uncovering possible lexical gaps within a language, elicitation can reveal knowledge gaps in endangered languages which can contribute to revitalisation efforts.

In addition to our analyses of lexical variation, we also investigated semantics. Our novel use of cluster analysis provided a framework for comparison across languages for the overall structure of body part terms (Figures 5.4 and 5.5). The cluster analysis uncovered a hierarchy in the body part lexicon which was implicit in the data. Previous studies have asked speakers to make judgements about the relations between terms using semantic elicitation—e.g., "Is the nail a part of the hand?", but such judgements are an unreliable basis for inferences about language (e.g., Dąbrowska 2010). Our method provides a different way to derive body partonomies, and uncovers important insights. Brown (1976) and Andersen (1978) both independently suggested a hierarchically structured lexicon for body parts that consists of up to six levels. Our data suggest the Japonic languages partonomy does not go deeper than three levels, which is considerably less than Brown and Andersen suggested, and more in line what has been found in recent cross-linguistic work (see Majid, Enfield & van Staden 2006).

It has been argued there is no single organisational principle for the body part lexicon as a whole, but that if principles exist "they are more likely to be limited to distinct sub-systems such as the face, internal organs, or limbs" (Majid & Enfield 2017). The sub-systems uncovered in the cluster analysis suggest a new method for uncovering cross-linguistic regularities. Some subsystems resemble previously proposed universal body part categories—e.g., the upper limb (Brown 1976; Andersen 1978). Second, the subsystems correspond to the end points of cross-linguistic tendencies in semantic shifts (Wilkins 1981, 1996). Third, there are some correspondences between the subgroups found here and those found in free-listing data from English speaking children (e.g., a cluster of digits—see Crowe & Prescott 2003). Future work on a broader sample of languages families will have to confirm whether the observations reported here are specific to Japonic or whether they have wider applicability.

If there are indeed cross-linguistically stable sub-systems within the body part domain, this raises the question whether some of these are more variable than others. In the introduction, we had suggested that parts of the face might show more lexical and semantic similarity across languages because of the psychological salience of the face, and its dedicated neural architecture (Kanwisher, McDermott & Chun 1997). However, our study did not find support for this conjecture. Some parts for both the face and body showed high stability, and others were more varied. A post-hoc analysis of the World Loanword Database (Haspelmath & Tadmor 2009), which has a measure that corresponds to lexical stability (called *Age score*), for the body part equivalents of our study also showed little difference between the face (mean Age score = .86) and body (mean Age score = .84). Instead, the results from the current study suggest that bounded parts of the face (e.g., nose, eyes, ears) are likely to show more stability than unbounded parts (e.g., cheek, jaw), and the same likely holds for bounded vs. unbounded body parts. Majid (2010) suggested that joints may provide a perceptually salient boundary for the segmentation of parts (see also Majid & van Staden 2015). So, perhaps bounded parts are more lexically and semantically stable than unbounded ones. Future studies could be designed to test this proposal more systematically.

Finally, this study is the first to combine a body part naming and a colouring in task to uncover the lexical and semantic structure of the body. Overall, the results from the two methods provide converging evidence. As shown by the face versus body comparison, the free naming task provides the researcher with a starting point for the colouring data, but the colouring-in task helps better understand the extension of terms. In this study, the colouring data confirmed the distinctness of the parts of the face and their relative uniformity across languages. It also provided further clarity about the exact extension of terms referring to the upper and lower back, and confirmed the semantic variability for terms referring to the limb parts. So, both tasks are valuable—and necessary—for our understanding of body part terminology.

5.5 Conclusion

Previous studies have suggested that body part semantics is negligibly different within a language family, although there may be cross-linguistic differences when comparing languages from different stocks. Our study of the Japonic language family shows that body part terminology can vary within a language family in substantive ways, suggesting that the similarity uncovered so far may be limited to the specific language family studied (i.e., Germanic). Lexical similarity for body part terminology is more differentiated between the Japanese and Ryukyuan varieties, and this is reflected in semantics too. Within body parts, we see different structuring principles for parts of the face and parts of the body, with the former having a relatively flat hierarchy while the later shows deeper hierarchical structuring. However, when examining specific parts, there are no broad differences between the face and body. Rather bounded parts seem to show more stability in lexical form and semantics than unbounded parts. Our multi-method exploration of body parts demonstrates that we cannot presume a universal conceptualisation of the body as is often assumed in cognitive linguistics. Instead, in-depth studies of under-studied languages are urgently required.

CHAPTER

6

6 Cross-linguistic constraints and lineage-specific developments in the semantics of cutting and breaking in Japonic and Germanic[1]

Abstract

Semantic variation in the cutting and breaking domain has been shown to be cross-linguistically constrained in a previous typological study, but it was unclear whether Japanese was an outlier in this domain. Here we revisit cutting and breaking in the Japonic language area by collecting new naming data for 40 videoclips depicting cutting and breaking events in Standard Japanese, the highly divergent Tohoku dialects, as well as four related Ryukyuan languages (Amami, Okinawa, Miyako and Yaeyama). We found that the Japonic languages recapitulate the same semantic dimensions attested in the previous typological study, confirming that semantic variation in the domain of cutting and breaking is indeed cross-linguistically constrained. We then compared our new Japonic data to previously collected Germanic data and found that, in general, related languages resemble each other more than unrelated languages, and that Japonic languages resemble each other more than Germanic languages do. Nevertheless, English resembles all of the Japonic languages more than it resembles Swedish. Together, these findings show that the rate and extent of semantic change can differ between language families, indicating the existence of lineage-specific developments on top of universal cross-linguistic constraints.

Keywords: cutting and breaking, semantics, semantic variation, Japonic, Germanic

[1] This chapter is based on Huisman, J. L. A., van Hout, R. & Majid, A. Cross-linguistic constraints and lineage-specific developments in the semantics of cutting and breaking. Submitted to *Linguistic Typology*. Formatting follows the journal's guidelines. I certify that I performed data collection, analysis and writing of the manuscript, with feedback from the co-authors.

6.1 Introduction

Every part of language is subject to variation and change, and meaning is no exception. There are numerous cross-linguistic studies exploring various semantic domains that demonstrate languages differ substantially in the number, boundaries, and foci of meaning categories they distinguish. While the domain of colour (Berlin & Kay 1969; Kay et al. 2009) is perhaps the most well-studied, there are many other examples, such as spatial relations (Levinson & Wilkins 2006), body parts (van Staden & Majid 2006), temperature (Koptjevskaja-Tamm 2015), and cutting and breaking events (Majid, Boster & Bowerman 2008)—for reviews see (Evans 2010) and (Koptjevskaja-Tamm, Rakhilina & Vanhove 2016).

The studies referred to above have been successful at elucidating the structure of various semantic domains, but one issue that plagues all such studies is how to understand cases that appear to form an exception to broader cross-linguistic patterns. In a comparative study of separation events, speakers of 28 typologically different languages were shown a set of videoclips depicting various cutting, breaking and tearing actions, and were asked to freely describe them in their native language (Majid, Boster & Bowerman 2008). The overall categorisation of separation events was found to be largely constrained cross-linguistically with a small number of semantic dimensions capturing the structure of the domain: one dimension that represents high versus low predictability of the point of separation, a second dimension distinguishing 'tearing' events, and another dimension that distinguishes 'snapping' from 'smashing' events[2] (Majid, Boster & Bowerman 2008). However, Japanese did not fit well within this common cross-linguistic semantic structure: while the overall mean correlation between languages was $r = 0.53$, Japanese showed the lowest correlation at $r = 0.04$. This seemed to be the case because many verbs were unique to only a single event in the cutting and breaking stimulus set (Majid, Boster, & Bowerman 2008, p. 245).

[2] Majid, Boster and Bowerman (2008; p. 242) describe 'snapping' as the breaking of one-dimensional rigid objects into two pieces by application of pressure to both ends, and 'smashing' as the breaking of rigid objects into many pieces by a blow.

However, the data in the original Majid, Boster and Bowerman (2008) cross-linguistic study was sparse—i.e., only one speaker contributed data from Standard Japanese[3]—and the authors suggested further work was required to determine whether Japanese does indeed "categorise strikingly differently from the other languages" or whether this was a sample artefact (Majid, Boster, & Bowerman 2008, p. 245). Therefore, the first aim of the current study was to re-examine the semantic structure of the cutting and breaking domain in Japanese with more data. Since then, Fujii, Radetzky, and Sweetser (2013) proposed several additional distinctions that Japanese makes, such as 'loss of functionality', but it is unclear how these distinctions interact with the three main dimensions. We ask whether Japanese is truly unique in its semantic organisation of separation events or whether it respects the dimensional structure found in other languages.

Another question that remains open in cross-linguistic studies of semantic categories is how to sample languages adequately to make generalizations. Many studies take a sample that is as diverse as possible, arguing that lineage-specific similarities can lead to biases in estimating cross-linguistic patterns (Dryer 1989; Perkins 1989; Rijkhoff & Bakker 1998). However, comparison of the cutting and breaking vocabulary in four Germanic languages showed that even closely related languages can differ substantially in how they categorise these events (Majid et al. 2007). Similar differences between related languages have also been demonstrated for other domains, including locomotion (Malt et al. 2014; Slobin et al. 2014), containers (Majid, Jordan, & Dunn 2015; Malt et al. 1999) and spatial relations (Gentner & Bowerman 2009; Majid, Jordan & Dunn 2015). Since typological studies tend to favour diverse language samples, the scope of semantic variation across related languages is not well understood.

[3] The contemporary de facto standard (*hyōjungo* 'standard language') is predominantly based on the Tokyo variety, and it is this variety that has been included in the cross-linguistic comparison of cutting and breaking events (Majid, Boster & Bowerman 2008), as well as several other semantic domains (Kita 2006; Malt et al. 2014; Shindo 2015).

Therefore, the second aim of the current study was to establish the amount of semantic variation across the Japonic language family by comparing newly collected Standard Japanese data with data from the highly divergent Tohoku dialects, and four related Ryukyuan languages—the *Methods* section (6.2) introduces these languages in more detail. We ask whether all languages within the Japonic language family categorise separation events in a similar way. This is a particularly interesting question given the previously reported uniqueness of Japanese. If there are more languages that do not fit the common cross-linguistic structure, a good place to start would be relatives of Japanese. We examine the semantic similarity of cutting and breaking events in varieties of the Japonic language family. We ask whether the potential Japanese uniqueness in this domain is a feature of the entire language family.

Since comparative studies rarely examine languages from the same family, it is unclear how the semantic variation in, e.g.,, the Germanic languages (Majid et al. 2007) fits within the broader cross-linguistic context. Therefore, the third aim of the current study was to compare the amount of semantic variation across the two language families, by making a direct comparison between the newly collected Japonic data and the previously reported Germanic data (Majid et al. 2007). Given that the two language families are approximately of the same age (see below), we ask whether the amount of variation within language families is comparable.

In the next section we discuss the methodology for collecting the data, starting with an introduction of the Japonic languages, followed by descriptions of the speakers, materials and procedure used in the study. The results are then presented in three parts, first focusing on the Japonic data to answer (1) whether Japanese is unique in its semantic organisation of the cutting and breaking domain (Section 6.3.1), and (2) whether languages or language varieties related to Japanese organise this domain in the same way (Section 6.3.2). The third part of the results section compares the variability in the Japonic and Germanic data to answer (3) how semantically similar the languages within the Japonic and Germanic are in the cutting and breaking domain (Section 6.3.3).

6.2 Methods

6.2.1 Languages and speakers

6.2.1.1 The Japonic languages

The Japonic language family is spoken across the Japanese archipelago and consists of two main branches: Japanese spoken across the main islands, and Ryukyuan spoken across the smaller islands in the south (see Figure 6.1). While the exact number of "languages" is under debate—UNESCO lists seven (Moseley 2010), whereas Ethnologue mentions eleven (Eberhard, Simons & Fennig 2020)—there is more consensus about which dialect areas are considered to be unintelligible for Standard Japanese speakers. Within the Japanese branch, the Hachijo dialects are considered the most divergent (Hattori 1976; Pellard 2011). In addition, the varieties in the northern periphery (Tohoku) and the southern periphery (Kyushu) are highly divergent subgroups (Shibatani 1990; see also Huisman, Majid, & van Hout 2019). For example, the Tsugaru dialect from the Tohoku region is unintelligible for speakers of Standard Japanese (Takubo 2018).

For the Ryukyuan branch, there is a general division into at least five "languages" that roughly correspond to the geographical island clusters: Amami and Okinawa in the north, and Miyako, Yaeyama and Yonaguni in the south (Pellard 2015; Shibatani 1990). Mutual intelligibility is generally considered impossible between these five subgroups (Pellard 2011), although intelligibility between varieties within a subgroup can also be limited—e.g., within Yaeyama (Aso 2015). All Ryukuan languages are listed as either definitely or severely endangered (Moseley 2010). Fluent speakers are generally in their 60s or 70s at least, depending on the specific variety, and intergenerational transmission of the languages has been disrupted (Anderson 2015; Heinrich 2009). The Japanese and Ryukyuan branches are estimated to have diverged from each other over 2000 years ago (Lee & Hasegawa 2011), which is comparable to—in fact, slightly further back than—what is generally accepted for the Germanic languages (Atkinson & Gray 2006; Bouckaert et al. 2012), providing ample time for linguistic change to occur.

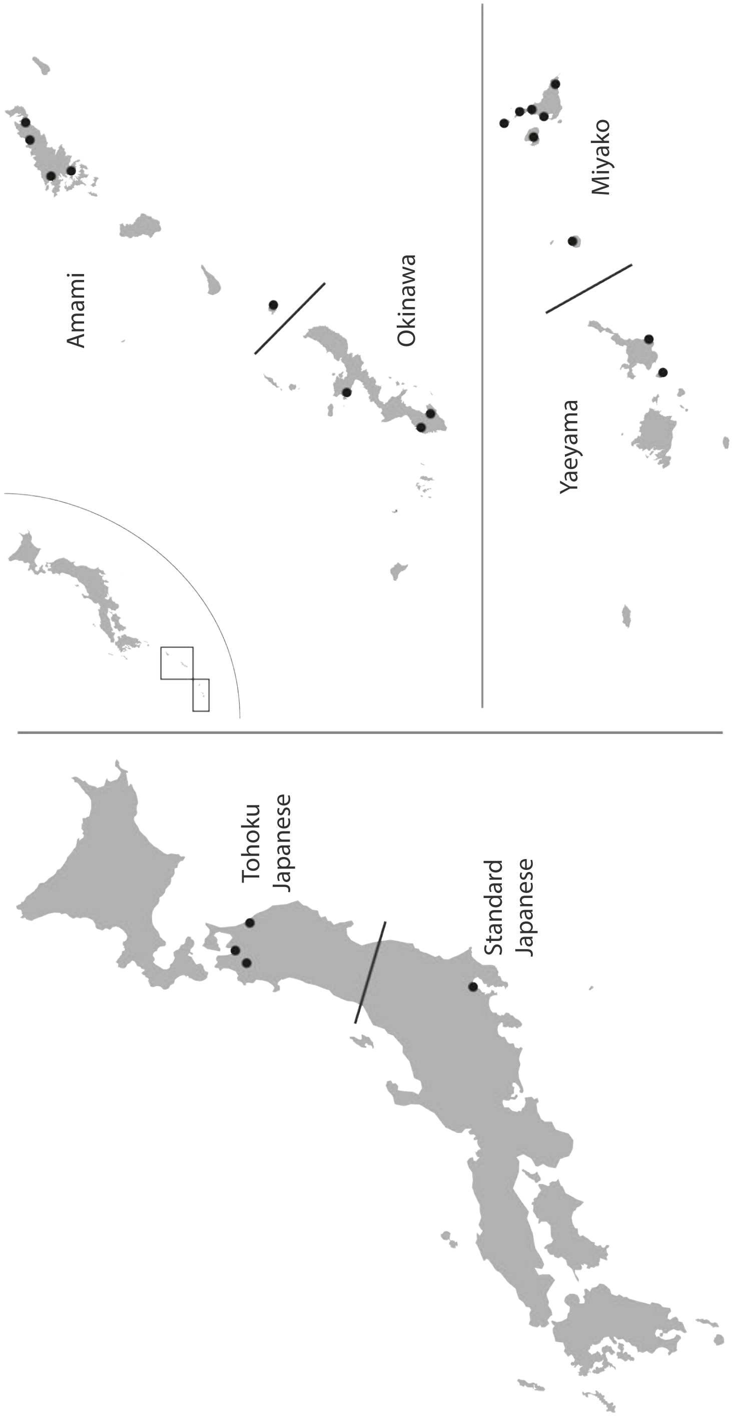


Figure 6.1. Map of Japonic language areas included in this study (left pane: Japanese, right pane: Ryukyuan) with fieldwork locations marked.

6.2.1.2 Speakers

Data was collected from 64 speakers in six areas (two Japanese and four Ryukyuan, see Table 6.1) during four fieldtrips conducted between 2017 and 2019. For all areas, data was collected from multiple localities, i.e. in multiple dialects—see also Figure 6.1. With the exception of Tokyo Japanese—which serves as the de factor national standard—there is no standardised variety of Tohoku Japanese or any of the Ryukyuan languages (Heinrich, Miyara & Shiomji 2015). As such, we will refer to Tokyo Japanese as "Standard Japanese", and use the term "language area" for the other five for the remainder of this paper, e.g., the Amami language area. Given the endangered status of Ryukyuan, the data was collected from elderly native speakers, some of which had little experience in performing abstract, reflective language tasks. As a result, some interview sessions were done with multiple speakers simultaneously. However, to minimise any effects arising from this, all analyses were conducted on sessions rather than speakers—see also the C*oding* paragraph.

Table 6.1
Speaker and session information per Japonic language area.

Japanese	
Tohoku	10 speakers (4 female) in 8 sessions
Tokyo	12 speakers (11 female) in 10 sessions
Ryukyuan	
Amami	12 speakers (4 female) in 10 sessions
Okinawa	9 speakers (5 female) in 5 sessions
Miyako	15 speakers (8 female) in 12 sessions
Yaeyama	6 speakers (3 female) in 6 sessions

6.2.2 Materials and procedure

The Japanese and Ryukyuan data were collected using a set of 40 videoclips depicting different cutting, breaking and tearing events. A standardised set of non-linguistic stimuli provides a frame of reference against which similarities and differences across languages and their varieties can be compared (Majid 2011). The set of videoclips consisted of the 28 test items from the *Kids' Cut & Break* set (Bowerman & Majid 2003), supplemented with four clips from the original *Cut and Break Clips* (Bohnemeye, Bowerman & Brown 2001), four rerecorded clips based on this original set, as well as four new videoclips. The set was designed to better represent the distinctions made in Japanese, based on the findings presented in (Majid, Boster & Bowerman 2008; Fujii, Radetzky & Sweetser 2013) and several Japanese descriptive studies (e.g., Kaetsu 1979; Kunihiro 1970). The videoclips were presented in two pseudorandom orders (A and B), one being the reverse of the other. Table 6.2 gives an overview of the videoclips and their sources. Data collection was in the speakers' native languages. Speakers saw the videoclips one by one on a tablet or laptop and were asked to describe the event depicted in the clip in their own language variety. Responses of any length were accepted, and speakers were free to give multiple descriptions of each event. All sessions were audio (and sometimes video) recorded for later transcription. The data was collected under the Ethics Assessment Committee of the Centre for Language Studies at Radboud University.

Table 6.2

Descriptions and sources of the video clips used in the naming task, as well as the two orders (A and B) in which they were presented. Scenes that were included in the Germanic data are marked*.

Order			
A	B	Action	Source
1	40	cut paper with scissors	Kids' C&B
2	39	break a twig with hands	Kids' C&B
3	38	break a mirror with a hammer*	New, rerecord
4	37	cut bread with a knife*	Kids' C&B
5	36	tear paper using a knife	Kids' C&B
6	35	tear cloth with hands*	Kids' C&B
7	34	break a glass by throwing it	New
8	33	cut a watermelon with a machete*	Original C&B
9	32	cut fingernails with a clipper	Kids' C&B
10	31	break a chocolate bar with hands	Kids' C&B
11	30	cut a piece of pie with a shard	Kids' C&B
12	29	hack off branch with an axe*	New, rerecord
13	28	cut an egg with a slicer	Kids' C&B
14	27	break a pot with a hammer*	Kids' C&B
15	26	cut off a branch with a knife*	Kids' C&B
16	25	cut grass with a sickle	New
17	24	cut a branch with a saw*	Original C&B
18	23	cut a nail with pliers	Kids' C&B
19	22	cut cardboard with a knife	Kids' C&B
20	21	tear bread with hands	Kids' C&B
21	20	cut into a tree with a knife	New
22	19	cut hair with scissors*	Kids' C&B
23	18	tear a bag with hands*	Kids' C&B
24	17	cut a banana with a knife*	Kids' C&B
25	16	break off a branch with hands	New, rerecord
26	15	cut into a watermelon with a knife*	Original C&B
27	14	tear a bread roll with hands	Kids' C&B
28	13	tear a banana peel with pliers	Kids' C&B
29	12	tear out a page with hands	Kids' C&B
30	11	cut cloth with scissors*	Kids' C&B

31	10	cut wood with an axe*	New, rerecord
32	9	cut breads using scissors	Kids' C&B
33	8	cut scallops with a knife	Kids' C&B
34	7	cut off a branch with an axe	Kids' C&B
35	6	break a glass by accident	Kids' C&B
36	5	tear up paper using hands	New
37	4	break a twig partially*	Original C&B
38	3	cut a rope with a chisel*	Kids' C&B
39	2	cut a banana with scissors	Kids' C&B
40	1	cut a rope with a knife*	Kids' C&B

Kids' C&B: Bowerman & Majid (2003).
Original C&B: Bohnemeyer, Bowerman & Brown (2001).

6.2.3 Coding

For each videoclip, we extracted the main descriptor(s) that encoded the target event depicted. Across the Japonic language family, this is typically done through a verbal construction (Example 1a). We coded verbs in their citation form (non-negative non-past; Example 1b). For V-V compound verbs[4], we coded the first and the second verb individually (Example 1c). For light verb constructions with the verb *suru* 'to do', we coded the element combining with *suru*, as that part of the expression that carries semantic content. Verbal nouns were coded as such (Example 1d), adverbial nouns were coded without their particle (Example 1e), and adverbialised adjectives were coded in their citation form (non-negative non past; Example 1f). For syntagms, we coded all elements without any particles (Example 1g).

[4] Kageyama (2016) divides compound verbs into two classes: lexical and syntactic compound verbs. In syntactic compound verbs, the second verb takes on a purely aspectual meaning through semantic bleaching or metaphorical extension. The verb *kiru* (lit. 'to cut') is one of the few verbs that are used in the V2 position of such constructions: *V-kiru* 'to V completely'—see Kageyama (2016; p. 283). There was only one such response in the data of the current study: *ori-kire-nakatta* [break-cut:POT-NEG:PST] 'could not break completely', which was used for the videoclip in which a twig was broken without separation. In this case, we coded the first verb as the main response, because that was the verb encoding the target event.

		Standard Japanese			Coded responses
(1)	a.	*kagami=o*	*waru*		WARU
		mirror=ACC	break:NPST		
		'They break the mirror.'			
	b.	*kagami=o*	*watta*		WARU
		mirror=ACC	break:PST		
		'They broke the mirror.'			
	c.	*kagami=o*	*tataki-waru*		TATAKU, WARU
		mirror=ACC	strike:INF-break:NPST		
		'They break the mirror (by hitting it).'			
	d.	*tamago=o*	*suraisu*	*suru*	SURAISU
		egg=ACC	slice	do:NPST	
		'They slice an egg.'			
	e.	*tamago=o*	*barabara=ni*	*suru*	BARABARA
		egg=ACC	pieces-ADV	do:NPST	
		'They cut an egg into pieces.'			
	f.	*tamago=o*	*komaka-ku*	*suru*	KOMAKAI
		egg=ACC	fine-ADV	do:NPST	
		'They finely cut an egg.'			
	g.	*kirikomi=o*	*ireru*		KIRIKOMI, IRERU
		incision=ACC	insert:NPST		
		'They cut into (something).'			

As speakers were allowed to produce multiple responses, we coded all responses. In sessions where multiple speakers were present, we coded all unique responses produced for a scene, i.e., if two speakers used different verbs, we coded both.

6.3 Results

6.3.1 Semantic dimensions of cutting and breaking in Standard Japanese

We first asked whether we could uncover the semantic dimensions that organise the domain of cutting and breaking in Standard Japanese. Following (Majid, Boster & Bowerman 2008), we coded, per interview session, for each pair of videoclips whether the speaker(s) described those two clips with the same verb (coded as 1) or different verbs (coded as 0). This created a videoclip-by-videoclip similarity matrix. Majid et al. created binary matrices in their study, but since we had data from multiple speakers/sessions, we were able to take into consideration speaker variation using weighted matrices. To do this, individual matrices were summed to create a Standard Japanese matrix with frequency counts representing how often speakers used the same verbs for videoclips. We then performed correspondence analysis in R (R Core Team 2019; `CA` function in the *FactoMineR* package—Husson et al. 2008), using the videoclip-by-videoclip matrix of Standard Japanese as input. Correspondence analysis calculates distances between rows and columns of a contingency table based on chi-squared distances. As such, these distances correspond to the strength of association between rows and columns, which can then be visualised in a (series of) plot(s)—(see e.g., Baayen 2008). The more similar columns or rows are to each other, the closer together they will be in the plot(s). In this case, we focus on the similarity of videoclips.

The correspondence analysis showed that the first four dimensions accounted for approximately 79% of the variance in the data. Three of these dimensions correspond to the main dimensions uncovered by Majid, Boster and Bowerman (2008)—i.e., predictability of the locus of separation, 'tearing', and 'snapping'—albeit they appear in a different order and with more distinctions in the tearing subdomain. It is most likely that the differences between the studies can be accounted for by stimulus sampling (the original study only included two tearing videoclips), although they could also reflect the differential salience of 'tearing' within Standard Japanese relative to other languages.

Looking at each of the dimensions in the order uncovered by the analysis, we find that the first dimension, accounting for 28% of the variance, distinguished a single videoclip (carving/making a cut into a tree trunk) from all others, which was described with two verbs not used for any other videoclips (*kizu (o) tsukeru* [wound (ACC) attach] 'to scratch' and, *hiQkaku* [pull-scratch] 'to scratch'). The videoclip distinguished by this dimension was not included in the original Cut & Break videoclips (Bohnemeyer, Bowerman & Brown 2001).

The second dimension, accounting for 19% of the variance, largely corresponds to the dimension representing predictability of the point of separation (the first dimension in Majid, Boster, & Bowerman 2008): it puts on one end events with a high predictability of separation, which were generally described with the generic 'cutting' verb *kiru* (e.g., scene number 1, cutting a piece of paper with scissors—the location of the separation of the paper is predictable from where the scissor intersects it) and at the other end events with low predictability of separation (e.g., scene 23, tearing a shopping bag). The tearing events loaded more distinctly from the cutting events on this dimension than the breaking events, unlike the original Majid et al. study. In addition, the tearing events were into divided into 'destructive tearing'—generally described with the verbs *yabuku* and *yaburu* [5]—and 'non-destructive tearing'. This ties in with the 'loss of functionality' distinction argued for by Fujii, Radetzky and Sweetser (2013), showing that this distinction interacts with (at least one of) the main cross-linguistic semantic dimensions. We come back to this in the next section.

The third dimension, accounting for 17% of the variance, distinguished 'snapping' events—which were generally described with the verb *oru*—from all others. In contrast to the findings in Majid, Boster and Bowerman (2008), where 'snapping' events were distinguished from 'smashing' events within the subgroup of events with low predictability of

[5] The two verbs are related in that *yabuku* is a blend of *yaburu* + *saku* (Shogakukan 2002).

the locus of separation, 'snapping' events in Japanese were distinct from all other scenes (but see also below).

The fourth dimension, accounting for 15% of the variance, distinguished the 'functional tearing' of e.g., edible objects—generally described with the verb *chigiru*—from 'neutral tearing', which was generally described with the verb *saku*. Both these types of 'tearing' contrast with 'destructive tearing' distinguished along the first dimension of the correspondence analysis. In addition, this dimension directly contrasted 'snapping' events with 'smashing' events—generally described with the verb *waru*—which corresponds to the third dimension in Majid, Boster and Bowerman (2008).

Figure 6.2 shows dimensions 2 and 4 of the correspondence analysis, with labels for the main categories that Standard Japanese distinguishes. As the figure shows, most videoclips cluster together closely, indicating that the semantic space of the cutting and breaking domain is well structured in Standard Japanese.

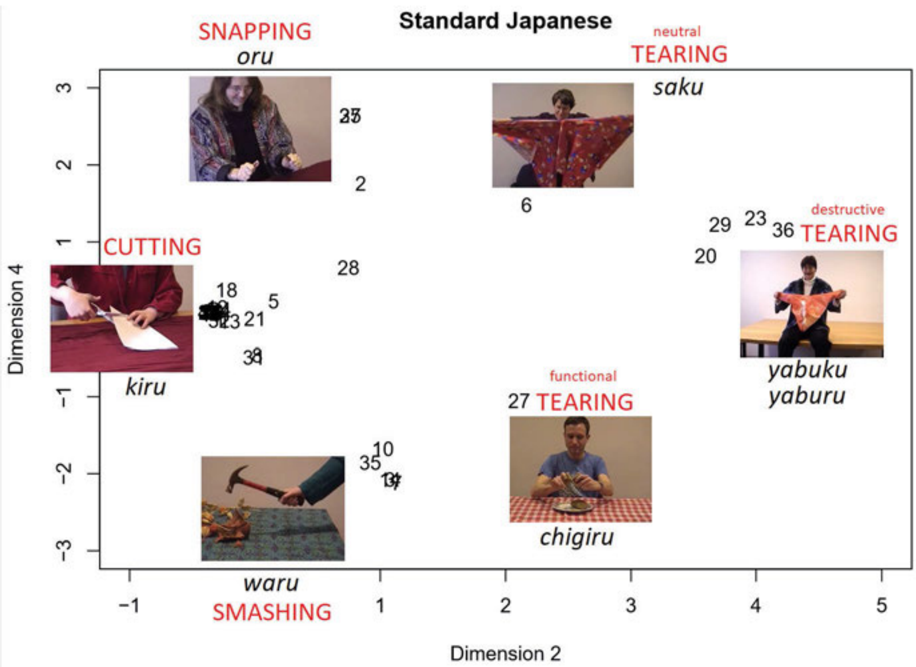


Figure 6.2. Dimensions 2 and 4 of the correspondence analysis for Japanese.

Comparing our findings to those of Majid, Boster and Bowerman (2008), who found the Japanese data in their study was an outlier with respect to other languages, we suggest that Japanese does not violate the dimensional structure that has been found in the cross-linguistic sample. One of the unique characteristics of Japanese in the cross-linguistic comparison was that a relatively large number of verbs was used uniquely for a single videoclip. While this could indicate that the Japanese language fundamentally partitions things differently, it could also be that the single speaker was partitioning things with unusually high specificity. In this study, where we included more participants, we see that high specificity is not a general characteristic of Japanese and that the language in fact has a small class of (main) categories. Moreover, these categories partition the semantic space of cutting and breaking events along similar dimensions to previous cross-linguistic studies. First, a dimension representing the predictability of the locus of separation emerged with events such as cutting paper with scissors or knife on one side, events like as smashing a pot with a hammer one the other, and chopping a melon with a machete in the middle of the dimension. Secondly, snapping events appear to be categorically distinct from smashing in Japanese. Finally, tearing events appear as completely distinct from cutting, smashing and snapping events. These three distinctions are the same found by Majid, Boster and Bowerman, and as such, is consistent with the claim that the Japanese semantic system follows the cross-linguistic constraints on the categorisation of cutting and breaking events. We provide additional evidence for this conclusion below.

While there are some noteworthy differences, such as the fine-grained distinctions within the tearing subdomain, these differences could be due to differences in the stimulus sampling—the current study included more tearing events, for example. However, it could also be the case that Japanese is distinct in this respect. To address this, we compare the Standard Japanese data to the other varieties for which we collected data in order to test whether these differences are present elsewhere; and more generally to examine the similarity of cutting and breaking event categories in this language family.

6.3.2 The cutting and breaking domain across the Japonic languages

We set out to compare Standard Japanese to its relatives, using three different measures to investigate the categories of cutting and breaking across the Japonic languages. For a general overview of the structure of the domain, we used correspondence analysis. Then, we used Mantel correlations to examine how similar the language areas were overall in their categorization of cutting and breaking. Finally, we also examined naming consensus between speakers through a measure based on Simpson's Diversity Index.

To explore the semantic space of cutting and breaking events in the Japonic language family, we created videoclip-by-videoclip matrices for each language, using the procedure described above, i.e. constructing aggregate similarity matrices. We then used all six matrices (two Japanese; four Ryukyuan) as input for a second correspondence analysis as described above, to uncover how well the semantic dimensions of Standard Japanese reflect the language family as a whole.

The correspondence analysis showed that the first four dimensions of the solution accounted for approximately 91% of the variance in the data. Again, three of the dimensions resemble those described by Majid, Boster and Bowerman (2008). The first dimension, accounting for 31% of the variance, distinguished the three snapping events from all others. The second dimension, accounting for 24% of the variance, corresponds to the dimension representing predictability of the point of separation: on one side we find events such as cutting with scissors or knife (high predictability), while on the side we find events such as smashing a pot or mirror with a hammer (low predictability)—see Figure 6.3. In the middle of the dimension are events of hacking a branch tree with an axe, or a melon with a machete. The continuity of this dimension when taking several languages into account resembles the first dimension in Majid, Boster and Bowerman to a higher degree then when only comparing Standard Japanese.

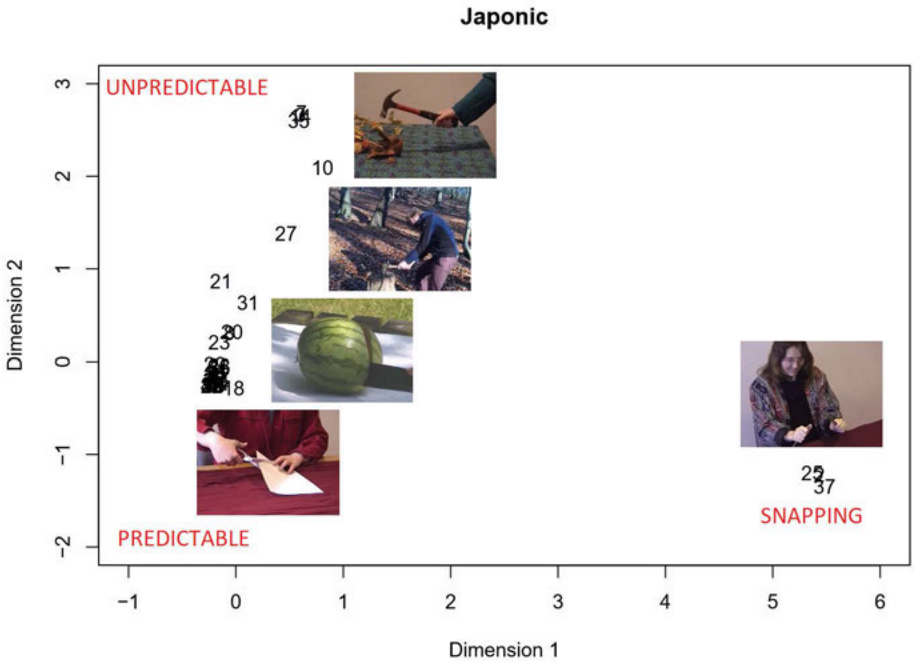


Figure 6.3. Dimensions 1 and 2 of the correspondence analysis for the Japonic language family.

The third dimension, accounting for 20% of the variance, distinguished a single event (carving/making a cut into a tree trunk) from all others. However, rather than pointing to a unique verb-videoclip combination, like in Standard Japanese (see above), this videoclip was described using many different verbs—including the generic cutting verb—which might be a further indication that its distinctness in Standard Japanese might be language-specific.

The fourth dimension, accounting for 16% of the variance, picked out the tearing events from the others—see Figure 6.4—which appeared along a continuum. At first glance, this dimension could be interpreted as representing the 'cleanness' of the tear as a result of the thickness and density of the objects, with thicker and less dense objects such as bread and banana peels on one side (less clean), and thinner objects such as cloth and plastic bags on the other (more clean). In the middle we find ripping up a piece of paper (less clean) and tearing a page from a notebook (more clean). However, a closer look at the individual languages suggests that each makes slightly different distinctions between tearing actions. Generally, the dimension distinguishes destructive tearing (e.g., of a plastic bag) from functional tearing (e.g., of a bread roll)—we come back to this in the discussion. To summarize, the overall semantic organisation of the cutting and breaking domain of the Japonic language family is similar to the cross-linguistic sample, albeit with differences in the order of the dimensions.

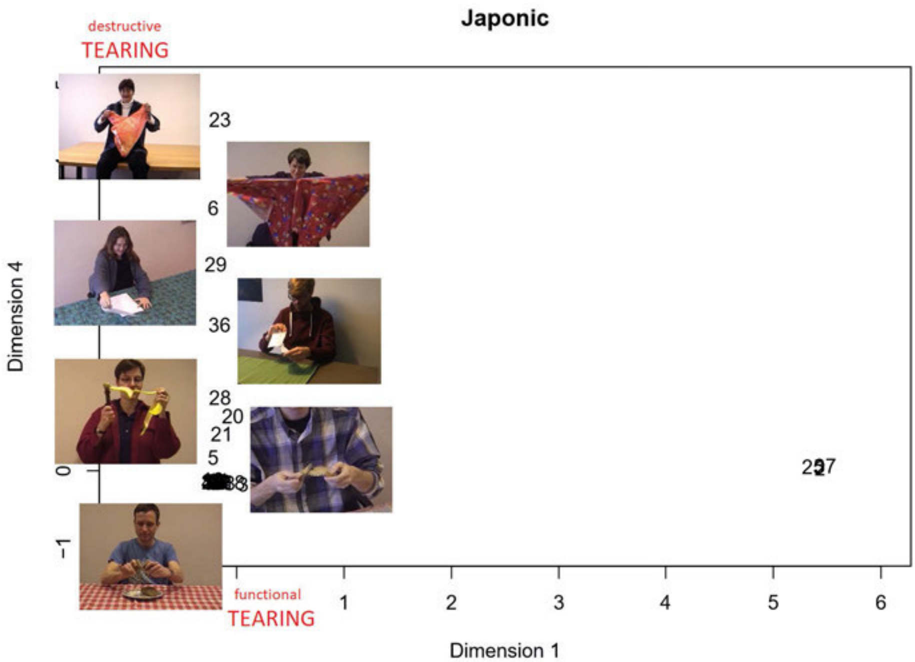


Figure 6.4. Dimensions 1 and 4 of the correspondence analysis for the Japonic language family.

To further investigate how similar individual languages were to each other in their semantic partitioning of cutting and breaking events, we compared how they grouped videoclips together. Using the six videoclip-by-videoclip matrices, we calculated pairwise Mantel correlations (in R; `mantel` function in the *ecodist* package—Goslee & Urban 2007) between them, using 10,000 permutations and 1,000 bootstrap iterations on 95% confidence intervals. The Mantel correlation tests showed that, overall, there was substantive similarity in the grouping of the cutting and breaking events across the Japonic language family. Average Mantel correlation between the six language areas was $r = .84$ ($SD = .06$), ranging between $r = .71$ and $r = .92$ (all p's < .001)—see Table 6.3, indicating that the semantic system of cutting and breaking across the Japonic languages is fairly homogenous. There is no clear division between the Japanese and Ryukyuan varieties, as might have been expected from the overall dissimilarity of the languages (see, e.g., Huisman, Majid & van Hout 2019). With an average Mantel correlation of $r = .77$, Okinawa Ryukyuan was the least similar to all other language areas, although this could also be the result of greater data sparsity for Okinawa Ryukyuan (fewer sessions, and more missing values).

Table 6.3
Mantel correlations between the six Japonic language areas. All languages show high positive correlations indicating the languages are very similar in how they partition the cutting and breaking domain.

	Tohoku	Amami	Okinawa	Miyako	Yaeyama
Japanese	.872	.878	.758	.907	.825
Tohoku		.862	.705	.845	.835
Amami			.816	.922	.871
Okinawa				.809	.742
Miyako					.912

As a final comparison between the individual languages, we examined whether there are differences in the codability of cutting and breaking events. Majid et al. (2007) found that among the Germanic languages, English showed lower between-speaker naming consensus than, for example, Swedish. This difference was found to be a result of the structure of the cutting and breaking lexicon. The English cutting and breaking lexicon was found to be more hierarchical, with two superordinate verbs (*cut* and *break*) divided into several subordinate verbs (*slice*/*chop*; *snap*/*smash*), meaning that the same videoclip can be described with several different verbs (at varying levels of specificity). In contrast, Swedish lacked this level of hierarchy, as there were more mandatory distinctions in the cutting and breaking lexicon, resulting in more constrained verb choice. To uncover whether such differences exist in the Japonic languages as well, we followed Majid et al. (2007) and calculated naming consensus across interview sessions using Simpson's Diversity Index (Simpson 1949, see also Majid et al. 2018). We calculated $\sum n(n-1)/N(N-1)$ per videoclip, in which lowercase n stands for the frequency of each unique verb, and uppercase N stands for the total number of responses for that videoclip. This produces a number between 0 (no consensus, where every speaker uses a different verb to describe a video clip) and 1 (complete consensus, where every speaker uses the same verb to describe a video clip).

Figure 6.5 shows the distribution of naming consistency scores for all the videoclips per language. Average naming consistency across speakers ranged between M= .54 for Miyako and M = .67 for Okinawa. This is comparable to what was found for the Germanic languages—which range between M= .4 (English) and M= .7 (Swedish). A one-way analysis of variance comparing naming consistency for the 40 cutting and breaking videoclips across the six language areas revealed no significant differences between languages: $F(5, 234) = 1.17$, $p = .323$.

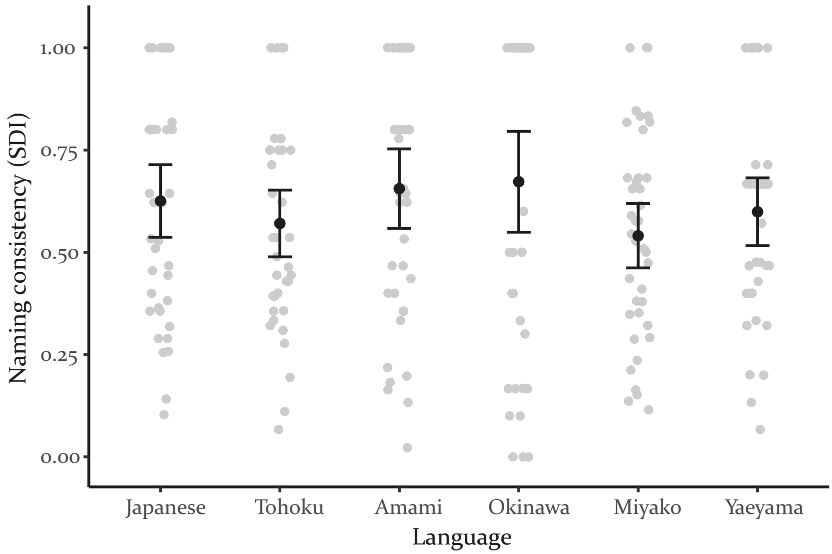


Figure 6.5. Plot of naming consistency scores for the 40 videoclips for each Japonic language area, with mean values per language represented by the black circle, error bars represent two times the standard error, grey dots represent individual scores for each videoclip.

In sum, the correspondence analysis revealed that the dimensions found in Standard Japanese are reflected in the entire language family. Moreover, the results correspond to what was found for a cross-linguistic sample, with some minor differences in the order and strength of the dimensions. The high Mantel correlations indicate a large degree of homogeneity across the language family. Naming consensus was similar across the six language areas, and towards the high end of what has been previously reported. In short, all three measures point to the same conclusion: the Japonic languages partition the cutting and breaking domain in similar ways.

6.3.3 Comparing the cutting and breaking domain in Japonic and Germanic

Finally, we aimed to put the results of the Japonic language family into cross-linguistic perspective. For this, the newly collected Japonic data described above was compared to existing data from four Germanic languages: English, Dutch, German and Swedish—see Table 6.4 for a speaker overview, and Majid et al. (2007) for a full description. The Germanic language family is originally spoken in north-western Europe but has—mainly through English—expanded across the globe. The language family consists of three branches: North Germanic mainly spoken across the Nordic countries, West Germanic mainly spoken across the north-western part of the European mainland and the British Isles, and the now extinct East Germanic.

Table 6.4
Speaker information per Germanic language.

West Germanic	
English	5 speakers in 5 sessions
Dutch	7 speakers in 7 sessions
German	5 speakers in 5 sessions
North Germanic	
Swedish	5 speakers in 5 sessions

The Japanese and Ryukyuan languages are estimated to have split from a common ancestor over 2000 years ago (Lee & Hasegawa 2011), whereas the Germanic languages are thought to have split from a common ancestor a few centuries later than that (Bouckaert et al. 2012), meaning these two language families are of a comparable age. Since diverging and developing along their own paths, how much similarity is there still between related languages within a language group? Given the cross-linguistic constraints on the categorisation of cutting and breaking events, we can expect meaning similarity to remain high, but is the extent of divergence the same across language families?

While the original Germanic data was collected with a different set of videoclips—the original *Cut and Break Clips* (Bohnemeyer, Bowerman & Brown 2001)—there was an overlapping subset of 17 videoclips (see Table 6.2) covering all major dimensions discussed in Majid, Boster and Bowerman (2008) and including scenes from all major clusters discussed in Majid et al. (2007). To ensure the comparison based on this reduced set reflects the results for the complete set of videoclips, we first compared the language-by-language similarities for these two sets of videoclips. If the correlations are high between the similarity matrices comprised of the overlapping subset of videoclips and the full videoclips, then we can confidently go on to compare the Germanic and Japonic languages.

For the Japonic languages, we had already created videoclip-by-videoclip matrices and calculated Mantel correlations between them (see above). We followed the same procedure for the subset of 17 videoclips that we have comparable naming data for Germanic and Japonic languages. We coded, per session, whether two videoclips were described with the same verb, and then summed all individual session matrices from each language. During this process, we found that the Okinawan data had data gaps for this subset of videoclips (5 out of the 17 videoclips were missing responses), so the Okinawan data was excluded for further comparison with Germanic. We followed the same procedure for the Germanic languages, creating one language-by-language similarity matrix based on the videoclip subset, as well as a similarity matrix based on the full set of 43 original core cutting and breaking clips (see Majid, Boster & Bowerman 2008). In the end, we had two language-by-language similarity matrices per language family, i.e. two sets of pairwise similarities—one based on the 17 shared videoclips and one based on the respective full sets of videoclips (40 for Japonic, 43 for Germanic).

To test whether these pairwise similarities based on the shared videoclips reflected similarities based on the full sets, we calculated—for each language family—Mantel correlations (R; *ecodist* package) between the similarity matrix based on the shared subset of videoclips and the matrix based on the full set of videoclips for each language, using 10,000 permutations and 1,000 bootstrap iterations on 95% confidence intervals. For the Japonic languages, the Mantel correlation between full and subset of videoclip similarity matrices was r = .862, p = .018; for the Germanic languages, the correlation was r = .909, p = .040. The similarities based on the smaller set of videoclips correlated highly with results based the full sets for both Japonic and Germanic, meaning the smaller shared set can be considered an adequate sample of the overall cutting and breaking domain, and thus the shared subset is suitable for comparing the two language families.

6.3.3.1 Measuring semantic similarity

To compare semantic similarity across the two languages families, we calculated the Mantel correlation between all language pairs for the 17 shared videoclips (in R; *ecodist* package), using 10,000 permutations and 1000 bootstrap iterations on 95% confidence intervals). This allowed us to compare similarity in three ways: (1) within-Japonic similarity versus within-Germanic similarity, (2) within-family similarity as opposed to between-family similarity, and (3) the overall similarity between languages from different families (see Figure 6.6).

First, we compared the correlations within the Japonic languages with those within the Germanic languages to see whether language-specific developments can lead to overall differences in the amount of semantic variation across language families. Mantel correlations ranged between r = .803 and r = .946 for the Japonic languages, and between r = .467 and r = .820 for the Germanic languages. Levene's test indicated that the amount of variability across the Germanic languages was significantly larger than across Japonic, $F(1,14)$ = 9.415, p = .008. In addition, the Japonic languages (M = .883, SD = .048) were, on average, more similar to each other than the Germanic languages were to other Germanic languages (M = .636, SD = .140), $t(5.17)$ = 4.19, p = .008, Cohen's d = 2.37.

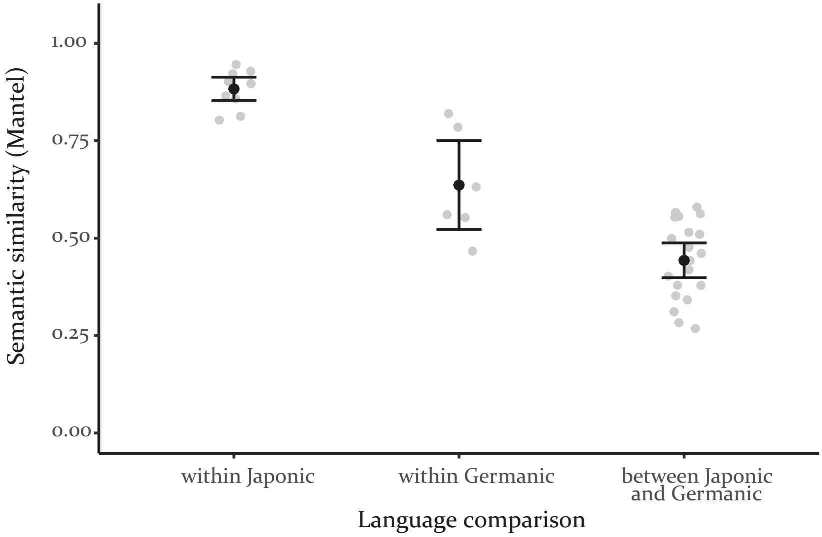


Figure 6.6. Plot of Mantel correlations for each language group, with mean values per group represented by the black circle, error bars represent two times the standard error, grey dots represent each Mantel correlation between a language pair.

Next, if there are strong lineage-specific patterns, then languages should correlate more highly with their own language family than with the other language family. However, as there are cross-linguistic constraints on the semantic variation in the cutting and breaking domain, it could be that the amount of variation found within language families does not differ from that between unrelated languages. To test this, we compared meaning similarity within language families (M = .791, SD = .152) with similarity between language families (M = .443, SD = .100), and found that overall within-family similarity was significantly larger, $t(34)$ = 8.25, p < .001, Cohen's d = 2.70. Separate tests showed that this was the case for both variation within the Japonic language family vs. between the language families, $t(27.99)$ = 16.31, p < .001, Cohen's d = 5.61, and for Germanic vs. between the language families, $t(24)$ = 3.97, p = .001, Cohen's d = 1.59.

Finally, given these cross-linguistic constraints on semantic variation in the cutting and breaking domain, we expected positive correlations between the Japonic and Germanic languages. Indeed, Mantel correlations between the Japonic and Germanic languages were all positive and statistically significant, ranging between r = .268 and r = .580, all p's < .01. A one-sample t-test showed that the average correlation between Japonic and Germanic languages was significantly larger than zero, $t(19) = 19.80$, $p < .001$, Cohen's $d = 4.43$.

These findings show that even though cross-linguistic constraints on semantic variation lead to similarity between unrelated languages, there can still be differences between language families as a result of lineage-specific developments, with some families showing more variability and a higher degree of similarity than others. Nevertheless, as Figure 6.6 shows, there are some Japonic-Germanic language pairs that are more similar to each other than some of the Germanic-Germanic pairs were. Looking at individual examples, English was more similar to any of the Japonic languages than it was to Swedish, highlighting the cross-linguistic constraints that apply to this domain.

6.3.3.2 Visualising semantic similarity

In addition to measuring semantic similarity, we also visualised the categorisation of the shared videoclips to further elucidate the patterns of variation. Figure 6.7 shows how the 17 videoclips can be grouped together, with each rectangle representing a category based on the modal, i.e. the most frequent, response they received. Intersecting rectangles with dashed lines represent cases where there were high frequency two responses with only a small frequency difference between them[6] (e.g., the 12 Miyako speaker sessions, 7 used *kˢɨː* and 5 used *sui* to describe the 'cut hair' videoclip). When there is no rectangle, it means there was no clear

[6] The criterion we used was a difference N ≤ 2 for language data coming from more than 10 sessions (Japanese, Amami and Miyako), and N = 1 for language data coming from less than 10 sessions (all other languages). In addition, the two most frequent responses had to account for 80% or more of all responses.

preference for any verb—either speakers gave varying responses or no response at all, so that even the most frequent response was used by less than 50% of speakers. The figure thus serves as a visualisation of the number and distribution of semantic categories in the cutting and breaking domain across the Japonic and Germanic languages.

Overall, the figure highlights the broader similarity in the semantic structure of the cutting and breaking domain across the two language families, which reflects the dimensions along which separation events are categorised (Majid, Boster & Bowerman 2008). Tearing events are a distinct subgroup listed on the left. The remaining videoclips are roughly ordered along the dimensions of predictability of the locus or separation, with high predictability towards the left, and low predictability towards the right. Within the low predictability scenes, snapping events are a distinct subgroup on the right.

At the same time, the figure also illustrates how individual languages differ from each other. First, the number of semantic categories that are distinguished differs per language—represented by individual rectangles. For example, the Germanic languages distinguish some cutting events based on how an instrument is used. In addition, the boundaries of the categories vary as well, this is particularly notable around videoclips labelled with “hack” and “chop”. Finally, Figure 6.7 illustrates the amount of similarity can differ per language family. The number and distribution of the categories (the rectangles) show a higher degree of variability across the Germanic languages, in contrast to the uniformity across the Japonic languages.

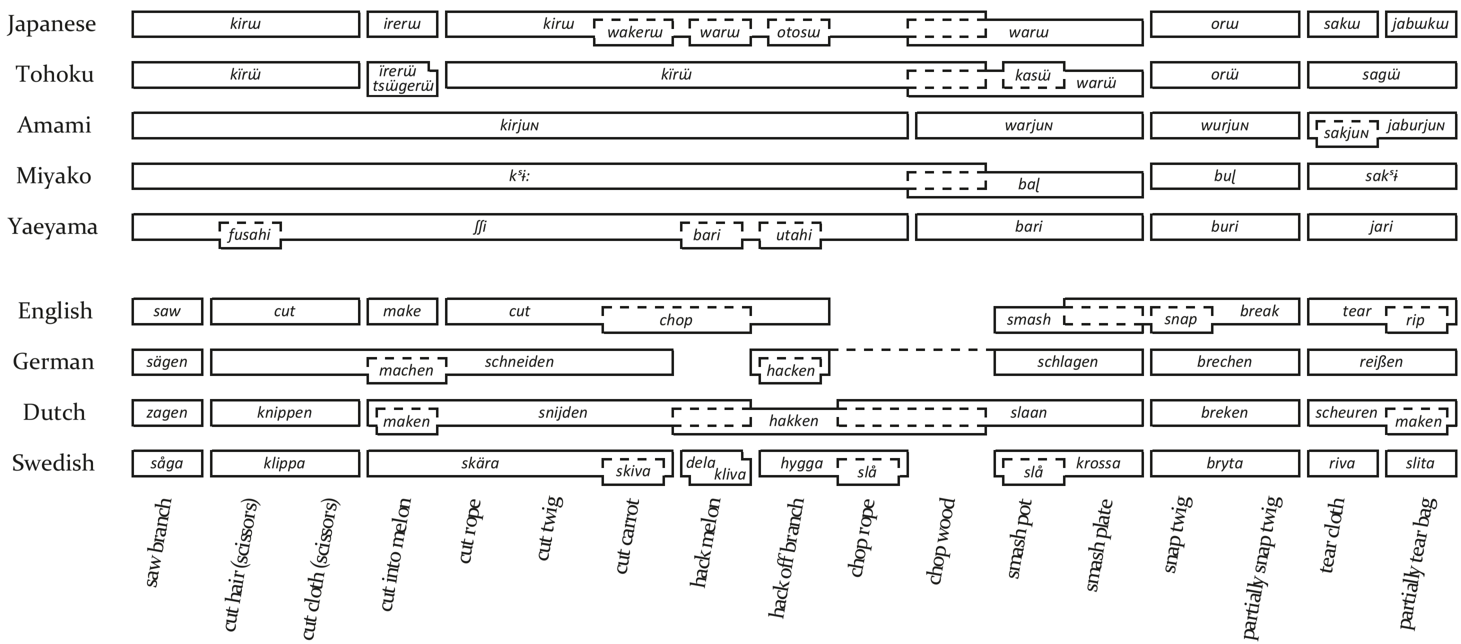


Figure 6.7. Semantic categories in the cutting and breaking domain across the 17 shared videoclips in Japonic and Germanic, based on the modal response (i.e., most frequent). Intersecting categories (dashed lines) represent videoclips where two high frequency verbs. Missing rectangles indicate that there was no clear preference for a particular response.

6.4 Discussion

Previous research was unclear as to whether Japanese was an outlier in how it categorised cutting and breaking events (Majid, Boster & Bowerman 2008), so we set out to investigate the semantic structure of the cutting and breaking domain in Japanese. We found that the three most important dimensions found in the cross-linguistic comparison ('predictability of locus of separation', 'tearing', and 'snapping') apply to Japanese as well. In addition to examining Standard Japanese, we also collected data in one highly divergent dialect area (Tohoku) and four Ryukyuan languages to investigate whether these languages categorise cutting and breaking events in similar ways and found they do. There were minor differences in the order and weighting of the dimensions—e.g., in addition to the three main dimensions, there was a 'carving' dimension represented by a single videoclip—but stimulus sampling undoubtedly played a role in this. Given the high correlations we found between all the languages it is highly likely that this differential weighting of the dimensions is specific to the Japonic language family. Overall, our findings suggest that Japanese and the Ryukyuan languages are not outliers, and further confirm that the semantic variability in the cutting and breaking domain is constrained cross-linguistically.

After showing that the overall dimensions that constrain variation in the cutting and breaking domain cross-linguistically also apply to Japanese and the other Japonic languages, we directly compared our results with the findings from a previous study on the Germanic languages to investigate the amount of variation within and between language families. We found that the correlations between Japonic and Germanic languages were positive, confirming that the cross-linguistic constraints on semantic variability in the cutting and breaking domain means that languages will always resemble each other. If Japanese was indeed an outlier, we would have expected zero or even negative correlations between the Japonic and Germanic languages, but this was not the case. Secondly, we found that similarities within each language family were larger than between language families, hinting at lineage-specific developments in addition to these broader cross-linguistic constraints. Finally, we found that the Japonic languages were more similar to each other than the Germanic languages, showing that the rate and extent of such lineage-specific developments can differ between language families. Some domains in some languages—in this case, cutting and breaking across the Japonic languages—appear to be very stable.

While this study reveals that the Japonic languages have very similar semantic categories, it is not known whether semantic stability is a general feature or whether it is domain-specific. Close examination of other semantic domains is required to adjudicate. For body parts, for example, Standard Japanese and other mainland dialects do not distinguish between the 'foot' and 'leg', while (at least some) varieties of Ryukyuan do (Huisman, personal field notes; see also Hirayama 1992). Similarly, the exact extensional ranges of the terms for 'arm' and 'hand' appear to differ between Japanese and Ryukyuan (Huisman, personal field notes), and even within Japanese (see Majid & van Staden 2015; p. 577). Systematic comparative work on semantic differences in other domains is needed to reveal the extent of such variation across the Japonic language family, and how this compares to the variation we find in other language families.

That is not to say that there is no semantic variation across the Japonic languages for the cutting and breaking domain. The tearing dimension seems to be particularly differentiated in Japanese and Ryukyuan. Such differences have not been addressed before as there were only two videoclips for tearing actions in the original *Cut and Break* stimuli, but the inclusion of additional tearing videoclips revealed interesting further distinctions. There are either one (Miyako), two (the other Ryukyuan languages) or three subcategories (Standard Japanese) of tearing events attested. Moreover, the specific distinctions also differ: while some languages distinguish tearing events based on the object (thick vs. thin), others appear to distinguish based on the type of separation (clean vs. messy), or the functionality of the object after tearing (destructive vs. non-destructive tearing). This final distinction has also been highlighted by

Fujii, Radetzky and Sweetser (2013), in the context of 'breaking' events. Our data shows the distinction to be of importance to 'tearing' as well. Further research could reveal further points of comparison across other languages as well—see e.g., *rip* vs. *tear* (Fujii, Radetzky & Sweetser 2013).

The apparent stability of this domain across the Japonic languages brings us back to the question of how to sample languages for cross-linguistic comparisons of semantics. Previous research showed that the closely related Germanic languages differ considerably in their semantic categories of cutting and breaking (Majid et al. 2007). Since the overall linguistic distance has been said to be comparable to what is found across Germanic, similar differences could have been predicted within the Japonic language family as well. However, the results show that there is remarkably little variation in the Japonic language family for this semantic domain. In addition, even though semantic variation in the cutting and breaking domain occurs within cross-linguistic constraints, we found a general pattern that languages within a family are more similar to each other than they are to members of a different family. Having said that, we did find that English was more similar to all the Japonic varieties than it was to Swedish, showing that even lineage-specific patterns can be irregular, and that the typological studies need not avoid or exclude related languages a priori. Additional comparisons within and between language families can give us further insight into which patterns of semantic variation are common across the world's languages and at what scale they occur. A study comparing the lexicon as opposed to syntactic features across Austronesian showed a higher level of stability in the lexicon (Greenhill et al. 2017), indicating that different parts of the languages change at different rates, but it remains unclear how patterns of semantic variation compare.

There are also language-specific features to consider. It is likely that the semantic stability of the cutting and breaking verbs in Japanese and Ryukyuan is a result of the broader issue of how "semantic choices made in one subsystem affect those in others" (Evans 2010; p. 508). Verbs are generally semantically underspecified in Japanese (Hamano 1998), but this is often compensated for by mimetics (ideophones), which have high referential specificity (see Akita 2012). For example, combining mimetics with the Standard Japanese verb *oru* 'to snap' allows for further specification of the characteristic of the snapping event, e.g., *pokiri to oru* 'to snap' vs. *pokiQ to oru* 'to snap suddenly', or the object being snapped, e.g., *pokiri to oru* 'to snap smaller objects' vs. *bokiri to oru* 'to snap bigger objects' (see Yamaguchi 2003). Similar observations have been made in a study of human locomotion, where the number of verbs used by speakers of Japanese was lower than Dutch and English, because mimetics were used to further differentiate specific ways of moving (Malt et al. 2014). This provides new opportunities to further investigate semantic differences in mimetics used to describe separation events.

6.5 Conclusion

To conclude, the overall findings suggest neither Japanese nor any of the related Ryukyuan language is an outlier for the domain of cutting and breaking, confirming that semantic variability in this domain is cross-linguistically constrained. In addition, a comparison between the Japonic and Germanic language families revealed that despite cross-linguistic constraints, lineage-specific semantic developments cause related languages to resemble each other more than unrelated languages. In addition, the rate and extent of such lineage-specific developments differs between language families, but unrelated languages can sometimes be more similar to each other than related languages. So, while there are cross-linguistic constraints, there is still much to learn about the forces leading to semantic diversity between language communities.

CHAPTER

7

7 Summary and conclusion

7.1 Introduction

In the opening chapter of this thesis, I introduced Sapir's dictum "everyone knows that language is variable" (Sapir 1921, p. 147), but the full extent of this variation and how it emerges is still not completely understood. This thesis contributed to the study of linguistic variation through an investigation of the Japonic languages (Chapter 2 presented an overview, while Chapter 3 provided an empirical demonstration of linguistic variation using basic vocabulary), with emphasis on the highly endangered Ryukyuan languages. There was also an in-depth exploration of three semantic domains—colour in Chapter 4, body parts in Chapter 5, and cutting and breaking events in Chapter 6—which filled a gap in the literature given that semantics is both understudied in the Ryukyuan languages specifically (Shigeno et al. 2015) and underrepresented in work on endangered languages generally (Seifart et al. 2018). In this chapter, I summarise the main results of the empirical work of this thesis in light of the issues raised in the introductory chapter, put the findings into broader theoretical perspective, and discuss directions for future research. The chapter closes with a general conclusion.

7.2 Results

7.2.1 Geography and linguistic diversity

In Chapter 3, I compared overall linguistic diversity across the Japanese mainland with overall linguistic diversity across the Ryukyu Islands using basic vocabulary. Combining methodology from dialectometry with approaches from population genetics, I found that differences between Japanese language varieties gradually increase over geographic distance, i.e., an isolation-by-distance pattern. In contrast, variation across the Ryukyuan language varieties mostly reflects the time since varieties diverged, i.e. an isolation-by-colonisation pattern. Moreover, the analyses showed that overall variation across the Ryukyuan varieties is larger than across Japanese mainland varieties. These two factors combined show that the relative isolation of Ryukyu Islands has affected linguistic diversity in that area differently from the Japanese mainland.

Chapter 3 also provided a classification of the Japonic language family based on dialectometric methods applied to basic vocabulary. This classification was largely in line with both traditional classifications in Japanese dialectology (Kindaichi 1955; Shibatani 1990; Tōjō 1927) and with a classification based on phylogenetic methods (Lee & Hasegawa 2011)—the main difference being that Tohoku varieties in the north were classified as distinct from all other varieties. Important in light of the discussion of geography is that the new analysis showed that Japanese and Ryukyuan are clearly distinct, and that the subgrouping of the Ryukyuan varieties was largely along geographic patterns—i.e., by island cluster.

This overall geographical pattern of linguistic differences, i.e., the main difference between Japanese and Ryukyuan, was also reflected in semantic variation. In Chapter 4 on colour, aggregate comparisons over four core colour categories demonstrated Ryukyuan colour semantics was distinct from Japanese. In Chapter 5 on body parts, sessions with Ryukyuan speakers showed more overall variation in the naming task than sessions with Japanese speakers. The body colouring data also showed that semantic categories differ between Ryukyuan and Japanese, and sometimes between Northern and Southern Ryukyuan as well. Cutting and breaking events (Chapter 6) showed a different pattern. There was no clear geographic pattern of variation. This was likely due to the large degree of homogeneity for this particular semantic domain in the Japonic languages. This shows the importance of studying more than one domain in order to make reliable generalisations about the factors that condition variation.

In sum, across chapters we see that the Ryukyuan languages pattern distinctly from the mainland Japanese languages, and that the amount of variation across the Ryukyuan languages is generally larger than across Japanese.

7.2.2 Semantics in the Japonic language family

Semantic features are not often covered in descriptive work on the Ryukyuan languages (Shigeno et al. 2015), especially not from a comparative perspective. Apart from an extensive study of colour in the 1960s, Chapters 4, 5 and 6 provide the first systematic and comparative work on the semantics of colour, body parts, and cutting and breaking events across contemporary Ryukyuan using stimulus-based elicitation tasks. In addition to data from Ryukyuan, I collected data from the highly divergent Tohoku varieties spoken in the north of the main island Honshu to explore semantic variation across the Japanese branch of the language family as well.

For colour and body parts, the data showed that there are clear semantic differences between Ryukyuan and Japanese, and sometimes even between some of the Ryukyuan languages. Comparison with previously collected data on colour (Kusakabe 1964) showed that even though Ryukyuan varieties are more similar to Japanese nowadays than in the past, they are still clearly distinct. For body parts, a domain with considerable lexical variation (see, e.g., National Insitute for Japanese Language and Linguistics 1968), the colouring data showed semantic differences within the Ryukyuan branch of the language family for several body part categories. Even though the Tohoku varieties were classified as being distinct from other mainland varieties, they are relatively closer to Standard Japanese than any of the Ryukyuan varieties, which is likely why the Tohoku varieties were shown to be relatively similar to Standard Japanese for these two domains—nevertheless, some small differences were found. Finally, the cutting and breaking data showed a remarkable level of semantic homogeneity across all varieties of Japonic—which differs from the Germanic languages, where there is considerable variation in this domain (Majid et al. 2007).

7.2.3 Semantic change and variability across domains

While the study of semantic variation can be conducted based on data from multiple languages of any relationship, collecting data in related languages allows for the study of semantic change. The characteristics of the Japonic languages made it possible to investigate the influence of three specific factors on semantic variation and change: contact and exposure to a majority or world language, characteristics of specific parts of a semantic domain, and the interaction between cross-linguistic constraints and lineage-specific developments.

Chapter 4 on colour is one of the few studies that aimed to directly compare how the semantic organisation of a domain can be affected by interaction with a standard language as well as the influence of English in today's globalised world. It is possible that the colour lexicon of Japanese is fairly unique with its complete adoption of a set of English colour terms whose semantics supplement rather than replace the existing native vocabulary. While the colour lexicon of three Ryukyuan language areas in the 1960s had fewer basic colour terms than Japanese at the time (compare Kusakabe 1964, and Berlin & Kay 1969), Chapter 4 shows that contemporary Ryukyuan speakers have adopted several Japanese and English terms. At the same time however, the Ryukyuan languages are still distinct from Japanese—albeit less so than before—especially for a core set of colour terms, pointing to the apparent stability of language-specific developments in semantics.

The body part domain provided an excellent opportunity to study differences in variability between subfields of a domain, given the characteristics and organisation of the domain. As such, in Chapter 5 I compared different subfields within the domain, initially comparing parts of the face and parts of the body, expecting less variation in the face because of its evolutionary importance (Kanwisher, McDermott & Chun 1997). However, there was little difference between face parts and body parts, and instead, both the naming data and the colouring data revealed that the boundedness of a part determines variability more, with bounded parts more stable across languages than unbounded parts (see also Majid & van Staden 2015).

Finally, Japanese was previously found to be an outlier for cutting and breaking, a domain where variability is cross-linguistically constrained. Chapter 6 shows that the semantic of cutting and breaking in Japanese does in fact follow these cross-linguistic constraints, and that other Japonic languages do so as well. So, the previous aberrant status of Japanese was likely due to the smaller number of participants that were tested, and the large amount of missing data. By testing a larger number of participants, I was able to provide a more reliable estimate of the variation in Japonic. In addition, I compared the Japonic data with previously collected data from the Germanic language family to investigate how lineage-specific developments interact with these cross-linguistic constraints on semantic variation in the categorisation of separation events. Through this comparison, I showed that the Japonic languages are, on average, more similar to each other than the Germanic languages, that the amount of overall variability was larger in Germanic than in Japonic, and that average semantic similarity within the language families was higher.

In sum, these findings show that there can be considerable semantic variation within a single language family, that variability can differ across domains, and that this variation is influenced by both lineage-specific and cross-linguistic factors.

7.3 Contributions of this thesis

One of the aims of this thesis was to use linguistic data from endangered languages to address theoretical issues in linguistics. Through this, I hope to make the Ryukyuan data I collected more valuable beyond its direct value in describing these endangered language varieties (Hale 1992; Seifart et al. 2018). In analysing the data, I applied new techniques which led to theoretical innovations. I will briefly summarise these novel analysis techniques before discussing how the findings in this thesis have benefited from their application.

7.3.1 Methodological contributions

Chapter 3 investigated the influence of geographic features on linguistic diversity and is the first study to apply methods from dialectometry to do so. In a previous study, Gavin and Sibanda (2012) based their measure of linguistic diversity on the *number* of languages spoken on an island, but 195 of the 264 islands in the study's sample supported only a single language. Counting languages does not capture variation *within* each language, which can be significant if one studies differences between related languages—as is the case with Japonic in this thesis. Chapter 3 improved on this by using dialectometric methods to measure pairwise linguistic differences in basic vocabulary between all island varieties, to calculate a more continuous measure of linguistic diversity across the islands, and thus provide a more fine-grained analysis of linguistic variation. In addition, the application of mixed-effects modelling made it possible to control for the distinctiveness of each language variety, which is not possible in multiple regression over distance matrices (MRM) to provide a more precise assessment of linguistic diversity across different geographic settings.

For Chapter 4 on colour, I applied correspondence analysis as used by Jäger (2012) for the World Color Survey data. However, while his analyses were mainly aimed at uncovering patterns of basic colour categorisation across diverse languages, I used correspondence analysis to directly compare colour terms in the same language family—see also Jäger (2012, p. 526–531). Results from these comparisons were then used to create a classification of the languages through cluster analysis. Chapter 4 also consolidated different data sets—my own and the earlier Kusakabe (1964) data—to directly compare stages of the same languages at different points in time, which made it possible to track changes in Ryukyuan colour semantics over several decades.

Chapter 5 was the first to collect body part naming and body colouring-in data in tandem, allowing for a holistic analysis of body part semantics in the Japonic languages. In addition, the study presented a novel way to study the semantic organisation of the body part domain—which has received a lot of attention (e.g., Andersen 1978; Brown 1976; Majid, Enfield & van Staden 2006)—by applying cluster analysis, for the first time, to body part naming data (see Crowe & Prescott 2003, for a similar use on free-listing data, or Majid et al. 2007, for a similar analysis in the cutting and breaking domain). Application of this method uncovered a hierarchy in the body part lexicon that was implicit in the data, without needing to rely on traditional qualitative methods—e.g., asking speakers to make judgements about the relations between terms, which can be an unreliable basis for inferences about language (e.g., Dąbrowska 2010).

Chapter 6 included the first direct comparison between multiple languages from distinct families for the cutting and breaking domain. Previous studies on semantic variation have compared either a broad sample of mostly unrelated language (Majid, Boster & Bowerman 2008), or a sample of related languages from a single family (Majid et al. 2007), but my method of data collection made it possible to compare the Japonic data with previously collected data in the Germanic languages.

Finally, where traditional dialectology in Japan focuses heavily on data compiled through translation, this thesis departed from that tradition by collecting data for the three semantic domains using stimulus-based elicitation tasks. This approach does not assume semantic equivalence and provides a neutral frame of reference against which linguistic similarities and differences can be systematically compared. An additional advantage of this method for research on endangered languages is that subtle semantic nuances can be studied without directly asking speakers to make linguistic judgements. This can be valuable in cases were speakers may have little experience (or confidence) in performing such meta-judgement tasks, which can be particularly relevant for endangered languages.

7.3.2 Theoretical contributions

In Section 6 of Chapter 1, I defined the aims and scope of this thesis. I also raised several broader issues through four research question, the findings for which I will describe in the following paragraphs.

7.3.2.1 Are there differences in the patterns of linguistic variation between connected land and islands?

Although geographic features have always been important in discussions on diversity, the island configuration is not often studied specifically in linguistics (although see e.g., Gavin & Sibanda 2012). A recent review asked whether there is anything "particular and peculiar about languages spoken on islands as compared to languages spoken on mainlands and continents" (Nash et al. 2020, p. 82), showing that the role of islands is not fully understood. The geographic configuration of the Japonic language area, comprising both a larger mainland and several smaller island clusters, made it an ideal location to examine the role of "islandness" on linguistic diversity. This type of direct comparison between islands and a mainland is a new approach within linguistics—even within biology, it is a recent development (Patiño et al. 2017). The results presented in Chapter 3 showed that the relative isolation of islands compared to a more connected mainland changes the patterns of diversity and increases overall diversity, confirming that languages spoken across islands do in fact show specific characteristics unique to that geographic configuration. The distinctiveness of the Ryukyuan languages spoken on islands was also evident in Chapters 4 and 5 where they patterned distinctively from languages spoken on the Japanese mainland. This calls for further systematic and in-depth work on island languages and the patterns of diversity found across them.

7.3.2.2 *To what extent are semantic categories in endangered languages affected by contact with majority languages?*

Data from all three semantic domains highlighted the importance of studying endangered languages sooner rather than later. Small indigenous languages are changing as a result of outside forces. In Chapter 4, I showed that contact with the standard language (Japanese) and a world language (English) has introduced both new colour terms and new colour categories to speakers of Ryukyuan. Data from other semantic domains showed similar patterns of influence. For example, a small group of Ryukyuan speakers was grouped together with speakers from the Japanese mainland because of their use of mainland-specific body part terms in the naming task in Chapter 5. In addition, during elicitation interviews for all three domains, Ryukyuan speakers displayed meta-linguistic awareness along the lines of: "Well, this is called X in Japanese and there is a word for it in [my variety], but I don't remember it now". Such findings are significant given the endangered status of many of the world's languages.

As discussed in previous work on language loss (e.g., Dorian 1977; Hale et al. 1992; Swadesh 1948), bilingualism can have strong effects on endangered languages and studies in the domain of colour have shown that several languages spoken in the Americas use English or Spanish borrowings to describe colour categories that traditionally emerge later in the evolutionary sequence (see e.g., MacLaury 1986, 2001). The colour domain stands out in this respect, as data collection started in the 1960s and 1970s. However, with recent developments in interconnectedness and the number of speakers dwindling for many languages, it is important to collect comparative data on semantic systems (see also Seifart et al. 2018), given that we lack a good understanding of many of them.

Nevertheless, analysis of the colour and body part naming data showed that the contemporary semantic systems of the Ryukyuan varieties in these domains resemble each other more than they resemble Standard Japanese—and they are also clearly distinct from Standard Japanese—which hints at the apparent stability of language-specific developments in semantics even in situations with majority language contact.

7.3.2.3 *Are some subparts of a semantic domain more variable than others?*

This thesis has shown that *within* a semantic domain, there can be different levels of variability. The body part lexicon has been addressed for many decades in studies of semantics (Andersen 1978; Brown 1976), but recent work has questioned the existence of a single overarching organisational principle for this domain (Majid & Enfield 2017). The cluster analysis in Chapter 5 revealed a structure of several subgroups of body parts that correspond to previously proposed universals in body part categories, e.g., the upper limb (see Brown 1976; and Andersen 1978). Using two different tasks (body part naming, and body colouring-in), I showed there are differences in variability between subsystems of the body part lexicon, i.e. between bounded and unbounded parts. Similarly, analysis of four colour cognate sets showed considerable stability in the Ryukyuan languages, and a series of cluster analyses grouped the Ryukyuan languages differently for each cognate, suggesting that the meanings of colour terms developed along distinct trajectories in each language. Finally, the cutting and breaking naming data showed that the Japonic languages all distinguish 'cutting', 'smashing', and 'snapping' along the same lines, but that the individual languages differ in their specific distinctions between different tearing actions. These findings demonstrate that semantic variability is not necessarily homogenous within a domain. Further work on other languages and additional domains will be needed to confirm this.

7.3.2.4 *How do language-specific developments interact with cross-linguistic constraints?*

The final contribution of the thesis is to the study of semantic variation *within* a language family. Where most cross-linguistic work samples data from a wide range of unrelated languages (see Dryer 1989; Perkins 1989; Rijkhoff & Bakker 1998), I focused on the variation within a single language family, a level at which semantic variation is understudied—and thus not fully understood. For example, within Germanic, Majid, Jordan and Dunn

(2015) found that languages resemble each other most for colours and body parts. However, whether this high degree of semantic similarity is peculiar to these two domains—and thus stable across all languages—remained an open question due to the lack of similar studies in other language families. In Chapter 4, we saw that in the 1960s, Standard Japanese had 11 basic colour terms, whereas Ryukyuan had between 4 and 8. In Chapter 5, I found that the average similarity for body parts in Japonic was considerably lower than what has been reported for Germanic (Majid et al. 2007). As such, it is therefore more likely the findings for these domains were specific to the Germanic languages and not necessarily cross-linguistically prevalent.

Further evidence of lineage-specific, rather than domain-specific, development was shown in Chapter 6, where there was higher semantic similarity across the Japonic languages than the Germanic languages for cutting and breaking. Cross-linguistic constraints that apply to some domains—such as the three main dimensions that distinguish separation events (Majid, Boster & Bowerman 2008)—limit the amount of variation between languages. Languages are thus limited in how this domain can develop and such parallel evolution can cause unrelated languages to resemble each other more than one would expect. For example, Chapter 6 showed that English was more similar to any of the Japonic languages than it was to Swedish in its cutting and breaking semantics. These results combined showed that it is far from obvious that all closely-related languages are equally similar across all semantic domains, which calls for more work to be done on semantic variation across related languages and highlights the need for more studies comparing variation within, as well as between languages families.

7.4 Future directions

Having discussed several new theoretical insights in the paragraphs above, I propose new avenues of research that can be explored in future work to build on the findings of this thesis.

7.4.1 The role of geography in linguistic diversification

Chapter 3 showed that languages diverge in different ways on islands and mainlands. As this thesis focused on the Japonic language family, it will be important to confirm these patterns in different language families. Obvious candidates would be different parts of the Austronesian languages, for which considerable linguistic data is available, and which have been studied using similar approaches (i.e. punctual vs. gradual diversification, Atkinson et al. 2008). The use of dialectometric methods to determine linguistic diversity (see also Section 7.2.2) a means of studying differences in language areas where there are multiple non-hierarchical subgroups—e.g., the Polynesian languages (see Blust 2013). For Japonic itself, a more detailed analysis using additional data collated from existing sources could explore how variation *within* the Ryukyuan languages is shaped by geographic factors, such as variability in the sizes of the islands as well as the distances between them. Given that settlement history is an important contributor to linguistic diversity across island languages, findings from such studies also offer new perspectives on the origin and spread of the Ryukyuan languages.

In addition, while this thesis focused on island characteristics, other geographical features could also be examined in future work. As described in Chapter 2, the Japanese archipelago is very mountainous, which is bound to influence patterns of variation as well. For example, on the larger scale of the Himalayas, valleys have been shown to increase linguistic convergences, whereas the mountain between them increases linguistic diversity (Post 2015, p. 232–233). A similar pattern seems to emerge in the Tohoku region in northern Japan. However, future work focused on such characteristics is needed to better understand patterns across the entire archipelago.

Perhaps unsurprisingly highly distinct varieties of Japonic are generally found in the mountains, and these varieties often exhibit conservative features—as expected for mountain languages (Pronk 2009; Steinicke et al. 2012). Examples include the negative auxiliary *-noo* in Narada dialect (Yamanashi Prefecture) and Ikawa dialect (Shizuoka

Prefecture)—see Iitoyo, Hino and Satō (1983)—which is theorised to be a remnant of Eastern Old Japanese (see *-(a)nap-* in Frellesvig 2010, p. 154); a retained distinction between historic long vowels *ɔɔ* and *oo* e.g., in Ōtori dialect (Yamagata Prefecture; Shibata 1953) and Akiyama dialect (Nagano Prefecture; Hirayama 1992); and retention of initial /p-/ in Ikawa dialect, which is virtually non-existent across the Japanese mainland outside the class of ideophones. In addition to these retentions, some mountain dialects exhibit unusual features—also typical of mountain languages (Urban 2020)—such as the raising of close-mid vowels in Ikawa dialect, or the lowering of close vowels in Akiyama dialect.

For the future, integrated analyses including data on both island characteristics and mountain characteristics can further increase our understanding of the influence of geography on linguistic diversification. The Japanese archipelago is good starting point to explore these factors on a fairly shallow time-scale (±2,000 years, see Lee & Hasegawa 2011), but studies can expand into other areas, such as Indonesia, which provide the opportunity to look at language diversity on deeper time-scales.

7.4.2 Semantics in the Japonic language family

The semantic variation in Ryukyuan found in this thesis calls for further data collection, both to confirm the patterns presented here, and to expand the data to ensure good and equal coverage of the different language areas before there are too few fluent speakers. For example, the high lexical diversity of the body parts domain made the colouring-in task challenging for varieties with distinct lexical items, but for which only a few speakers were interviewed (e.g., the Tarama variety of Miyako). Also, the cutting and breaking data failed to elicit some highly specific—and perhaps lower frequency—verbs previously reported in dictionary data (e.g., *badzauɴ~battso:ɴ* in Yaeyama, used for cutting meat or fish in preparation for cooking, Miyara 1930), so further work with additional speakers and different methods could clarify their semantics. Additionally, studying the use of colour terms, body part vocabulary and cutting and breaking verbs in old Ryukyuan texts can improve our understanding of the semantic developments that have taken place in the wider historical context.

For the Japanese branch of the language family, it is important not to discard the lack of major differences between the Tohoku varieties and Standard Japanese as indicative of an overall lack of semantic variation across mainland Japanese. One thing to keep in mind is that while Tohoku varieties are very divergent, they are also often classified as Eastern Japanese (Shibatani 1990), which puts them in the same subgrouping as Tokyo Japanese from which the standard language developed. Thus, additional data collection in other divergent Japanese varieties, such as those spoken in Kyushu (closer to Western Japanese, Shibatani 1990) or on Hachijo Island (a highly divergent variety, see Iannucci 2019), is required to explore how semantic variation patterns across mainland varieties.

Another avenue for future work is to study semantic differences on the individual speaker level, which can be particularly important in endangered language situations. As Chapter 4 (colour) showed, influence from the standard language has affected colour categorisation, and individual speaker analyses make it possible to determine which speakers adhere to a more conservative semantic system that better reflects the original endangered language—which is valuable for revitalisation efforts. In addition to determining the original semantic system, individual speaker analyses can help reconstruct past changes in this system, which helps understand how variation emerges across different domains, and which semantic categories are most likely to change. Finally, individual speaker variation should be studied in tandem with sociolinguistic background. Sociolinguistic work on endangered languages is often lacking (Seifart et al. 2018), even though the importance of such factors was one of the first points made in early discussions on language shift and language loss (e.g., Swadesh 1948). Especially for colour, where direct influence from contact with other languages was shown, this provides an excellent opportunity to study the role individual of speakers in language change.

7.4.3 Mutual intelligibility

Chapter 1 discussed the difference between dialects and languages, and how mutual intelligibility testing can be used to distinguish the two. Chapter 2 introduced the dialect vs. language issue for the Japonic language family, where the exact number of languages is still under debate as there are still few studies that specifically test mutual intelligibility (cf. Takubo 2018). That there are multiple distinct languages should be beyond doubt at this point. This thesis has provided an overview of semantic variation within the Japonic languages and shown it to resemble the level of variation found between e.g., the Germanic languages (cf. Majid et al. 2007, 2015).

Knowledge about semantic variation can be used to test mutual intelligibility. It is especially relevant for cases where shared cognates have different meanings, or where shared conceptual categories are named with different lexemes. Such information can help improve intelligibility testing that includes questions in which target items are presented in a context, as it allows for more careful manipulation of target items. For example, differences in the semantic ranges of colour terms, body part terms, and cutting and breaking verbs can be used to control the predictability of items such “The colour of this [object] is [colour term]” or “I used a [tool] to [cutting and breaking verb] the [object]”—items to be manipulated are in square brackets. This introduces a semantic dimension to mutual intelligibility testing which has largely focused on translation tasks up until now (see Gooskens 2013).

7.5 Conclusion

The findings presented in this thesis show that the geographical configuration of a language area is an important component in the process of language diversification and should thus be included in models that explain language variation and change. The multi-methods exploration of three semantics domains demonstrates that semantic systems can vary within a language family in substantive ways. Moreover, data from the Japonic languages shows that lineage-specific semantic developments can cause related languages to semantically resemble each other more than unrelated languages, even when the overall variation between languages is limited by cross-linguistic constraints. In addition, data from the Ryukyuan languages in particular shows that language-specific semantic developments are stable even in contexts of interaction between standard and (endangered) minority languages and against the backdrop of globalisation. Together, these facts show that there is still much to learn about the forces leading to semantic diversity between linguistic varieties, and that in-depth investigations of diverse (including understudied) languages are urgently required to better understand the evolution of words and their meanings.

REFERENCES

References

Aikhenvald, A. Y. & Dixon, R. M. W. (Eds.). (2017). *The Cambridge Handbook of Linguistic Typology*. Cambridge University Press.

Akita, K. (2009). *A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of mimetics* [PhD thesis]. Kobe University.

Akita, K. (2012). Toward a frame-semantic definition of sound-symbolic words: A collocational analysis of Japanese mimetics. *Cognitive Linguistics*, *23*(1), 67–90.

Akita, K. & Tsujimura, N. (2016). 4. Mimetics. In T. Kageyama & H. Kishimoto (Eds.), *Handbook of Japanese Lexicon and Word Formation* (pp. 133–160). De Gruyter Mouton.

Andersen, E. S. (1978). Lexical universals of body-part terminology. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of Human Language—Volume 3: Word Structure* (pp. 335–368). Stanford University Press.

Anderson, M. (2015). Substrate-influenced Japanese and code-switching. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages. History, Structure, and Use* (pp. 481–510). De Gruyter Mouton.

Aoi, H. (2015). Tarama Miyako grammar. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages* (pp. 405–422). De Gruyter Mouton.

Asato, S., Takara, K., Dana, M., Tomiyama, K., Nishizato, K. & Maehira, F. (2004). *Okinawa-ken no Rekishi [The History of Okinawa Prefecture]*. Yamagawa shuppan.

Aso, R. (2010). Hateruma (Yaeyama Ryukyuan). In M. Shimoji & T. Pellard (Eds.), *An Introduction to Ryukyuan languages* (pp. 189–227). ILCAA.

Aso, R. (2015). Hateruma Yaeyama grammar. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages. History, Structure, and Use* (pp. 423–448). De Gruyter Mouton.

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism: Language and Cognition*, *12*(1), 83–95.

Atkinson, Q. D. (2011). Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science*, *332*(6027), 346–349.

Atkinson, Q. D. & Gray, R. D. (2006). How old is the Indo-European language family? Illumination or more moths to the flame. In P. Forster & C. Renfrew (Eds.), *Phylogenetic methods and the prehistory of languages* (pp. 91–109). McDonald Institute for Archaeological Research.

Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J. & Pagel, M. (2008). Languages Evolve in Punctuational Bursts. *Science*, *319*(5863), 588.

Axelsen, J. B. & Manrubia, S. (2014). River density and landscape roughness are universal determinants of linguistic diversity. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1788), 20133029.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Bakker, F. & van Hout, R. (2012). De afgrenzing tussen de noordelijke en zuidelijke dialecten in Noord-Limburg. *Taal En Tongval*, *64*(2), 159–193.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Benedict, P. K. (1990). *Japanese/Austro-Tai*. Karoma.

Bentley, J. R. (2008). *A Linguistic History of the Forgotten Islands: A Reconstruction of the Proto-language of the Southern Ryūkyūs*. Kent: Global Oriental.

Berendt, E. A. & Tanita, K. (2011). The 'Heart' of Things: A Conceptual Metaphoric Analysis of Heart and Related Body Parts in Thai, Japanese and English. *Intercultural Communication Studies*, *20*(1), 65–78.

Berlin, B. & Kay, P. (1969). *Basic Color Terms: Their University and Evolution*. California University Press.

Blair, D. & Collins, P. (Eds.). (2001). *English in Australia*. Amsterdam: John Benjamins Publishing.

Blank, A. (1999). Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. In P. Koch & A. Blank (Eds.), *Historical Semantics and Cognition* (pp. 61–90). De Gruyter Mouton.

Bloch, B. (1942). Studies in colloquial Japanese IV: Phonemics. *Language*, *26*, 86–125.

Bloomfield, L. (1933). *Language*. Henry Holt and Company.

Blust, R. (2013). *The Austonesian Languages*. Australian National University.

Blust, R. (2014). Some Recent Proposals Concerning the Classification of the Austronesian Languages. *Oceanic Linguistics*, *53*(2), 300–391.

Boberg, C., Nerbonne, J. & Watt, D. (Eds.). (2017). *The Handbook of Dialectology*. Wiley.

Bohnemeyer, J. (2007). Morpholexical Transparency and the argument structure of verbs of cutting and breaking. *Cognitive Linguistics*, *18*(2), 153–177.

Bohnemeyer, J., Bowerman, M. & Brown, P. (2001). Cut and break clips. In S. C. Levinson & N. J. Enfield (Eds.), *Manual for the field season 2001* (pp. 90–96). Max Planck Institute for Psycholinguistics.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., Gray, R. D., Suchard, M. A. & Atkinson, Q. D. (2012). Mapping the Origins and Expansion of the Indo-European Language Family. *Science*, *337*(6097), 957–960.

Bowerman, M. (1996). Learning How to Structure Space for Language: A Crosslinguistic Perspective. In P. Bloom, M. F. Garrett, L. Nadel & M. A. Peterson (Eds.), *Language and Space* (pp. 385–436). The MIT Press.

Bowerman, M. & Majid, A. (2003). Kids' cut & break. In *Field research manual 2003, part I: Multimodal interaction, space, event representation* (pp. 70–71). Max Planck Institute for Psycholinguistics.

Bréal, M. J. A. (1897). *Essai de Sémantique: Science des Significations [Semantics: Studies in the Science of Meaning]*. Hachette.

Britain, D. (2010). Language and space: The variationist approach. In P. Auer & J. E. Schmidt (Eds.), *Language and Space: An International Handbook of Linguistic Variation—Theories and Methods* (pp. 142–163). De Gruyter Mouton.

Broca, P. (1862). La linguistique et l'anthropologie. *Bulletin de La Société d'anthropologie de Paris*, *3*(2), 264–319.

Brown, C. H. (1976). General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. *American Ethnologist*, *3*(3), 400–424.

Brown, P. (2007). 'She had just cut/broken off her head': Cutting and breaking verbs in Tzeltal. *Cognitive Linguistics*, *18*(2), 319–330.

Burenhult, N. (2006). Body part terms in Jahai. *Language Sciences*, *28*(2–3), 162–180.

Caskey-Sirmons, L. A. & Hickerson, N. P. (1977). Semantic shift and bilingualism: Variation in the color terms of five languages. *Anthropological Linguistics*, *19*(8), 358–367.

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

Chambers, J. K. (1995). *Sociolinguistic Theory*. Blackwell.

Chambers, J. K. & Trudgill, P. (1998). *Dialectology* (2nd ed.). Cambridge University Press.

Chang, W., Cathcart, C., Hall, D. & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, *91*(1), 194–244.

Clegg, S. M. & Phillimore, A. B. (2010). The influence of gene flow and drift on genetic and phenotypic divergence in two species of Zosterops in Vanuatu. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1543), 1077–1092.

Conlan, F. (2005). *Searching for the semantic boundaries of the Japanese colour term 'ao'* [PhD thesis]. Edith Cowan University.

Conway, B. R., Ratnasingam, S., Jara-Ettinger, J., Futrell, R. & Gibson, E. (2020). Communication efficiency of color naming across languages provides a new framework for the evolution of color terms. *Cognition*, *195*, 104086.

Crawford, G. W. (2008). The Jomon in early agriculture discourse: Issues arising from Matsui, Kanehara and Pearson. *World Archaeology*, *40*(4), 445–465.

Crawford, G. W. (2011). Advances in Understanding Early Agriculture in Japan. *Current Anthropology*, *52*(S4), S331–S345.

Crowe, S. & Prescott, T. (2003). Continuity and change in the development of category structure: Insights from the semantic fluency task. *International Journal of Behavioral Development*, *27*(5), pp. 467–479.

Cysouw, M., Dediu, D. & Moran, S. (2012). Comment on "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa". *Science*, *335*(6069), 657.

Cysouw, M. & Good, J. (2013). Languoid, Doculect, and Glossonym: Formalizing the Notion 'Language'. *Language Documentation & Conservation*, *7*, 331–359.

Czekanowski, J. (1928a). Na marginesie recenzji P. K. Moszydiskieg o ksigice: Wstep do historji Slowian [On P. K. M.'s marginal criticism of the book: Introduction to the history of the Slavs]. *Lud*, *2*(7).

Czekanowski, J. (1928b). *Róznicowanie sie dialectów prastowinskich w swietle kryterjum ilociowego [Differentiation of ancient Slavic dialects]*. First Congress of Slavic Dialectology, Prague.

Dąbrowska, E. (2010). Naive v. Expert Intuitions: An Empirical Study of Acceptability Judgments. *The Linguistic Review*, *27*(1), pp. 1–23.

De Meester, L., Gómez, A., Okamura, B. & Schwenk, K. (2002). The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica*, *23*(3), 121–135.

Dorian, N. C. (1977). The Problem of the Semi-Speaker in Language Death. *Linguistics*, *15*(191), 23–32.

Dryer, M. S. (1989). Large Linguistic Areas and Language Sampling. *Studies in Language*, *13*(2), 257–292.

Dumont d'Urville, J. (1834). *Voyage de découvertes de l'Astrolabe. Exécuté par ordre du Roi pendant les années 1826-1827-1828-1829, sous le commandement de M. J. Dumont d'Urville. Philologie, seconde partie. Les autres Vocabulaires des Langues ou Dialectes océaniens recueillis durant le voyage, et le vocabulaire comparatif des Langues françaises, madekass, malaïo, mawi, taïti et hawaii, suivis de quelques considérations générales sur ces langues*. Ministère de la marine.

Dunn, M., Greenhill, S. J., Levinson, S. C. & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*, 79–82.

Dunn, M., Terrill, A., Reesink, G., Foley, R. A. & Levinson, S. C. (2005). Structural Phylogenetics and the Reconstruction of Ancient Language History. *Science*, *309*(5743), 2072–2075.

Eberhard, D. M., Simons, G. F. & Fennig, C. D. (2020). *Ethnologue: Languages of the world* (23rd ed.). SIL International.

Elmer, P. (2019). *On the Origins of the Japanese Language* [Masters]. University of Vienna.

Ember, M. (1978). Size of Color Lexicon: Interaction of Cultural and Biological Factors. *American Anthropologist*, *80*(2), 364–367.

Enfield, N. J. & Wierzbicka, A. (2002). Introduction: The body in description of emotion. *Pragmatics & Cognition*, *10*(1–2), pp. 1–25.

Etler, D. A. (1996). The fossil evidence for human evolution in Asia. *Annual Review of Anthropology*, *25*(1), 275–301.

Evans, N. (2010). Semantic typology. In Jae J. Song (Ed.), *The Oxford Handbook of Linguistic Typology* (pp. 504–533). Oxford University Press.

Evans, N., Majid, A., Levinson, S. C. & Gaby, A. (2011). *Reciprocals and Semantic Typology*. John Benjamins.

Evans, N. & Sasse, H.-J. (2007). Searching for meaning in the Library of Babel: Field semantics and problems of digital archiving. *Archives and Social Studies: A Journal of Interdisciplinary Research*, *1*, 63–123.

Frellesvig, B. (2010). *A history of the Japanese language*. Cambridge University Press.

Fujii, S., Radetzky, P. & Sweetser, E. (2013). A Multi-Frame Analysis of Separation Verbs. In M. Borkent, B. Dancygier & J. Hinnell (Eds.), *Language and the creative mind* (pp. 137–154). CSLI Publications.

Gauchat, L. (1905). *L'unité phonétique dans le patois d'une commune*. Niemeyer.

Gavin, M. C., Botero, C. A., Bowern, C., Colwell, R. K., Dunn, M., Dunn, R. R. et al. (2013). Toward a Mechanistic Understanding of Linguistic Diversity. *BioScience*, *63*(7), 524–535.

Gavin, M. C. & Sibanda, N. (2012). The island biogeography of languages. *Global Ecology and Biogeography*, *21*(10), 958–967.

Gentner, D. & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. Guo, E. Lieve, N. Budwig, S. Ervin-Tripp, K. Nakamura & Ş. Õzçalişkan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 465–480). Pyschology Press.

Gerritsen, M. & van Hout, R. (2006). Sociolinguistic Developments as a Diffusion Process. In U. Ammon, N. Dittmar, K. J. Mattheier & P. Trudgill (Eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society / Soziolinguistik: Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft* (2nd ed., Vol. 3). De Gruyter Mouton.

Goddard, C. (2001). Lexico-Semantic Universals: A Critical Overview. *Linguistic Typology*, *5*(1), 1–65.

Goddard, C. & Wierzbicka, A. (Eds.). (1994). *Semantic and Lexical Universals: Theory and empirical findings*. John Benjamins.

Goebl, H. (1981). Eléments d'analyse dialectométrique (avec application à l'AIS). *Revue de Linguistique Romane*, *179–180*, 349–420.

Goebl, H. (1983). „Stammbaum" und „Welle". Vergleichende Betrachtungen aus numerisch-taxonomischer Sicht. *Zeitschrift Für Sprachwissenschaft*, *2*(1), 3–44.

Goebl, H. (1984). *Dialektometrische Studien: Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. Niemeyer.

Gooskens, C. (2006). Linguistic and extra-linguistic predictors of inter-Scandinavian intelligibility. *Linguistics in the Netherlands*, *23*(1), 101–113.

Gooskens, C. (2013). Methods for measuring intelligibility of closely related language varieties. *Handbook of Sociolinguistics*, 195–213.

Gooskens, C. & Schneider, C. (2016). Testing mutual intelligibility between closely related languages in an oral society. *Language Documentation & Conservation*, *10*, 278–305.

Gooskens, C., van Heuven, V. J., Golubović, J., Schüppert, A., Swarte, F. & Voigt, S. (2018). Mutual intelligibility between closely related languages in Europe. *International Journal of Multilingualism*, *15*(2), 169–193.

Goslee, S. C. & Urban, D. L. (2007). The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *Journal of Statistical Software*, *22*(7), 1–19.

Gray, R. D., Drummond, A. J. & Greenhill, S. J. (2009). Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science*, *323*(5913), 479–483.

Gray, R. D. & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, *426*(6965), 435–439.

Greenhill, S. J. (2011). Levenshtein Distances Fail to Identify Language Relationships Accurately. *Computational Linguistics*, *37*(4), 689–698.

Greenhill, S. J., Blust, R. & Gray, R. D. (2008). The Austronesian basic vocabulary database: From bioinformatics to lexomics. *Evolutionary Bioinformatics*, *4*, EBO–S893.

Greenhill, S. J., Heggarty, P. & Gray, R. D. (2020). Bayesian phylogenetics. In R. D. Janda, B. D. Joseph & B. S. Vance (Eds.), *The Handbook of Historical Linguistics, Volume II* (pp. 226–253). Wiley-Blackwell.

Greenhill, S. J., Wu, C.-H., Hua, X., Dunn, M., Levinson, S. C. & Gray, R. D. (2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, *114*(42), E8822–E8829.

Grootaers, W. A. (1967). Dialectology. In C. Yuen Ren, R. B. Noss & Joseph. K. Yamagiwa (Eds.), *Linguistics in East Asia and South East Asia* (pp. 585–607). De Gruyter Mouton.

Grootaers, W. A. (1982). Dialectology and sociolinguistics: A general survey. *Lingua*, *57*(2–4), 327–355.

Haarmann, H. (1989). *Symbolic Values of Foreign Language Use: From the Japanese Case to a General Sociolinguistic Perspective*. De Gruyter Mouton.

Haegyoung, K. A. (2009). Shintaigo o fukumu nikkanryougengo no kotowaza no imi no hikakukenkyuu—"kao" o megutte [A comparative study on the meaning of Japanese and Korean proverbs containing body parts terms: Concerning 'the face']. *Kyouiku Jissengaku Ronshuu [Journal for the Science of Schooling]*, *10*, 89–102.

Hale, K. (1992). Language endangerment and the human value of linguistic diversity. *Language*, *68*(1), 35–42.

Hale, K. & Keyser, S. J. (1987). *A view from the middle*. MIT Center for Cognitive Science.

Hale, K., Krauss, M., Watahomigie, L. J., Yamamoto, A. Y., Craig, C., Jeanne, L. M. & England, N. C. (1992). Endangered languages. *Language*, *68*(1), 1–42.

Hamano, S. (1998). *The Sound-Symbolic System of Japanese*. CSLI Publications.

Hammarström, H. (2008). Counting Languages in Dialect Continua Using the Criterion of Mutual Intelligibility. *Journal of Quantitative Linguistics*, *15*(1), 34–45.

Hammarström, H., Forkel, R., Haspelmath, M. & Bank, S. (2020). *Glottolog* (4.2.1). Max Planck Institute for the Science of Human History.

Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M. & Horai, S. (2006). Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *Journal of Human Genetics*, *51*(1), 47–58.

Han, M. S. (1962). The feature of duration in Japanese. *Study Sounds*, *10*, 65–80.

Hanihara, K. (1991). Dual structure model for the population history of the Japanese. *Japan Review*, *2*, 1–33.

Hasada, R. (2002). 'Body part' terms and emotion in Japanese. *Pragmatics & Cognition*, *10*(1–2), 107–128.

Hashimoto, H. (2007). Heian/Kamakura-jidai ni okeru dōshi kohoru/hohotsu no imi [The meaning of the verbs kohoru/kohotsu in the Heian and Kamakura periods]. *Bungei Kenkyuu*, *164*, 28–38.

Hashio, N. (1992). Hyōgo-ken Kōbe-shi Suma hōgen ni okeru hasaidōshi 'megu', 'kaku', 'kudaku', 'waru', 'kowasu' no imi [The meaning of the verbs 'megu', 'kaku', 'kudaku', 'waru', 'kowasu' in the Suma dialect of Kobe City in Hyogo Prefecture]. *Nihongo Kenkyū*, *13*, 69–82.

Haspelmath, M. & Tadmor, U. (2009). *Loanwords in the world's languages: A comparative handbook*. Walter de Gruyter.

Hattori, S. (1954). On the Method of Glottochronology and the Time-Depth of Proto-Japanese. *Gengo Kenkyū, Journal of the Linguistic Society of Japan*, *26–27*, 29–77.

Hattori, S. (1961). A Glottochronological Study on Three Okinawan Dialects. *International Journal of American Linguistics*, *27*(1), 52–62.

Hattori, S. (1976). Ryūkyū hōgen to hondo hōgen [The Ryukyu dialects and the mainland dialects]. In Iha Fuyū 100th Anniversary Commemoration Association (Ed.), *Okinawagaku no reimei [The dawn of Okinawan Studies]* (pp. 7–55). Okinawa Bunka Kyõkai.

Haynie, H. J. & Bowern C. (2016). Phylogenetic approach to the evolution of color term systems. *Proceedings of the National Academy of Sciences*, *113*(48), 13666–13671.

Heeringa, W. & Nerbonne, J. (2001). Dialect areas and dialect continua. *Language Variation and Change*, *13*(3), 375–400.

Heine, B. (1997). *Cognitive foundations of grammar*. Oxford University Press.

Heinrich, P. (2005). What Leaves a Mark Should no Longer Stain. *Refereed Papers from the 1st International Small Islands Cultures Conference*, 61–72.

Heinrich, P. (2009). The Ryukyuan languages in the 21st century global society. *Human Migration and the 21st Century Global Society. Okinawa: University of the Ryukyus*, 16–27.

Heinrich, P., Miyara, S. & Shimoji, M. (2015). Introduction: Ryukyuan languages and Ryukyuan linguistics. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages. History, Structure, and Use* (pp. 1–10). De Gruyter Mouton.

Hellwig, B. (2006). Field semantics and grammar-writing: Stimulibased techniques and the study of locative verbs. In F. K. Ameka, A. Dench & N. Evans (Eds.), *Catching Language: The Standing Challenge of Grammar Writing*. Mouton de Gruyter.

Hendrick-Wong, E. (2013). Purple Mania: How a Snail Created the Most Sought After Colour of All Time. *Textiles Asia*, *4*(3), 7–13.

Hill, J. & Hill, K. (1977). Language Death and Relexification in Tlaxcalan Nahuatl. *Linguistics*, *15*(191), 55–70.

Hinds, J. V. (1974). Make mine burakku. *Language Research*, *10*(2), 92–108.

Hirayama, T. (1960). *Zenkoku Akusento Jiten [Countrywide Pitch Accent Dictionary]*. Tōkyōdō.

Hirayama, T. (1965). *Izu-shotō hōgen no kenkyū*. Meijishoin.

Hirayama, T. (1966). *Ryūkyū hōgen no sōgōteki kenkyū [A comprehensive study of the Ryukyuan dialects]*. Meijishoin.

Hirayama, T. (1992). *Gendai Nihongo Hōgen Daijiten [Dictionary of Contemporary Japanese Dialects]*. Meiji-shoin.

Hojo, H. (1983). A maximum likelihood method for additive clustering and its applications. *Japanese Psychological Research*, *25*(4), 191–201.

Hojo, H. (1991). An analysis of three Japanese verbs of breaking: An empirical study of word meaning. *Japanese Psychological Research*, *33*(2), 64–74.

Hojo, H. (1993). A new nonmetric multidimensional scaling method for sorting data. *Japanese Psychological Research*, *35*(3), 129–139.

Hōjō, T. (1948). Hachijō hōgen no kenkyū-tokuni jōdaisei no izon ni tsuite-(1) [A study of Hachijo dialect-focusing on retentions from Old Japanese-(1)]. *Nihon No Kotoba*, *6*, 84–88.

Hōjō, T. (1966). *Jōdai Azuma hōgen no kenkyū [Study of the Eastern Old Japanese dialect]*. Nihon gakujutsu sjinkōkai.

Hong, W. (2005). Yayoi Wave, Kofun Wave, and Timing: The Formation of the Japanese People and Japanese Language. *Korean Studies*, *29*, 1–29.

Honkola, T., Ruokolainen, K., Syrjänen, K. J. J., Leino, U.-P., Tammi, I., Wahlberg, N. & Vesakoski, O. (2018). Evolution within a language: Environmental differences contribute to divergence of dialect groups. *BMC Evolutionary Biology*, *18*(132), 1–15.

Hudson, M. J. (2002). Agriculture and language change in the Japanese Islands. In P. Bellwood & C. Renfrew (Eds.), *Examining the Farming/Language Dispersal Hypothesis* (pp. 311–318). McDonald Institute.

Hudson, M. J., Nakagome, S. & Whitman, J. B. (2020). The evolving Japanese: The dual structure hypothesis at 30. *Evolutionary Human Sciences*, *2*, e6.

Hughes, J. (2018). *reghelper: Helper Functions for Regression Analysis*. Version 0.3.4. Available from: https://github.com/jeff-hughes/reghelper/.

Huisman, J. L. A., Majid A. & van Hout, R. (2019). The geographical configuration of a language area influences linguistic diversity. *PLoS ONE*, *14*(6), e0217363.

Hundertmark, K.J. & Van Daele, L. J. (2010). Founder effect and bottleneck signatures in an introduced, insular population of elk. *Conservation Genetics*, *11*(1), 139–147.

Hutchison, D. W. & Templeton, A. R. (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, *53*(6), 1898–1914.

Iannucci, D. J. (2019). *The Hachijo Language of Japan: Phonology and Historical Development* [PhD thesis]. University of Hawai'i.

Iijima, T., Wenning, W. & Zollinger H. (1982). Cultural Factors of Color Naming in Japanese: Naming Tests with Japanese Children in Japan and Europe. *Anthropological Linguistics*, *24*(2), 245–262.

Iitoyo, K., Hino, S. & Satō, R. (1983). *Kōza Hōgengaku 6: Chūbu-chihō no Hōgen [A Course on Dialectology 6: Dialects of the Central Japan]*. Kokusho Kankōkai.

Inoue, F. (1996a). Computational dialectology (1). *Area and Culture Studies*, *52*, 67–102.

Inoue, F. (1996b). Computational dialectology (2). *Area and Culture Studies*, *53*, 115–134.

Inoue, F. (2004a). Multivariate Analysis, Geographical Gravity Centers and the History of Standard Japanese Forms. *Area and Culture Studies*, *68*, 15–36.

Inoue, F. (2004b). Geographical Factors of Communication on the Basis of Usage Rate of the Standard Japanese Forms and Railway Distance. *The Japanese Journal of Language in Society*, 7(1), 19–29.

Inoue, F. (2019). Common developments of Japanese and Dutch-Flemish dialectology: Computational dialectology in Japan. *Dialectologia: Revista Electrònica*, 35–64.

Inoue, F. & Kasai, H. (1989). Dialect classification by standard Japanese forms. In S. Mizuntani (Ed.), *Japanese Quantitative Linguistics* (pp. 220–235). Studienverlag Dr. N. Brockmeyer.

Itou, T. (2006). Butsugo couper to Nihongo waru/kiru no imironteki kōsetsu [A semantic study of French couper and Japanese waru/kiru]. *Nagoya University of Foreign Studies Journal of School of Foreign Languages*, *30*, 83–105.

Iwasaki, N., Sells, P. & Akita, K. (Eds.). (2017). *The grammar of Japanese Mimetics: Perspectives from structure, acquisition, and translation*. Routledge.

Iwasaki, S. (2013). *Japanese*. Amsterdam/Philadelphia: John Benjamins.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, *37*, pp. 547–579.

Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New Phytologist*, *11*(2), pp. 37–50.

Jäger, G. (2012). Using Statistics for Cross-linguistic Semantics: A Quantitative Investigation of the Typology of Colour Naming Systems. *Journal of Semantics*, *29*(4), 521–544.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlỳ, P. & Suchomel, V. (2013). The tenten corpus family. *7th International Conference on Corpus Linguistics*, 125–127.

Jameson, K. A. & Alvarado, N. (2003). Differences in color naming and color salience in Vietnamese and English. *Color Research & Application*, *28*, 113–138.

Jeszenszky, P., Hikosaka, Y., Imamura, S. & Yano, K. (2019). Japanese Lexical Variation Explained by Spatial Contact Patterns. *ISPRS International Journal of Geo-Information*, *8*(9), 400.

Jodaigo Jiten Henshu Iinkai (Ed.). (1967). *Jidaibetsu kokugo daijiten Joudai hen*. Sanseido.

Kaetsu, M. (1979). Yaburu, yabuku, saku, kiru, waru. *Nihongo Kenkyu*, *2*, 57–61.

Kageyama, T. (2016). Verb-compounding and verb-incorporation. In T. Kageyama & H. Kishimoto (Eds.), *Handbook of Japanese Lexicon and Word Formation* (pp. 273–310). De Gruyter Mouton.

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Kanwisher, N., McDermott, J. & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), pp. 4302–4311.

Karimata, S. (1999). Onsei no men kara mita Ryūkyū shohōgen [The Ryukyuan dialects seen from the point of view of phonetics]. In Gengogaku Kenkyūkai (Ed.), *Kotoba no Kagaku [The Science of Language]* (pp. 13–85). Mugi Shobō.

Kawaguchi, Y. & Inoue, F. (2002). Japanese Dialectology in Historical Perspectives. *Revue Belge de Philologie et d'Histoire*, *80*(3), 801–829.

Kawamoto, T. (1977). Toward a Comparative Japanese-Austronesian. *Bulletin of the Nara University of Education*, *26*(1), 23–49.

Kay, P., Berlin, B., Maffi, L. & Merrifield, W. R. (1997), Color naming across languages. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language* (21–56). Cambridge: Cambridge University Press.

Kay, P., Berlin, B. & Merrifield, W. (1991). Biocultural Implications of Systems of Color Naming. *Journal of Linguistic Anthropology*, *1*(1), 12–25.

Kay, P. & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, *100*(15), 9085–9089.

Kay, P. (2015). Universality of color categorization. In A. J. Elliot, M. D. Fairchild & A. Franklin (Eds.), *Handbook of Color Psychology* (pp. 243–244). Cambridge University Press.

Kay, Paul, Berlin, B., Maffi, L., Merrifield, W. R. & Cook, R. S. (2009). *The World Color Survey*. CSLI Publications.

Kerswill, P. (1996). Dialect levelling, koinéisation and the speech of the adult migrant. *University of Pennsylvania Working Papers in Linguistics*, *3*(1), 237–246.

Kerswill, P. (2002). Koineization and accommodation. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 669–702). Blackwell.

Kessler, B. (1995). Computational dialectology in Irish Gaelic. *EACL '95: Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics*, 60–66.

Kibe, N. (Ed.). (2013). *Hachijō hōgen chōsa hōkokusho [Research Report on Hachijō Dialect]*. National Institute for Japanese Language and Linguistics.

Kibe, N. (Ed.). (2016). *Yoron hōgen / Okinoerabu hōgen chōsa hōkokusho [Research Report on Yoron and Okinoerabu Dialects]*. National Institute for Japanese Language and Linguistics.

Kibe, N. & Aoi, H. (Eds.). (2020). *Aomori-ken Mutsu-shi hōgen chōsa hōkokusho [Research Report on Mutsu City dialect in Aomori Prefecture]*. National Institute for Japanese Language and Linguistics.

Kindaichi, H. (1955). Nihongo (hōgen) [Japanese (dialects)]. In S. Ichikawa & H. Kōzu (Eds.), *Sekai Gengo Gaisetsu* (pp. 212–238). Kenkyūsha.

Kindaichi, H. (1978). *The Japanese Language* (U. Hirano, Trans.). Tuttle Publishing.

Kishimoto, H. & Uehara, S. (2016). 2. Lexical categories. In T. Kageyama & H. Kishimoto (Eds.), *Handbook of Japanese Lexicon and Word Formation* (pp. 51–92). De Gruyter Mouton.

Kita, S. (2006). A grammar of space in Japanese. In S. C. Levinson & D. P. Wilkins (Eds.), *Grammars of space: Explorations in cognitive diversity* (pp. 437–474). Cambridge University Press.

Kopecka, A. & Narasimhan, B. (2012). *Events of putting and taking: A crosslinguistic perspective*. John Benjamins.

Koptjevskaja-Tamm, M. (Ed.). (2015). *The Linguistics of Temperature*. John Benjamins.

Koptjevskaja-Tamm, M., Rakhilina, E. & Vanhove, M. (2016). The semantics of lexical typology. In N. Riemer (Ed.), *The Routledge Handbook of Semantics* (pp. 434–454). Routledge.

Kortmann, B. (2004). *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*. Mouton de Gruyter.

Kövecses, Z. (2003). *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.

Koyama, S. & Thomas, D. H. Affluent foragers: Pacific Coasts East and West. *American Anthropologist, 86*(1), 188–190.

Kraska-Szlenk, I. (2014). Semantic extensions of body part terms: Common patterns and their interpretation. *Language Sciences*, *44*, pp. 15–39.

Kristol, A. M. (1980). Color Systems in Southern Italy: A Case of Regression. *Language*, *56*(1), 137–145.

Kroeber, A. L. & Chrétien, C. D. (1937). Quantitative Classification of Indo-European Languages. *Language*, *13*(2), 83–103.

Kroeber, A. L. & Chrétien, C. D. (1939). The Statistical Technique and Hittite. *Language*, *15*(2), 69.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27.

Kunihiro, T. (1970). *Imi no Shosō [Aspects of Meaning]*. Sanseidō.

Kuriki, I., Lange, R., Muto, Y., Brown, A. M., Fukuda, K., Tokunaga, R., Lindsey, D. T., Uchikawa, K. & Shioiri, S. (2017). The modern Japanese color lexicon. *Journal of Vision*, *17*(3), 1.

Kusakabe, F. (1964). Aka to Ao—Okinawa, Miyako, Yaeyama o megutte [Red and blue around Okinawa, Miyako and Yaeyama]. *Kokugogaku*, *58*, 60–75.

Kuzmin, Y. V. (2006). Chronology of the earliest pottery in East Asia: Progress and pitfalls. *Antiquity*, *80*(308), 362–371.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26.

Kwon, G.-S. (2013). Nikkan ryoukokugo no setsudandōshirui no kōsatsu [A study of cutting verbs in Japanese and Korean]. *Japanese Language and Literature Studies*, *58*, 5–26.

Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.

Labov, W. (1994). *Principles of Linguistic Change 1: Internal Factors*. Blackwell.

Labov, W. (2001). *Principles of Linguistic Change 2: Social Factors*. Blackwell.

Labov, W. (2010). *Principles of Linguistic Change 3: Cognitive and Cultural Factors*. Wiley-Blackwell.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic Books.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Lê, S., Josse, J. & Husson, F. (2008). *FactoMineR: An R Package for Multivariate Analysis*. *25*(1), 1–18.

Lee, S. & Hasegawa, T. (2011). Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1725), 3662–3669.

Lee, S. & Hasegawa, T. (2014). Oceanic barriers promote language diversification in the Japanese Islands. *Journal of Evolutionary Biology*, *27*(9), 1905–1912.

Lees, R. B. (1953). The Basis of Glottochronology. *Language*, *29*(2), 113–127.

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*(5), 573–579.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady*, *10*(8), 707–710.

Levinson, S. C. (2000). Yélî Dnye and the Theory of Basic Color Terms. *Journal of Linguistic Anthropology*, *10*(1), 3–55.

Levinson, S. C. (2007). Cut and break verbs in Yélî Dnye, the Papuan language of Rossel Island. *Cognitive Linguistics*, *18*(2), 207–218.

Levinson, S. C. & Wilkins, D. P. (Eds.). (2006). *Grammars of space: Explorations in cognitive diversity*. Cambridge University Press.

Levisen, C. (2015). Scandinavian semantics and the human body: An ethnolinguistic study in diversity and change. *Language Sciences*, *49*, pp. 51–66.

Li, Z. K. (2018). Setsudandoushi kiru to jareuda no taishoukenkyuu [A contrastive study of the cutting verbs kiru and jareuda]. *Japanese Language and Literature Studies*, *77*, 193–210.

Lillo, J., González-Perilli, F., Prado-León, L., Melnikova, A., Álvaro, L., Collado, J. A. & Moreira, H. (2018). Basic Color Terms (BCTs) and Categories (BCCs) in Three Dialects of the Spanish Language: Interaction Between Cultural and Universal Factors. *Frontiers in Psychology*, *9*, 761.

Lindsey, D. T. & Brown, A. M. (2004). Sunlight and 'Blue': The Prevalence of Poor Lexical Color Discrimination Within the 'Grue' Range. *Psychological Science*, *15*(4), 291–294.

Lucy, J. A. (1997). The linguistics of color. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language* (320–346). Cambridge: Cambridge University Press.

Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers*, *28*(2), 203–208.

MacLaury, R. E. (1986). *Color in Meso-America: Vol I. A theory of composite categorization* [PhD thesis]. Berkeley: University of California.

MacLaury, R. E. (1991), Social and Cognitive Motivations of Change: Measuring Variability in Color Semantics. *Language*, *67*(1), 34–62.

MacLaury, R. E. (2001). Color terms. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language Typology and Language Universals* (Vol. 1). Walter de Gruyter.

MacLaury, R. E., Hewes, G. W., Kinnear, P. R., Deregowski, J. B., Merrifield, W. R., Saunders, B. A. C., Stanlaw, J., Toren, C., Van Brakel, J. & Wescott, R. W. (1992). From Brightness to Hue: An Explanatory Model of Color-Category Evolution [and Comments and Reply]. *Current Anthropology*, *33*(2), 137–186.

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M. & Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, *48*(2), 345–371.

Majid, A. (2011). A Guide to Stimulus-Based Elicitation for Semantic Categories. In N. Thieberger (Ed.), *The Oxford Handbook of Linguistic Fieldwork* (pp. 54–71). Oxford University Press.

Majid, A. (2019). Mapping words reveals emotional diversity. *Science*, *366*(6472), pp. 1444–1445.

Majid, A., Boster, J. S. & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, *109*(2), 235–250.

Majid, A. & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*(2), pp. 266–270.

Majid, A. & Enfield, N. J. (2017). Body. In H. Burkhardt, J. Seibt, G. Imaguire & S. Gerogiorgakis (Eds.), *Handbook of Mereology* (pp. 100–103). Philosophia.

Majid, A., Enfield, N. J. & van Staden, M. (Eds.). (2006). Parts of the body: Cross-linguistic categorisation. *Language Sciences*, *28*(2–3), 137–360.

Majid, A., Gullberg, M., van Staden, M. & Bowerman, M. (2007). How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. *Cognitive Linguistics*, *18*(2), 179–194.

Majid, A., Jordan, F. & Dunn, M. (2011). *Evolution of Semantic Systems: Procedures Manual*. Nijmegen: Max Planck Institute for Psycholinguistics.

Majid, A., Jordan, F. & Dunn, M. (2015). Semantic systems in closely related languages. *Language Sciences*, *49*, 1–18.

Majid, A. & Levinson, S. C. (2011). The senses in language and culture. *The Senses and Society*, *6*(1), 5–18.

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., Woll, B., LeLan, B., de Sousa, H., Cansler, B. L., Shayan, S., de Vos, C., Senft, G., Enfield, N. J., Razak, R. A., Fedden, S., Tufvesson, S., Dingemanse, M., Ozturk, O., … Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369–11376.

Majid, A. & van Staden, M. (2015). Can Nomenclature for the Body be Explained by Embodiment Theories? *Topics in Cognitive Science*, *7*(4), 570–594.

Malt, B. C., Ameel, E., Imai, M., Gennari, S. P., Saji, N. & Majid, A. (2014). Human locomotion in languages: Constraints on moving and meaning. *Journal of Memory and Language*, *74*, 107–123.

Marck, J. C. (1986). Micronesian dialects and the overnight voyage. *The Journal of the Polynesian Society*, *95*(2), 253–258.

Martin, S. E. (1966). Lexical evidence relating Korean to Japanese. *Language*, *42*(2), 185–251.

Martin, S. E. (1991). Morphological clues to the relationships of Japanese and Korean. In P. Baldi (Ed.), *Patterns of Change—Change of Patterns: Linguistic Change and Reconstruction Methodology* (pp. 235–62). De Gruyter Mouton.

Matsumori, A. (1995). Ryūkyuan: Past, present, and future. *Journal of Multilingual and Multicultural Development*, *16*(1–2), 19–44.

Mayr, E. (1943). *Systematics and the Origin of Species*. New York: Columbia University Press.

McLean, B. (in press). Revising an implicational hiearchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology*.

McNeill, N. B. (1972). Colour and Colour Terminology. *Journal of Linguistics*, *8*(1), 21–33.

McVeigh, B. (1996). Standing stomachs, clamoring chests and cooling livers: Metaphors in the psychological lexicon of Japanese. *Journal of Pragmatics*, *26*(1), 25–50.

Merriam, C. H. (1907). Distribution and Classification of the Mewan Stock of California. *American Anthropologist*, *9*(2), 338–357.

Miyara, S. (2015). Shuri Okinawan grammar. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages* (pp. 379–404). De Gruyter Mouton.

Miyara, T. (1930). *Yaeyama Goi [Yaeyama Vocabulary]*. Tōyō Bunko.

Miyara, T. (1980). *Miyara Toso Zenshu [Complete Works of Toso Miyara]*. Daiichi Shobo.

Moseley, C. (2010). *Atlas of the World's Languages in Danger*. UNESCO Publishing.

Mufwene, S. S. (1996). The founder principle in creole genesis. *Diachronica*, *13*(1), 83–134.

Mufwene, S. S. (2001). *The Ecology of Language Evolution*. Cambridge University Press.

Musgrave, S. (2016). Language Endangerment. In K. Allan (Ed.), *The Routledge Handbook of Linguistics* (pp. 385–400). Routledge.

Nakagawa, R., Doi, N., Nishioka, Y., Nunami, S., Yamauchi, H., Fujita, M., Yamazaki, S., Yamamoto, M., Katagiri, C. & Mukai, H. (2010). Pleistocene human remains from Shiraho-Saonetabaru Cave on Ishigaki Island, Okinawa, Japan, and their radiocarbon dating. *Anthropological Science*, *118*(3), 173–183.

Nakama, M. (1978). Setsudandōshi sui, ki:, batsi ni tsuite—Okinawa Miyako Nishihara hōgen no rei [On the cutting verbs sui, ki:, batsi—Examples from the Nishihara dialect in Miyako, Okinawa]. *Nihongo Kenkyu*, *1*, 156–164.

Nakama, M. (1984). Hasaidōshi no imironteki kenkyū-1-Okinawa no goikyōiku ni kansuru shiron [A Semantic Study of Destruction Verbs 1—Essays on Okinawan Vocabulary Education]. *Bulletin of the College of Education*, *27*, 87–94.

Nakama, M. (1985). Hasaidōshi no imironteki kenkyū-2-Okinawa no goikyōiku ni kansuru shiron [A Semantic Study of Destruction Verbs 2—Essays on Okinawan Vocabulary Education]. *Bulletin of the College of Education*, *28*, 59–67.

Nakamoto, M. (1981). *Zusetu Ryukyugo Jiten [Illustrated Dictionary of the Ryukyuan Languages]*. Rikitomi Shobō.

Naroll, R. (1970). What Have We Learned from Cross-Cultural Surveys? *American Anthropologist*, *72*(6), 1227–1288.

Nash, J., Bakker, P., Bøegh, K. F., Daval-Markussen, A., Haberland, H., Kedwards, D., Ladhams, J., Levisen, C., Markússon, J. S., Robbe, J. & Willemsen, J. (2020). On languages on islands. *Acta Linguistica Hafniensia*, *52*(1), 81–116.

National Insitute for Japanese Language and Linguistics. (1963). *Okinawago Jiten*. Zaimushō Insatsukyoku [Printing Bureau of the Ministry of Finance].

National Insitute for Japanese Language and Linguistics. (1966). *Nihon Gengo Chizu [Linguistic Atlas of Japan]*. Ōkurashō Insatsukyoku [Printing Bureau of the Ministry of Finance].

National Insitute for Japanese Language and Linguistics. (1968). *Nihon Gengo Chizu 3 [Linguistic Atlas of Japan 3]*. Ōkurashō Insatsukyoku [Printing Bureau of the Ministry of Finance].

National Insitute for Japanese Language and Linguistics. (1978). *Hōgen Danwa Shiryō [Texts of tape-recorded conversations in Japanese dialects]*. Shūei Shuppan.

National Insitute for Japanese Language and Linguistics. (1989). *Hōgen Bunpō Zenkoku Chizu [Grammar Atlas of Japan]*. Zaimushō Insatsukyoku [Printing Bureau of the Ministry of Finance].

National Insitute for Japanese Language and Linguistics. (2001). *Nihon no Furusato Kotoba Shūsei [Collection of Japanese Dialects]*. Kokusho Kankōkai.

Nerbonne, J. (2010). Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1559), 3821–3828.

Nerbonne, J. (2013). How much does geography influence language variation? In Auer, P., Hilpert, M., Stukenbrock, A. & Szmrecsanyi, B. (Eds.). *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, 222–239. Berlin: De Gruyter.

Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P. & Leinonen, T. (2011). Gabmap – A web application for dialectology. *Dialectologia*, *Special Issue II*, 65–89.

Nerbonne, J. & Heeringa, W. (2007). Geographic distributions of linguistic variation reflect dynamics of differentiation. In Featherston, S. & Sternefeld, W. (Eds.). *Roots: Linguistics in Search of its Evidential Base*, 267–298. Berlin: De Gruyter.

Nettle, D. (1998). Explaining global patterns of language diversity. *Journal of Anthropological Archaeology*, *17*(4), 354–374.

Nettle, D. (1999). *Linguistic Diversity*. Oxford: Oxford University Press.

Niinaga, Y. (2015). Amami grammar. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages* (pp. 323–344). De Gruyter Mouton.

Ogawa, A. (1984). Nihongo Eigo no goikenkyuu—dageki/hakai/secchi doushi [A study of Japanese and English vocabulray—hitting, destruction and placement verbs]. *Essays from the Yokohama College of Commerce*, *17*(1–2), 71–88.

Okinawa Kogo Daijiten Henshu Iinkai. (1995). *Okinawa Kogo Daijiten*. Kadokawa Shoten.

Onishi, T. (2010). Analyzing dialectological distributions of Japanese. *Dialectologia: Revista Electrònica*, *Special Issue 1*, 123–135.

Ōno, S. (1970). *The Origin of the Japanese Language*. Kokusai Bunka Shinkōkai.

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J. & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, *22*(24), 5983–5999.

Özgen, E. & Davies, I. R. L. (1998). Turkish color terms: tests of Berlin and Kay's theory of color universals and linguistic relativity. *Linguistics*, *36*(5), 919–956.

Paramei, G. V. (2005). Singing the Russian Blues: An Argument for Culturally Basic Color Terms. *Cross-Cultural Research*, *39*(1), 10–38.

Patiño, J., Whittaker, R. J., Borges, P. A. V., Fernández-Palacios, J. M., Ah-Peng, C., Araújo, M. B., Ávila, S. P., Cardoso, P., Cornuault, J., de Boer, E. J., de Nascimento, L., Gil, A., González-Castro, A., Gruner, D. S., Heleno, R., Hortal, J., Illera, J. C., Kaiser-Bunbury, C. N., Matthews, T. J., … Emerson, B. C. (2017). A roadmap for island biology: 50 fundamental questions after 50 years of The Theory of Island Biogeography. *Journal of Biogeography*, *44*(5), 963–983.

Pawley, A. & Green, R. Dating the dispersal of the Oceanic languages. *Oceanic Linguistics*, *12*(1/2), 1–67.

Pearson, R. (2013). *Ancient Ryukyu: An archaeological study of island communities*. University of Hawai'i Press.

Pellard, T. (2009). *Ōgami: Éléments de description d'un parler du Sud des Ryūkyū*. Ecole des Hautes Etudes en Sciences Sociales.

Pellard, T. (2010). Ōgami (Miyako Ryukyuan). In M. Shimoji & T. Pellard (Eds.), *An Introduction to Ryukyuan languages* (pp. 113–166). ILCAA.

Pellard, T. (2011). The historical position of the Ryukyuan Languages. *Historical Linguistics in the Asia-Pacific Region and the Position of Japanese*, 55–64.

Pellard, T. (2015). The linguistic archeology of the Ryukyu Islands. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages. History, Structure, and Use* (pp. 13–38). De Gruyter Mouton.

Perkins, R. D. (1989). Statistical techniques for determining language sample size. *Studies in Language*, *13*(2), 293–315.

Port, R. F., Dalby, J. & O'Dell, M. (1987). Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America*, *81*(5), 1574–1585.

Post, M. W. (2013). The Siyom River Valley: An essay on intra-subgroup convergence in Tibeto-Burman. *North East Indian Linguistics*, *5*, 60–90.

Post, M. W. (2015). Morphosyntactic reconstruction in an arealhistorical context. In N. J. Enfield & B. Comrie (Eds.), *Languages of Mainland Southeast Asia: The State of the Art* (pp. 209–265). De Gruyter Mouton.

Pronk, T. (2009). *The Slovene Dialect of Egg and Potschach in the Gailtal, Austria*. Rodopi.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Regier, T., Kay, P. & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*(4), 1436–1441.

Regier, T., Kay, P. & Khetarpal, V. (2009). Color naming and the shape of color space. *Language*, *85*(4), 884–892.

Riemer, N. (2006). Reductive Paraphrase and Meaning: A Critique of Wierzbickian Semantics. *Linguistics and Philosophy*, *29*(3), 347–379.

Rijkhoff, J. & Bakker, D. (1998). Language sampling. *Linguistic Typology*, *2*(3), 263–314.

Robbeets, M. I. (2005). *Is Japanese Related to Korean, Tungusic, Mongolic and Turkic?* Harrassowitz Verlag.

Ronce, O. (2007). How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics*, *38*, 231–253.

Sapir, E. (1921). *An introduction to the study of speech*. Harcourt Brace.

Satō, T. (2010). Ainugo no fukkō to Ainugo kenkyū [Revival of the Ainu Language and its study]. In Hokkaido University Center for Ainu and Indigenous Studies (Ed.), *Ainu kenkyū no genzai to mirai [The present and future of Ainu Studies]* (pp. 278–303). Hokkaidō Daigaku Shuppankai.

Sato, T., Nakagome, S., Watanabe, C., Yamaguchi, K., Kawaguchi, A., Koganebuchi, K. et al. (2014). Genome-Wide SNP Analysis Reveals Population Structure and Demographic History of the Ryukyu Islanders in the Southern Part of the Japanese Archipelago. *Molecular Biology and Evolution*, *31*(11), 2929–2940.

Séguy, J. (1971). La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane*, *35*(138), 335–357.

Séguy, J. (1973). La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de Linguistique Romane*, *37*(145), 1–24.

Seifart, F., Evans, N., Hammarström, H. & Levinson, S. C. (2018). Language documentation twenty-five years on. *Language*, *94*(4), e324–e345.

Seol, H. (2016). Kankokujin Nihongo gakushūsha ni yoru setsudan-hakaidōshi shūtoku [Acquisition of cutting and breaking verbs in Korean learners of Japanese]. *Nagoya University Course on Japanese Education*, 30–33.

Serafim, L. A. (2003). When and from where did the Japonic language enter the Ryukyus?–A critical comparison of language, archaeology, and history. In A. Vovin & T. Osada (Eds.), *Perspectives on the Origins of the Japanese Language* (pp. 463–476). International Research Center for Japanese Studies.

Serva, M. & Petroni, F. (2008). Indo-European languages tree by Levenshtein distance. *EPL (Europhysics Letters)*, *81*(6), 68005.

Shibata, T. (1953). Yamagata-ken Ōtori-hogen no onsobunseki [Analysis of the phonemes in Ōtori dialect in Yamagata Prefecture]. In *Gengo Minzoku Ronsō—Kindaichi-hakushi koki kinen [Essay on Language and Folklore—in celebration of Professor Kindaichi's 70th birthday]*. Sanseidō.

Shibatani, M. (1990). *The Languages of Japan*. Cambridge University Press.

Shigeno, H. (2010). Ura (Amami Ryukyuan). In M. Shimoji & T. Pellard (Eds.), *An Introduction to Ryukyuan languages* (pp. 15–34). ILCAA.

Shigeno, H., Shimoji, K., Matayoshi, S. & Nishioka, S. (2015). A selected bibliography of Ryukyuan dialectology. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages* (pp. 705–720). De Gruyter Mouton.

Shindo, M. (2015). Subdomains of temperature concepts in Japanese. In M. Koptjevskaja-Tamm (Ed.), *The Linguistics of Temperature* (pp. 639–665). John Benjamins.

Shinoda, K. & Adachi, N. (2017). Ancient DNA Analysis of Palaeolithic Ryukyu Islanders. In P. J. Piper, H. Matsumura & D. Bulbeck (Eds.), *New Perspectives in Southeast Asian and Pacific Prehistory* (p. 51). Australian National University Press.

Shogakukan (2002). *Nihon Kokugo Daijiten [Unabridged Dictionary of the Japanese Language]*. Tokyo: Shogakukan

Simpson, E. H. (1949). Measurement of Diversity. *Nature*, *163*(4148), 688–688.

Slobin, D. I., Ibarretxe-Antuñano, I., Kopecka, A. & Majid, A. (2014). Manners of human gait: A crosslinguistic event-naming study. *Cognitive Linguistics*, *25*(4), 701–741.

Song, Jae Jung (Ed.). (2010). *The Oxford Handbook of Linguistic Typology*. Oxford University Press.

Spurgin, L. G., Illera, J. C., Jorgensen T, H., Dawson, D. A. & Richardson, D. S. (2014). Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology*, *23*(5), 1028–1039.

Stanford, J. N. (2012). One size fits all? Dialectometry in a small clan-based indigenous society. *Language Variation and Change*, *24*(2), 247–278.

Stanlaw, J. (1987). *Colors, culture, and contact: English loanwords and problems of color nomenclature in modern Japanese* [PhD dissertation]. University of Illinois.

Stanlaw, J. (1997). Two observations on culture contact and the Japanese color nomenclature system. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language* (pp. 240–260). Cambridge University Press.

Stanlaw, J. (2007). Japanese color terms, from 400 CE to the present: Literature, orthography, and language contact in light of current cognitive theory. In R. E. MacLaury, G. V. Paramei & D. Dedrick (Eds.), *Anthropology of Color: Interdisciplinary multilevel modeling* (pp. 295–318). John Benjamins.

Stanlaw, J. (2010). Language, contact, and vantages: Fifteen hundred years of Japanese color terms. *Language Sciences*, *32*(2), 196–224.

Steinicke, E., Walder, J., Löffler, R. & Beismann, M. (2012). Autochthonous Linguistic Minorities in the Italian Alps: New Legislation – New Identifications – New Demographic Processes. *Revue de Géographie Alpine | Journal of Alpine Research*, *99*(2).

Stepp, J. R., Castaneda, H. & Cervone, S. (2005). Mountains and biocultural diversity. *Mountain Research and Development*, *25*(3), 223–227.

Stern, G. (1931). *Meaning and change of meaning*. Indiana University Press.

Summerhayes, G. R. & Anderson, A. (2009). An Austronesian Presence in Southern Japan: Early Occupation in the Yaeyama Islands. *Bulletin of the Indo-Pacific Prehistory Association*, *29*, 76–91.

Svorou, S. (1993). *The Grammar of Space*. Amsterdam/Philadelphia: John Benjamins.

Swadesh, M. (1948). Sociologic Notes on Obsolescent Languages. *International Journal of American Linguistics*, *14*(4), 226–235.

Swadesh, M. (1950). Salish Internal Relationships. *International Journal of American Linguistics*, *16*(4), 157–167.

Swadesh, M. (1952). Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society*, *96*(4), 452–463.

Swadesh, M. (1954). Perspectives and Problems of Amerindian Comparative Linguistics. *WORD*, *10*(2–3), 306–332.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, *21*, 121-137.

Tadmor, U. (2009). Loanwords in the world's languages: Findings and results. In Haspelmath, M. & Tadmor, U. (Eds.). *Loanwords in the World's Languages: A Comparative Handbook* (pp. 55–75).

Takahashi, T. & Ihara, Y. (2020). Quantifying the spatial pattern of dialect words spreading from a central population. *Journal of The Royal Society Interface*, *17*(168), 20200335.

Takamiya, H. (2001). The transition from foragers to farmers on the island of Okinawa. *Bulletin of the Indo-Pacific Prehistory Association*, *21*, 60–67.

Takamiya, H. (2005). *Shima no Senshigaku [The Prehistory of the Okinawa Archipelago]*. Border Ink.

Takamiya, H., Hudson, M. J., Yonenobu, H., Kurozumi, T. & Toizumi, T. (2016). An extraordinary case in human history: Prehistoric hunter-gatherer adaptation to the islands of the Central Ryukyus (Amami and Okinawa archipelagos), Japan. *The Holocene*, *26*(3), 408–422.

Takeuchi, F., Katsuya, T., Kimura, R., Nabika, T., Isomura, M., Ohkubo, T. et al. (2017). The fine-scale genetic structure and evolution of the Japanese population. *PLoS ONE*, *12*(11), e0185487.

Takubo, Y. (2018). Mutual Intelligibility as a measure of linguistic distance and intergenerational transmission. *Approaches to Endangered Languages in Japan and Northeast Asia: Description, Documentation and Revitalization*.

Tang, C. & Van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, *119*(5), 709–732.

Terrill, A. (2006). Body part terms in Lavukaleve, a Papuan language of the Solomon Islands. *Language Sciences*, *28*(2–3), 304–322.

Thorpe, M. (1983). *Ryukyuan Language History* [PhD thesis]. University of Southern California.

Tōjō, M. (1927). *Kokugo no Hōgen Kukaku [Japanese Dialect Divisions]*. Ikuei Shoin.

Tōjō, M. (1938). *Hōgen to Hōgengaku [Dialects and Dialectology]*. Shun'yodo.

Tōjō, M. (1954). *Nihon Hōgengaku [Japanese Dialectology]*. Yoshikawa Kōbunkan.

Tokugawa, M. (1981). *Kotoba—Nishi to Higashi [Language—West and East]*. Chūokōronsha.

Tranter, N. (2012). *The Languages of Japan and Korea*. Routledge.

Trudgill, P. (1974). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*, *3*(2), 215–246.

Tsuhako, T. (1992). Okinawa Chūnanbu Hōgen [Dialects of central and southern Okinawa]. In T. Kamei, R. Kono & E. Chino (Eds.), *Nihon Rettō no Gengo [Languages of the Japanese archipelago]*. Tokyo: Sanseido.

Uchikawa, K. & Boynton, R. M. (1987). Categorical color perception of Japanese observers: Comparison with that of Americans. *Vision Research*, *27*(10), 1825–1833.

Uemura, Y. (1997). Amami hōgen [Amami dialects]. In T. Kamei, R. Kōno & E. Chino (Eds.), *Nihon rettō no gengo [The languages of the Japanese Archipelago]* (pp. 311–354). Sanseidō.

Ullmann, S. (1951). *The Principles of Semantics*. Blackwell.

Unger, J. M. (2001). Layers of Words and Volcanic Ash in Japan and Korea. *Journal of Japanese Studies*, *27*(1), 81–111.

Unger, J. M. (2009). *The role of contact in the origins of the Japanese and Korean languages*. University of Hawai'i Press.

Unger, J. M. (2014). No Rush to Judgment: The Case against Japanese as an Isolate. *NINJAL Project Review*, *4*(3), 211–230.

Urban, M. (2020). Mountain linguistics. *Language and Linguistics Compass*, e12393.

van Bezooijen, R. & van den Berg, R. (1999). Word Intelligibility of Language Varieties in the Netherlands and Flanders under Minimal Conditions. *Linguistics in the Netherlands*, *16*(1), 1–12.

van der Lubbe, G. & Tokunaga, A. (2015). Okinoerabu grammar. In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages* (pp. 345–378). De Gruyter Mouton.

van Staden, M. & Majid, A. (2006). Body colouring task. *Language Sciences*, *28*(2–3), pp. 158–161.

Vovin, A. (2013). From Koguryŏ to T'amna—Slowly riding to the South with speakers of Proto-Korean. *Korean Linguistics*, *15*(2), 217–235.

Vovin, A. (2016). On the linguistic prehistory of Hokkaidō. *Studia Orientalia Electronica*, *117*, 29–38.

Wang, C. C., Ding, Q. L., Tao, H. & Li, H. (2012). Comment on "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa". *Science*, *335*(6069), 657.

Ward, J. H. Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Weinreich, M. (1945). Der YIVO un di problemen fun undzer tsayt [The YIVO and the problems of our time]. *YIVO Bleter*, *25*(1), 3–18.

Whitman, J. (2011). Northeast Asian Linguistic Ecology and the Advent of Rice Agriculture in Korea and Japan. *Rice*, *4*(3–4), 149–158.

Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press.

Wierzbicka, A. (2005). There Are No 'Color Universals' but There Are Universals of Visual Semantics. *Anthropological Linguistics*, *47*(2), 217–244.

Wierzbicka, A. (2007). Bodies and their parts: An NSM approach to semantic typology. *Language Sciences*, *29*(1), pp. 14–65.

Wierzbicka, A. (2008). Why there are no 'colour universals' in language and thought. *Journal of the Royal Anthropological Institute*, *14*(2), 407–425.

Wilkins, D. P. (1981). *Towards a Theory of Semantic Change*. Australian National University.

Wilkins, D. P. (1996). Natural tendencies of semantic change and the search for cognates. In M. Durie & M. Ross (Eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change* (pp. 224–304). Oxford University Press.

Wild, F. (2015). *lsa: Latent semantic analysis*. R Package Version 0.73.

Wnuk, E., Levinson, S. C. & Majid, A. (submitted), Color technology is not necessary for linguistic and efficient codability of color.

Wolfram, W. & Schilling-Estes, N. (2017). Dialectology and Linguistic Diffusion. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 713–735). Blackwell.

Woo, D. (2010). Hasaigoi koboru/kobotsu/yaburu/kudaku ni no gogi to Konjaku Monogatari-shū ni okeru 壊 no kunshakushoko [Shome thoughts on the meaning of the breaking verbs koboru, kobotsu, yaburu and kudaku and the meaning of 壊 in the Konjaku Monogatari]. *Gakushuin University Journal of Language and Literature*, *53*, 46–68.

Wright, R. (1990). Semantic change in Romance words for "cut". In H. Andersen & E. F. K. Koerner (Eds.), *Papers from the 8th International Conference on Historical Linguistics, Lille, August 30-September 4, 1987* (p. 553). John Benjamins.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138.

Wulff, D.U., De Deyne, S., Jones, M.N., Mata, R. & The Aging Lexicon Consortium. (2019). New Perspectives on the Aging Lexicon. *Trends in Cognitive Sciences*, *23*(8), pp. 686–698.

Yamada, M., Pellard, T. & Shimoji, M. (2015). Dunan grammar (Yonaguni Ryukyuan). In P. Heinrich, S. Miyara & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages* (pp. 449–478). De Gruyter Mouton.

Yanagita, K. (1930). *Kagyūkō (Reflections on Snail)*. Tōkō Shoin.

Yarimizu, K. (2009). *Kyoutsuugoka Katei no Keiryouteki Bunseki [Quantitative Analysis of Standardisation Process]* [PhD thesis]. Tokyo University of Foreign Studies.

Yoshida, N. & Zhi, H. (1999). Shintaigo o fukumu kanyōku ni tsuite no nichichū taishō kenkyū—Atama no baai [A contrastive study of Japanese and Chinese idiomatic phrases containing body part terms: The case of 'head']. *Bulletin of the Okayama University Faculty of Education*, *110*(3), 105–109.

Zaslavsky, N., Kemp, C., Regier, T. & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942.

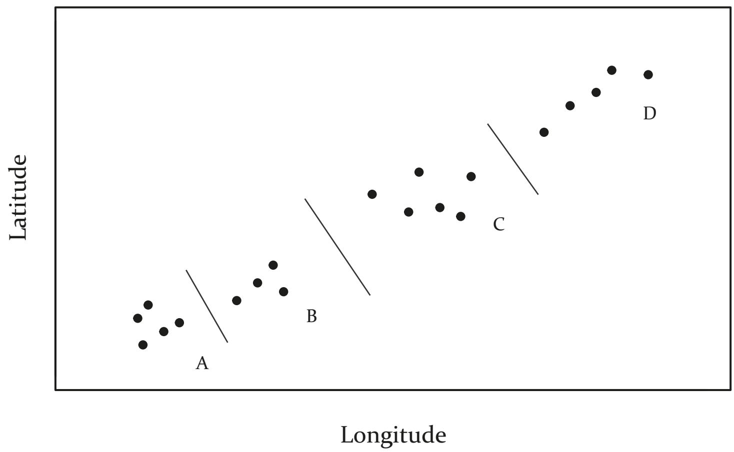
Zollinger, H. (1988). Categorical color perception: Influence of cultural factors on the differentiation of primary and derived basic color terms in color naming by Japanese children. *Vision Research*, *28*(12), 1379–1382.

APPENDIX

Appendix
Simulation of geographic and linguistic distances

The simulation modelled 20 locations spread cross four subgroups lined up in an archipelago-like fashion (see Figure A.1). First, we calculated geographic distances between them using Euclidean distance. Next, we simulated linguistic distances based on the characteristics of prototypical isolation-by-distance and isolation-by-colonisation patterns as described by Orsini et al. (2013—see their Figure 1, p. 5987).

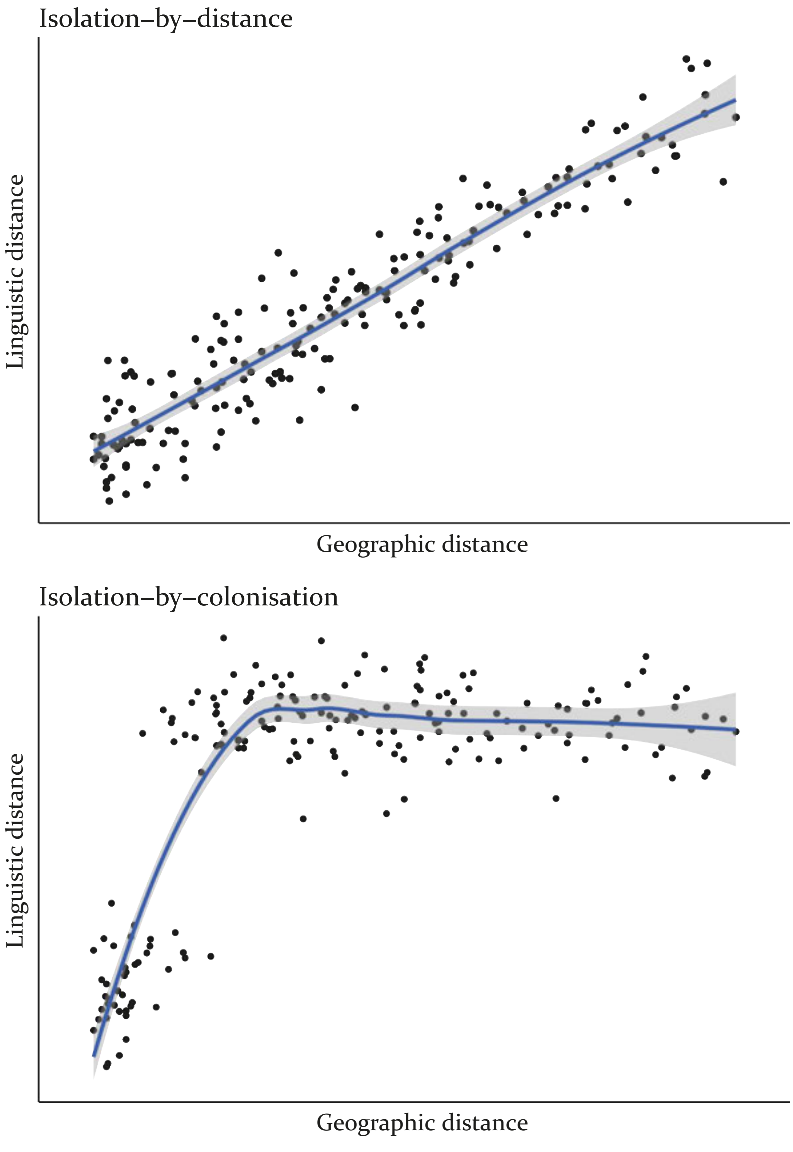
Figure A.1. Twenty simulated locations divided into four subgroups.



In the isolation-by-distance scenario, there is a direct relationship between geographic and linguistic distance across the whole area. As such, simulated linguistic distance was determined as the geographic distance between two points divided by the maximum geographic distance across the sample, with the addition of some normally distributed random noise ($M = 0$, $SD = 0.1$).

In the isolation-by-colonisation scenario, there is no contact beyond the subgroup and linguistic distance is the result of drift, without any relationship to geographic distance. As such, simulated linguistic distance between locations belonging to different subgroups was determined at a fixed level ($d = 0.75$), with the addition of normally distributed random noise (M = 0, SD = 0.1). Simulated linguistic distance within each subgroup was determined following the isolation-by-distance pattern described above. Figure A.2 shows the simulation results for the two prototypical scenarios.

Figure A.2. Plots of simulated geographic and linguistic distances in isolation-by-distance (top panel) and isolation-by-colonisation (bottom panel) scenarios.



Then, we simulated two scenarios in which one of the peripheral subgroups was isolated from the other three. In that case, simulated linguistic distances between subgroup A (or D) and the other three groups was determined by the isolation-by-colonisation pattern (IBC). The other three subgroups would maintain mutual contact and as such, simulated linguistic distances between them were determined by the isolation-by-distance pattern (IBD). Patterns that determined simulated linguistic distances within and between subgroups are shown in Table A.1.

Table A.1. Diversification patterns in peripheral isolation scenarios.

Isolation of subgroup A					Isolation of subgroup D				
	A	B	C	D		A	B	C	D
A	*IBD*	**IBC**	**IBC**	**IBC**	A	*IBD*	*IBD*	*IBD*	**IBC**
B		*IBD*	*IBD*	*IBD*	B		*IBD*	*IBD*	**IBC**
C			*IBD*	*IBD*	C			*IBD*	**IBC**
D				*IBD*	D				*IBD*

In addition, we simulated a scenario in which a split between the two "northern" and two "southern" subgroups would result in a situation in which there is only contact between subgroups A and B, and between C and D. The patterns determining simulated linguistic distance within and between subgroups in this scenario are shown in Table A.2.

Table A.2. Diversification patterns in a "north" vs. "south" scenario.

	A	B	C	D
A	*IBD*	*IBD*	**IBC**	**IBC**
B		*IBD*	**IBC**	**IBC**
C			*IBD*	*IBD*
D				*IBD*

The simulations show that linguistic continuity is disrupted in any scenario in which one (Figure A.3), several (Figure A.4), or all (Figure A.2, bottom panel) subgroups are isolated. The isolation of a subgroups results in higher than expected linguistic distances for small geographic distances, which subsequently leads to a sublinear trend as observed in several dialects areas (Nerbonne, 2010).

Figure A.3. Plots of simulated geographic and linguistic distances in scenarios where the two peripheral subgroups are isolated.

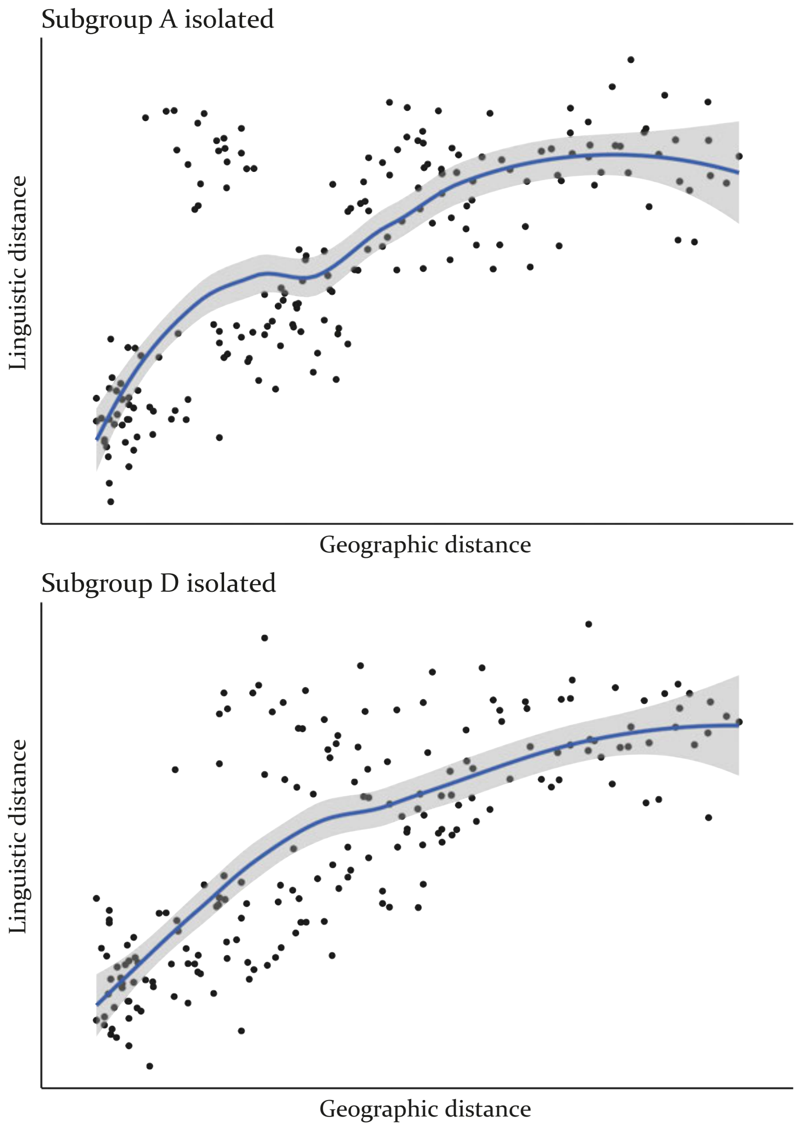
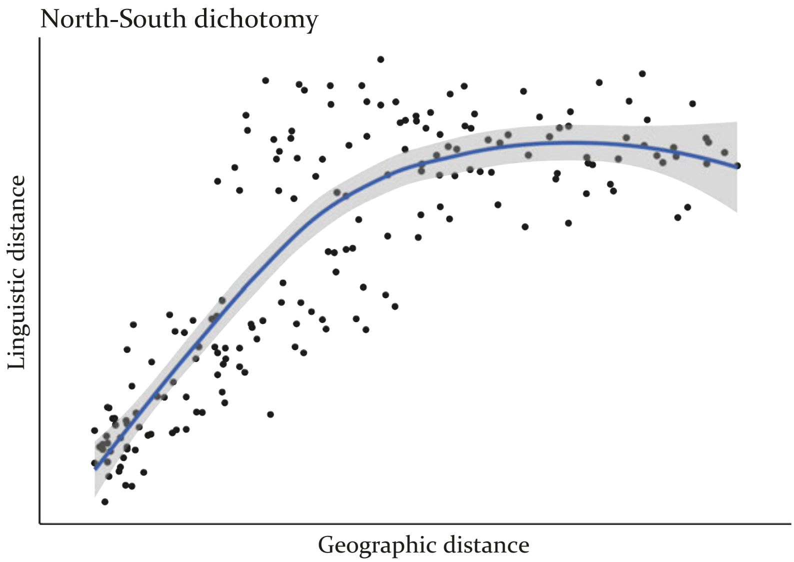


Figure A.4. Plot of simulated geographic and linguistic distances in scenarios where the northern (A+B) and southern subgroups (C+D) are isolated.



SAMENVATTING

Nederlandse samevatting

In de duizenden talen die verspreid over de wereld worden gesproken, vinden we variatie in de klanken die ze gebruiken, in de woorden die door de klanken gevormd worden, en in de grammatica waarmee de woorden gecombineerd worden. Hoewel er binnen de linguïstiek veel aandacht wordt besteed aan variatie in deze taalvormen—bijvoorbeeld in fonologische en lexicale vergelijkingen in de dialectologie en historische taalkunde—is dit voor betekenis beduidend minder het geval. Het creëren van betekenis is echt een fundamenteel aspect van taal en reflecteerd de manier waarop de sprekers van een taal de wereld om hen heen conceptualiseren. Op basis van overeenkomsten en verschillen categoriseren en groeperen we onze ervaringen, zodat ze kunnen worden uitgedrukt in taal. Ervaringen die veel op elkaar lijken, behoren tot dezelfde categorie en worden uitgedrukt met hetzelfde label. De verscheidenheid aan culturen heeft geleid tot een wijd scala aan conceptualiseringen en betekenissen die woorden kunnen uitdrukken. Het bestuderen van deze semantische variatie verbetert ons begrip van taal als geheel.

Neem bijvoorbeeld de kleurterm *blauw*. In het Nederlands kunnen we verschillende tinten onderscheiden door constructies als *lichtblauw*, *donkerblauw* of *hemelsblauw* te zeggen, maar de basisterm blijft *blauw*. Het Japanse woord voor 'blauw' is *ao*, maar de betekenis van *ao* is breder en omvat ook groentinten, zodat 'groenten' ook wel *aomono* (mono betekent 'dingen') genoemd worden. Daarnaast heeft het Japans aparte woorden voor zowel lichte (*mizuiro*) als donkere (*kon*) blauwtinten. Daartegenover staat dat we in het Nederlands spreken van een *been* en een *voet*, maar het Japans beiden beschrijft met *ashi*. En waar Nederlanders *snijden* met een mes, maar *knippen* met een schaar, zullen Japanners voor beide handelingen *kiru* noemen. Hoewel er overtuigend bewijs is voor dit soort verschillen tussen talen, is het onduidelijk in hoeverre deze verschillen terug te vinden zijn bij vergelijking van *verwante* talen. In dit proefschrift onderzoek ik daarom vorm- en betekenisvariatie binnen één taalfamile, de Japanse taalfamilie. Hoofdstuk 1 geeft een algemene inleiding. Ik bespreek eerder onderzoek naar de processen en oorzaken van taalverandering, eerdere toepassingen van kwantitatieve methodes in linguïstische afstanden, en eerder werk over semantiek en de studie van semantische variatie. Vervolgens introduceer ik de drie semantische velden die gedetailleerde aandacht krijgen in dit proefschrift: (1) kleur, (2) het menselijk lichaam, en (3) snij- en breekhandelingen. Ten slotte formuleer ik specifieke onderzoeksvragen en beschrijf ik de gebruikte methoden. Dit proefschrift combineert nieuwe data verzameld tijdens mijn veldwerk in Japan met bestaande data uit dialectwoordenboeken.

Hoofdstuk 2 dient als een inleidend overzicht van de Japanse taalfamilie. Aan bod komen geografie en bevolkingsgeschiedenis van Japan, de geschiedenis en hoofdindeling van de taalfamilie. De taalfamilie bestaat uit twee hoofdtakken: de Japanse talen en de Riukiu-talen. Deze laatste zijn bedreigd, maar weinig bestudeerd—met name voor de semantiek. Dit proefschrift legt daarom de nadruk op deze talen, waarmee niet alleen wordt bijgedragen aan hun beschrijving, maar hun beschrijving heeft tegelijkertijd een toegevoegde waarde: het beantwoorden van bredere vragen binnen de taalkunde. Ik sluit het hoofdstuk af met een taaltypologisch overzicht van de Japanse taalfamilie, gevolgd door een samenvatting van hun opname in eerder werk in de drie semantische velden die later in dit proefschrift aan bod komen.

In hoofdstuk 3 ga ik in op de vraag hoe de geografische samenstelling van een taalgebied van invloed is op taalvariatie. In Japan vinden we een contrast tussen een groot, verbonden ‘vasteland’ bestaande uit de vier hoofdeilanden Hokkaido, Honshu, Shikoku en Kyushu, en de kleine, meer verspreide eilandengroepen van de Riukiu-eilanden in het zuiden. Aan de hand van variatie in de basiswoordenschat—woorden waarvan wordt aangenomen dat zij in vrijwel elke taal voorkomen, zoals “rood”, “hoofd” en “snijden”—bereken ik de linguïstische afstanden tussen 90 taalvariëteiten verspreid over deze twee gebieden, en probeer deze vervolgens te verklaren aan de hand van een aantal taalexterne factoren. Hoewel in zowel vastelandstalen als eilandtalen een grotere ruimtelijke afstand leidt tot een grotere linguïstische afstand, is dit effect aanzienlijk kleiner in eilandtalen—waar verschillen tussen talen eerder een weerspiegling zijn van historische kolonisatiepatronen. Deze resultaten suggeren dat geografie en tijd verschillen in hun invloed op de ontwikkeling van vastelandstalen en eilandtalen.

In hoofdstuk 4 onderzoek ik het kleurenlexicon van hedendaagse Riukiu-sprekers aan de hand van een taak waarin sprekers een reeks kleurchips benoemen. Ik leg daarbij de nadruk op de mogelijke impact van taalcontact met het Standaardjapans en het Engels, en vergelijk de nieuwe data met gelijksoortig materiaal uit de jaren 60 van de vorige eeuw. Deze vergelijking toont dat hedendaagse Riukiu-sprekers meer gebruik maken van zowel Standaardjapanse kleurtermen (bijv. *midori* 'groen'), als Engelse leenworden *pinku* (Engels *pink*, 'roze') en *orenji* (Engels *orange*, 'oranje'). Het gebruik van deze nieuwe kleurtermen heeft ook geleid tot een groter aantal kleurcategorieën dat men onderscheidt: waar 60 jaar geleden het chromatisch deel van het kleurspectrum doorgaans nog werd opgedeeld in vier categoriën, gebeurt dit nu steeds vaker in acht tot negen categoriën net zoals in het Standaardjapans. Verdere analyse bevestigt dat de kleursystemen van de Riukiu-talen inmiddels meer lijken op het Standaardjapans dan op hun eigen systeem uit de jaren 60. Dit geeft aan dat contact met een standaardtaal en een wereldtaal (Engelse) een aanzienlijke invloed kan hebben op het kleurenlexicon en de kleurcategorieën van (bedreigde) minderheidstalen. Ondanks deze veranderingen lijken Riukiu-talen echter nog steeds meer op elkaar dan op het Japans. Dit is het sterkst te zien in een groep kleurwoorden die reeds meer dan 1000 jaar in gebruik zijn (cognaten van *aka*, *ao*, *ki*, en *murasaki*). Deze uitkomst suggereert dat de kern van een semantisch veld ondanks invloed van buitenaf stabiel kan blijven.

Hoofdstuk 5 behandelt in twee deelstudies de semantiek van de lichaamsdelen. Het eerste deel betreft het benoemen van lichaamsdelen. Om de verschillende termen te verzamelen is gebruik gemaakt van een taak waarin sprekers een aantal lichaamsdelen benoemen aan de hand van een gemarkeerde tekening van het menselijk lichaam. Een aanzienlijk aantal termen voor lichaamsdelen wordt gerekend tot de basiswoordenschat. Variatie in dit deel van het lexicon zou daarom een goede weerspiegeling moeten zijn van de algemene relaties tussen verwante talen, hetgeen wordt bevestigd door de verzamelde benoeminsgsdata. Daarnaast wordt vaak aangenomen dat dit semantisch veld hiërarchisch gestructureerd is vanwege meronymie (de relatie tussen deel en geheel), waarbij de vingers bijvoorbeeld deel uitmaken van de hand, die weer deel is van de arm. De benoemingsdata tonen echter aan dat delen van het gezicht minder een hiërarchische structuur vertonen dan de rest van het lichaam. In het tweede deel van dit hoofdstuk bespreek ik de precieze plaats en omvang van de lichaamsdelen (het extensionele bereik) door sprekers deze te laten inkleuren in een blanco tekening van het mensenlijk lichaam. Hierbij is er een duidelijk verschil tussen weinig variatie in bereik voor delen die ‘natuurlijk afgebakend’ zijn, zoals bijvoorbeeld de oren en mond, tegenover meer variatie voor delen zonder duidelijke afbakening, zoals bijvoorbeeld de wangen, borst, en rug. Deze bevindingen suggereren dat de hoeveelheid variatie binnen een semantisch veld kan verschillen.

In hoofdstuk 6 bespreek ik de semantiek van zogenaamde snij- en breekhandelingen. Eerder onderzoek in dit semantisch veld wees uit dat een klein aantal semantische dimensies gedeeld wordt door alle talen, en dat de variatie tussen talen hierdoor beperkt wordt. In deze eerdere studies nam het Japans crosslinguïstisch gezien een uitzonderlijke positie in, maar deze bevinding was gebaseerd op een beperkt aantal gegevens. In dit hoofdstuk onderzoek ik op basis van nieuwe, uitgebreidere data of dit inderdaad het geval is en, zo ja, of dit ook van toepassing is op de andere talen uit de Japanse taalfamilie. Aan de hand van data verzameld in een taak waarbij sprekers één voor één verschillende videoclips beschrijven, laat ik zien dat de semantische categorieën in zowel het Japans als de Riukiu-talen binnen het gedeelde crosslinguïstisch kader vallen. Het laatste deel van het hoofdstuk vergelijkt de data voor de Japanse taalfamilie met eerder verzamelde data in vier Germaanse talen: Engels, Nederlands, Duits en Zweeds. Deze vergelijking laat zien dat er binnen de Japanse talen minder variatie is dan binnen de Germaanse talen, en dat talen uit eenzelfde familie doorgaans meer op elkaar lijken dan op talen uit andere families. Opvallend was echter dat het Engels meer lijkt op elk van de Japanse talen dan op het Zweeds. Deze bevindingen suggeren dat verwantschap een belangrijke rol speelt in semantische gelijkenis, maar dat een gedeeld crosslinguïstische kader ervoor kan zorgen dat zelfs niet-verwante talen semantisch sterk op elkaar lijken.

In het laatste hoofdstuk (7) vat ik de bevindingen uit voorgaande hoofdstukken samen, bespreek ik de theoretische implicaties en kaart ik enige mogelijkheden aan voor toekomstig onderzoek. In mijn proefschrift laat ik zien dat geografie een belangrijke rol speelt in diversificatie en daarom moet worden meegewogen in verklaringen van taalvariatie en -verandering. Voor de drie semantische velden besproken in dit proefschrift geldt dat er aanzienlijke variatie kan zijn binnen een taalfamilie, maar dat deze variatie beperkter is dan die tussen niet-verwante talen. De bevindingen laten verder zien dat er nog veel te leren valt over taaldiversiteit in de wereld en dat diepgaand onderzoek naar diverse (bedreigde) talen nodig is om veranderingen in woorden en hun betekenissen beter te begrijpen. Toekomstig onderzoek zal nog moeten uitwijzen wat dit soort semantische variatie voor gevolgen heeft voor de onderlinge verstaanbaarheid van talen en hun variëteiten.

Biographical note

John Huisman (1987, Kerkrade) majored in Japanese through the bachelor Oriental Languages and Communication at Zuyd University of Applied Sciences in Maastricht. After that, he completed the Research Master in Linguistics and Communication Sciences at Tilburg University. He wrote his MA thesis on the role of psycholinguistic factors in odour naming, which earned him the prize for best Research Master thesis across all faculties of Tilburg University. In 2016, he was awarded an NWO PhDs in the Humanties grant and started his PhD research on form and meaning variation in the Japonic language family at the Centre for Language Studies at Radboud University, carrying out a total of around 10 months of fieldwork in Japan. He currently works as a postdoctoral researcher at Uppsala University through a Niels Stensen Fellowship, investigating patterns of semantic variation and change in the Indo-European language family.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*
17. Moving eyes and naming objects. *Femke van der Meulen*

18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*
25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*
27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Broersma*
35. Retrieving semantic and syntactic word properties. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*
37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*
38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*
39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*
40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*
41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*
42. The acquisition of auditory categories. *Martijn Goudbeek*
43. Affix reduction in spoken Dutch. *Mark Pluymaekers*
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*
45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*
50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*
54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*
60. Processing casual speech in native and non-native language. *Annelie Tuinman*
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Robinson*
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*
63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*
64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*
66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*
67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*

69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*
70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*
71. Gaze behavior in face-to-face interaction. *Federico Rossano*
72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Conny de Vos*
73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*
74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*
75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*
76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*
77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*
78. Contributions of executive control to individual differences in word production. *Zeshu Shao*
79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*
80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*
81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*
82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*
83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*
84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*

85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*
86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*
87. The acquisition of morphophonological alternations across languages. *Helen Buckler*
88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*
89. Rediscovering a forgotten language. *Jiyoun Choi*
90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*
91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*
92. Information structure in Avatime. *Saskia van Putten*
93. Switch reference in Whitesands. *Jeremy Hammond*
94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*
95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. *Beyza Sümer*
96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*
97. Incrementality and Flexibility in Sentence Production. *Maartje van de Velde*
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*
99. The request system in Italian interaction. *Giovanni Rossi*
100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*
101. Assessing birth language memory in young adoptees. *Wencui Zhou*
102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*

103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*
104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition**.** *Rósa Signý Gisladottir*
105. Modelling Multimodal Language Processing. *Alastair Smith*
106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*
107. Situational variation in non-native communication. *Huib Kouwenhoven*
108. Sustained attention in language production. *Suzanne Jongman*
109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntatic, and orthographic effects. *Malte Viebahn*
110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*
111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*
112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*
113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*
114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*
115. Music and language comprehension in the brain. *Richard Kunert*
116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language**.** *Nietzsche H.L. Lam*
117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*
118. Language processing in a conversation context. *Lotte Schoot*
119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*
120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*

121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*
122. Sensorimotor experience in speech perception. *Will Schuerman*
123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*
124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*
125. Lapse organization in interaction. *Elliott Hoey*
126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*
127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*
128. The role of voice and word order in incremental sentence processing. *Sebastian Sauppe*
129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*
130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*
131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*
132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*
133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*
134. Indirect request comprehension in different contexts. *Johanne Tromp*
135. Envisioning Language - An Exploration of Perceptual Processes in Language Comprehension. *Markus Ostarek*
136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*

137. Let the agents do the talking: on the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*
138. Age and hearing loss effects on speech processing. *Xaver Koch*
139. Vocabulary knowledge and learning: Individual differences in adult native speakers. *Nina Mainz*
140. The face in face-to-face communication: Signals of understanding and non-understanding. *Paul Hömke*
141. Person reference and interaction in Umpila/Kuuku Ya'u narrative. *Clair Hill*
142. Beyond the language given: The neurobiological infrastructure for pragmatic inferencing. *Jana Bašnáková*
143. From Kawapanan to Shawi: Topics in language variation and change. *Luis Miguel Rojas-Berscia*
144. On the oscillatory dynamics underlying speech-gesture integration in clear and adverse listening conditions. *Linda Drijvers*
145. Understanding temporal overlap between production and comprehension. *Amie Fairs*
146. The role of exemplars in speech comprehension. *Annika Nijveld*
147. A network of interacting proteins disrupted in language-related disorders. *Elliot Sollis*
148. Fast speech can sound slow: Effects of contextual speech rate on word recognition. *Merel Maslowski*
149. Reason-giving in everyday activities. *Julija Baranova*
150. Speech planning in dialogue - Psycholinguistic studies of the timing of turn taking. *Mathias Barthel*
151. The role of neural feedback in language unification: How awareness affects combinatorial processing. *Valeria Mongelli*
152. Exploring social biases in language processing. *Sara Iacozza*
153. Vocal learning in the pale spear-nosed bat, *Phyllostomus discolor*. *Ella Lattenkamp*
154. The effect of language contact on speech and gesture: The case of Turkish-Dutch bilinguals in the Netherlands. *Elif Zeynep Azar*

155. Language and society: How social pressures shape grammatical structure. *Limor Raviv*
156. The moment in between: Planning speech while listening. *Svetlana-Lito Gerakaki*
157. How speaking fast is like running: Modelling control of speaking rate. *Joe Rodd*
158. The power of context: How linguistic contextual information shapes brain dynamics during sentence processing. *René Terporten*
159. Neurobiological models of sentence processing. *Marvin Uhlmann*
160. Individual differences in syntactic knowledge and processing: The role of literacy experience. *Saoradh Favier*
161. Memory for speaking and listening. *Eirini Zormpa*
162. Masculine generic pronouns: Investigating the processing of an unintended gender cue. *Theresa Redl*
163. Properties, structures and operations: Studies on language processing in the brain using computational linguistics and naturalistic stimuli. *Alessandro Lopopolo*
164. Investigating spoken language comprehension as perceptual inference. *Greta Kaufeld*
165. What was that Spanish word again? Investigations into the cognitive mechanisms underlying foreign language attrition. *Anne Mickan*
166. A tale of two modalities: How modality shapes language production and visual attention. *Francie Manhardt*
167. Why do we change how we speak? Multivariate genetic analyses of language and related traits across development and disorder. *Ellen Verhoef*
168. Variation in form and meaning across the Japonic language family with a focus on the Ryukyuan languages. *John Huisman*