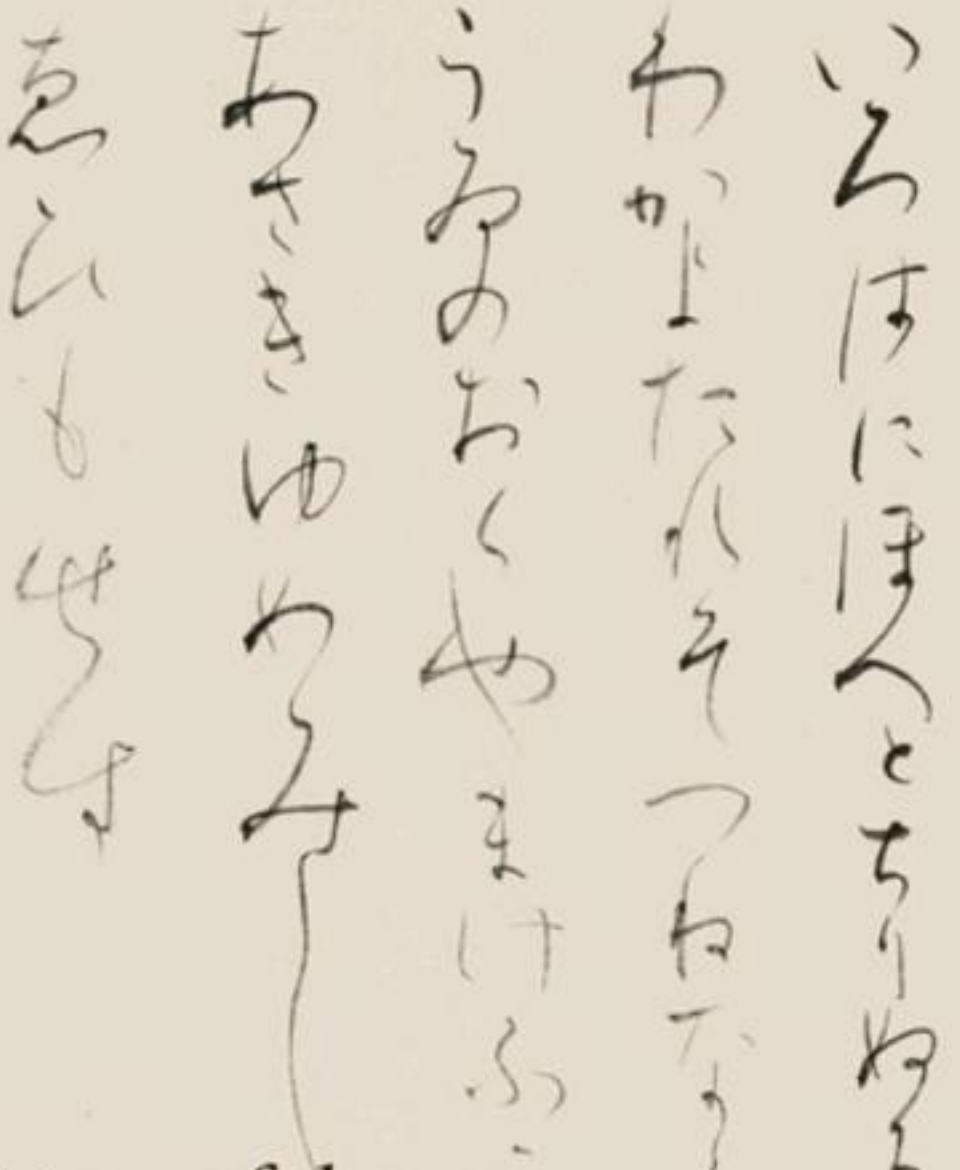


A History of the JAPANESE Language

BJARKE FRELLESVIG

CAMBRIDGE

This page intentionally left blank

A History of the Japanese Language

Bjarke Frellesvig describes the development of the Japanese language from its recorded beginnings until the present day as reflected by the written sources and historical record. Beginning with a description of the oldest attested stage of the language, Old Japanese (approximately the eighth century AD), and then tracing the changes which occurred through the Early Middle Japanese (800–1200), Late Middle Japanese (1200–1600) and the Modern Japanese (1600 onwards) periods, a complete internal history of the language is examined and discussed. This account provides a comprehensive study of how the Japanese language has developed and adapted, providing a much-needed resource for scholars. *A History of the Japanese Language* is invaluable to all those interested in the Japanese language and also students of language change generally.

Bjarke Frellesvig is Professor of Japanese Linguistics, Director of the Research Centre for Japanese Language and Linguistics at the University of Oxford, and a Fellow of Hertford College, Oxford. He is the author of *A Case Study in Diachronic Phonology: The Japanese Onbin Sound Changes* (1995) and the co-editor of *Japan and Korea: Contemporary Studies* (1997), *Current Issues in the History and Structure of Japanese* (2007) and *Proto-Japanese: Issues and Prospects* (2008).

A History of the Japanese Language

Bjarke Frellesvig

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521653206

First published 2010

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-65320-6 Hardback

For my children and my wife

Contents

Part II Early Middle Japanese

TABLES

MAPS

FIGURE

Acknowledgements

I would like to acknowledge the support, both funding for research travel and research assistance and support in the form of sabbatical leave, that I have received in the course of writing this book from the Faculty of Oriental Studies and Hertford College in the University of Oxford. I also want to express my gratitude to the University's Sasakawa Fund and John Fell Fund for their financial support. Without the generous support of these institutions, I would not have been able to complete this project.

A number of present and former Oxford students provided research and editorial assistance and valuable comments on parts of the book: Madeleine Brook, Marion Cossin, Laurence Mann, Dan Millichip, Nina Mirnig, Jenny Moore, Cecily Nowell-Smith, Thu Phuong Nguyen, Muneto Ozaki, James Stone-Lunde, James Webb, and Asa Yoneda. On numerous occasions Greg Jennings and his staff at the IT department in Hertford solved hardware and software problems for me.

Anton Antonov, Stephen Wright Horn, Kerri L. Russell, Matt Shibatani, Antoni Üçerler, Uwano Zendō, and an anonymous reader offered important comments and suggestions on earlier drafts of the book. Among the many colleagues from whom I have learned much of what is reflected in this book, I must also in particular mention Kinsui Satoshi, Alexander Vovin, and Janick Wrona. Janick also provided invaluable comments and suggestions for improvement on the final draft. Above all, John Whitman has been instrumental in shaping my views on general and specific problems in the history of the Japanese language. John read and commented extensively and incisively on several draft versions and much of the value this book has in its present form is due to his input. Naturally, all insufficiencies that remain, despite the best efforts of those mentioned above and many others as well, are entirely of my own making.

I would also like to thank Andrew Winnard, Sarah Green, Joanna Garbutt and Rosina Di Marzo of Cambridge University Press for their guidance and help in producing the book. Anna Oxbury was an amazing copy-editor.

I am grateful to my father-in-law for the calligraphy of the *Iroha-uta* on the cover of the book. Finally, for emotional and practical support through the long process of putting together and completing this book, I have my children and my wife to thank.

Abbreviations

SYMBOLS

.	syllable boundary
/.../	enclose phonemes or phonemic representation
//.../	enclose underlying representation
=>	generative rules ('becomes in the course of derivation')
>	diachronic correspondence ('becomes through time')

GRAMMATICAL TERMS

ABL	ablative
ACC	accusative
ACOP	adjectival copula
ADJ	adjective
ADN	adnominal
ALL	allative
AUX	auxiliary
CAUS	causative
COM	comitative
COMP	complementizer
CONC	concessive
COND	conditional
CONCL	conclusive
CONJ	conjectural
CONT	continuative
COP	copula
DAT	dative
DESID	desiderative
EMPH	emphatic
ETOP	emphatic topic
EVID	evidential
EXCL	exclamatory
FOC	focus

GEN	genitive
GER	gerund
H	high (tone or pitch)
HON	honorific
HUM	humble
IMP	imperative
INF	infinitive
INT	intentional
intr.	intransitive
k-irr	*k*-irregular (*ka*-hen)
L	low (tone or pitch)
LB	lower bigrade (*shimo nidan*)
LM	lower monograde (*shimo ichidan*)
MPST	modal past
MVR	mid vowel raising
NCONJ	negative conjectural
NEC	necessitive
NEG	negative
n-irr	*n*-irregular (*na*-hen)
NMLZ	nominalizer
NMNL	nominal
NOM	nominative
NONPST	nonpast
OPT	optative
PASS	passive
PCONJ	present conjectural
PERF	perfective
POL	polite
PRES	presumptive
PROV	provisional
PST	past
PSTCONJ	past conjectural
PURP	purposive
Q	interrogative
QD	quadrigrade (*yodan*)
RESP	respect
r-irr	*r*-irregular (*ra*-hen)
s-irr	*s*-irregular (*sa*-hen)
SPST	simple past
STAT	stative
SUBJ	subjunctive

TOP	topic
tr.	transitive
UB	upper bigrade (*kami nidan*)
UM	upper monograde (*kami ichidan*)
VOL	volitional

LANGUAGES

cNJ	contemporary Modern Japanese
EMC	Early Middle Chinese
EMJ	Early Middle Japanese
EOJ	Eastern Old Japanese
J-Ch	Japano-Chinese
LMC	Late Middle Chinese
LMJ	Late Middle Japanese
MC	Middle Chinese
MJ	Middle Japanese
MK	Middle Korean
NJ	Modern ('new') Japanese
OC	Old Chinese
OJ	Old Japanese
pJ	proto-Japanese
pK	proto-Korean
SJ	Sino-Japanese
Skt	Sanskrit

TEXTS

Ars gr.	*Ars grammaticae iaponicae linguae*
Arte	*Arte da lingoa de Iapam* (page references are to Doi 1955)
Arte breve	*Arte breve da lingoa Iapoa*
Bussoku	*Bussokuseki-ka*
EN	(*Engishiki*) *Norito*
Esopo	*Esopono fabulas*
Feiqe	*Feiqe monogatari*
Genji	*Genji monogatari*
Ise	*Ise monogatari*
KK	*Kojiki kayō* (songs/poems in the *Kojiki*)
Kokin	*Kokinwakashū*
MYS	*Man'yōshū*
NSK	*Nihon shoki kayō* (songs/poems in the *Nihon shoki*)

Ochikubo	*Ochikubo monogatari*
SM	(*Shoku nihongi*) *Senmyō*
Taketori	*Taketori monogatari*
Tosa	*Tosa nikki*
Vocabulario	*Vocabulario da lingoa de Iapam*

Introduction

This book describes the development of the Japanese language from its written beginning until the present day as it is reflected in the written sources; that is to say, its internal history. This is accomplished by first giving an overall description of the oldest attested stage of the language, Old Japanese, and then tracing changes since then, as they are reflected in the written sources and in the present-day language. The possible cognation of Japanese to other languages, its external pre-history, is not considered. Nor is its internal pre-history discussed to any significant extent, except where relevant to understanding its attested history. And nothing is said about dialects, except where they are prominently reflected in the written sources and where they have contributed to the formation of the modern standard language.

The periodization of Japanese shown in (1) is adopted, which overlaps with the main political periods.

(1)

Linguistic periods		Political periods
Old Japanese (OJ)	700–800	Nara, 712–794
Early Middle Japanese (EMJ)	800–1200	Heian, 794–1185
Late Middle Japanese (LMJ)	1200–1600	Kamakura, 1185–1333 Muromachi, 1333–1573
Modern Japanese (NJ)	1600–	Edo, 1603–1868 Meiji, 1868–1912 Taishō, 1912–1926 Shōwa, 1926–1989 Heisei, 1989–

Where necessary, early is distinguished from late within both Early Middle Japanese (early: 794–1086; late: the *Insei* period, 1086–1185) and Late Middle Japanese (early: the Kamakura period; late: the Muromachi period). Modern Japanese is abbreviated as 'NJ' (for 'new Japanese') to avoid

confusion with Middle Japanese (MJ). By 'cNJ' is meant contemporary NJ, the Japanese of the twentieth century. 'Classical Japanese' is the fossilized, relatively fixed written norm which arose largely out of the language of the twelfth century and which thereafter remained the dominant form of writing in Japanese until the beginning of the twentieth century. This book is not to any great extent concerned with the fossilized Classical Japanese written norm.

Extensive attestation of Japanese goes back to the eighth century. Old Japanese is mainly the language of the Nara period, although it also comprises earlier texts which are included in sources compiled or completed in the eighth century, but their age is difficult to assess. It is sometimes said that Japanese has not changed greatly since Old Japanese. However, it is possible to identify two large sets of internal change, largely coinciding with EMJ (phonological change) and LMJ (grammatical change), respectively, as well as two waves of contact induced change which took place during OJ/EMJ (sinification) and late NJ (westernization), which together transformed the language from its Old Japanese to its modern form.

In the transition between OJ and EMJ and within EMJ, the language underwent significant phonological changes, both in syllable structure and in segmental phonology. At the end of the EMJ period, the phonological structure of Japanese was largely as it is today and phonological changes since then have by comparison been minor. Most major phonological changes were complete by the end of the eleventh century and some scholars do not include the last century of the Heian period, the so-called *Insei* period (1086–1185), as a part of EMJ, but instead as part of LMJ. It was also during the EMJ period that the contact with Chinese manifested its influence in the texts, although we suspect that this influence was present already in OJ.

During the LMJ period by contrast, major grammatical changes took place, which affected both morphology and syntax. Some of these changes are initiated or anticipated in the twelfth-century materials, but they are mainly reflected in the written sources of the LMJ period. The dating of the end of LMJ is difficult to determine on linguistic grounds. The main issue is whether to include the language stage reflected in the Christian materials from the end of the sixteenth century and early years of the seventeenth century in LMJ or in NJ. The sweeping grammatical changes which took place during LMJ were complete by then and in many respects the language of those sources is similar to NJ. However, it is also very similar to the language in an earlier set of sources, the so-called *shōmono* which date from the middle of the fifteenth to the middle of the sixteenth century. The Christian sources, alongside the *shōmono*, are accordingly included in the LMJ period, and we will consider them to constitute the end of LMJ.

NJ is thus the language from then on. Once the phonological changes of EMJ and the grammatical changes of LMJ were complete, Japanese did not

change very much structurally. However, NJ was affected by significant external changes brought about by extensive contact with European languages in the course of the modernization of Japan and Japanese in the late nineteenth and early twentieth centuries.

Attestation of forms

OJ, EMJ and LMJ are dead languages and the sources are closed and limited text corpora. Naturally, many specific forms of individual words are not attested in those texts. The facts and state of attestation are of paramount importance when investigating the system of a language, but of less interest when explaining or exemplifying the system. Recall Winston Churchill's famous experience, recounted in the first chapter of his *My Early Life*, of being set the task upon arrival at his boarding school at the age of seven of learning the singular of the noun of the first declension in Latin. He managed to memorize the paradigm and reproduce it to the satisfaction of his teacher, but he did not understand what it meant and asked about it:

'But,' I repeated, 'what does it mean?'
'Mensa means a table,' he answered.
'Then why does mensa also mean O table,' I enquired, 'and what does O table mean?'
'Mensa, O table, is the vocative case,' he replied.
'But why O table?' I persisted in genuine curiosity.
'O table, – you would use that in addressing a table, in invoking a table.' And then seeing that he was not carrying me with him, 'You would use it in speaking to a table.'
'But I never do,' I blurted out in honest amazement.
'If you are impertinent, you will be punished, and punished, let me tell you, very severely,' was his conclusive rejoinder.

An inflectional paradigm is a pattern of relations; the combination of mutually exclusive inflectional endings with stems of words; or, put differently, the morphologically possible forms of a word. It is possible that no one ever produced the vocative of *mensa* in actual speech or writing in classical Latin. But if anyone wished to address a table, the vocative case was available for that purpose and even if it was never spoken, the vocative form of that noun existed in Latin as a systemic possibility. In OJ, the imperative of the verb *kog-* 'row' is not attested. There is no systemic reason that *kogye* 'row!' should not exist, much less so than for the vocative of Latin *mensa*. The non-occurrence of *kogye* may be regarded as an accident of attestation. Given the existence of OJ *kog-*, which is amply attested in various forms, we know that its imperative would be *kogye*. On the other hand there are restrictions on the use of the imperative of certain types of verbs, e.g. those which signify spontaneous, nonvolitional action. It is debatable whether this is a morphological restriction (that form does not exist) or a syntactic/pragmatic restriction

(that form cannot be used, it does not make sense for that form to be used), but it is a grammatical restriction in the language, not an accident of attestation. Throughout the book morphological paradigms, particularly verbs, are exemplified. Some of the forms of some of the verbs are not attested, but this is generally not remarked, except for small, irregular verb classes or grammatical auxiliaries where the situation of attestation may be of significance.

Conventions

Throughout the book a phonemic transcription of cited forms is generally employed, appropriate to the period of citation. For example, the word for 'front', which is NJ *mae*, will be written as shown in (2), reflecting its phonemic shape at different stages of the language (exemplified by texts from those periods):

(2)	OJ	early EMJ	mid EMJ	late EMJ onwards
	Man'yōshū	*Tosa nikki*	*Genji monogatari*	*Sarashina nikki*
	mapye	*mape*	*mawe*	*mae*

Japanese editions of pre-modern texts, by contrast, employ a historical spelling and will spell 'front' as まへ ('mahe') regardless of the period from which the text dates. From late EMJ until the second half of NJ this word was, like all words of the shape /Ve/, pronounced with a palatal onglide before the /e/, [ma^{j}e], but that will not be noted in the phonemic transcription. Nor will other allophonic features of pronunciation, such as the prenasalization of /b, d, g, z/ (which from OJ into the LMJ period were pronounced [^{m}b, ^{n}d, ^{ŋ}g, ^{n}z]). However, when transcribing forms with a moraic consonant (which in EMJ and early LMJ was phonemically underspecified for nasality in morpheme internal position), we will use a semi-allophonic transcription and write *punde* 'brush', not *puCde*, although strictly speaking the phonemic shape was /puCde/; see further 7.1.3 about this.

When citing words or passages from OJ texts italics are used for phonographically written text portions and normal type for logographically written text. For example, when citing the word *suru*, the adnominal form of *se-* 'to do', it is transcribed in accordance with the writing in the source as exemplified in (3), with different writings of *suru* cited from different poems in the *Man'yōshū*. In (3a), 須 and 流 are used phonographically for the syllables /su/ and /ru/. In (3b), 為 is used logographically to write part of a form of a verb meaning 'to do' and 流 is used phonographically for the syllable /ru/. Finally, in (3c), 為 is used logographically for a form

of a verb meaning 'to do' which reading tradition interprets as the adnominal form.

(3)	*MYS* poem no.	Writing	Transcription
a.	17:3932	須流	*suru*
b.	19:4198	為流	su*ru*
c.	2:164	為	suru

Words from cNJ will be noted in the modified Hepburn transcription, for example *fuji* '(Mount) Fuji' or *chōshoku* 'breakfast'. Japanese script will generally not be used, except when discussing script and writing. However, where relevant the Chinese characters (*kanji* 漢字) used to write Chinese or Sino-Japanese words are given.

Japanese proper names will usually be given in their usual cNJ form, e.g. *Hōjōki* (not the shape this word would have had at the beginning of the thirteenth century: *Paũdyaũ-ki*) and personal names are given in the traditional Japanese order (surname, given name), e.g. Hattori Shirō. Modern Chinese words will be transcribed in *pinyin*. Early Middle Chinese reconstructions follow Pulleyblank 1991.

When citing verbs, the basic stem is used, rather than the OJ/EMJ/early LMJ conclusive and the late LMJ/NJ nonpast form, which are the forms usually used as citation forms, i.e. the forms used to cite or talk about a verb, for example in dictionaries.[1] Citing the basic stem in the majority of cases unambiguously identifies the conjugation class of a verb, especially when comparing quadrigrade and bigrade verbs, e.g. OJ *ok-* 'put' (quadrigrade), *ake-* 'dawn' (lower bigrade), *okwi-* 'arise' (upper bigrade), whereas that is not the case with the OJ/EMJ/early LMJ conclusive: *oku*, *aku*, *oku*.

Almost all verb suffixes attach directly to the basic stem of regular vowel base verbs, but for some irregular vowel base verbs and for all consonant base verbs, most verb suffixes attach to one of several *derived stems* (see further 3.4.4 and 8.1.4). When citing verbal suffixes a morphophonemic notation is used which shows which, if any, derived stem the suffix attaches to, by using bracketed prefixes: (a), (i), (e), (I). This is exemplified in (4), using the verb *sak-* 'to come into bloom'. The *a-* stem corresponds to the *mizenkei* of Japanese school grammar, the infinitive to the *ren'yōkei*, and the exclamatory to the *izenkei* (see 3.4.6).

[1] Notable exceptions, which list verbs under their infinitive, are Ohno's dictionary of pre-modern Japanese (1990) or dictionaries published by the Jesuit missionaries in the early seventeenth century, e.g. Rodrigues (1603–4).

(4)

Prefix	Stem	Verb	Suffix	Verb + suffix
(a)	*a-* stem	*saka-*	*-(a)n-* (Negative)	*saka-n-*
(i)	Infinitive	*saki-*	*-(i)n-* (Perfective)	*saki-n-*
(e)	Exclamatory	*sake-*	*-(e)do* (Concessive)	*sake-do*
(I)	*onbin* stem	*sai-*	*-(I)ta* (Past)	*sai-tar-*

This notation does not mean that the prefixes are part of the (synchronic) phonemic shape of the suffixes. The prefixes only show which stem a suffix selects. For example, for both the negative, *-(a)n-*, and the perfective, *-(i)n-*, the basic stem shape of the suffix is /-n-/, but they select different stems, as shown in (4).

In examples, verb forms will be segmented as in (5), i.e. noting (by '-') morpheme boundaries between verb stems and auxiliaries, but not between flectives and the immediately preceding verb or auxiliary stem. In glosses, the inflected form will be noted as part of the gloss for a verb or auxiliary, separated by '.', showing that *saku* is the conclusive form of the verb *sak-* 'bloom', *sakedo* is the concessive form of that verb, and *-kyeri* is the conclusive form of the modal past auxiliary *-(i)kyer-*. This notation shifts the focus away from individual morphemes to the actual inflected word forms. See sections 3.1 and 3.4 for the analysis underlying this notation.

(5) a. *saku*
bloom.CONCL
'it blooms'

b. *sakedo*
bloom.CONC
'although it blooms'

c. *maywopi-ki-ni-kyeri*
fray-come-PERF-MPST.CONCL
'it had become frayed'

Where possible, examples from Japanese texts are cited from Iwanami's critical edition *Nihon koten bungaku taikei* (Iwanami 1957–69). Examples from Nara period *Senmyō* (Imperial edicts, see 1.2.3.2) are cited from Kitagawa (1982) and texts originally published in print, such as the Christian materials from the end of the sixteenth and beginning of the seventeenth centuries (see 10.2.2), are cited from the originals (or photographic reproductions). References to Rodrigues's *Arte da lingoa de Iapam* (1604–8) are to Doi's (1955) translation into Japanese, which is more easily accessible for most readers and which gives page references to the original.

As few references as possible are given in the body of the text, but a short list of references is provided at the end of most chapters; where possible references are given to scholarship in English. Overall, factual information or generally accepted descriptions which are available in common handbooks, overviews and dictionaries, or by looking at the texts, are not referenced. Handbooks frequently consulted include: *The Japanese language through time* (Martin 1987), *Jidai-betsu kokugo daijiten: Jōdaihen* (Omodaka *et al.* 1967), *Kokugogaku daijiten* (Kokugogakkai 1980), *Kokugogaku jiten* (Kokugogakkai 1955), *Nihon kokugo daijiten* (Shōgakukan 2000–2), *Kokugogaku kenkyūjiten* (Satō 1977), *Nihon bunpō daijiten* (Matsumura 1971), *Nihongo bunpō daijiten* (Yamaguchi and Akimoto 2001), *Nihongogaku kenkyūjiten* (Hida *et al.* 2007), *Nihongo hyakka daijiten* (Kindaichi *et al.* 1988). Of these, the *Nihongogaku kenkyūjiten* has an exhaustive listing and description of available textual sources (pp. 629–1129).

Part I

Old Japanese

1.1 Writing

1.1.1 Introduction of writing in Japan

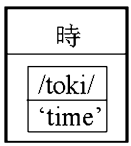
The Japanese were exposed to written matter as early as the late Yayoi period (*c.* ?1000 BC – 300 AD). Thus, inscribed Chinese coins have been unearthed in excavations of sites dating from the first century AD. There is no evidence of any awareness of the function of writing at that early stage, however, and it is likely that the characters which appear on mirrors and other artefacts produced in Japan through the third and fourth centuries were also simple ornaments, in imitation of those found on articles from the continent. To all appearances, writing as such, in the form of Chinese Classics, was introduced into Japan early in the fifth century as part of the great cultural influx from Paekche. The *Kojiki* and the *Nihon shoki* recount this event as the advent of the scribes Wani and Akichi in the years Ōjin 15 and 16 (thought to be early in the fifth century, possibly 404–5; the traditional dating puts this at as early as 284–5, two 60-year cycles earlier). For some time, writing remained in the hands of hereditary professional scribes (*fubito*) who were of continental heritage. Through the sixth and seventh centuries Sinitic culture, including Chinese Buddhism, flowed into Japan through Paekche. In the course of this, written Chinese assumed enormous importance in matters of state, philosophy and religion. Any serious engagement with such matters required knowledge of written Chinese and for some time writing was equivalent with writing in Chinese. Also composition of Chinese poetry became highly regarded and remained so long into the medieval period. Thus, the oldest surviving poetry anthology in Japan is the *Kaifūsō* (懐風藻; *c.* 751) a compilation of Chinese poetry written in Japan. Reading and interpretation of Chinese canonical texts came to assume great importance, both within Buddhism and in government administration. Chinese texts were read in two ways: either (a) reading them out in a form of Chinese (*ondoku* 音読 'sound reading'); or (b) translating or rendering the texts into Japanese (*kundoku* 訓読 'gloss reading'). These two practices have exerted great influence both on the Japanese language itself and on the way it is written. This will be discussed in detail in 9.1.

1.1.2 Writing in Japanese

The earliest attestation of writing in Japanese dates from the fifth century, but it is not until the middle of the seventh century that writing in Japanese became widespread. Throughout the OJ period Japanese was written entirely in *kanji* which were used logographically or phonographically.

1.1.2.1 Logographic versus phonographic writing

Writing is a *representation* of language: elements of writing represent elements of language. Depending upon the nature of the linguistic elements that elements of writing stand for, there are in principle two types of writing. First, writing which represents those elements of language which *carry* meaning: words or morphemes. This is *logographic* writing. Second, writing which represents those elements of language which *distinguish* among elements carrying meaning: phonemes or phonological units of greater or smaller extent. This is *phonographic* writing. Below, these two types of writing are illustrated with examples from NJ. (1a) shows logographic writing, with 時 standing for the word which has the sound shape /toki/ and the meaning 'time'; 時 does not stand primarily for the meaning or the sound shape, but for the word, the linguistic sign, which comprises both. (1b) exemplifies phonographic writing, with と standing for the syllable /to/; と can thus be used to write any occurrence of /to/ regardless of the word of whose sound shape /to/ forms part.

(1) a. 

b. 

Actual orthographic systems and practices rarely, if ever, limit themselves to one of these types of writing. For example, Chinese, which is the stock example of logographic writing, has a strong phonographic element, DeFrancis (1984) arguing that this is more prominent than the logographic element. Conversely, most writing systems have a logographic element. This includes alphabet writing as used to write English; for example, *red* and *read* (past tense of the verb 'to read') are written differently although they are homophonous. Also spaces between words, capitalization of some words, and punctuation all contribute an element of logography to alphabet writing.

1.1.2.2 Adaptation of Chinese script

It is not known specifically when or how the Chinese script began to be used to write Japanese. Nor is it clear by what stages this took place. It *is* clear, though, that it makes little sense to consider this from a purely Japanese per-

spective. This development should be seen within the wider context of the adaptation within the Sinitic cultural sphere of Chinese script to write local languages, which took place earlier on the Korean peninsula than in Japan. There is ample evidence, in the form of orthographic 'Koreanisms' in the early inscriptions in Japan, that the writing practices employed in Japan were modelled on continental examples; later texts with Koreanisms include the epitaph from the grave of Ō no Yasumaro, the compiler of the *Kojiki*. Scholars from the Korean peninsula, particularly Paekche, played an essential role in the development of writing in Japanese by introducing, practising (as scribes) and teaching methods of writing already in use on the Korean peninsula. It is important to appreciate, however, that the presence of identifiable 'Koreanisms' in a text simply means that it incorporated writing practices developed on the continent; it says nothing about where or by whom the text was written.

Chinese script may be adapted to write other languages either logographically or phonographically. Pre-alphabetic writing in Korean comprises the following three main types: (a) pure logographic writing, with *kanji* used for lexical words, but with no indication of grammatical particles or morphology; (b) logographic writing, with conventionalized logographic and phonographic writing of some grammatical elements (Korean *idu* 'clerk readings'); (c) logographically written lexical items supplemented by phonographically written grammatical elements (*hyangch'al*). These types are all found in the OJ text corpus. It is likely that the correspondences in specific types of writing on the continent and in Japan reflect a common continental source rather than parallel developments and there is therefore little sense in trying to reconstruct an independent course of evolution of adapting Chinese script to write Japanese. However, writing extensive text passages entirely or mostly phonographically, reflected in the widespread use of *man'yōgana* (1.1.2.5), is a practice not attested in Korean sources which therefore seems to be an independent development which took place in Japan.

1.1.2.3 Logographic writing of Japanese

The principle for logographic adaptation of Chinese script is that you represent a word with a character which stands for a Chinese word which is semantically or referentially similar to the word you wish to write. In a sense, you translate the word you want to write into Chinese and use the character which stands for the resulting Chinese word. Simplistically put, you take the meaning associated with a character with no regard to its sound value. For example, 犬, which in modern Chinese stands for the word *quǎn* 'dog', could be, and today is, used to write the Japanese word *inu* 'dog'; or it could be used to write the English word *dog*. In an extreme type of logographic writing of Japanese, known as *hentai kanbun* 'deviant Chinese text', stretches of text longer than

single words are represented by strings of *kanji* in an order different from the word order in Japanese, making the texts appear more like Chinese than Japanese, but they are in fact complex representations of Japanese (see further 9.1.2.1).

1.1.2.4 Phonographic writing of Japanese

Phonographic writing makes possible a precise and unambiguous recording of linguistic forms, something which is not possible to the same extent in logographic writing. The principle for phonographic use of *kanji* is that a sound sequence – usually a syllable – is represented by a character otherwise used logographically to write a word whose sound shape is similar to the syllable you want to write. Simplistically, you take the sound value of a *kanji* with no regard to its meaning. For example, 弥 which stands for the word *mí* 'full, fill, more' in Chinese, could be used to write the syllable /mí/ in Chinese regardless of the word in which it occurs. Or it could be borrowed to write the sound sequence /mi:/ in English, or /mi/ in Japanese. Thus for example, in both Chinese and Japanese the loan word meaning '(Catholic) mass' borrowed in the late sixteenth century from Portuguese *missa* is written 弥撒, standing for Chinese *mísa*, Japanese *misa*. In China itself there is a long tradition for transcribing foreign names and words whose sound shape is thought to be important in this way. That was for example the case with the religious terminology of Buddhism when it was introduced into China,[1] or with the recording of foreign words in Chinese histories.

1.1.2.5 Man'yōgana

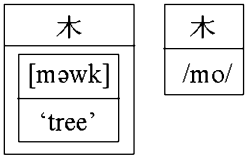
Kanji used to write Japanese phonographically are termed '*man'yōgana*' after their extensive use in the poetry anthology *Man'yōshū* (1.2.3.1). Sometimes the term '*man'yōgana*' is mistakenly used to refer to the entire orthography of the *Man'yōshū*, but this is wrong on two counts: First, *'man'yōgana'* refers exclusively to phonographic writing and not to the logographically written parts of the *Man'yōshū*. Second, *man'yōgana* are used also in other texts than the *Man'yōshū* and the use of *man'yōgana* continued long into the Middle Japanese period, also after the emergence of the simplified *kana* scripts (cf. 6.1.2).

It is customary to distinguish between two basic types of *man'yōgana*, depending on the provenance of their sound values: (a) *ongana*, which were used as phonograms on the basis of their (Japano-)Chinese readings; and (b) *kungana*, which were used on the basis of their *kun*-readings. This is illustrated

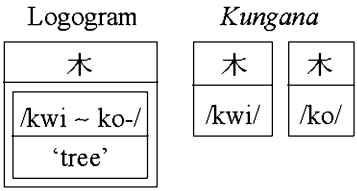
[1] Buddhist texts were translated into Chinese, but important concepts and terminology in them were rendered phonographically because their sound shape was thought to be related to their function, or because concepts such as *nirvana* were so alien as to make transliteration the only sensible option.

below with 木 which was used both as an *ongana* and as a *kungana*. Logographically, this character wrote words meaning 'tree' and the like. It was used as a *kungana* for the syllables /ko/ and /kwi/ because the OJ word meaning 'tree' had the variants *ko- ~ kwi*. It was used as an *ongana* for the syllable /mo/ because the phonetic manifestation of the Chinese word meaning 'tree' was similar to the phonetic manifestation of the OJ syllable /mo/.

(2)



(3)



Below are given some more examples of *man'yōgana*:

(4)

Kanji	***Ongana***	Early Middle Chinese	***Kungana***	OJ word
比	**/pi/**	*pjiʰ/bjiʰ; pji'; bji		
肥	**/pwi/**	*buj		
咩	**/mye/**	*mjiə'		
売	**/mye/**	*maɨjʰ, mɛ:jʰ		
米	**/me/**	*mɛj'		
古	**/kwo/**	*kɔ'		
許	**/ko/**	*xɨə̆'		
子	**/si/**	*tsɨ'/tsi'	**/kwo/**	*kwo* 'child'
木	**/mo/**	*məwk	**/ko, kwi/**	*ko- ~ kwi* 'tree'
八	**/pa/**	*pəɨt/pɛ:t	**/ya/**	*ya* 'eight'
田	**/de/**	*dɛn	**/ta/**	*ta* 'paddyfield'

In a very general sense, *kanji* were used as *ongana* on the basis of a perceived phonetic similarity between the OJ syllables they represented and their

Chinese pronunciation. The adapted form of Chinese pronunciation used in Japan will be referred to as 'Japano-Chinese'; we will discuss this term and its implications in 9.2.1. There are two points to be made in that regard: (a) The phonetics of J-Ch is not known. Early Middle Chinese (*c.* 601; see 9.2.1.2) is available in a number of reconstructions (here we follow Pulleyblank's (1991) reconstruction), but the relation between EMC and J-Ch is not direct (see 9.2.1.2). (b) In addition to phonetic and phonological considerations, it is likely that scribal tradition brought along from the continent by scribes, as well as any number of other 'extra-phonological' factors, played an important part in the choice of *kanji* to act as *ongana*. Consider for example the sound values (including the NJ *on*-readings) of 売 (EMC: *maɨj^{h}/mɛ:j^{h}; Go-on: *me*; Kan-on: *bai*), 米 (*mɛj'; *mai*; *bei*), and 咩 (not in Pulleyblank 1991, but reconstructed by Miyake (p.c.) as EMC *myiəq, which corresponds to Pulleyblank *mjiə'; *mi*; *bi*). It is clear that 売 has sound values far more similar to 米 than to 咩, yet 売 is used, like 咩, as an *ongana* for the syllable /mye/, as opposed to 米 which is used to write /me/.

1.1.2.6 Senmyō-gaki

The majority of OJ texts are written in a mixture of phonographic and logographic writing. Whereas some mixed texts have no clear functional differentiation between phonograms and logograms, others mainly use phonograms to write grammatical elements and logograms to write lexical words. A distinctive way of writing which has become known as *senmyō-gaki* 'edict-writing' (after its use in the *Senmyō*, see 1.2.3.2), is a mixture of logographic and phonographic writing in which some grammatical elements were written phonographically in smaller size characters than the rest of the text. *Senmyō-gaki* is a refined writing system, showing a sophisticated grammatical understanding of the language. It may well have been inspired by the Korean *idu* (see 1.1.2.2). *Senmyō-gaki* is similar to the mixed writing of modern Japanese, in the sense that both exhibit a high degree of orthographic distinction between lexical words and grammatical elements.

Senmyō-gaki is usually associated with *Senmyō* and *Norito* which are the main texts written in this way. It is, however, not exclusive to them, but is found in other texts as well. Both wooden tablets *(mokkan)* (see 1.2.2) and archival records from the mid eighth century show that the practice of distinguishing in size between characters used phonographically for grammatical markers and others was well established by that time, but it is not clear how far back this type of writing dates, and thus if all *Senmyō* and *Norito* were originally recorded in this way. Thus *mokkan* from the second half of the seventh century include texts in writing which otherwise is very similar to *senmyō-gaki*, but without a distinction in size.

1.1.3 Problems of decipherment

Writing is a representation of language and reading consists in reconstructing a linguistic specimen, a text, from its written representation, i.e. decoding the text. It thus requires knowledge of the 'code of transmutation' employed when a text was written down, i.e. encoded in writing. Although reading is commonplace and we tend to think little of it, it is in principle a complicated process. This is clear when we approach texts representing an older language stage or an unknown language with the purpose of establishing what the texts might mean or what the language might be like. This is very different from mapping our knowledge of a language onto a written representation of a text in order to recover it. Individual OJ texts exhibit some degree of consistency in their orthography, but viewed as a whole OJ writing is enormously complicated. This holds in particular for the writing in the main source of the OJ language, the *Man'yōshū,* which is not a single text, but a compilation of texts. The complexity of OJ writing means that deciphering the OJ text corpus is no simple matter and there are in fact still many obscure points despite a long philological effort.

1.1.3.1 Polyvalence and equivalence

The main regular problems of decipherment are posed by the polyvalence and equivalence of *kanji*. (a) *Kanji* used to write OJ were *polyvalent* in several respects. First of all, they could be used as logograms or as phonograms. As shown in the examples above, *kanji* used as phonograms could be used as *ongana* or as *kungana* and some were used as *kungana* for different syllables (this mostly confined to variant forms of one word). When used as logograms, *kanji* could be used for different OJ words. For example, 去 (EMC *$k^hiə^h$ 'go away, depart') could stand for the words *sar-* 'leave' or *yuk-* 'go'. (b) On the other hand, some *kanji* were *equivalent* in the sense that one linguistic unit could be represented by several different *kanji*. For example, the word *yuk-* 'go' could be represented logographically with, amongst others, 行 (*ɣaɨjŋ/ɣɛːjŋ 'walk, go'), 逝 (*dʐiaj^h 'pass away, die'), 去 (*$k^hiə^h$ 'go away, depart'), or 往 (*wuaŋ' 'go'). The origin of such usage is the translational inequivalence between Chinese and Japanese: several OJ words could be used to translate 去, and vice versa several Chinese words could be translated by OJ *yuk-*. Finally, each syllable could be written phonographically by a number of different *kanji*; for example, in the *Man'yōshū* close to twenty different *kanji* are used as *man'yōgana* for the syllable /ka/.

The problems of decipherment posed by polyvalence and equivalence cannot be solved on the basis of the texts alone. However, the main texts were read and copied through time and annotated in order to facilitate reading when in later periods the language and the orthography became obsolete. There is

thus a tradition concerning these texts which formed the point of departure for a long and thorough philological effort of decipherment; and today the phonographically written portions of the OJ texts are deciphered to general satisfaction on most points of significance.

1.1.4 Reading tradition

Reading tradition gives voice to the logographically written portions of the OJ texts. Undoubtedly the traditional readings are in many cases correct, in the sense that they reflect the words and word forms which were originally intended in a text. Strictly speaking, however, the readings constitute hypotheses about the texts; often probable hypotheses, well founded on sound knowledge of the language and on tradition, but all the same hypotheses which mostly remain unverifiable. Generally, we cannot *know* which words are represented in an older logographically written text. On this background it is obvious, but still enormously important to make clear, that no argument about the phonology or morphology of OJ can be based on logographically written text portions and their traditional readings. They can provide valuable clues in matters of syntax, but also in this area caution must be exercised.

1.1.5 Rebus writing

A special and intriguing problem is posed by obscure rebus writings, employed in particular in the *Man'yōshū*. Two well-known examples will suffice here. The first is found in *MYS* 9.1787 where a form of the word *ide-* 'emerge, come out' is written by the five characters 山上復有山 which stand for words meaning 'mountain top again exist mountain', in Chinese forming a sentence 'there is a mountain on top of another mountain', or 'there is a 山 on top of another 山', in reference to the graph 出 which is more conventionally used to write *ide-* and which does look like one 山 on top of another. The second, from *MYS* 11.2542, has *nikuku* (the infinitive of the adjective *niku-* 'hard') written by 二八十一, where 二 is a regular *ongana* for /ni/, but 八十一, '8-10-1' here standing for a word meaning '81', is used for the syllables /kuku/ because the character for the word meaning 'nine' was used as *ongana* for the syllable /ku/ and 81 is equal to 9×9. The orthographic playfulness evident in these writings highlights that writing in Japanese in many cases was no practical matter of communication, but a leisure activity.

1.1.6 Examples

The following are three examples of OJ writing, the first poems in the *Kojiki* and in the *Man'yōshū*, respectively, and the first sentence in the first *Engishiki*

norito. The texts are transcribed using CAPITALS for logograms, *plain italics* for *kungana* and ***bold italics*** for *ongana*. (This is different from the transcription of OJ examples adopted through the rest of the book where we use italics for phonographic text and normal type for logographic text.) The *Kojiki* text is written entirely in *ongana*; it consistently uses the same *kanji* for each syllable, e.g. 都 for /tu/ in *tatu* 'rise', *tuma* 'wife', *tukuru* 'make'. The *Man'yōshū* text is written in a complicated mixture of logographic and phonographic writing, using different ways of writing the same words, e.g. *moti* 'holding' written once phonographically as 母乳, with an *ongana* and a *kungana*, and once logographically as 持. The *Norito* text is written almost exclusively logographically, with only the complementizer *to* written in a small size *ongana*.

(5) a. *Kojiki* (KK 1)

Text: 夜久毛多都伊豆毛夜幣賀岐都麻碁微爾夜幣賀岐都久流曾能夜幣賀岐袁

Interpretation:

夜 久毛	多都	伊豆毛 夜 幣 賀岐	都麻 碁微 爾
ya-kumwo	***tatu***	***idumwo ya-pye-gaki***	***tuma-gomwi ni***
eight-cloud	rise.ADN	Izumo eight-fold-fence	wife-enclosing DAT

夜幣賀岐	都久流	曾 能 夜 幣 賀岐 袁
ya-pye-gaki	***tukuru***	***so no ya-pye-gaki wo***
eight-fold-fence	make.CONCL	that GEN eight-fold-fence EXCL

'The many-fenced palace of Idumo Of the many clouds rising – To dwell there with my spouse Do I build a many-fenced palace: Ah, that many-fenced palace!' (Philippi 1968:91)

(NJ reading: *yakumo tatsu Izumo yaegaki tsumagomi ni yaegaki tsukuru sono yaegaki o*)

b. *Man'yōshū* (*MYS* 1.1)

Text: 篭毛與美篭母乳布久思毛與美夫君志持此岳尓菜採須兒家吉閑名告紗根

Interpretation:

篭 毛 與	美 篭	母乳	布久思 毛 與	美夫君志
KWO ***mo yo***	***mi***KWO	***moti***	***pukusi mo yo***	***mibukusi***
basket ETOP EMPH	HON-basket	hold.INF	shovel ETOP EMPH	HON-shovel

持
MOTI
hold.INF

此	岳 尓	菜	採 須	兒
KO NO	WOKA ***ni***	NA	TUMA-***su***	KWO
this GEN	hill DAT	greens	pick-RESP.ADN	child

家	吉閑名	告 紗根
IPYE	***kikana***	NORA-***sane***
home	ask.OPT	tell-RESP.OPT

'Girl with your basket, with your pretty basket, with your shovel, with your pretty shovel, picking greens on this hillside, I want to ask your home. Please tell me!'

(NJ 'reading': *ko mo yo miko mochi fukushi mo yo mibukushi mochi kono oka ni na tsumasu ko ie kikana norasane*)

c. *Norito* (*EN* 1)

Text: 集侍神主祝部等諸聞食登宣

Interpretation:

集 侍		神主·	祝部 等、	諸
UGWONAPAR-ERU		KAMUNUSI,	PAPURI-RA,	MOROMORO
gather-STAT.ADN		kamunusi	hafuri	all

聞食	登	宣
KIKOSI-MYESE	***to***	NORU
hear.RESP-RESP.IMP	COMP	say.CONCL

'Hear me, all of you assembled *kamunusi* and *hafuri*. Thus I speak.' (Philippi 1990:17)

(NJ 'reading': *ugonawareru kannusi, hafurira, moromoro kikoshimese to noru*)

1.2 Sources

1.2.1 Japanese words in foreign sources

The Chinese history *Wèi zhì* (魏志 *Chronicles of Wei*, Japanese *Gishi*; a history of the Wei state (220–65), compiled towards the end of the third century) has a section, *Wō rén chuán* (倭人伝 *Account of the dwarfs*; Japanese *Wajinden*), which describes people living on the Japanese archipelago, also citing 53 phonographically transcribed words from the language spoken by these people. These words are often thought to be from an earlier stage of Japanese. Almost all are proper nouns or titles, with little known about their meaning. Among them are 邪馬台 and 卑弥呼 which have entered Japanese folklore in the cNJ forms *Yamatai* (usually thought to represent an earlier form of *Yamato* (OJ

Yamato), an autochthonous name for Japan) and *Himiko* (the name of a mythical early female ruler which presumably would have been OJ *Pimikwo, but which has no OJ attestation). It is unlikely that the words in *Wō rén chuán* are anything but ad hoc transcriptions made by Chinese; they should not be thought to represent conventional writings of these names and words. Indeed, the Japanese are not thought to have been literate in the third century. The transcriptions have not been deciphered, i.e. their sound values are not known, and they are not easily decipherable. The material against which they must be checked are the latest reconstructions of OC, which are, however, in themselves hypothetical. Even assuming sound values close to those which the characters had when used to write OJ words more than four centuries later, only in some cases do the words resemble known OJ words. Miller (1967:12–27) is an optimistic attempt to identify these transcriptions with Japanese vocabulary, representing the mainstream of traditional scholarship on the subject. It is not even evident that all of these words represent a language related to OJ. It is, in short, not possible to draw any conclusions about 'Japanese' on the basis of the transcriptions in *Wō rén chuán*.

1.2.2 Early inscriptions, wooden tablets and archival records

The oldest sources of Japanese in Japan are inscriptions made on stone and metal (swords and mirrors), the earliest thought to date back to the fifth century. The Japanese in these inscriptions is limited to proper names in texts otherwise written in Chinese or *hentai kanbun*. The following three are the most important early inscriptions:

Inariyama tumulus sword inscription. ?471 (/?531). Excavated from a tumulus in Saitama prefecture. Thought to be the oldest inscription made in Japan. The inscription is dated as a cyclical year which corresponds to 471; another possibility is 531. It is the longest early inscription, consisting of 115 characters, 46 of which are used phonographically to write Japanese personal and place names. The inscription employs features which are characteristic of early writing in Korea in terms of the shape of characters and the choice of characters used as phonograms; it also contains a significant grammatical Koreanism (中 used as a locative marker).

Eda Funayama tumulus sword inscription. Late fifth or early sixth century. Kumamoto prefecture. The text is thought to have been seventy-five characters long, but because of lacunae only sixty-four are legible. It is in Chinese, but with the same grammatical Koreanism as the Inariyama tumulus sword inscription. It has Japanese names written phonographically. However, the writer is identified as 張安 (Zhang An), someone of continental origin.

Yakushi nyorai statue inscription. Probably second half of seventh century. Hōryūji Temple in Nara. This is ninety characters long. It is the oldest extant

inscription written in *hentai kanbun*; it has both some Japanese word order and some logographic expression of Japanese grammatical elements.

There also exists a large corpus of more than 10,000 early archival records on paper (*komonjo* 古文書) from the eighth century; out of these, however, only two are written phonographically in Japanese.

Recently, a large number of wooden tablets with writing on them (*mokkan* 木簡) dating from mid seventh to mid eighth century have been unearthed in different parts of Japan, the majority, however, around the old capital areas (Nara, Fujiwara). The existence of wooden tablets has been known for a while, but it is only in the post-war period, and particularly through the 1980s and 1990s, that large numbers have come to light. The total of these wooden tablets is now in excess of 150,000. They are usually short, often written in Japanese, with both logographic and phonographic writing represented. The writing ranges from labels through writing practice and scribbles to letters. *Mokkan* constitute the earliest evidence of popular writing in Japanese, showing that writing in Japanese was widespread in the second half of the seventh century and that literacy was not as confined as has previously been thought.

Inscriptions, archival records and wooden tablets are valuable for their authenticity. They are, of course, enormously important to historians. They also provide important information about early use of writing in Japan and about the development of writing in Japanese. However, on the whole, these materials contribute relatively little to our understanding of the Old Japanese language as such when compared with the texts of the eighth century. It should be mentioned, though, that *mokkan* offer glimpses of language use which seems spontaneous and informal. However, the material is also very limited and *mokkan* may at best serve as a corrective to our knowledge of the Old Japanese gleaned from other sources.

1.2.3 Eighth-century texts

The bulk of our sources of OJ are texts from the eighth century. They have been handed down in copy and many competing manuscripts exist, all of which date from later periods. A long philological tradition, however, has made it possible to arrive at critical editions which are widely accepted as representing fairly well the texts of the eighth century.

As mentioned, the texts comprise both logographic and phonographic writing. It goes without saying that it is the phonographically recorded texts or text portions which are the most important for the study of the language. It is important to keep in mind that no argument about the phonology or morphology of OJ can be based on the traditional readings of logographic

Table 1.1 *Important OJ sources*

Kojiki (712)
Harima fudoki (*c.* 715)
Hitachi fudoki (714–718)
Nihon shoki (720)
Izumo fudoki (733)
Bungo fudoki (730s)
Hizen fudoki (730s)
Bussokuseki-ka (after 753)
Man'yōshū (after 759)
Senmyō (697–791)
Kakyō hyōshiki (772)
Kogoshūi (807)
Engishiki norito (dates unknown; compiled 927)

passages. They can provide valuable clues in matters of syntax, but also in this area caution must be exercised. There are two main genres of text in the material: poetry and prose. Stylistically, these texts are either in highly formal, ritualistic prose or in a poetic form, ranging from folk-songs which were handed down (and thus subject to some editing both in that process and in that of recording) to elaborate poems. In any case, the language of these texts is probably in some aspects quite far removed from contemporary spontaneous and informal spoken language. In addition, there are Japanese vocabulary items and proper names in texts written in Chinese or in *hentai kanbun*, in the form of phonographically written items inserted directly into the texts, or explanatory notes written as part of the original text (as opposed to later additions). Needless to say, this in the main provides information about the OJ lexicon, not its grammar. Notes and glosses added onto Chinese texts in order to facilitate their interpretation and rendition into Japanese, the so-called *kunten* texts, constitute important material for the study of EMJ (see 6.2.2, 9.1.1). Although the practice probably had caught on already towards the end of the Nara period, surviving materials from that time are extremely scarce. Table 1.1 is a chronological list of main sources for the OJ period.

1.2.3.1 Poetry

The main corpus for OJ is the poetry in the *Kojiki*, the *Nihon shoki* and the *Man'yōshū*. This constitutes the material on which most of our knowledge

about OJ rests. The poetry comprises large portions written phonographically, thus making possible a comprehensive study of the phonology and morphology of OJ. Because of the general nature of poetry and the specific rhythmic constraints of Japanese poetry, with its dominant five- to seven-syllable metre, these materials provide less information about syntax.

The *Kojiki* (古事記 'Record of ancient matters') is a history, compiled in 712. The preface is written in Chinese and the main text is written in *hentai kanbun*, but it also contains 112 phonographically written songs, as well as proper names and vocabulary. The *Nihon shoki* (日本書紀 'Chronicles of Japan') is also a history, compiled in 720. It is written in Chinese, but contains 128 songs, as well as proper names and vocabulary, written phonographically. The *Man'yōshū* (万葉集 'Collection of myriad leaves') is the major source of the OJ language. It is a poetry anthology of more than 4,500 poems in 20 volumes. It was compiled in late Nara or early Heian. The latest poem is dated 759 and the earliest poem is usually said to date back to the middle of the fifth century. The poems are in Japanese, written both phonographically and logographically. Most poems present a mixture of phonographic and logographic writing, but the proportions differ enormously. The poems in the anthology are drawn from several sources, both older and contemporary. It is thus not a single text, but a collection of texts from different times, and consequently there is no overall orthographic consistency. There is, however, a substantial, identifiable portion representing early eighth-century OJ and another representing mid eighth-century OJ. Eastern OJ dialect is also represented in volumes 14 (*azuma uta* 'eastern songs') and 20 (*sakimori uta* 'border-guard songs').

1.2.3.2 Prose

The prose corpus consists of two sets of texts: (a) *Norito* (祝詞 'liturgies') comprise ritual prayers and blessings. Twenty-seven *Norito* are recorded in volume 8 of the *Engishiki* (延喜式 'Procedures of the Engi Era', completed 927) and are, despite the late date of compilation, thought to retain their OJ form and to reflect fairly accurately the OJ language. (b) *Senmyō* (宣命) are imperial edicts. Sixty-two edicts are recorded in the *Shoku nihongi* (続日本紀 797; a history covering 697–791). Both *Norito* and *Senmyō* are written in *senmyō-gaki*, i.e., mainly logographically, but with grammatical items and some vocabulary noted phonographically. Although smaller than and very different from the corpus of poetry, these prose texts are extremely valuable for the study of some aspects of OJ syntax, especially the use of case particles. They also contribute to lexical studies and give an insight into ritual uses of language, as well as evincing clear evidence of influence from Chinese. The *Norito* in particular are thought to reflect quite early features of the language. Both *Norito* and *Senmyō* are as yet inexhaustively studied.

1.2.3.3 Others

Fudoki (風土記 'Records of wind and earth') are topographies compiled on the order given in 713 by the empress Genmei to the governors of the provinces to provide information about local products and produce, the fertility of the land, and folklore. Five *fudoki* have survived: *Hitachi fudoki* (常陸風土記 (the north-eastern part of present-day Ibaraki prefecture), compiled between 714 and 718); *Harima fudoki* (播磨風土記 (in Hyōgo prefecture), compiled around 715), *Bungo fudoki*, *Hizen fudoki* (豊後風土記 (in Oita prefecture), 肥前風土記 (in Saga and Nagasaki prefectures), compiled in the 730s), and *Izumo fudoki* (出雲風土記 (in Shimane prefecture), completed 733), which is the only *fudoki* to have survived in its entirety and which is written in *hentai kanbun*, whereas the rest are written in Chinese. They contain phonographically written poems (some twenty), place names and local vocabulary.

The *Bussokuseki-ka* (仏足石歌 'Footprints of the Buddha poems') is a series of twenty-one poems inscribed on a stone at the Yakushi-dera in Nara. The provenance is unknown, as is the date of inscription, but it is thought to date after 753. The poems are written entirely phonographically and are valuable because of their authenticity as a primary source. The *Kakyō hyōshiki* (歌経標式) is the earliest known poetic treatise written in Japan, from 772 and by Fujiwara no Hamanari. It is written in Chinese, but has phonographically written poems in Japanese. The *Kogoshūi* (古語拾遺 'Gleanings in Old Words') is a history, compiled in 807 by Imbe no Hironari (then in his seventies). It is mostly in Chinese, but contains two phonographically written poems and some vocabulary.

REFERENCES

Writing in Korea: Sohn 1999: 124–8. Early Korean–Japanese contact: Lewin 1976, Bentley 2000. Introduction of writing and adaptation of Chinese script in Japan: Miyake 2003a, Seeley 1991. *Man'yōgana*: Lange 1973, Miyake 2003a. Overview over sources: Vovin 2005: 1–20. Inscriptions: Murayama and Miller 1979, Seeley 1991: 16–40. Reproductions and transcriptions of most inscriptions: Kariya *et al.* 1912. Wooden tablets: Farris 1998: 201–32, Kotani 1986. *Wèi zhì*: Farris 1998: 10–14, Miller 1967: 12–27, Tsunoda and Goodrich 1951. *Norito* and *Senmyō*: Bentley 2001a.

Early Old Japanese had the distinct syllables shown in Table 2.1, illustrated with one representative *man'yōgana* for each syllable.

2.1 *Kō-rui* and *otsu-rui* syllables

OJ kept distinct so-called *kō-rui* (type A) and *otsu-rui* (type B) syllables, which merged in the transition to EMJ. The difference is here noted by subscript '1' and '2'. Thus, the two distinct OJ syllables represented by, amongst others, 売 and 米 respectively are termed *me*$_1$ (*kō-rui me*) and *me*$_2$ (*otsu-rui me*) because they merged and are reflected as EMJ (and NJ) *me*. The following are examples of minimal pairs.

(1)	OJ		EMJ	NJ
	pi$_1$	'sun'	*pi*	*hi*
	pi$_2$	'fire'		
	me$_1$	'woman'	*me*	*me*
	me$_2$	'eye'		
	ko$_1$	'child'	*ko*	*ko*
	ko$_2$	'this'		

Phonemically, the difference between the *kō* and *otsu* syllables is generally agreed to pertain to the post-consonantal part of each syllable. However, despite the subscript convention used here and elsewhere, and despite the way the orthographic *kō–otsu* distinction is often talked about, it is important to appreciate that OJ did not have 'two kinds of' /i/, /e/ or /o/, any more than Classical Greek had many kinds of '*i*'. EMJ /i/, /e/ and /o/ each represent the outcome of a merger between two distinct sounds of OJ, just like Modern

Table 2.1 *Orthographically distinct syllables in early OJ*

阿 .a	加 ka	我 ga	左 sa	射 za	多 ta	陀 da	奈 na	波 pa	婆 ba	麻 ma	夜 ya	良 ra	和 wa
伊 .i	支 ki_1 貴 ki_2	祇 gi_1 疑 gi_2	之 si	自 zi	知 ti	遲 di	爾 ni	比 pi_1 肥 pi_2	鼻 bi_1 備 bi_2	美 mi_1 未 mi_2		利 ri	為 wi
宇 .u	久 ku	具 gu	須 su	受 zu	都 tu	豆 du	奴 nu	布 pu	夫 bu	牟 mu	由 yu	流 ru	
衣 .e	家 ke_1 気 ke_2	牙 ge_1 義 ge_2	勢 se	是 ze	天 te	田 de	尼 ne	平 pe_1 戸 pe_2	弁 be_1 倍 be_2	売 me_1 米 me_2	延 ye	礼 re	恵 we
於 .o	古 ko_1 許 ko_2	呉 go_1 其 go_2	蘇 so_1 曾 so_2	俗 zo_1 叙 zo_2	刀 to_1 止 to_2	度 do_1 特 do_2	努 no_1 乃 no_2	富 po	煩 bo	毛 mo_1 母 mo_2	用 yo_1 余 yo_2	路 ro_1 呂 ro_2	乎 wo

Greek /i/ reflects the merger of a number of distinct sounds of Classical Greek (including υ /y:/, ει /e:/, η /ɛ:/, υι /yi/, ηι /ɛ:i/, οι /oi/).

The distinction between the *kō-* and *otsu-rui* syllables disappeared from the language in the transition to EMJ and is not reflected in the *kana* writing of EMJ or later periods. As each OJ syllable could be represented by different *man'yōgana*, the distinction was not conspicuous to later generations of readers and in fact it was only finally discovered at the beginning of the twentieth century by Hashimoto Shinkichi, one of the great Japanese linguists and philologists. This is Hashimoto's own account of his discovery (*c.* 1915, cited from Ohno 1980: 139ff.; my translation).

While in February 1909 I was researching developments in the syntax of Japanese at the request of the Japanese Language Investigative Committee [*Kokugo Chōsa Iinkai*], I noticed in the eastern songs [*azuma uta*] in volume 14 of the *Man'yōshū* that the character 家 in not a few cases was written in places which ought to have the particle *ga*. This gave rise to the suspicion that the eastern dialect of that time had a particle *ke* which was used in the same way as *ga*. As a way of solving this I felt it necessary to check up on every single 家. First I examined volume 14, but to no avail. Because there are examples in volume 5 of 家 used in the meaning of *ga* I decided also to examine volume 5. This time I did not limit myself to the character 家, but collected every single *kana* for *ke*. I then found – and this had no bearing on the problem I was investigating – that forms such as *ni-keri* 'perfective-modal.past', *ke-mu* 'past-conjectural', *kerasi* 'past.presumptive', and *kepu* 'today' were written exclusively with characters from one set, comprising e.g. 家 and 計, whereas words such as *take* 'bamboo', *sake* 'saké', and *take* 'mountain, peak', *nageki* 'sigh', *sigesi* 'thick, dense (of growths)' were written exclusively with characters from another set, comprising e.g. 気 (*ke*) and 既 (*ke*), 宜 (*ge*). I discovered that among words written with *ke* some characters could be used interchangeably, while others could not, and that *kana* for *ke* accordingly fell in two groups; the distinction between these two groups was maintained consistently. I felt strange discovering this. Experiencing great interest I decided to proceed and examine the other volumes. I wanted to begin with the eastern dialect and examined volume 20, but in the eastern dialect border-guard songs [*sakimori uta*] I was not able to verify this distinction, which made me feel greatly disappointed. However, in the poems in the central dialect in that volume, I found that this distinction clearly existed, and I thus came to the conclusion that this was a distinction which existed in the central dialect, but not in the eastern dialect.

Next, I examined those volumes of the *Man'yōshū* which are written almost entirely phonographically, volumes 15, 17, 18. There would occasionally be a few exceptions, but I found that in almost all instances this distinction was there. When I looked at the remaining volumes of the *Man'yōshū* as well as the songs in *Nihon shoki* and *Kojiki*, apart from volume 14 of the *Man'yōshū*, the distinction between these two types was there, without a single exception. This led me to believe that this distinction generally held for the Nara period – excepting the eastern dialect. Going on to check the exceptions I had previously found in the *Man'yōshū* against an old copy of the *Man'yōshū* in the possession of Ōya Tōru [1850–1928], I learned that almost all were not real exceptions, but mistakes in the wood-block printed editions, and I arrived at an even

firmer belief in the existence of this distinction. (In the following year, 1910, I was able to take a look at a *Genryaku*-period (1184) manuscript of the *Man'yōshū*, which agrees with Ōya's. It was increasingly clear that the printed editions were mistaken.) When I next examined texts from the Nara period such as *Jōgū Shōtoku-hōō taisetsu* [late seventh century], texts in the *Dai-Nihon komonjo* [a compilation of handwritten materials, published 1901–40 by Tokyo University Press] and the *Senmyō* in the *Shoku nihongi* there was not a single exception. I wanted to look at the traces of the changes in this *kana*-use and therefore examined texts from the Heian period written in *man'yōgana* such as edicts and songs contained in the official histories from *Nihon kōki* [840] onwards, *Nihongi kyōenka* [compilation from three occasions (882, 906, 943) of poems composed at the court], *Shinsen jikyō* [a dictionary, compiled *c.* 898–901], and the *Nihon ryōiki* [a collection of tales, written mainly in *kanbun*, completed after 822]. In most it was not possible to observe this distinction. Thus, I was able to learn that this distinction between two types of *ongana* for *ke* existed in the Nara period, but had broken down in the Heian period.

While pursuing the investigation of *ongana ke*, I also examined the *k*-initial syllables in volume 5 of the *Man'yōshū* in order to check if there was such a distinction for other *kana*. I was able to infer that besides *ke* there must have been also two types of *ki* and *ko*. I had previously gathered all the verbs in the *Nihon shoki* and *Kojiki* according to inflected forms. Looking at these in the light of my research on *ke*, I was furthermore able to learn that there must have been two kinds of *pi*, *mi*, *pe*, and *me*.

This kind of *kana*-use was something no one else had hitherto explained, and just when I thought that this was one of the few original discoveries since the time of Keichū [famous Edo period philologist, 1640–1701] and had decided to proceed with my investigation of each *kana* in larger materials, I accidentally looked at *Kogen betsu'on shō* [*c.* 1849, by Kusakado Nobutaka, 1818–69] which had just been acquired by the Japanese Language Department at Tokyo Imperial University. Thereby I learnt that this kind of research had been done by someone else previously, and when I looked at the *Kana-zukai no Oku no yamamichi* [1798, by Ishizuka Tatsumaro, 1764–1823] on which that book is based, I realized that it shows that there was a distinction between two types of *kana* on a much grander scale than what I had researched and been able to infer. My discovery was really a rediscovery. Thus, this discovery of mine was a discovery in a two-fold sense. One was the rediscovery of this special *kana*-use, the other was the discovery of Ishizuka Tatsumaro's unknown studies into *kana*-use. If I had not on my own discovered this *kana*-use, I surely would not have been able to understand the importance of *Kana-zukai no Oku no Yamamichi*, or to appreciate its value. Thus, only because I had investigated this independently, I grasped the merits of the studies of that man of old.

Ever since Hashimoto's discovery of this orthographic distinction, the linguistic interpretation of it has been the dominant topic of research in Old Japanese phonology. It is a matter which continues to attract attention and different views. As there is no convenient way of representing the distinction in the present-day Japanese writing system, it remains widely ignored outside linguistics. This is probably to some extent a consequence of the lack of consensus about its phonemic interpretation, but it is regrettable and makes difficult a full appreciation of the sound texture of OJ poetry.

As shown in Table 2.1, the earliest attested Old Japanese distinguished eighty-eight syllables. The distinction between *mo*$_1$ and *mo*$_2$ is found in the *Kojiki* (712), but had merged in later sources. Sometimes a distinction is also posited between *po*$_1$ ≠ *po*$_2$ and *bo*$_1$ ≠ *bo*$_2$. It is questionable whether the existence of such an orthographic difference can be demonstrated within the OJ text corpus, but it is almost certain that **po*$_1$ ≠ **po*$_2$ and **bo*$_1$ ≠ **bo*$_2$ were phonologically distinct in slightly earlier Japanese. The syllable inventory in Table 2.1 represents a system in change and the gradual breakdown of the distinctions is evident through the OJ text corpus. The last syllable pair to be distinguished was *ko*$_1$ ≠ *ko*$_2$, which were kept distinct into the early tenth century, but by then probably merely as an orthographic convention.

2.1.1 Co$_1$ *versus Co*$_2$

There are not many minimal pairs distinguished by the difference between *Co*$_1$ and *Co*$_2$ and it has been proposed (Matsumoto 1984) that *Co*$_1$ and *Co*$_2$ in fact did not represent distinct syllables, but were allographic variants, possibly standing for allophonic variants. This is a consequence of the following distributional differences between *Co*$_1$ and *Co*$_2$: Most occurrences of *Co*$_1$ are in morpheme final position and *Co*$_2$ did not occur in a root morpheme with *Ca*, *Cu*, *Co*$_1$, see below. Also, there are not many monosyllabic lexical morphemes with *Co*$_2$. However, while the scope for opposition is fairly limited, there are minimal pairs, e.g. *ko*$_1$ 'child' versus *ko*$_2$ 'this', making clear that the two sets of graphs did stand for distinct syllables.

2.1.2 Phonetic reconstruction and phonemic interpretation

We face two distinct, but interrelated, problems concerning the phonological understanding of the OJ syllable inventory: phonetic reconstruction and phonemic interpretation, in particular segmentation. Phonetic reconstruction is based mainly on the following two types of evidence: (a) *external*: comparison with the phonetics of other languages, first of all comparison with the (Early Middle) Chinese sound values of the *kanji* used as *on-gana*; (b) *internal*: comparison with the phonetics of later stages and dialects of Japanese, and sound changes within Japanese, either documented changes in EMJ (or later) or reconstructed changes in pre-OJ. Such internal and external evidence can provide valuable clues, but it also has important difficulties of interpretation and includes a number of unknown factors. Arguments based on internal evidence run the evident risk of circularity. Regarding external evidence, as mentioned above (1.1.2.5), the choice of *kanji* to act as phonograms was not exclusively based on phonological considerations; furthermore, the actual relation between EMC and OJ is anything but direct, as Chinese came to Japan

by way of the Korean peninsula and was very likely based on a different stage, if not variety, of Chinese; and finally, EMC is itself a reconstruct, i.e. hypothetical. For this type of reconstruction, Miyake (2003a) is an extremely useful and well-documented study with an impressive command of all the relevant materials.

2.1.3 Sound values

Miyake's broad reconstruction posits the basic sound values shown in (2). Some of these sound values are more similar to what we reconstruct for pJ (2.7.2) than for OJ, but overall they are compatible with the phonemic interpretation adopted below.

(2)

i_1	[i]
i_2	[ɨ]
e_1	[e]
e_2	[əy]
a	[a]
o_1	[o]
o_2	[ə]
u	[u]

More narrowly, the syllable pairs seem to have differed phonetically as follows:

(3)

			Examples	
Ci_1	more palatal	[C$_{\underline{\text{a}}}$i]	pi_1	[p$_{\underline{\text{a}}}$i] 'sun'
Ci_2	more labial; a falling diphthong	[C$_{\underline{\text{u}}}$i, Cu$_{\underline{\text{a}}}$]	pi_2	[p$_{\underline{\text{u}}}$i, pu$_{\underline{\text{a}}}$] 'fire'
Ce_2	a falling diphthong	[Ce$_{\underline{\text{a}}}$, Ca$_{\underline{\text{a}}}$]	me_2	[me$_{\underline{\text{a}}}$, ma$_{\underline{\text{a}}}$] 'eye'
Ce_1	more palatal; a rising diphthong	[C$_{\underline{\text{a}}}$e]	me_1	[m$_{\underline{\text{a}}}$e] 'woman'
Co_2	a monophthong	[Co]	ko_2	[ko] 'this'
Co_1	more labial; a rising diphthong	[C$_{\underline{\text{u}}}$o]	ko_1	[k$_{\underline{\text{u}}}$o] 'child'

2.1.4 Phonemic interpretation

The syllabic writing gives no clues regarding the organization into segments of distinctive phonic qualities and that makes the phonemic interpretation difficult. Even so, the difference between the *kō* and *otsu* syllables is, as

Table 2.2 *Transcription systems for OJ*

Syllable type	Index notation	Ohno	Modified Mathias–Miller	Yale	Frellesvig & Whitman
Kō-rui	i_1	i	î	yi	i
Otsu-rui	i_2	ï	ï	iy	wi
neutral	i	i	i	i	i
Kō-rui	e_1	e	ê	ye	ye
Otsu-rui	e_2	ë	ë	ey	e
neutral	e	e	e	e	e
Kō-rui	o_1	o	ô	wo	wo
Otsu-rui	o_2	ö	ö	o̲	o
neutral	o	o	o	o	o
	u	u	u	u	u
	a	a	a	a	a

mentioned above, generally agreed to pertain to the post-consonantal part of each syllable. Traditionally, the difference was thought to be a distinction in vowel quality, leading to the hypothesis that OJ had *eight* distinct vowel phonemes, with many different proposals concerning the structural organization of these eight vowels. However, since Lange (1973), it is common to interpret the difference in terms of sequential diphthongs, i.e. as being due to the presence of a palatal or labial glide, /y/ or /w/, in one or the other member of the syllable pairs. The phonemic interpretation and transcription which will be adopted here is shown in Table 2.2, based on Frellesvig and Whitman (2008a). For reference, the Yale system of transcription (in Martin 1987), Ohno's system (1990), and modified Mathias–Miller are also included. Table 2.3 gives some examples.

On this interpretation, Old Japanese had the five vowels also found in Middle and Modern Japanese: /i, e, a, o, u/. In addition to identifying the properties distinguishing *kō* and *otsu* syllables, the assignment of neutral syllables, i.e. syllables with undistinguished *Ci*, *Ce*, or *Co*, is an issue. As shown, they are here phonemically identified with Ci_1, Ce_2, and Co_2, respectively. Table 2.4 gives the inventory of distinct syllables in this phonemic notation.

All diphthongal interpretations suffer from some awkwardness from a structural point of view in the distribution of glides, mainly because they usually incorporate diachronic considerations. The one adopted here has the advantage of positing only sequential diphthongs that are also found as free syllables. For some of the syllables involving the labial glide, in particular /mwo, nwo, ywo, rwo; pwi, bwi, mwi/, it seems reasonable to assume a pronunciation like [u̯o̯] and [u̯i̯], that is, with a shift in phonetic sonority peak from the vowel to the glide.

Table 2.3 *Examples of transcribed OJ forms*

Gloss	cNJ	Frellesvig & Whitman	Index notation	Yale	Modified Mathias–Miller	Ohno
'sun'	hi	pi	pi_1	pyi	pî	pi
'fire'	hi	pwi	pi_2	piy	pï	pï
'blood'	chi	ti	ti	ti	ti	ti
'woman'	me	mye	me_1	mye	mê	me
'eye'	me	me	me_2	mey	më	më
'hand'	te	te	te	te	te	te
'child'	ko	kwo	ko_1	kwo	kô	ko
'this'	ko	ko	ko_2	ko	kö	kö
'ear (of rice)'	ho	po	po	po	po	po

Table 2.4 *Phonemic syllables in OJ*

.a	ka	ga	sa	za	ta	da	na	pa	ba	ma	ya	ra	wa
.i	ki	gi	si	zi	ti	di	ni	pi	bi	mi		ri	wi
	kwi	gwi						pwi	bwi	mwi			
.u	ku	gu	su	zu	tu	du	nu	pu	bu	mu	yu	ru	
.e	ke	ge	se	ze	te	de	ne	pe	be	me	ye	re	we
	kye	gye						pye	bye	mye			
.o	ko	go	so	zo	to	do	no	po	bo	mo	yo	ro	wo
	kwo	gwo	swo	zwo	two	dwo	nwo			mwo	ywo	rwo	

2.1.5 Neutralization

It is notable that *-i* and *-wi*, and *-ye* and *-e* only were kept distinct after /p, b, m, k, g/; after glides and alveo-dental consonants the distinction was neutralized, cf. the imperative and exclamatory forms of consonant base verbs in (4) and the gerund forms of quadrigrade (5a) and upper bigrade (5b) verbs:

(4)	Base	Imperative	Exclamatory
	kak- 'write'	/kakye/	/kake/
	mat- 'wait'	//matye// => /mate/	/mate/

(5)		Base	Gerund
	a.	*ok-* 'put'	/okite/
		or- 'weave'	/orite/
	b.	*ori-* 'descend'	/orite/
		okwi- 'arise'	/okwite/

It is likely that there earlier was a distinction /i/ ≠ /wi/ and /ye/ ≠ /e/ after all true consonants and that the state exhibited by OJ reflects a system in collapse. On the other hand, /p, b, m, k, g/ constitute the class of *grave* consonants and the environment of distinction is therefore general. It is thus also quite possible that the OJ state reflects a stable stage of some standing.

The distinction between /mwo/ and /mo/ is thought to have been found only in the *Kojiki*, although some scholars posit it for portions of other texts. In the later sources /Cwo/ and /Co/ were only kept distinct where C was not a labial consonant. This, however, does not reflect systematic neutralization, but simply the short-term course of the merger of the distinction. Thus we assume that there was distinction between /*pwo/ ≠ /*po/ and /*bwo/ ≠ /*bo/ in slightly earlier Japanese.

2.2 Consonants

OJ had the following inventory of consonant phonemes. The mediae (/b, d, g, z/) and the liquid (/r/) did not occur word initially, see below. There is nothing remarkable about the phonetics and phonology of the nasals, glides, and the liquid. The liquid was a flap, like its NJ reflex: [ɾ], in the following simply noted by [r]. Note that the palatal glide /y/ will be noted by IPA [j] in phonetic transcription.

(6)

	Labial	Alveo-dental	Palatal	Velar
Tenuis	**p**	**t, s**		**k**
Media	**b**	**d, z**		**g**
Nasal	**m**	**n**		
Liquid		**r**		
Glide	**w**		**y**	

2.2.1 Obstruents

The obstruents, on the other hand, exhibited several interesting features. The main phonetic variants may be thought to have been as in Table 2.5, noted in a broad transcription. Phonologically, the phonetic system manifested in Table 2.5 may be understood in terms of four overlapping sets of obstruents: tenuis versus media, and sibilant versus non-sibilant obstruent, which were distinguished by two distinctive feature categories, *tenseness* (+/–*tense*) and *stridency* (+/–*strident*), and which displayed phonetic variation with regard to *voicing*, *nasality*, and *continuousness*.

2.2.2 Tenues versus mediae; medial voicing and prenasalization

The traditional Japanese terms for tenues and mediae are *sei'on* (清音 'clear sounds') and *daku'on* (濁音 'muddy sounds'), respectively. The mediae

Table 2.5 *Phonetic realization of OJ obstruents*

Phoneme	Phonetic realization: Word initial	Phonetic realization: Word medial	Phonological classification	
/**p**/	[p ~ ɸ	b ~ β]		
/**t**/	[t	d]	non-sibilant	tenuis (*sei'on*)
/**k**/	[k ~ x	g ~ ɣ]		
/**s**/	[s ~ ˈs	z ~ ᵈz]	sibilant	
/**b**/		[ᵐb ~ ᵐβ]		
/**d**/		[ⁿd]	non-sibilant	media (*daku'on*)
/**g**/		[ᵑg ~ ᵑɣ]		
/**z**/		[ⁿz ~ ⁿᵈz]	sibilant	

(/b, d, g, z/) were phonetically *prenasalized*. That is to say, they had a nasal onset or onglide. This feature of pronunciation is generally thought not to have been lost until early in the NJ period. We will refer to this simply as *prenasalization*. Evidence comes both from EMC sound values for the *ongana* for syllables with initial media and from a number of sound changes in which nasals and mediae behave alike, e.g. the *onbin* sound changes which took place in the transition from OJ to EMJ (see 7.1.4), cf. OJ *yomite* 'read.GER' > EMJ *yoNde*, OJ *ywobite* 'call.GER' > EMJ *yoNde*. Note in this connection that to all appearances vowels before nasals and mediae were allophonically nasalized.

The tenues (/p, t, k, s/) were distinguished from the mediae in terms of the phonetic parameter of *tenseness*: the tenues being *tense* and the mediae *lax*. The tenues were allophonically *voiced* in word medial position. In a narrow phonetic transcription this might be noted [p̆, t̬, k̬, s̬], but for the sake of convenience [b, d, g, z] will be used here. *Voicing*, then, was subject to variation and was assigned by a redundancy rule; it was not a distinctive phonetic feature in OJ. This is the reason that we prefer to speak of /p, t, k, s/ as *tenues* and of /b, d, g, z/ as *mediae*, rather than as unvoiced and voiced obstruents. Voicing of tense obstruents in word medial position will be referred to in the following simply as *medial voicing*. The hypothesis of medial voicing is due to Wenck (1959). Evidence for it comes mainly from sound changes, first of all /-p-/ [b ~ β] merging with /-w-/ in EMJ (see 7.3.1.1), secondarily from the kind of syllable weakening involved in the *onbin* sound changes.

Thus, the distinction between tense and lax obstruents in word medial position was phonetically manifested primarily as one of prenasalization, cf. *pata* 'flag', *pada* 'skin', and *pana* 'flower', (7):

(7)

Tenuis	Media	Nasal
/pa**t**a	pa**d**a	pa**n**a/
[pada	pã^{n}da	pãna]

Similar phonetic realizations of tenuis and media are retained in some dialects (for example in northern Japan or in Shikoku), but mostly prenasalization and medial voicing were lost in the course of LMJ (see 11.1). That phonetic change led to major differences in sound texture between OJ and NJ. We illustrate these differences in (8), where we also assume that OJ word initial tenues were not aspirated, or at least not as aspirated as NJ tenues. The word /tanabata/ 'Vega, the Weaver' has not changed phonemically over more than a millenium between OJ and NJ, but its phonetics have changed considerably. Because of prenasalization and medial voicing, OJ sounded very different from NJ.

(8)

	/tanabata/
OJ	[tãnã^{m}bada]
NJ	[t^{h}anabat^{h}a]

2.2.3 Non-sibilant versus sibilant obstruents

The sibilants (/s, z/) were pronounced with a sibilant (high intensity) noise and were distinguished thereby from the non-sibilant obstruents (/p, t, k, b, d, g/) which were pronounced with no such noise. However, both sibilant and non-sibilant obstruents exhibited variation with regard to *continuousness*. That is to say, both continuant (fricative) and abrupt (occlusive and affricative) sound types are found among both classes of phonemes. This is the reason we prefer to speak of /s, z/ versus /p, t, k, b, d, g/ as *sibilant* versus *non-sibilant* obstruents, rather than as fricatives versus stops. The sibilants had both fricative, [s-, -z; -^{n}z], and affricated, [^{t}s-, -^{d}z; -ndz], variants. The distribution seems to have been conditioned by the following vowel (here following Kobayashi 1981 and extending her findings to /-s-, -z-/):

(9)

	Word initial	Word medial	
	/s-	-s-	-z-/
/-a/	[^{t}sa]	[^{d}za]	[ndza]
/-o/	[^{t}so]	[^{d}zo]	[ndzo]
/-u/	[su ~ ^{t}su]	[zu ~ ^{d}zu]	[^{n}zu ~ ndzu]
/-wo/	[swo]	[zwo]	[^{n}zwo]
/-i/	[ʃi]	[ʒi]	[^{n}ʒi]
/-e/	[ʃe]	[ʒe]	[^{n}ʒe]

The phonetic reconstruction of /s, z/ remains debated, with proposals ranging from fricatives only (e.g. Sandness 1986, Miyake 2003a) to affricates only (e.g. Ogura 1998). If any, the consensus view is probably the one adopted here with variation between [ᵗs] and [s]. The debate is to some extent obscured by a lack of clear distinction between phonetic and phonological reconstruction. It should be noted that common to all proposals, though not explicitly addressed, is that /s, z/ are manifested as sibilant sound types. The evidence concerning the reconstruction of /s, z/ comes from two main sources: (a) Reconstructed EMC sound values for the *ongana* for syllables with initial /s, z/. Miyake's material clearly shows that among the *ongana* for *sa*, *so*, and *su*, many, some, and a few, respectively, are reconstructed with initial fricative. For *zV*, many of the *ongana* are reconstructed with initial affricate, though not as neatly distributed as for *sV*. (b) The priest Ennin's *Zaitōki* (858), in which he described Sanskrit sounds by means of illustration with phonograms used for their Japanese or Chinese sound values. Ennin seems to equate the pronunciation of Sanskrit *ča* ([tʃa]) with Japanese *sa* and that is often taken as evidence for an affricative sound value for /s/, at least before /a/. Apart from the fact that *Zaitōki* primarily concerns EMJ and not OJ, the interpretation of the text is far from straightforward, as pointed out by e.g. Sandness (1986). Finally, Ainu *časi* [tʃasi] borrowed from OJ *sasi* 'castle' (in turn borrowed from a Korean language, reflected as MK *cas*) also points towards an affricative pronunciation of OJ /s/, at least before /a/.

/p/ had both occlusive (abrupt), [p-, -b], and fricative (continuant), [ɸ-, -β] variants. It is not possible to identify any conditioning phonological environment for this variation, so it is possible that [p-, -b] and [ɸ-, -β] may have been in free variation. Earlier it was thought that /p/ had already shifted completely to [ɸ] in OJ, but there is little to support that view. It seems that similar variation was displayed by other non-sibilant obstruents; in particular /k, b, g/ seem to have had both abrupt and continuant variants. Some of this variation was likely subject to stylistic variation, with abrupt variants being more characteristic of careful and continuant variants more characteristic of casual speech. Evidence for continuousness variation for the non-sibilant obstruents comes from sound changes, first of all /-p-/ [b ~ β] merging with /-w-/ in EMJ (7.3.1.1), secondarily from the kind of syllable weakening involved in the *onbin* sound changes (7.1.4).

In this way, *continuousness* was subject to variation and both of its values were represented in the realization of members of both sibilant and non-sibilant obstruents. This phonetic variation will be referred to as *continuousness variation*.

(10)

	Non-sibilant									Sibilant		
	/p-	t-	k-	-p-	-t-	-k-	-b-	-d-	-g-	s-	-s-	-z-/
Abrupt	[p	t	k	b	d	g	^{m}b	^{n}d	^{ŋ}g	^{t}s	^{d}z	ndz]
Continuant	[ɸ	(x)		β		(ɣ)	(^{m}β)		(^{ŋ}ɣ)	s	z	^{n}z]

2.3 Other allophonic variation

Two fairly low-level phonetic allophonic rules may be posited, (11) and (12), exemplified in (13). Nasalization probably also to some smaller extent applied to vowels *after* nasals.

(11) *Nasalization of vowels*: Vowels *before* nasals or mediae were phonetically nasalized (reflected for example in a variant spelling of *uma* 'horse' as 牟麻 'mu.ma').

(12) *Tonality adjustment*: Consonants before /i, y/ were phonetically palatalized (sharpened), and consonants before /u, w/ were labialized (flattened).

(13)

uma	*kimi*	*ywobi*	*kubi*	*pikye*
'horse'	'lord'	'calling'	'head'	'pull!'
[ũma	k$_{j}$ĩm$_{j}$ĩ	ju̯õ^{m}b$_{j}$i	k^{w}ũ^{m}b$_{j}$i	p$_{j}$igje]

Palatalization probably also to some extent applied to consonants before /e/, see (14). That is to say, 'neutral /Ce/' was phonemically like *Ce*$_2$, but phonetically more like *Ce*$_1$. Presumably, the gradual emergence of uniform phonetic palatalization before /e/ is what led to the disappearance of a distinctive phonemic glide in *Ce*$_1$ (/Cye/) syllables.

(14)

Ce$_1$	*me*$_1$ 'woman'	/Cye/	/mye/	[Cje]	[mje]
Ce	*te* 'hand'	/Ce/	/te/	[C$_{j}$e]	[t$_{j}$e]
Ce$_2$	*me*$_2$ 'eye'	/Ce/	/me/	[Ce]	[me]

2.4 Phonetic transcription of a text

On this background it is now possible to illustrate what OJ probably sounded like, giving a broad, but still reasonably detailed, phonetic transcription of a short text (the first poem in the *Kojiki*, cf. 1.1.6), with a. illustrating more careful and b. more casual diction.

(15)

	夜久毛 多都	伊豆毛 夜幣賀岐	都麻碁微 爾
	/yakumwo tatu	**idumwo yapyegaki**	**tumagomwi ni**
a.	[jag^{w}ũmuo̯ tad^{w}u	ĩ^{n}d^{w}ũmuo̯ jabjẽ^{ŋ}gag$_{j}$i	t^{w}ũmã^{ŋ}gõmui̯ n$_{j}$i
b.	[jaɣ^{w}ũmuo̯ tad^{w}u	ĩ^{n}d^{w}ũmuo̯ jaβjẽ^{ŋ}ɣaɣ$_{j}$i	t^{w}ũmã^{ŋ}ɣõmui̯ n$_{j}$i

	夜幣賀岐	都久流	曾 能	夜幣賀岐 袁
	yapyegaki	**tukuru**	**so no**	**yapyegaki wo/**
a.	jabjẽ^{ŋ}gag$_{j}$i	t^{w}ug^{w}uru	^{t}sõ no	jabjẽ^{ŋ}gag$_{j}$i wo]
b.	jaβjẽ^{ŋ}ɣaɣ$_{j}$i	t^{w}uɣ^{w}uru	sõ no	jaβjẽ^{ŋ}ɣaɣ$_{j}$i wo]

'The many-fenced palace of Idumo Of the many clouds rising – To dwell there with my spouse Do I build a many-fenced palace: Ah, that many-fenced palace!'
(Philippi 1968: 91)

2.5 Syllable and word structure

As seen in Table 2.4 above, OJ syllables have the following structure: **(C)(G)V**. Syllables consisting of a single vowel are generally restricted to word initial position, except for a few exceptions, including a single noun, *ka.i*, and inflected forms in the bigrade conjugations (cf. 3.4.1.2), such as *ku.i* 'regret. INF'(<= //kuyi//, cf. conclusive *kuyu*) and *ku.u* (<= //kuwu//, cf. infinitive *kuwe*-), i.e. verbs whose bases end in a syllable with initial glide. Thus all syllables are open and short, surface forms having to conform to repetition of the simple (C)(G)V pattern.

A number of distributional facts about OJ sounds are best understood in diachronic terms and will be explained in the section on proto-Japanese: (a) absence of word initial mediae or liquid (2.7.1.2); (b) Arisaka's Law (2.7.2.1); (c) limited distribution of *Cwo*, *Cye*, *Ce*, and *Cwi* (2.7.2.4); (d) noun apophony (2.7.2.2).

2.6 Morphophonemics

2.6.1 Vowel deletion

When in morphological derivation or word formation two vowels come together without an intervening consonant, one is elided. Depending on the number of syllables (one or more) in the two morphemes, the following rules apply. The second vowel is only elided when a monosyllabic morpheme is followed by a vowel initial polysyllabic morpheme, i.e. (16a); elsewhere the first vowel is elided (16b):

(16) Synchronic rules of vowel deletion

a. **$V_1 + V_2 \Rightarrow V_1$**

/-CV_1 + V_2CV/	=>	/-CV_1CV/
wa + ***ga*** + ***ipye*** 'I + GEN + house'	=>	***wagapye*** 'my house'

b. **$V_1 + V_2 \Rightarrow V_2$** elsewhere

/CVCV_1 + V_2/	=>	/CVCV_2/
ake + ***u*** 'redden + CONCL'	=>	***aku*** 'redden.CONCL'
/CVCV_1 + V_2CV/	=>	/CVCV_2CV/
myesi + ***age*** 'see.RESP + raise'	=>	***myesage***- 'summon'
waga + ***ipye*** 'my + house'	=>	***wagipye*** 'my house'
toko + ***ipa*** 'eternal + rock'	=>	***tokipa*** 'eternal rock; everlasting'
/CV_1 + V_2/	=>	/CV_2/
ko + ***i*** 'come + INF'	=>	***ki*** 'come.INF'

It should be noted that there are apparent counterexamples to (16a), such as the competing form *wagipye* in (b). However, in such cases a different constituent structure may be assumed for the underlying form, here *waga-ipye* with univerbation between pronoun and genitive particle. Another similar set of forms are those where the existential verb *ar-* seemingly fuses with a preceding grammatical monosyllabic morpheme: the periphrastic stative *-te-ar-* giving *-tar-*, the extended negative *-(a)zu-ar-* giving *-(a)zar-*, and the extended adjectival copula *-ku-ar-* giving *-kar-*. In these cases, however, *ar-* must be thought to fuse with the full inflected form, not just the suffix: *kakite-ar-* => *kakitar-*, *kakazu-ar-* => *kakazar-*, *akaku-ar-* => *akakar-*. In OJ all such forms were simple phonological fusions (3.4.2.1.2), but in EMJ they were reanalysed to give rise to the morphemes *-tar-*, *-zar-*, and *-kar-* (8.2.1). Similarly, the EMJ inflected copula *nar-* seems to be from *ni-ar-*, but also here an intermediary derivational step must be posited: *ikusa-ni ar-* => *ikusani-ar-* => *ikusanar-* (*ikusa* 'army'). The same holds for the copula *tar-* which is found from EMJ and which is formed from the copula infinitive *to* and *ar-*.

2.6.2 Rendaku

In the morphophonemic process known as '*rendaku*', an initial tenuis in the second component of a compound changed to a media to express close compounding. '*Rendaku*' is usually rendered 'sequential voicing' in English, but as voicing was not the phonetic feature which distinguished tenues from mediae, that is an inappropriate translation, so here '*rendaku*' will simply be

retained. It is thought that this process originated in the reduction of a particle with initial nasal, usually identified with OJ genitive *no* or dative *ni*. Thus *sakurabana* is thought etymologically to derive from *sakura-no-pana*, or *yamadori* from *yama-no-tori*. However, examples such as (18) show that *rendaku* already in OJ was established as a morphophonemic process, for they cannot be etymologized with *no* or *ni*: *na* 'you' invariably takes *ga* as genitive particle, never *no*, and *yomwi ni kapyer-* would mean 'return *to* Hades', not *from*. Examples such as *tuma-gomwi* 'wife-enclosing' and *tuma-gwopwi* 'wife-loving' incorporate a direct object, again not directly derivable from *no* or *ni* or any other *n*-initial particle.

(17) *sakura* 'cherry' + *pana* 'flower' => *sakurabana* 'cherry-flower'
yama 'mountain' + *tori* 'bird' => *yamadori* 'copper-pheasant'

(18) *na* 'you' + *tori* 'bird' => *nadori* 'your bird'
yomwi 'Hades' + *kapyer-* 'return' => *yomwigapyer-* 'revive (intr.)'

As forms resulting from *rendaku* were lexicalized, a number of alternations resulted, shown in (19), some of which are not phonologically automatic because of later sound changes (/p/ > /f/ > /h/ (see 11.3, 14.3) and /d/ > /z/ before /i, u/ (see 14.1).

(19)

p- > LMJ f > NJ h-	~	-b-
t-	~	-d-
t-	~	-d- > NJ -z- (/ __ {i, u})
k-	~	-g-
s-	~	-z-

Rendaku is not phonologically predictable. It is, however, blocked when the second component contains an internal media, e.g. *kamu-* 'spirit' + *kaze* 'wind' => *kamukaze* 'divine wind', not *kamugaze. This regularity is known as *Lyman's Law*, named after B. S. Lyman (1894, see Vance 2007). It continues to hold in NJ where *rendaku* is still very active.

2.7 Proto-Japanese

It is possible to reconstruct a simple pJ phoneme inventory and some of the sound changes that took place between pJ and OJ. The main point of interest is that some phonemes and diphthongs of OJ arose through contraction of sequences of segments in pre-OJ. Table 2.6 gives an overview.

Table 2.6 *Phonemic Correspondences between OJ and pJ*

Primary OJ consonants		pJ
p		*p
t		*t
k		*k
s		*s
m		*m
n		*n
r		*r
w		*w
y		*y
Secondary OJ consonants	**Pre-OJ sequence**	
b	*mVp, *nVp; *np	
d	*mVt, *nVt; *nt	
g	*mVk, *nVk; *nk	
z	*mVs, *nVs; *ns	
Primary OJ vowels and diphthongs		**pJ**
i		*i
i, ye		*e
a		*a
u, wo		*o
u		*u
o		*ɨ
o		*ə
Secondary OJ vowels and diphthongs	**Pre-OJ sequence**	
wi	*uy, *ɨy	
e	*ay, *əy	
ye	*ia (*iə)	
(*wo*	*uo, *ua)	

2.7.1 *Consonants*

2.7.1.1 *Secondary origin of OJ mediae (/b, d, g, z/)*

The OJ mediae reflect contractions of a nasal with a following obstruent. Where recoverable, such sequences arose through weakening and loss of a syllable, often where *univerbation* of a complex form took place; this was also the origin of *rendaku* (2.6.2). It should be noted that the syllable loss involved is sporadic, that is, not part of regular sound changes.

(20) *a**mi**-**pi**ki 'net-pull' > *abiki* [ã^{m}b$_{j}$ig$_{j}$i] 'trawling'
*yama-**mit**i 'mountain-path' > *yamadi* [jãmã^{n}d$_{j}$i] 'mountain-path'
*mura-**nus**i 'village-master' > *murazi* [murã^{n}ʒi] proper name
*yama-**ni**-**tə**ri 'mountain-GEN-bird' > *yamadori* [jãmã^{n}dori]

2.7.1.2 Distribution of OJ mediae and liquid

The mediae and the liquid (/r/) did not occur word initially. There were, however, suffixes with initial media or liquid, for example, *be-* (verb extension; necessitive), *dani* (restrictive particle; 'at least; even'), *ga* (case particle; genitive), *ra* (noun suffix; plural). For the mediae, this does not seem to reflect an original phonotactic restriction, but rather the secondary origin of the mediae. It is on the other hand a feature common with the so-called Altaic languages not to allow word initial liquids. SJ loanwords were readily taken into the language with initial media or liquid (7.2), and there were a few examples of them already in OJ (4.2.2); there is also a single example of a mimetic with initial /b/ in OJ, *bisibisi* 'sniffling'.

2.7.1.3 PJ syllable final nasals

For many OJ forms with a media it is not possible to recover a source with a vowel between nasal and tenuis. Some such forms may reflect contraction of a tenuis with a directly preceding syllable final nasal:

(21) *tu**np**u 'grain' > *tubu*
*ka**nt**i 'rudder' > *kadi*
*pi**ns**a 'knee' > *piza*

It is not possible to recover internally the nature of these nasals. For the forms above, /*-n/ is posited, but there is no reason to suppose that pre-OJ did not have /*-m/ or /*-ŋ/ in addition to /*-n/. There is nothing to say that *piza* is not from *pimsa or *piŋsa.

Recently it has been proposed that pJ in addition to /m, n/ had syllable *initial* *ŋ, reflected as OJ /Ø/ or in some cases /g/.

2.7.1.4 PJ glides

Table 2.6 shows one of two main views on the pre-history of the OJ syllable initial glides, /w, y/, viz. that they go back to pJ /*w, *y/. On this view pJ did not have voiced stops, but a competing hypothesis reconstructs pJ /*b, *d, *g, *z / which subsequently weakened and are reflected as OJ glides and zero: /w, y, Ø, Ø/. Positive evidence for this competing view is scarce.

2.7.2 Vowels and diphthongs

Until recently the main view held that there were only four primary pJ vowels: /*i, *a, *u, *ə/ reflected as OJ /i, a, u, o/. This view was adopted, for example, in Martin (1987), but recently more vowels are reconstructed. We here show the seven primary vowels reconstructed by Frellesvig and Whitman (2008b), adapting the account there of developments between pJ and OJ.

2.7.2.1 Arisaka's Law; distribution of primary vowels

Arisaka's Law (named after Arisaka Hideyo) is a restriction on the shape of OJ root morphemes: /Co/ (= Co_2) did not occur in a root morpheme with /Ca, Cwo, Cu/. That is to say, there were (almost) no simple words of the structure /CoCa, CoCwo, CoCu, CaCo, CwoCo, CuCo/. Arisaka's Law is summarized in (22) for the direct reflexes of the pJ vowels:

(22)	Neutral	A	B
	-i (<*i, *e)		-u (< *u, *o-)
	-ye (< *e)	-o (<*ɨ, *ə)	-wo (< *-o)
			-a (< *a)

This has been taken to mean that Japanese earlier had some form of 'vowel harmony', but such a restriction on the shape of root morphemes is different from the kind of vowel harmony which applies to the concatenation of stem and affix in 'vowel harmony languages' such as for example Turkish. Structurally the primary pJ vowels were organized as in (23), with restrictions holding on the co-occurrence of central and back vowels within a root morpheme:

(23)	Neutral	Central	Back
	*i	*ɨ	*u
	*e	*ə	*o
			*a

2.7.2.2 Secondary vowels and diphthongs: OJ /-wi, -e/

OJ /-wi, -e/ are thought to be secondary and to reflect contraction of falling diphthongs /*Vy/ or /*Vi/:

(24)	Pre-OJ		OJ
	*uy	>	wi
	*ɨy	>	wi
	*əy	>	e
	*ay	>	e

The most instances of OJ /e, wi/ are found in alternations which have come to be known as 'apophonic'. For example, a number of nouns have alternating shapes with variation in the final syllable: One variant, the free form (known as *roshutsukei* (露出形) 'exposed form'), occurs in word final position, while the other, the compound form (*hifukukei* (被覆形) 'covert form'), usually occurs in compounds or derived forms. Traditionally the covert form is thought to represent an older or more original shape of the word. Similar alternations are found among verbs, reflecting similar contractions, see (3.5.1).

(25)		Exposed form		Covert form
	wi ~ u	*mwi* 'body'	~	*mu-kapari* ('substitute') 'hostage'
		kamwi 'spirit'	~	*kamu-kaze* ('wind') 'divine wind'
	wi ~ o	*kwi* 'tree'	~	*ko-dati* ('tree-stand') 'grove'
		yomwi 'Hades'	~	*yomo-tu-kuni* ('Hades-GEN-land') 'land of Hades'
	e ~ o	*se* 'back'	~	*so-muku* 'turn'
		me 'bud'	~	*moyasi* 'bud, sprouting forth'
	e ~ a	*me* 'eye'	~	*ma-pye* ('side, direction') 'front'
		sake 'saké'	~	*saka-duki* ('cup') 'saké cup'

In such cases the exposed form ends in /-e, -wi/, which are thought to reflect contraction of /*Vy/ or /*Vi/ as shown above in (24). The source of the /*-y/ in the diphthongs that were contracted has been a matter of debate, with proposals including the ancestor of the particle *i* (see 3.7.1.3), but the apophonic nouns are now generally thought to go back to consonant final shapes, with the final consonant being weakened and reduced to *y. In some cases it is not possible to identify the consonant, hence *səC, *maC.

(26)	'body'	*mwi*	< *muy	< *mum (cf. MK *mom*)
	'tree'	*kwi*	< *kɨy	< *kɨr (cf. MK *kɨrɨh* 'stump'; counter for trees)
	'back'	*se*	< *səy	< *səC
	'eye'	*me*	< *may	< *maC

The final consonant of such forms is thought to have been deleted in composition before another consonant, e.g. *mum + kapari => *mukapari*. That is the origin of the covert form which more directly reflects the original root vowel:

(27)

	'body'	*mu-*	< *mu-C...	< *mum-C...
	'eye'	*ma-*	< *ma-C...	< *mar-C...

As shown in (25), OJ /o/ takes part in two apophonic alternations, /o ~ -wi/ and /o ~ e/. This is thought to reflect contraction of two different vowels with /*-y/ (or /*-i/) and that is the reason for reconstructing two central vowels, pJ /*ɨ, *ə/, which later, after the lexicalization of the contracted forms, merged, most likely as /*ə/, which later backed to give OJ /o/. Thus, OJ /o/ has two different pJ sources, /*ɨ/ and /*ə/, and the only reason we recognize this is that the two sources gave different outcomes when contracted with /*y/, with the result that OJ /o/ takes part in two alternations:

(28) pJ *ə, *ɨ > pre-OJ *ə > OJ o

The two types of covert forms in OJ /o/ are derived as in (29):

(29)

	'tree'	*ko-*	< *kɨ-C...	< *kɨr-C...
	'back'	*so-*	< *sə-C...	< *səC-C...

More generally, the source of the apophonic alternations may be set out as in (30):

(30)

	u ~ wi	*u > *u*	~	*wi* < *uy
	o ~ wi	*ɨ > *ə > *o*	~	*wi* < *ɨy
	o ~ e	*ə > *o*	~	*e* < *əy
	a ~ e	*a > *a*	~	*e* < *ay

In addition to apophonic alternations there are a few cases of fairly transparent lexical contractions which involve /*Vi/ sequences, shown in (31). Such cases were important in formulating the hypothesis that OJ /-e, -wi/ generally reflect contractions of diphthongs, but in fact there are very few solid examples of such lexical contractions.

(31)

	*ui > wi	*waku-iratukwo 'young-honourable.male' > *wakwiratukwo*
	*ɨi > wi	*ɨpɨ-isi 'big-stone' (cf. OJ *opo-* 'big') > *opwisi* proper name
	*əi > e	*tənə-iri 'palace-enter' (cf. OJ *tono* 'palace') > *toneri* 'palace servant'
	*ai > e	*taka-iti 'high-market' > *taketi* proper name

Also here OJ /o/ is involved in two outcomes of contraction with /*i/:

(32)	'big'	*ɨpɨ > *opo-*	~	*opwisi* < *ɨpɨ-isi*
	'palace'	*tənə > *tono*	~	*toneri* < *tənə-iri*

2.7.2.3 Mid vowel raising

The pJ mid vowels *e and *o raised as in (33):

(33) *e > *-ye* in final position (partial MVR)
i elsewhere (full MVR)
*o > *-wo* in final position (partial MVR)
u elsewhere (full MVR)

That is to say, in nonfinal position pJ *e, *o merged with pJ *i, *u as OJ /i/ and /e/, respectively, but were preserved as distinct only in final position. It appears that there were dialectal differences in what constituted 'final position' for these sound changes, ranging between root final, morpheme final, and word final. Mid vowel raising seems to have followed the course of a gradual phonetic diphthongization, eventually giving a high vowel, except in final position where the diphthongal realization was phonemicized as a diphthong, see (34):

(34)	*o >	[^{w}o] >	/wo/ (partial MVR)
	*o >	[^{w}o > u] >	/u/ (full MVR)
	*e >	[^{j}e] >	/ye/ (partial MVR)
	*e >	[^{j}e > i] >	/i/ (full MVR)

A few examples which show both outcomes are given in (35). The different developments of *yo(-)ri reflect different interpretations of an internal boundary.

(35)	*me 'woman' >	OJ *mye* ~ *womina* (< *wo-me-na 'small-female-person')
	*e 'placenta' >	OJ *ye* ~ *iro-* 'of same mother' (< 'placenta-GEN')
	*yo(-)ri 'behind' >	OJ *ywo(-)ri* ~ *yuri* 'from; behind'
	*moko 'bridegroom' >	OJ *mwokwo* ~ *mukwo* > EMJ *muko*

Note that the effect of *partial MVR* was neutralized in OJ for *e after coronals and glides, for *o after labials, and eventually in EMJ after all consonants, because of the merger of /Ce/ and /Cye/, and of /Co/ and /Cwo/ (see 7.3.2.1).

(36)	pJ	pre-OJ	OJ	EMJ
a.			Cye > / C = {p, b, m, k, g}	Ce
	*Ce >	*Cye >		
			Ce > / C = {t, d, s, z, n, r, w (y)}	Ce
b.			Cwo > / C = {t, d, k, g, s, z, n, r, y}	Co
	*Co >	*Cwo >		
			Co > / C = {p, b, (m) w}	Co

In addition to reflecting directly pJ /*o/ and /*e/, some instances of OJ /-wo/ and /-ye/ are secondary in the sense that they derive from contraction of sequences of vowels, e.g. (37):

(37)		
	*ia > ye	*saki-ari 'bloom-is' > *sakyeri* 'is blooming'
	*iə > ye	*pi-əki 'sun-put' > *pyeki* proper name
	*uə > wo	*situ-ori* 'native.weaving-weave' ~ *sitwori* 'type of cloth'
	*ua > wo	*kazu-ap- 'number-join' > *kazwope-* 'count'

However, there are very few good examples of this, apart from a small number of fairly transparent grammatical morphemes involving contraction of /*-i-a-/:

(38) *-yer-* stative < *-i-ar- 'INF-exist'
-kyeku simple past nominal < *-ki-aku 'SPST-nominalizer'
-kyeku adjectival copula nominal < *-ki-aku 'ACOP.ADN-nominalizer'
-kyem- simple past conjectural < *-ki-am- 'SPST-CONJ'
-kyer- modal past < *-ki-ar- 'SPST-be' (or 'come'-exist)
kyer- 'be wearing' (lexicalized stative of *ki-* 'wear') < *ki-ar- 'wear-exist'
kyes- wear, respectful < *ki-as- 'wear-RESP'
myes- see, respectful < *mi-as- 'see-RESP'
kyesa this morning < *ki-asa 'this-morning'

2.7.2.4 Lexical distribution of /Cwo, Cye, Ce, and Cwi/

Reflecting their pre-history, the syllables /Cwo, Cye, Ce, Cwi/ are restricted in the OJ lexicon. This is shown by the following result of a count of the text occurrence of different syllable types in the *Man'yōshū* (Ohno 1980: 151ff.), here maintaining the orthographic three-way distinction between *kō*, *otsu*, and neutral syllables.

(39)

Ci$_1$	*Ci*	*Ci*$_2$
3,160	6,103	370
Ce$_1$	*Ce*	*Ce*$_2$
686	2,299	853
Co$_1$	*Co*	*Co*$_2$
1,030	3,631	5,280
Ca	*Cu*	
12,120	6,415	

A frequency count in running text gives no real picture of lexical distribution, but these figures do indicate that /Cwi/ (= *Ci*$_2$), /Cye/ (= *Ce*$_1$), /Ce/, and /Cwo/ (= *Co*$_1$) were relatively infrequent compared to /Ci/ (= *Ci*$_1$), /Ca/, /Co/ (= *Co*$_2$), and /Cu/. In terms of distribution in the lexicon (as opposed to running text), /Cwi/ is infrequent and is almost exclusively found in morpheme final position. /Cye/ and /Ce/ are also rather infrequent. By far the most occurrences are in morpheme final position, but while simple words with nonfinal /Cye/ and /Ce/ are rare, they are not exceptional. /Cwo/ is not lexically infrequent, making up approximately 25 per cent of occurrences of *Co*$_1$ and *Co*$_2$ in the words listed in Omodaka *et al.* 1967, but most occurrences of /Cwo/ are in morpheme final position (see Martin 1987: 60–2 for a list of words with /Cwo/).

These distributional facts reflect (a) that all OJ /Cwi, Ce/ are etymologically secondary, and (b) that /*o/ and /*e/ merged with /u/ and /i/ in nonfinal position and only gave /wo/ and /ye/ in root, morpheme, or word final position.

2.7.2.5 Changes between proto-Japanese and Old Japanese

An approximate relative chronology of the changes which took place between pJ and OJ may be outlined as in (40) overleaf:

(40) a. Contraction of falling diphthongs

Pre-OJ	>	OJ
*ui		wi
*ɨi		wi
*əi		e
*ai		e

b. Merger of central vowels
pJ {*ɨ, *ə} > *ə

c. Mid vowel raising
pJ *e > OJ i, -ye

d. Contraction of rising diphthongs
pre-OJ *ia > *ya > OJ -ye
pre-OJ *iə > OJ -ye

e. Mid vowel raising
pJ *o > OJ u, -wo

f. Backing of central vowel
pre-OJ *ə > OJ o

g. Contraction of rising diphthongs
pre-OJ *ua > *wa > OJ -wo
pre-OJ *uo > OJ -wo

REFERENCES

General: Mabuchi 1971, Martin 1987, Wenck 1959. Greek vowels: Horrocks 1997: 102ff. Phonetic reconstruction based on Chinese sound values for *ongana*: Lange 1973, Miyake 2003a. *Kō-otsu* distinctions: Lange 1973, Miyake 2003a, Unger 2007; Co_1 vs Co_2: Hayata 1998, Matsumoto 1984. Consonants: Frellesvig 1995: 58–86, Martin 1987: 1–36. Prenasalization and medial voicing: Wenck 1959. /s, z/: Kobayashi 1981, Ogura 1998, Sandness 1986. Proto-Japanese: Frellesvig and Whitman 2008a, Whitman 1985.

3.1 Verbs

In a segmentational model, Japanese verb forms may be described in terms of the following structure, with five main morphemic layers:

(1) 1 2 3 4 5
root – derivative – auxiliary verb – auxiliary – flective

(2)–(4) are examples of OJ verb forms:

(2) a. *tuku* '(it) attaches'

b.

1	5
tuk-	*u*
attach-	CONCL

(3) a. *tukeme* 'can/will attach it!'

b.

1	2	4	5
tuk-	*e-*	*m-*	*e*
attach	TRANSITIVIZER	CONJ	EXCL

(4) a. *maywopikinikyeri* 'had become frayed!'

b.

1	3	4	4	5
maywopi-	*ki-*	*ni-*	*kyer-*	*i*
fray	come	PERF	MPST	CONCL

The central inflectional morphology is realized by *flectives*, expressing obligatory inflectional categories for which any verb inflects (3.1.3), and by *auxiliaries* expressing optional categories for which verbs can inflect (3.1.4). As seen in (4), more than one auxiliary can be attached to a verb, in which case the auxiliaries appear in the order given in Table 3.2 and (21) below (3.1.4.1 and 3.1.4.5). Of the four layers, or positions, in (1) after the root, only flective has obligatory expression, reflecting that any instantiation of a verb appears in an inflected form.

3.1.1 Derivatives

A number of pre-OJ verbal derivatives may be reconstructed, most of which were, however, fully lexicalized and unproductive by the time of OJ. The verbal derivatives shown in (5) may be assumed to have been fairly productive in OJ; they mostly specify or switch transitivity, but also include an iterative/continuative formant. In addition there are a number of derivatives which form verbs from other parts of speech, e.g. *-bwi-* and *-sabwi-* mentioned in 3.4.1.2 below.

(5) *-(a)s-* transitive; *-(a)r-* intransitive; *-e-* opposite transitivity (cf. 3.5.1); *-(a)p-* activity verb.

3.1.2 Auxiliary verbs

As mentioned, root + derivative constitute the lexical base. More than one lexical base can combine to form a compound. Lexical composition must, however, be distinguished from affixation of auxiliary verbs, which to some extent correspond to, or at least overlap with, what have been termed 'syntactic compounds' (e.g. Kageyama 1999). Grammaticalized auxiliary verbs attach to the infinitive of a main verb, specifying amongst other things various motional, directional and aspectual meanings, but including also expression of social deixis. They are set off from auxiliaries (3.1.4) by generally being used as free lexical verbs in addition to their grammaticalized use as auxiliary verbs. The degree of grammaticalization varies and it is therefore difficult in some cases to distinguish between verbs commonly used in lexical compounds and fully grammaticalized auxiliary verbs. Among the fully grammaticalized auxiliary verbs are those in (6):

(6) *-ko-* [< 'come'] 'come to . . .', *-yuk-* [< 'go'] 'continuative, go on . . .ing; gradually become . . . more and more', *-ide-* [< 'emerge'] 'VERB out'; *-ap-* [< 'meet'] 'reciprocal'; *-e-* [< 'get'] 'be able to'

-tamap- [< 'receive'] 'respect',[1] *-mas-* [< *imas-* 'exist, respect'] 'respect'; *-mawos-* [< 'say to/tell the emperor/a superior; ask for permission (of a superior)'] 'humble', *-matur-* [< 'offer to spirit or emperor'] 'humble'

[1] The auxiliary verb *-tamap-* follows rather than precedes the respect auxiliary -(a)*s-* (cf. 3.1.4.3), e.g. *oka-si-tamapite* (*MYS* 5.813) 'put-HON-HON'. This is an exception to the sequence set out in (1) above. The number of attested examples is very small, but this indicates that *-tamap-* should perhaps be interpreted as an auxiliary rather than an auxiliary verb; also in EMJ, *-tamap-* follows supporting respect auxiliaries.

3.1.3 Inflected verb forms: obligatory categories

In the segmentational model given above in (1), inflectional categories are represented by 5, flective. Table 3.1 lists the inflected forms for each of the eight OJ conjugational verb classes, noting the verb *base* (basic stem) at the top of each column. These are the paradigmatically opposed forms for which any lexical verb inflects (except that some verbs will not form an imperative). See 3.4 about the conjugation classes.

The first thing to note is that inflection in OJ is almost exclusively for *syntactic*, *modal* and *conjunctional* categories (as opposed to NJ which also inflects, for example, for tense and aspect). The primary distinction is between finite forms, which can conclude a main clause (3.1.3.1), and non-finite forms, which conclude nonfinal clauses (3.1.3.2); there is also a nominal form (3.1.3.3). The following summarizes main functions of each category and gives for reference the forms of the verb *sin-* 'dies'.

3.1.3.1 Finite verb forms

Conclusive (*sinu*) This form has two functions: (a) The conclusive is used for concluding declarative main clauses, see (7b). This is the function from which its name derives. (b) It is also used before most extensions (3.6), final particles (3.7.5), and with the concessive conjunctional particle *to(mo)* (3.7.4). It should be noted, however, that it is quite rare to find simple verbs (without auxiliaries attached) in the conclusive form concluding a main clause within the OJ corpus; most occurrences of simple conclusive verb forms are found with extensions or particles. The conclusive form is unspecified for – neutral with regard to – tense, aspect, or mood; thus all auxiliaries, including those expressing tense, aspect, and mood, have a conclusive form. Labels such as 'indicative' or 'non-past' which are sometimes used about this form in OJ (and EMJ and early LMJ) are therefore misleading.

Adnominal (*sinuru*) This form had several uses: First, the function after which the form is named is to modify a noun, that is to say, as the verb in an adnominal clause: 'who dies; dying'. In this function it contrasts with the conclusive form:

(7) a. *tama-no-ura ni asari* ***suru*** *tadu*
Tama-no-ura DAT forage.INF do.ADN crane
'A crane fishing in the Tama-no-ura' (*MYS* 15.3598)

b. nwosima *ga saki ni ipori* ***su,*** *ware pa*
Noshima GEN cape DAT hut do.CONCL, I TOP
'Me, I make a hut on the cape of Noshima' (*MYS* 15.3606)

Table 3.1 *Inflected forms for the eight OJ verb classes*

	QD	*r*-irr	*n*-irr
Base	***kak-*** 'paint, write'	***ar-*** 'exist, be'	***sin-*** 'die'
Finite			
Conclusive	*kaku*	*ari*	*sinu*
Adnominal	*kaku*	*aru*	*sinuru*
Exclamatory	*kake*	*are*	*sinure*
Imperative	*kakye*	*are*	*sine*
Neg. conject.	*kakazi*	*arazi*	*sinazi*
Optative	*kakana*	*arana*	*sinana*
Prohibitive	*na kaki so*	*na ari so*	*na sini so*
Non-finite			
Infinitive	*kaki*	*ari*	*sini*
Gerund	*kakite*	*arite*	*sinite*
Continuative	*kakitutu*	*aritutu*	*sinitutu*
Conditional	*kakaba*	*araba*	*sinaba*
Provisional	*kakeba*	*areba*	*sinureba*
Concessive	*kakedo*	*aredo*	*sinuredo*
Nominal	*kakaku*	*araku*	*sinuraku*

	LB	UB	UM	*k*-irr	*s*-irr
Base	***ake-*** 'open'	***okwi-*** 'rise'	***mi-*** 'see'	***ko-*** 'come'	***se-*** 'do'
Finite					
Conclusive	*aku*	*oku*	*(mi)*	*ku*	*su*
Adnominal	*akuru*	*okuru*	*miru*	*kuru*	*suru*
Exclamatory	*akure*	*okure*	*mire*	*kure*	*sure*
Imperative	*ake(yo)*	*okwi(yo)*	*mi(yo)*	*ko*	*se(yo)*
Neg. conject.	*akezi*	*okwizi*	*mizi*	*kozi*	*sezi*
Optative	*akena*	*okwina*	*mina*	*kona*	*sena*
Prohibitive	*na ake so*	*na okwi so*	*na mi so*	–	*na se so*
Non-finite					
Infinitive	*ake*	*okwi*	*mi*	*ki*	*si*
Gerund	*akete*	*okwite*	*mite*	*kite*	*site*
Continuative	*aketutu*	*okwitutu*	*mitutu*	*kitutu*	*situtu*
Conditional	*akeba*	*okwiba*	*miba*	*koba*	*seba*
Provisional	*akureba*	*okureba*	*mireba*	*kureba*	*sureba*
Concessive	*akuredo*	*okuredo*	*miredo*	*kuredo*	*suredo*
Nominal	*akuraku*	*okuraku*	*miraku*	*kuraku*	*suraku*

Second, the adnominal also functions to form headless nominalizations, meaning both (i) 'the one who dies'; and (ii) 'that someone dies'. In OJ, the former use is almost only found in pseudo-cleft constructions, e.g. (8). The use of the adnominal in complement clauses ('that . . .') is very rare in OJ, where generally the nominal form was used, see 3.1.3.3. However, in EMJ both of these nominalizing uses of the adnominal became frequent.

(8) [*kadi* *no* *oto* *suru*] *pa* *ama-wotomye* *kamo*
oar GEN sound do.ADN TOP fisher-girl Q
'the ones making the oar-sounds, is that the fisher-girls?' (*MYS* 15.3641)

Third, the adnominal form is also used as the predicate in exclamative or interrogative main clauses, often but not always in correlation with a focus or interrogative particle, *so*, *namo*, *ya*, *ka* (see 3.7.2 and in particular 8.9). Finally, it is used as the verb of a subordinate clause concluded by some conjunctional particles, e.g. *ni*, *wo*, and those derived from nouns (see 3.7.4). In both these last two functions, the adnominal may be regarded as being nominalized and it has also been proposed (Shibatani, p.c.) that the use of this form in adnominal clauses is a nominalizing use, which would make all uses of this form nominalizing. The functions of the OJ conclusive and adnominal largely remained unchanged in EMJ, but underwent important changes in LMJ, see 12.6.1.

Exclamatory (*sinure*) This form is mainly used to form the predicate of an exclamative main clause, sometimes on its own, but often in correlation with the focus particle *koso* (see 8.9.2). The exclamatory can also function as the predicate in a subordinate clause, often translated with one of a number of conjunctional meanings: 'if, when, although, because'. Usually it is accompanied by some modal or other particle. It also functions as a combining stem, selected by some suffixes, see 3.4.4.

Imperative (*sine*) This is used as the direct imperative and to form directive clauses when followed by the purposive conjunctional particle *to* ('in order that . . .').

Negative Conjectural (*sinazi*) The negative conjectural is semantically the negative counterpart of the conjectural auxiliary *-(a)m-* and has the same range of meanings. It generally means 'I don't want to, I shan't die; you shouldn't, mustn't die; he probably, surely won't die'. It is also often used to form directive clauses when followed by the purposive conjunctional particle *to* ('lest . . .'). Traditionally, the negative conjectural has, due to its semantic affinity with the conjectural auxiliary *-(a)m-* (3.1.4.10), been regarded as an auxiliary, *-(a)zi*, with the sole form *-zi* functioning as conclusive, adnominal and exclamatory (reflecting the traditionally recognized types of sentence conclusion, neutral and in correlation with a focus or interrogative particle).

However, nothing supports that view: the negative conjectural is simply an inflected word form.

***Optative** (sinana)* OJ has three sets of optative forms, see (10), of which only one is listed in Table 3.1. They express the wish of the speaker, predominantly (but not exclusively) about actions of a 1st, 2nd, or 3rd person, respectively. With the exception of *-(a)namu*, which is used throughout EMJ as a general optative, these forms disappear from the language in the transition from OJ to EMJ; *-(a)namo* is thought to be the older variant and the source (through mid vowel raising, see 2.7.2.3) of *-(a)namu* which was more frequent already in OJ.

(9)	*sinana*	'I would like to die'
	sinane(mo), sinani(mo)	'I wish you'd die/ please die'
	sinanamo, sinanamu	'I wish he'd die/ let him die'

***Prohibitive** (na-sini-so)* The prohibitive expresses a negative command 'don't ..!'. There are three variants, see (10), of which only *na . . . so* survived into EMJ. The prohibitive may well be thought to be a construction rather than an inflected word form, and in EMJ longer stretches than a single verb can occur between *na* and *so*, but this does not seem to have been possible in OJ. Including the prohibitive as a word form allows us to posit a *circumfixal* formant, the only one of its kind in pre-modern Japanese, *na-. . .-so*, which surrounds the infinitive used as stem. (With *se-* 'do', the base rather than the infinitive is used: *na-se-so* 'don't do!'.) Etymologically, *na* is probably a negative adverb and it may be thought that the use of *na* as a prohibitive prefix is the older construction which later came to be supported by *so* (cf. also 3.1.5); *so* is thought to reflect *sə, the historical root, used as imperative, of the verb *se-* 'do'. This finds some support in the long form *na-. . .-sone* which seems to involve an older fossilized optative form, *sone* 'I wish you'd . . .', of the ancestor of OJ *se-*, again involving the older root. Note also the prohibitive final particle *na* (3.7.5).

(10)	*na-sini-so*	'don't die!'
	na-sini-sone	'(please) don't die!'
	na-sini	'don't die!'

3.1.3.2 Non-finite verb forms

All stages of Japanese, including OJ, are characterized by the existence of a number of non-finite verb forms which are differentiated by the type of conjunction that holds between the clause they conclude and a following clause.

***Infinitive** (sini)* The non-finite verb forms include one coordinate form, the infinitive, whose main use is to form a nonfinal predicate coordinate with a

following predicate. A number of examples point to an earlier stage where the infinitive was subordinate,[2] a function which at the OJ stage had largely been taken over by the gerund (which is newer in the language). Derived from and segmentally identical with the infinitive is the substantive noun, e.g. *itupari* 'lie, deceit' (*itupar-* 'to lie'); in EMJ and in later periods, the infinitive and its derived noun are attested with different prosodic shapes (see 7.4.4.3) and this is also thought to have been the case for OJ. The infinitive is also used as a combining stem, both in composition and selected by some suffixes (see 3.4.4.1).

The remaining non-finite verb forms are subordinate.

***Gerund** (sinite)* The gerund is a neutral subordinate verb form, unspecified for the type of conjunction that holds between its own and a following higher clause. The use of the word 'gerund' for such subordinate adverbial verb forms is now customary in the description of many languages, although it is somewhat misleading to those more familiar with the use of 'gerund' to designate a nominalized verb form in for example Latin grammar. For Japanese this nomenclature is found first in the work of the Portuguese Jesuits around 1600 (e.g. Rodrigues's *Arte da lingoa de Iapam*, see 10.2.2.2), based, presumably, on the similarity in meaning with the Portuguese *gerúndio*, an adverbial form which historically derives from the ablative of the Latin gerund. Rodrigues, however, also classifies the form as 'participle'.

The other subordinate, non-finite forms are specified for some type of conjunction with the higher clause.

***Continuative** (sinitutu)* The continuative expresses continuation, repetition, contemporaneity, or by extension, concession; many of its meanings can be captured by English 'while'.

***Conditional** (sinaba)* The conditional concludes a conditional subordinate clause 'if'.

***Provisional** (sinureba)* The provisional concludes a provisional, temporal, or causal subordinate clause, most of which can be rendered by English 'as'.

***Concessive** (sinuredo)* The concessive concludes a concessive subordinate clause 'although, even though'; the concessive is often followed by the particle *mo* with no discernible difference in meaning.

3.1.3.3 Nominal

This form expresses nominalization: 'the fact of dying; that someone dies'. It was a very frequent and prominent form in OJ (also of adjectives, see 3.2.2.2)

[2] For example (i) which is found in *MYS* 2.85, 5.867, *K* 88:

(i)	*kimi*	*ga*	***yuki***	*ke*	*naga-ku*	*nari-nu*
	my.lord	GEN	go.INF	day	long-ACOP.INF	become-PERF.CONCL

'many days have passed since you, my lord, left'

whose basic function was abstract nominalization ('the fact that . . .'), (11a), often used in complement clauses ('that . . .'), (11b–c), but also for example as a nominalized exclamative predicate, (11d). See 3.4.4.4 about the formation of the nominal form.

(11) a. aki *tukeba* *momiti* *tiraku*
autumn arrive.PROV autumn.leaves scatter.NMNL
'The scattering of the autumn leaves when autumn comes' (*MYS* 19.4161)

b. [*nara no* *asuka wo* mi*raku*] *si* *yo-si* *mo*
Nara COP.ADN Asuka ACC see.NMNL EMPH good-ACOP.CONCL ETOP
'It is good to see Asuka which is in Nara' (*MYS* 6.992)

c. [wa ga kokoda *sinwopaku*] sira-*ni*
I GEN this.much yearn.NMNL know-NEG.INF
'Not knowing that I would yearn this much' (*MYS* 19.4195)

d. kimi ga mi-koto *wo* moti*te* *kaywopaku*
lord GEN HON-word ACC hold.GER arrive.NMNL
'It [the branch of jewel-pine] arrives bearing the words of you, my lord!' (*MYS* 2.113)

The nominal form disappears as a productive form in the transition to EMJ where it only survived in set phrases with a small number of verbs (cf. 9.1.7). From EMJ the adnominal form takes over the function of use in complement clauses, but in OJ there was clear functional differentiation between the nominal and the adnominal form, the latter not being used in abstract nominalizations and only very rarely in complement clauses, of which (12) is an example. Wrona (2008) details the uses of the nominal and adnominal forms in OJ, and more generally describes the OJ system of complementation, including the use of the formal noun *koto* 'thing, fact' both in abstract nominalization and in complement clauses.

(12) [yama-mori no ari-*kyeru*] sira-*ni*
mountain-guard GEN be-MPST.ADN know-NEG.INF
'not knowing that there was a mountain guard' (*MYS* 3.401)

3.1.4 Auxiliaries: optional categories

Auxiliaries are *inflecting suffixes*. Most auxiliaries belong to and inflect according to one of the eight verbal conjugation classes, but a few have idiosyncratic conjugations; none has the full range of inflected forms exhibited by lexical verbs, see below (3.1.4.2).

Whereas simple inflection (3.1.3) mainly expresses modal and conjunctional categories, auxiliaries express central morphological categories such as voice, tense, aspect, and mood. Thus the two sets of categories complement each other. However, an important difference is that inflection is *obligatory*, whereas the expression of the categories of the auxiliaries is *optional*. That is to say, any OJ verb form expresses one and only one of the set of paradigmatically opposed inflected forms in Table 3.1, but no auxiliary need be present. This means that a simple verb form, without an auxiliary, is *unmarked* (unspecified) with regard to the categories expressed by auxiliaries. For example, a simple verb form is unmarked with regard to tense and can well refer to deictic past time. Further, more than one auxiliary can be attached to a verb, in which case they occur in the order given in Table 3.2 (from the top down; cf. also (16) in 3.1.4.5). This affords the categories expressed by auxiliaries a different position within the morphological system from those expressed, obligatorily, by simple inflection. The combination of auxiliaries is subject to systematic restrictions, reflecting paradigmatic subsystems which are described in 3.1.4.5.

3.1.4.1 Formation on lexical verbs

Table 3.2 shows the OJ auxiliaries attached to verbs from the eight verb classes. As the auxiliaries themselves inflect, the resulting forms are shown in the basic stem, not in a word form.[3] As may be seen, there are some systematic restrictions on combinations between lexical verb classes and auxiliaries. The most important is that the bigrade verbs (which are diachronically secondary) do not combine with the respect, passive or stative auxiliaries; see Frellesvig (2008). Other non-formations concern small classes with few members and may not be significant. It is noteworthy, however, that *n*-irr verbs combine with no aspectual auxiliaries, and that no UM verb combines with the *-(i)n-* perfective. The latter is probably related to the fact that most (original) UM verbs are quite transitive; cf. 3.4.3.1. It is not surprising that *r*-irr verbs, which are all semantically stative, do not combine with the stative auxiliary.

3.1.4.2 Inflected forms

The auxiliaries have the inflected forms shown in Table 3.3, with the auxiliaries ordered in four subgroups: respect, voice (causative and passive), aspect/negation (perfective, stative, negative) and tense/mood (modal past, simple past, conjectural, subjunctive). It is significant that the auxiliaries do not have

[3] Note that the two perfective auxiliaries generally combine with different verbs (see 3.1.4.6.2); hence we have added *sak-* 'bloom', *tuke-* 'attach', and *yodi-* 'uproot' to the QD, LB, and UB columns. The only UM verb attested with the stative auxiliary is *ki-* 'put on', so *kyer-* 'be wearing' is added to the UM column. *k*-irr and *s*-irr verbs have limited and partly irregular formations with the simple past, see (3.4.3.2).

Table 3.2 *OJ auxiliaries attached to verbs from the eight verb classes*

	QD *kak-*	*r*-IIT *ar-*	*n*-IIT *sin-*
Respect	*kakas-*	–	–
Voice			
Causative	*kakasime-*	*arasime-*	*sinasime-*
Passive	*kakaye-*	*araye-*	*sinaye-*
Passive	*kakare-*	*arare-*	*sinare-*
Aspect/Negation			
Perfective	*kakite-*	*arite-*	–
Perfective	*(sakin-)*	*arin-*	–
Stative	*kakyer-*	–	–
Negative	*kakan-*	*aran-*	*sinan-*
Negative	*kakazu*	*arazu*	*sinazu*
Tense/Mood			
Modal past	*kakikyer-*	*arikyer-*	*sinikyer-*
Simple past	*kakiki*	*ariki*	*siniki*
Conjectural	*kakam-*	*aram-*	*sinam-*
Subjunctive	*kakamasi*	*aramasi*	*sinamasi*

	LB *ake-*	UB *okwi-*	UM *mi-*	*k*-IIT *ko-*	*s*-IIT *se-*
Respect	–	–	*myes-*	–	*ses-*
Voice					
Causative	*akesime-*	*okwisime-*	*misime-*	*kosime-*	*sesime-*
Passive	–	–	*miye-*	–	–
Passive	–	–	–	–	–
Aspect/Negation					
Perfective	*(tukete-)*	*(yodite-)*	*mite-*	*kite-*	*site-*
Perfective	*aken-*	*okwin-*	–	*kin-*	*sin-*
Stative	–	–	*(kyer-)*	*kyer-*	*ser-*
Negative	*aken-*	*okwin-*	*min-*	*kon-*	*sen-*
Negative	*akezu*	*okwizu*	*mizu*	*kozu*	*sezu*
Tense/Mood					
Modal past	*akekyer-*	*okwikyer-*	*mikyer-*	*kikyer-*	*siker-*
Simple past	*akeki*	*okwiki*	*miki*	*(kosi)*	*siki*
Conjectural	*akem-*	*okwim-*	*mim-*	*kom-*	*sem-*
Subjunctive	*akemasi*	*okwimasi*	*mimasi*	*komasi*	*semasi*

Table 3.3 *Inflected forms of OJ auxiliaries*

	Respect *-(a)s-*
Conclusive	*su*
Adnominal	*su*
Exclamatory	*se*
Imperative	*se*
Optative	*sane*
Prohibitive	*na VERB-si so*
Infinitive	*si*
Gerund	*site*
Conditional	*saba*
Provisional	*seba*
Concessive	*sedo*
Nominal	*saku*

	Causative *-(a)sime-*	Passive *-(a)ye-*	Passive *-(a)re-*
Conclusive	*simu*	*yu*	*ru*
Adnominal	*simuru*	*yuru*	*ruru*
Imperative	*sime(yo)*	–	–
Negative conjectural	–	*yezi*	*rezi*
Infinitive	*sime*	*ye*	*re*
Gerund	*simete*	*yete*	*rete*
Continuative	*simetutu*	*yetutu*	*retutu*
Conditional	–	*yeba*	*reba*
Nominal	–	*yuraku*	*ruraku*

	Perfective *-(i)n-*	Perfective *-(i)te-*	Stative *-yer-*	Negative *-(a)zu ~ -(a)n-*
Conclusive	*nu*	*tu*	*yeri*	*zu*
Adnominal	*nuru*	*turu*	*yeru*	*nu*
Exclamatory	*nure*	*ture*	*yere*	*ne*
Imperative	–	*teyo*	*yere*	–
Optative	*nana*	*tena*	–	–
Infinitive	*ni-*	*te-*	*yeri*	*zu ~ ni*
Gerund	*nite*	–	–	*zute ~ nito*
Continuative	*nitutu*	–	–	–
Conditional	*naba*	*teba*	*yeraba*	*zupa*
Provisional	*nureba*	*tureba*	*yereba*	*neba*
Concessive	*nuredo*	*turedo*	*yeredo*	*nedo*
Nominal	*nuraku*	*turaku*	*yeraku*	*naku*

Table 3.3 (*cont.*)

	Modal past *-(i)kyer-*	Simple past *-(i)ki*	Conjectural *-(a)m-*	Subjunctive *-(a)masi*
Conclusive	*kyeri*	*ki*	*mu*	*masi*
Adnominal	*kyeru*	*si*	*mu*	*masi*
Exclamatory	*kyere*	*sika*	*me*	–
Conditional	–	*seba* ~ *kyeba*	–	*maseba*
Provisional	*kyereba*	*sikaba*	–	–
Concessive	*kyeredo*	*sikado*	*medo*	–
Nominal	*kyeraku*	*siku* ~ *kyeku*	*maku*	–

the full range of inflected forms exhibited by verbs. They also differ a great deal among themselves with regard to the categories they inflect for, but note that the differences are greater between the four subgroups than within them.

3.1.4.3 Respect

The respect auxiliary belongs to the QD conjugation. It regularly combines with QD and *s*-irr (and according to reading tradition also *r*-irr verbs). There is a small number of lexicalized forms, (13a) with UM verbs, (13b) derived from roots underlying bigrade verbs, and finally (13c) with slightly irregular formation.

(13) a. *myes-* < *mi-as-; cf. *mi-* 'see' (also in the compound *-(a)si-myes-* 'RESP-RESP')
kyes- < *ki-as-; cf. *ki-* 'put on' (rare, only in the form *kyes-eru* (RESP-STAT.ADN 'which you are wearing') and in *mi-kyesi* (deverbal noun; 'honourable garment'))

b. *koyas-* < *koyV-as-; cf. *ko(y)i-* 'lie (down)' (UB)
nas- < *na-as-; cf. *ne-* 'sleep, lie' (LB)

c. *omopos-* <= *omop-* 'think (of), believe' (reading tradition also gives the regular form *omopas-*)
kikos- <= *kik-* 'hear, listen' (the regular form *kikas-* is also attested)
oros- <= *or-* 'weave'
siros- <= *sir-* 'rule; know' (cf. *siro* 'castle') (there is only one OJ attestation of *siros-*, in the form *sirosi-myes-* (*EN* 8); *siros-* gained in usage in EMJ. The regular OJ forms are *siras-*, *sirasimyes-*.)

The respect auxiliary is used to express 'subject exaltation', cf. 12.7.1, which however also has basic deictic functions, as illustrated in (14):

(14)	ko	no	woka	*ni*	na	tuma-***su***	kwo
	this	GEN	hill	DAT	greens	pick-RESP.ADN	child
	ipye	*kikana*				nora-***sane***	
	home	ask.OPT				tell-RESP.OPT	

'(You,) girl who picks greens on this hillside, I want to ask your home. Please tell me!' (*MYS* 1.1).

3.1.4.4 Voice; causative and passive

The causative and passive auxiliaries belong to the LB conjugation. The OJ causative is used to express causation and respect (see 12.7.1.2). It is formed on verbs from all conjugation classes. In EMJ, a LB causative *-sase-* was used alongside, and eventually replaced, *-(a)sime-* (see 8.4.1). Early formations involving a precursor of the new emerging causative are found already in OJ, in particular in the following forms:

(15) *kise-* <= *ki-* 'wear'; *mise-* <= *mi-* 'see'; *apase-* <= *ap-* 'meet'; *kikase-* <= *kik-* 'hear'; *sirase-* <= *sir-* 'know; rule'; *opose-* <= *op-* 'carry'.

The OJ passive functioned as a pure passive, a medium voice, and as a potential. OJ had two competing passive markers of which *-(a)ye-* was used rather more frequently than *-(a)re-* (their roles are reversed in EMJ into which *-(a)ye-* only survived in a number of lexicalized forms (*kikoye-*, *miye-*, *omopoye-*) and in reading glosses to Chinese texts, see 9.1.6). *-(a)ye-* was used with QD, *n*-irr, *r*-irr and UM verbs; *-(a)re-* with QD, *n*-irr, and *r*-irr verbs. Passives were not formed on verbs from other conjugation classes. *-(a)ye-* thus had a wider use than *-(a)re-*, both in terms of distribution and in terms of frequency. The straightforward interpretation of these facts is that *-(a)ye-* was the earlier passive marker and was replaced by innovative *-(a)re-*. There are a few lexicalized passive forms with a slightly irregular formation: *kikoye-* <= *kik-* 'hear'; *omopoye-* <= *omop-* 'think' (found alongside *omopaye-*, but was much more frequent).

The passive auxiliary variant *-rare-*, attaching to vowel base verbs, does not appear in the language until EMJ (see 8.4.1) and forms no part of OJ, although it is included in some grammars. Furthermore, an OJ passive auxiliary *-raye- is traditionally posited, said to attach to UB and LB and sometimes also to *s*-irr and *k*-irr verbs. This is based solely on four occurrences of the set phrase *i no nerayenu* (*i* 'sleep', *no* genitive) in *MYS* 15 (3665, 3678, 3680, 3684), whose *nerayenu* is traditionally analysed as *ne-raye-nu* sleep-PASS-NEG 'not

be able to sleep'. However, the non-formation of passives on bigrade verbs is entirely systematic (see 3.5.1). The singular form *neraye-* must be explained otherwise, e.g. as the passive of a verb *ner-*,[4] or as an idiosyncratic analogical formation; it cannot alone form the basis for positing a general variant *-raye- as part of the grammar of OJ.

Note that the causative and the passive were paradigmatically opposed in OJ (could not co-occur in a verb syntagm), as opposed to later stages of the language where they combined in some of their functions (see 8.4.1).

3.1.4.5 Aspect, tense, negation, and mood

The core aspectual and temporal morphology of OJ is expressed by the five auxiliaries *-(i)n-*, *-(i)te-*, *-yer-*, *-(i)ki*, *-(i)kyer-* (usually an auxiliary *-(i)tar-* is also posited in OJ grammars, but see 3.1.4.7.3). There is a remarkable lack of agreement about the basic function of each of these auxiliaries, which form part of the core verbal morphology of OJ, and about their mutual relations, probably related to the focus of most pre-modern Japanese grammars on a positive identification of the meaning of individual grammatical morphemes, instead of seeing them as parts of a (number of) system(s).

The morphological system of tense and aspect is inextricably interwoven with that of negation and mood. The full system also includes the negative *-(a)zu* ~ *-(a)n-*, the conjectural *-(a)m-*, and the subjunctive *-(a)masi*. The system exhibits the following paradigmatic relations within three categories. Thus, only one member from each category can occur in a verb syntagm. If more than one category is present, they occur in the order given here.

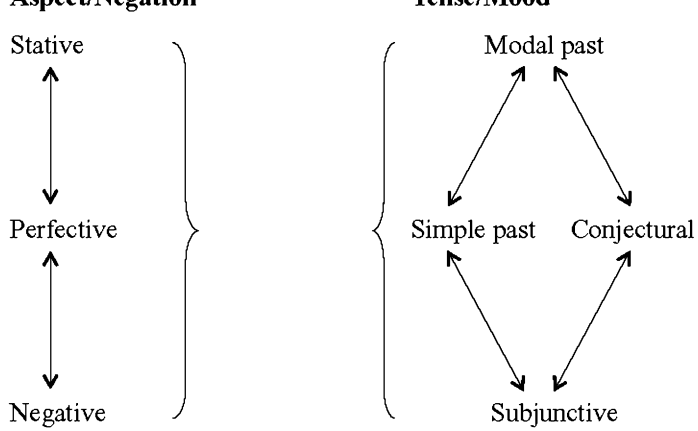
(16) a.	***Aspect/Negation***		Perfective *-(i)te-* ~ *-(i)n-*
			Stative *-yer-*
			Negative *-(a)zu* ~ *-(a)n-*
b.	***Tense/Mood*** (i)		Modal past *-(i)kyer-*
			Simple past *-(i)ki*
			Subjunctive *-(a)masi*
		(ii)	Modal past *-(i)kyer-*
			Conjectural *-(a)m-*
			Subjunctive *-(a)masi*

As indicated by the occurrence of the subjunctive and the modal past in two categories each, these three categories combine to form two subsystems which

[4] *Ner-* would in turn probably be best understood as a form, e.g. a lexicalized stative, of the root, *nV-, which diachronically underlies LB *ne-*. This root is also reflected in the lexicalized respectful verb *nas-* 'deign to sleep', and possibly also in the noun *i* 'sleep' ?< *ni (cf. Whitman 1985: 34 about palatalization and loss of /*d, *z, *n, *r/ before /*i/).

may be illustrated as follows; the horizontal dimension shows *combination*, the vertical dimension shows *selection* (mutual exclusion):

(17)



3.1.4.5.1 Aspect and negation Aspect and negation together form one subsystem with the perfective as the pivot: the perfective (*-(i)te-* ~ *-(i)n-*) forms an opposition with and never combines with the stative (*-yer-*). The perfective is also opposed to and never combines with the negative (*-(a)zu* ~ *-(a)n-*). Finally, the negative and the stative do not combine, but this seems, at least diachronically, to be a secondary opposition (cf. 3.1.4.6.1). See further 3.1.4.6 about the perfective, 3.1.4.7 about the stative and 3.1.4.8 about the negative.

3.1.4.5.2 Tense and mood Tense and mood form a more complex second subsystem: the modal past (*-(i)kyer-*) is opposed to and never combines with the simple past (*-(i)ki*) or the conjectural (*-(a)m-*). Likewise, the subjunctive (*-(a)masi*) is opposed to and never combines with the simple past or the conjectural. The simple past does, however, combine with the conjectural *-(a)m-* in the fairly autonomous combined shape *-(i)kyem-*. See 3.1.4.9 for the functions of the past tenses. Note that the modal past and the conjectural, as subjective modals (cf. 3.1.4.9, 3.1.4.10), do not form a conditional, as opposed to the simple past and the subjunctive; see Table 3.3(d) in 3.1.4.2.

3.1.4.5.3 Combination As implied by the illustration of the system in (17), aspect/negation combines relatively freely with tense/mood, as in the following examples:

(18)

		Aspect/Negation	Tense/Mood
'hear'	*kiki-**te-kye-mu***	PERF	SPST-CONJ
'bloom'	*saki-**ni-kyeri***	PERF	MPST
'meet'	*ap-**yeri-ki***	STAT	SPST
'stand'	*tat-**era-masi***	STAT	SUBJ
'be satisfied'	*aka-**ni-kye-mu***	NEG	SPST-CONJ
'be satisfied'	*aka-**zu-kyeri***	NEG	MPST

Note, however, that examples of the negative with the past tense auxiliaries are very rare and that the negative combines with the conjectural and subjunctive only in its analytic forms formed with the existential verb *ar-* (see 3.1.4.8.1), and that it thus strictly speaking does not directly combine with those auxiliaries.

3.1.4.6 Perfective

3.1.4.6.1 Functions The perfective combines two main functions: Aspectually, a perfective is a verb form that views a situation in its entirety, including beginning, middle, and end. This whole can be looked at from both ends, as it were, so as with perfectives in other languages both (a) *ingressive* (inceptive) and (b) *completive* uses are possible, as in (19). This is the origin of the opposition with the stative.

(19) a. *naki-**nu*** 'begins to sing', *naki-**tu*** 'begins to sing'
b. *tiri-**nu*** 'has fallen', *mi-**tu*** 'has seen'

The other main function of the perfective is to *assert or affirm* the state of affairs expressed by the verb. This is the origin of the opposition with the negative (and with negation in general: the perfectives do not form a negative conjectural). In this function the perfective often combines with the conjectural, *-te-m-*, *-na-m-*, e.g. (20a), or assumes a modal form, e.g. (20b) with the perfective in the optative. Diachronically this seems to be the primary function of *-(i)te-* ~ *-(i)n-*, reflecting the fact that they derive from the ancestors of the OJ copulas *to* ~ *ni*, see (3.5.2).

(20) a. miti-ki-***na-mu***
rise-come-PERF-CONJ
'(the tide) will surely rise' (*MYS* 2.121)

b. kari-***tena***
cut-PERF.OPT
'I want to cut (seaweed)!' (*MYS* 2.121)

3.1.4.6.2 Distribution of the variants *-(i)te-* and *-(i)n-* The distribution of the two perfective auxiliaries has traditionally been described as being based on the transitivity of the host verb:

(21) *-(i)te-* is used overwhelmingly with transitive verbs
-(i)n- is used mainly with intransitives

There are a fair number of exceptions to this generalization, however, and for that reason some scholars hold that *-(i)te-* and *-(i)n-* (already) in OJ were distinct morphemes expressing separate grammatical categories (for example Sandness (1999: 54ff.): *-(i)n-* 'punctuality', *-(i)te-* 'recent past'). However, a number of distributional facts make it clear that *-(i)te-* and *-(i)n-* do belong closely together in OJ; *-(i)te-* and *-(i)n-*:

(22) a. are mutually exclusive;
b. occupy the same position in a verb syntagm;
c. do not co-occur with the stative and with the negative;
d. exhibit mostly the same inflected forms.

On the other hand, the precise details of the differences in use between *-(i)te-* and *-(i)n-* remain unclear. There are clearly discernible, strong tendencies in their distribution in terms of semantico-syntactic properties of the host verb, particularly when refining this in terms of *split intransitivity*:[5]

(23)	Transitives	Intransitives	
		unergatives	unaccusatives
	-(i)te-	*-(i)te-*	*-(i)n-*

There are, however, also exceptions to this generalization, and in particular there are some verbs which occur with both perfective auxiliaries. First, the following *grammatical* verbs are found with both perfective auxiliaries: light verbs: *se-* 'do' and *ar-* 'exist'; auxiliary verbs: *-ko-* 'come to . . .', *-yuk-* 'continuative'. However, in such cases, the choice of perfective auxiliary does not depend on the grammatical verb, but rather on the lexical host verb. More importantly, there is a small number of *lexical* verbs which are attested with both suffixes:

[5] Though it appeared for the first time, unattributed, in print in Frellesvig (2001: 14), this suggestion is due to John Whitman (p.c.).

(24) *ip-* 'say, tell, call', *kakure-* 'hide (oneself); die (respect)', *ko-* 'come', *mi-* 'see, look', *miye-* 'see.PASS: be visible/seen, seem; come.RESP', *nak-* 'give voice, let out a cry, cry', *ne-* 'lie down, sleep/lie with, sleep', *ori-* 'go down, descend', *pum-* 'step (on), stamp (on); walk', *sinwop-* 'praise, long for', *yuk-* 'go'.

It should first of all not be overlooked that there may have been variation between OJ speakers in the use of the perfectives and in the definition of criteria for distribution. More importantly, however, as pointed out by Sorace (2000), auxiliary selection based on split intransitivity is not exclusively a matter of invariable lexical properties of host verbs, but proceeds along a hierarchical scale and includes semantic properties deriving from the clauses/predications in which the perfective occurs.[6] It is single argument (intransitive) verbs which have no or a low lexical specification of *telicity*, *affectedness*, *agentivity*, and *volitionality* which exhibit variation in the choice of perfective auxiliary. In particular, agentivity and volitionality seem to be relevant for OJ.

3.1.4.7 Stative

3.1.4.7.1 Function The stative presents a situation as a state. This includes the result of an action and the stative is thus in some respects similar to a perfect. The Japanese stative has been known by many different names, e.g. resultative, durative, progressive, perfect, or imperfect, but the term 'stative' adopted here captures better the basic meaning. A stative is not formed on *r*-irr verbs, which are inherently stative; in OJ the two lexical *n*-irr verbs, *sin-* and *in-*, did not form a stative. OJ had a morphological stative auxiliary and a periphrastic stative construction.

3.1.4.7.2 Morphological stative auxiliary: *-yer-* The stative auxiliary *-yer-* belongs to the *r*-irr conjugation. It has a systematically restricted distribution: It is regularly formed on QD, *s*-irr and *k*-irr verbs, and it is also found with the UM verb *ki-* 'put on' (*kyer-* 'be wearing, have on'), but is not formed on other verb classes, in particular not the lower and upper bigrade classes, reflecting that they are derived and secondary and most likely younger in the language (cf. 3.5.1). Etymologically, the morphological stative results from univerbation of an analytic construction consisting of the infinitive followed by the existential verb *ar-*, with contraction of the two contiguous vowels (see 2.7.2.2): *saki-ar- 'bloom-be' > *sakyer-* 'be in bloom, be blooming'.

[6] Washio (2002) is the first to point out explicitly that selection of perfect auxiliaries in Germanic and Romance involves similar issues to the selection of perfective auxiliaries in OJ.

3.1.4.7.3 Periphrastic stative: *-(i)te ar-, -(i)tar-* Whereas the morphological stative was not used with all verb classes, a periphrastic stative construction was available for that purpose, also formed with an existential verb, but following the gerund of the verb, e.g. ***panarete ar-*** 'be separate' (<= *panare-* 'grow/become separate'). This construction is synonymous with the morphological stative. It is likely that the periphrastic stative construction arose in complementation of *-yer-* to be used with the secondary (bigrade) verb classes as they emerged and later was generalized to occur with all classes. The periphrastic stative was also formed on the continuative + *ar-*: VERB-*tutu ar-*, and it was also formed using one of the exalted synonyms of *ar-*, for example *mas-*, *imas-*. The pattern of forming statives by gerund + existential verb is pervasive through the history of Japanese and it is still the way statives are formed in cNJ where the new existential verb *i-* is used (see 12.4).

In the later part of OJ we find phonological fusion of *-te ar-*, e.g. *nokorite ar-* 'remain' => *nokoritar-* (<= *nokor-* 'be left behind'). The auxiliary *-(i)tar-* is a prominent and important auxiliary in MJ; through EMJ it increased dramatically in frequency at the expense of *-yer-* which it eventually replaced in LMJ, and it is the ancestor of the NJ past tense flective *-(i)ta*. However, although *-(i)tar-* is included in most grammars of OJ, there is little positive basis for positing it as an independent grammatical form: OJ /*-tar-*/ is a simple phonological shortening of the periphrastic stative VERB-*te ar-* => VERB-*tar-*, derived by synchronic phonological rules (see 2.6.1, 3.4.2.1.2).

3.1.4.7.4 Analytic progressive Activity verbs formed an analytic *progressive* with the verb *wor-* 'be sitting' used directly after the infinitive: *tomosi wor-* 'be burning (something), be lighting a fire'; *wor-* is the opaque lexicalized stative form of *wi-* 'sit down', see further 3.4.2.1.

3.1.4.8 Negative

The negative exhibits a suppletive paradigm in OJ, with variation between two basic variants, *-(a)n-* and *-(a)zu-*, even within some forms, namely infinitive and gerund, column (a) in Table 3.4. Infinitive *-ni* and gerund *-nito* are rare already in OJ and both forms went out of use in the transition from OJ to EMJ, to give the suppletive paradigm in column (b) in Table 3.4 (which also includes a new gerund variant). The OJ paradigm reflects a transitional stage, with innovative *-(a)zu* based variants being in the process of replacing, or having replaced, some of the forms in an earlier QD type paradigm. See 3.1.4.8.3 about the reformation of the inflectional paradigm of the negative.

Morphophonologically, the negative paradigm is interesting. Its two basic variants belong to entirely different conjugational types: *-(a)n-* conjugates like a QD verb (with a notable irregular gerund). There are no lexical QD

Table 3.4 *OJ transitional paradigm and EMJ suppletive paradigm of the negative*

	(a) OJ transitional paradigm	(b) EMJ suppletive paradigm
Conclusive	*zu*	*zu*
Adnominal	*nu*	*nu*
Exclamatory	*ne*	*ne*
Infinitive	*ni ~ zu*	*zu*
Gerund	*nito ~ zute (~ zuni ~ zusite)*	*zute ~ de*
Concessive	*nedo*	*nedo*
Provisional	*neba*	*neba*
Conditional	*zupa*	*zupa*
Nominal	*naku*	*naku*

-n- bases and the negative auxiliary *-(a)n-* is thus the only QD type *-n* base in the language.

On the other hand, the infinitive and conclusive *-zu* is in several respects like *-ku*, the infinitive of the adjectival copula (cf. 3.2.2.4.1). Both serve as a stem in the formation of further forms, cf. gerund *-zute* and *-kute*, and conditional *-zupa* and *-kupa* (note also the variant gerund *-zuni* and the extended gerund *-zusite*). Finally, *-(a)zu* is syntactically like the existential verb *ar-* in using the infinitive in conclusive function.

3.1.4.8.1 Analytic forms Also like *-ku*, negative *-zu* may be extended with *ar-*, especially to combine with auxiliaries which never or only rarely attached directly to the negative auxiliary. For example, the negative combines with the conjectural *-(a)m-* and the subjunctive *-(a)masi-* only in the *ar-* extended form: *-zu ara-m-*, *-zu ara-masi*, and although direct combinations of the negative auxiliary with tense auxiliaries are attested (see 3.1.4.5.3), they are very rare and *-zu ari-kyer-* or *-zu ari-ki-* are more frequent. The analytic forms gave full morphophonological versatility to the negative auxiliary and use of analytic forms continued through LMJ. In some cases, *-zu ar-* phonologically fused to give *-zar-*, e.g. *-zara-m-*, *-zara-masi, -zariki* (see further 3.4.2.1.1 and 3.4.2.1.2). In the extended forms, negation could apply recursively: *miyeza-ranaku ni* <= *mi-ye-**zu** ara-**naku** ni* see-PASS-NEG exist-NEG.NMNL COP.INF 'although it is not (the case) that (she) is not visible'.

3.1.4.8.2 Negative rhetorical questions In negative rhetorical questions the negative could occur at the end of a verb syntagm, rather than in its usual position (preceding tense auxiliaries), taking wide sentential scope. Thus, there are a few examples with the modal past *-kyerazu ya*, e.g. (25). In this use, the

Table 3.5 *Reformation of the OJ paradigm of the negative*

(a) OJ QD *-(a)n-*		(b) OJ innovating forms	(c) EMJ
Conclusive	*(nu)*	*zu* < *ni-su	*zu*
Adnominal	*nu*		*nu*
Exclamatory	*ne*		*ne*
Infinitive	*ni*	*zu* < *ni-su	*zu*
Gerund	*nito* (*nite)	*zute* < *ni-su-te	*zute* ~ *de*
Concessive	*nedo*		*nedo(mo)*
Provisional	*neba*		*neba*
Conditional	*naba	*zupa* < *ni-su-pa	*zupa*
Nominal	*naku*		*naku*

negative could even combine with the perfective, in which case it was attached through extension with *ar-*, e.g. (26):

(25) *kadura* *ni* *su* *be-ku*
hair.decoration COP.INF do.CONCL NEC-ACOP.INF
nari-ni-kyera-zu *ya*
become-PERF-MPST-NEG.CONCL Q
'shouldn't it have been made into a hair-decoration?' (*MYS* 5.817)

(26) *saku* *be-ku* *nari-nite* *ara-zu* *ya*
bloom.CONCL NEC-ACOP.INF become-PERF.GER exist-NEG.CONCL Q
'shouldn't it have started to bloom?' (*MYS* 5.829)

3.1.4.8.3 Reformation of the paradigm of the negative As mentioned above, *-(a)zu* was in OJ close to completing the replacement of some of the variants in *-(a)n-* within the negative paradigm, cf. the EMJ paradigm (repeated here in column (c) of Table 3.5). As *-(a)n-* conjugates like a QD verb (with a notable irregular gerund) we may hypothetically posit an earlier regular QD conditional and a regular gerund, marked with '*', and we include a conclusive *-nu* in brackets (paralleling a regular QD verb), see Table 3.5, column (a). The *-(a)zu* forms involve a formant *su attached to the the QD type infinitive *-ni*, as shown in (33b). This *su is usually identified with the conclusive *su* of the verb 'to do', but that is doubtful given the syntactic and morphological peculiarities of *su which, as mentioned above, behaves much like the adjectival *-ku*; see further 3.5.2 about this.

In order to understand the motivation for the reformation of the forms of the negative it is revealing to compare the forms of the negative of QD, LB and UB verbs with the corresponding forms of the *-n-* perfective, see Table 3.6. We see that with QD verbs, the inflected forms of the perfective and negative are distinct throughout. However, with the bigrade verbs some of the inflected forms of the perfective are homophonous with forms of the *-(a)n-* negative; those forms are underlined. It is precisely those forms which were replaced by *-zu* based variants. Thus, the paradigm of the negative may be thought to have been reformed in order to avoid homonymy between the forms of the negative and the perfective in the bigrade conjugations. Recall (3.1.4.5.1) that the negative and the perfective took part in a paradigmatic opposition: it is not tolerable to have widespread homonymy between the forms of two members of a grammatical opposition. The reformation of the negative paradigm thus seems to have been motivated by the emergence and lexicalization of the bigrade conjugations (see 3.5.1).

3.1.4.9 Simple and modal past

Of the two OJ past tense auxiliaries, modal past *-(i)kyer-* belongs to the *r*-irr conjugation whereas simple past *-(i)ki* has its own idiosyncratic paradigm, with both *s-* and *k-* initial forms (see Table 3.7; see further 3.5.2). In the conclusive, conditional and nominal form there are variant forms of the simple past, of which *-kyeku* and *-kyeba* are very rare; the few examples are limited to the early poetry in the *Kojiki* and the *Nihon shoki* and there are no examples in the *Man'yōshū*. On the other hand, *-siku* and *-seba* are frequent and survive into EMJ. The simple past combines with the conjectural *-(a)m-*: *-kye-m-*, whereas the modal past does not; nor does it form a conditional. This difference highlights the basically modal function of the modal past.

3.1.4.9.1 Functions We will here briefly discuss the function(s) of the two auxiliaries *-(i)ki* and *-(i)kyer-* and the differences between them. Sometimes the term 'retrospective' is preferred instead of 'past', but an important function of both *-(i)ki* and *-(i)kyer-* is to refer to deictic past time. The modal past is more complex in usage than the simple past. It is widely acknowledged to involve some modal element, hence the name given it here, but this has proved difficult to pinpoint. Often a general difference is said to be one of directly experienced (*-(i)ki*) versus indirectly experienced, or evidential, past (*-(i)kyer-*). More specifically, the difference is traditionally set out as in (27). The simple past is usually said to refer to something the speaker has experienced himself, but on the other hand it can also be used in historical accounts. The modal past has a variety of uses: it is said to be used about *hearsay*, or about *sudden realization*, and also with some more general *emphatic or exclamatory force*, in which case it frequently has no past reference. Another

Table 3.6 *OJ* -n- *perfective and negative forms of QD, LB and UB verbs*

QD *sak-* 'bloom'	*Perfective* *-(i)n-*	*Negative* *-(a)n-*	*Negative* *-(a)zu*
Conclusive	*sakinu*	*sakanu*	*sakazu*
Adnominal	*sakinuru*	*sakanu*	
Exclamatory	*sakinure*	*sakane*	
Infinitive	*sakini*	*sakani*	*sakazu*
Gerund	*sakinite*	*sakanito*	*sakazute*
Concessive	*sakinuredo*	*sakanedo*	
Provisional	*sakinureba*	*sakaneba*	
Conditional	*sakinaba*	*sakanaba	*sakazupa*
Nominal	*sakinuraku*	*sakanaku*	

LB *toke-* 'melt (intr.)'	*Perfective* *-(i)n-*	*Negative* *-(a)n-*	*Negative* *-(a)zu*
Conclusive	*tokenu*	*tokenu*	*tokezu*
Adnominal	*tokenuru*	*tokenu*	
Exclamatory	*tokenure*	*tokene*	
Infinitive	*tokeni*	*tokeni*	*tokezu*
Gerund	*tokenite*	*tokenito*	*tokezute*
Concessive	*tokenuredo*	*tokenedo*	
Provisional	*tokenureba*	*tokeneba*	
Conditional	*tokenaba*	*tokenaba	*tokezupa*
Nominal	*tokenuraku*	*tokenaku*	

UB *okwi-* 'arise'	*Perfective* *-(i)n-*	*Negative* *-(a)n-*	*Negative* *-(a)zu*
Conclusive	*okwinu*	*okwinu*	*okwizu*
Adnominal	*okwinuru*	*okwinu*	
Exclamatory	*okwinure*	*okwine*	
Infinitive	*okwini*	*okwini*	*okwizu*
Gerund	*okwinite*	*okwinito*	*okwizute*
Concessive	*okwinuredo*	*okwinedo*	
Provisional	*okwinureba*	*okwineba*	
Conditional	*okwinaba*	*okwinaba	*okwizupa*
Nominal	*okwinuraku*	*okwinaku*	

Table 3.7 *OJ modal past and simple past auxiliaries*

	Modal past	Simple past	
	-(i)kyer-	*-(i)ki*	EMJ
Conclusive	*kyeri*	*ki*	*ki*
Adnominal	*kyeru*	*si*	*si*
Exclamatory	*kyere*	*sika*	*sika*
Conditional	–	*seba ~ kyeba*	*seba*
Provisional	*kyereba*	*sikaba*	*sikaba*
Concessive	*kyeredo*	*sikado*	*sikado*
Nominal	*kyeraku*	*siku ~ kyeku*	*(siku)*
Conjectural	–	*kye-m-*	*kyem-*

set of uses, however, is said to have no modal meaning, but rather to resemble a *perfect* 'has/had been . . .'.[7]

(27)

Simple past *-(i)ki*	Modal past *-(i)kyer-*
direct experience, historical accounts	hearsay, sudden realization, emphatic or exclamatory force; perfect

3.1.4.9.2 'Perfect' Examples where *-(i)kyer-* is said to function as a *perfect* on closer inspection turn out to involve an entirely different formation, namely the combination of the auxiliary verb *-ko-* 'come to', cf. (4) with the stative auxiliary *-yer-* giving the form *-(i)kyer-* 'has come to', which is homophonous with the modal past. This is for example clearly illustrated by (28):

[7] On the basis of the traditional description outlined in (27), the Turkish 'direct experience' versus 'indirect experience' past tenses are often cited as a close parallel to the OJ simple past versus modal past (as in Shinzato 1991), see (i). Although this parallelism is widely accepted (e.g. Takeuchi 1999: 101ff.), it turns out on closer inspection to be quite general and to consist in some modal distinction among past tenses. There is no close fit between the uses of the two OJ past tenses and the two Turkish past tenses.

(i)

OJ	*-(i)ki*	*-(i)kyer-*
Turkish	*-DI* Direct experience	*-mIş* Indirect experience

(28) kamwiyo *ywori* ipi-tute-***kuraku***
god.age ABL say-transmit-come.NMNL
sworaamitu yamato no kuni pa
soaring Yamato GEN land TOP
sumyekamwi *no* *itukusiki* kuni
ruling.deity GEN august.ADN land
kotodama *no* *sakipapu* kuni *to*
word.spirit GEN bless.ADN land COMP
katari-tugi *ipi-tugapi-**kyeri***
tell-continue.INF say-continue-come.STAT.CONCL

'It has been recounted down through time since the age of the gods: that this land of Yamato is a land of imperial deities' stern majesty, a land blessed by the spirit of words.' (Levy 1981) (*MYS* 5.894)

Viewed in its entirety, (28) involves the framing construction where reported speech is not only followed by a complementizer and a verb of reporting, but also introduced by the same, or a similar, verb in the nominal form, schematically shown in (29) with the verb *ip-* 'say':

(29) . . . *ipaku* 'reported speech' *to* *ip-*
say.NMNL COMP say

In (28), the verb introducing and concluding the reported speech unambiguously incorporates the auxiliary verb *-ko-* 'come to': *-kuraku* is the nominal form of the auxiliary verb *-ko-*, and not of the modal past *-(i)kyer-*, whose nominal form is *-kyeraku* (cf. Tables 3.1 and 3.3d). The form concluding the reported speech therefore also involves *-ko-*, here in combination with the stative auxiliary. Leaving aside the question of whether the modal past is etymologically related to *ko-* 'come', the auxiliary verb *-ko-* and the auxiliary *-(i)kyer-* were clearly different at the OJ stage, both in their relative positions within a verb syntagm, cf. ***maywopi-ki-ni-kyeri*** ((4) above) which contains both *-ko-* and *-(i)kyer-*, but on different sides of perfective *-(i)n-*, and in their inflected forms, cf. Tables 3.1 and 3.3d. But the combination of the auxiliary verb *-ko-* with the stative auxiliary *-yer-* is homophonous with the modal past auxiliary *-(i)kyer-*.

It turns out, then, that what have until now been understood as different uses or functions of one grammatical morpheme, *-(i)kyer-*, in fact are two entirely different formations which are homophonous. One is the auxiliary

verb *-ko-* combined with the stative auxiliary *-yer-*; the other is the modal past auxiliary, whose functions will be described immediately below. The main environment where a distinction between the two can be formally made is in combination with a perfective auxiliary: *(-ko- + -yer-) => -kyer-* does not further combine with a perfective auxiliary (as stative and perfective are paradigmatically opposed and do not combine), whereas the modal past does combine with both of the perfective variants: *-(i)te-kyer-* and *-(i)ni-kyer-* unambiguously involve the modal past. Clearly, more detailed research is called for in order to chart the differences between these two different formations in other environments and their further history within Japanese.

3.1.4.9.3 Speaker commitment There is little or no positive evidence for anything 'indirect' or 'evidential' about the OJ modal past *-(i)kyer-*. In particular, there are no good examples of the modal past used about *hearsay* in OJ: in examples where a hearsay interpretation is not simply imposed by translators any sense of reporting actually resides in a lexical verb of reporting and not in the auxiliary *-(i)kyer-*. The element which is common to the uses of the modal past (captured by 'sudden realization', 'emphasis' in the traditional description), may best be understood as *speaker commitment*, imparting a sense of 'I tell you', as in examples such as (30) and (31), which clearly have nothing to do with indirect experience. Such a modal use of a past tense is well known from other languages, e.g. Norwegian, (32). In both (31) and (32), the past tense component is semantically overridden by the modal component, but that does not make the Norwegian past tense, or the OJ modal past, less of a past tense. It simply means that they have modal uses which pragmatically can dominate.

(30) *wegusi* *ni* *ware* *wepi-ni-kyeri*
smiling.saké DAT I get.drunk-PERF-MPST.CONCL
'I became drunk on the saké of smiles!' (*KK* 49)

(31) miyabwiwo *ni* ware pa ari-*kyeri*
elegant.fellow COP.INF I TOP exist-MPST.CONCL
'what an elegant fellow I am!' (*MYS* 2.127)

(32) *så fin jeg var nå!*
so fine I was now
'how nice I look!'

By contrast, the simple past *-(i)ki* is a neutral, simple preterite which places a situation before the time of speaking, or some other temporal reference point.[8]

Positing explicitly an element of subjectivity as the defining characteristic of the modal past is nothing new. Takeuchi (1999: 101ff.) shows that that interpretation goes back to the eighteenth century. Two facts which underpin this interpretative intuition are, first, that the modal past does not, as opposed to the simple past, have a conditional form, cf. Table 3.3d above. This is a morphological property the modal past shares with the conjectural auxiliary *-(a)m-*; both are subjective modals, and truly subjective modals do not form conditionals. Second, the difference between the two past tenses is clearly brought out by a striking difference in their distribution in the two sets of OJ prose texts:

(33)

	Senmyō addresser focused: *from* an emperor/empress	*Norito* addressee focused: *to* the spirits
Simple Past	+	+
Modal Past	+	–

The modal past does not occur in the *Norito* at all, while it is used in the *Senmyō*. In narrative terms, the main difference between these two sets of texts is one of *address*: the *Norito* are addressee focused, first of all being addresses *to* the deities/spirits; *Senmyō* on the other hand are addresser focused, being addresses *from* an emperor or empress. There is thus less scope for subjectivity and speaker commitment in the *Norito* which focus on the addressee. In the *Senmyō*, on the other hand, a clear pattern of usage may be observed which utilizes the distinction: the modal past is used in direct or indirect speech with some element of speaker commitment, e.g. (34)–(35). Conversely, the simple past is used in frame descriptions, almost entirely about divine or imperial past action (and with a respectful verb form), e.g. (36):

(34) sikaredomo ima *pa* akiraka *ni* Nakamaro *ga*
however now TOP clear COP.INF Nakamaro GEN
itupari *ni* ari-*kyeri* *to* sirite . . .
deceit COP.INF exist-MPST.CONCL COMP know.GER
'However, knowing clearly now that it was Nakamaro's deceit'
(*SM* 28)

[8] The distinction between the simple and modal past in OJ is similar to the modal distinction proposed – if not wholly uncontroversially – by Durst-Andersen (1996) for the two past tenses in Italian, with the *passato prossimo* resembling the OJ modal past in expressing speaker involvement, as opposed to the *passato remoto* which is said simply to report neutrally, like the OJ simple past.

(35) sikaredomo sumyera *to* imasi*te* ame no
however emperor COP.INF exist.RESP.GER heaven GEN
sita no maturigoto *wo* kikosi-myesu
bottom GEN ruling ACC perform-RESP.ADN
koto *pa* itapasi-*ki* ikasi-*ki* koto *ni*
thing TOP laborious-ACOP.ADN hard-ACOP.ADN thing COP.INF
ari-*kyeri*
exist-MPST.CONCL
'However, ruling the land as emperor was laborious and hard!'
(*SM* 23)

(36) takama no para *yu* amori-masi-*si* sumyera
high.heaven GEN plain ABL descend-RESP-SPST.ADN emperor
'an emperor who descended from the high heavens' (*SM* 2)

The *Senmyō* are recited texts with several embedded layers of narration and they might thus be said to be all direct speech, but the point is that the simple past is used in what is *presented* as frame description, or *asserted* to be part of common knowledge, i.e. shared and generally accepted truth, whereas the modal past expresses speaker involvement and commitment. Thus, the distinction between the two OJ past tenses has nothing to do with directness of experience; instead they differ in the expression of speaker commitment.

3.1.4.10 Conjectural and subjunctive

The *conjectural* belongs to the QD conjugation. The *subjunctive* has its own conjugation, the few forms of which resemble the simple past.

The conjectural is the least specific of the modal auxiliaries and is very frequent in the OJ texts. Its uses fall into two overall categories: (a) *conjecture*: probability and necessity; (b) *volition*: intention and wish. Note that the combination of the simple past and the conjectural *-kye-m-* is used only to express conjecture 'probably was, probably did', not volition. Like modals in many languages, the conjectural can also be used to refer to future time with little or no modal meaning; the conjectural is therefore sometimes termed a 'future tense'. Followed by the conjunctional particles *to* or *ni* it is used to form purposive adjunct and complement clauses 'that, so that'.

The subjunctive expresses a counterfactual proposition, most often contingent on a counterfactual condition in the frame VERB_1-(SUBJ).COND, VERB_2-SUBJ 'if it were the case that VERB_1, then it would be the case that VERB_2'.

3.1.5 Verbal prefixes

Like later stages of Japanese, OJ used prefixes with both nouns and verbs. Nominal prefixes include honorific *mi-*, which is reflected in later stages of the language (see 12.7.1.1), or endearing *mutu-* which was no longer productive at the OJ stage, but only found in lexicalized forms such as *mutu-tama* 'friendly spirit'. Verbal prefixes include what are usually glossed as 'emphatic' prefixes, e.g. *uti-* which seems to derive from the infinitive of the lexical verb *ut-* 'strike'. Use of such deverbal emphatic prefixes is prevalent also in later stages of the language.

Quite different from such prefixes which are found throughout the history of Japanese, we find in OJ what seem to be remnant uses of grammatical prefixes expressing functions which from OJ onwards were regularly expressed by suffixes: (a) reciprocal *api-* (*api-makura-mak-* 'pillow-spread (= sleep) together') < *ap-* 'meet'; cf. the derivational suffix *-(a)p-* (3.1.1) and the auxiliary verb *-ap-* (3.1.2). (b) Stative *ari-* (*ari-mat-* 'be waiting, keep waiting'); < *ar-* 'exist'; cf. the morphological stative *-yer-* < *-i ar- (3.1.4.7). (c) Potential *e-* (*e-yuk-* 'can go'); < *e-* 'get'; cf. both the use of *e* as a freer adverb introducing negative potentiality and the potential auxiliary verb *-e-* (3.1.2). (d) Prohibitive *na-* has no transparent lexical source and was in OJ usually used in combination with the suffix *-so* (see 3.1.3.1); cf. also the prohibitive final particle *na* (3.7.5).

The use of these prefixes which were lost in later stages of the language suggests that prefixation perhaps was used more widely in pre-OJ to express grammatical functions. This is not typical of verb final languages and could be taken to suggest that a change from preverbal to postverbal grammatical affixation could have been part of a wider change in syntactic typology towards verb final syntax. Less transparent, but perhaps part of the same phenomenon, are two verbal prefixes, *i-* and *sa-*. These two prefixes were frequent in OJ, but later disappeared from the language. Their function is not agreed upon (and they are often said to be 'emphatic'), but it has recently been suggested that *i-* and *sa-* are related to the expression of the transitivity or agentivity of the verbs they prefix to, so that *i-* attaches to active verbs (*i-yuk-* 'go') and *sa-* to inactive verbs (*sa-ne-* 'sleep'), see Yanagida and Whitman (2009); Russell points out that *i-* is attested mainly with verbs of movement (2006: 141ff.).

3.2 Adjectives

Adjectives are nominal roots or stems. The stem was usually followed by an auxiliary (see 3.2.1), but it could also be used with a fair amount of independence. The bare stem could be used exclamatorily, usually reinforced by an

interjection or an interjectional particle, e.g. (37). This use is still found in NJ, often with an elongated final vowel, *takaa* 'that's expensive!'

(37) a. *ana* ***omosirwo*** 'how wonderful!' (*Kogoshūi*)
b. ***oso ya***, kono kimi 'he is stupid, this man!' (*MYS* 9.1741)

The stem was also used attributively and adverbially, first by *compounding*, (38a), (39a); second by *juxtaposition*, (38b), (39b); third with a form of a *copula*, adnominal *no*, *tu* in (38c), or infinitive *ni* in (39c); there are no examples of the stem of a *shiku* adjective adnominalized with *tu*. (38b) also exemplifies the adjective stem forming a nexus with a preceding noun, either as a compound *pa-biro* 'wide-leaved', or as a phrase *yo no topo* 'of old age'. In such cases, the resulting structure was used to modify a following noun.

(38) a. ***yasu****-i* 'good, sound sleep'
pasi*-duma* 'dear, beloved wife'

b. [*yo no* ***topo***] *pito* 'person of old age; [age GEN distant] person' (*NSK* 62)
[*mi ga* ***posi***] *kuni* 'the country I long to see; [see GEN want] country' (*KK* 58)
*pa-****biro*** *kumakasi* 'wide-leaved great oak' (*KK* 91)

c. ***topo no*** *kuni* 'distant country' (*MYS* 15.3688)
ikasi no *miyo* 'abundant reign' (*EN* 27)
topo tu *pito* 'distant person' (*MYS* 17.3947)

(39) a. ***tika****-duku* 'approach; close/near-touch/stick to'

b. ***paya*** *ko* 'come quickly!' (*MYS* 15.3636)

c. ***ko ni*** *kaki* 'paint thickly' (*KK* 42)

3.2.1 Inflectional forms; adjectival copula

However, adjectives were usually predicated, adnominalized, or adverbialized by means of particular formants which attached to the adjective stem and inflected for many, but not all of the categories of the verbal inflection, see Table 3.8. These formants were suppletive and composite in their inflection. Because of their function (to adverbialize, adnominalize and predicate adjectives) and shared morphological development with the copula (12.2.3) we interpret these formants as forms of a restricted copula ('adjectival copula') and gloss examples accordingly, e.g. *samu-ku* cold-ACOP.INF 'coldly'. This

Table 3.8 *OJ adjectival copula forms*

Finite	
Conclusive	*si*
Adnominal	*ki*
Exclamatory-1	*sa*
Exclamatory-2	*kyere*
Non-finite	
Infinitive-1	*ku*
Infinitive-2	*mi*
Gerund-1	*kute*
Gerund-2	*mito*
Conditional-1	*kyeba*
Conditional-2	*kupa*
Provisional-1	*kyeba*
Provisional-2	*kyereba*
Concessive-1	*kyedo*
Concessive-2	*kyeredo*
Nominal	*kyeku*
Negative nominal	*kye-naku*
Conjectural	*kye-m-*

interpretation of the formants attaching to the adjective stem finds further support in its relations to some other grammatical forms (see 3.5.2).

There are two slightly different classes of adjective. Adjectives with stem final *-si* use a zero allomorph of the conclusive formant *-si*, i.e. *kwopwisi* and not *kwopwisi-si, see (50) which shows the combination of the conclusive, adnominal and infinitive forms of the adjectival copula with these two classes of adjective. Traditionally the two classes are termed '*ku* adjectives' and '*shiku* adjectives' in reference to the infinitive, *topoku* versus *kwopwisiku*, although in fact the *si* of the *shiku* adjectives is part of the stem and the infinitive ending in both cases is *-ku*. *Shiku* adjectives involve an adjectival derivational formant *-si*, see 3.2.4. A small subclass of *shiku* adjectives had stems in *-zi* rather than *-si*, see 3.2.4.2.

(40)

	Ku adjective *topo-* 'far, distant'	*Shiku* adjective *kwopwisi-* 'dear, beloved'
Conclusive	*topo-si*	*kwopwisi*
Adnominal	*topo-ki*	*kwopwisi-ki*
Infinitive-1	*topo-ku*	*kwopwisi-ku*

In OJ the syntactic specialization which we find in EMJ between the conclusive, *-si*, and adnominal, *-ki*, forms of the adjectival copula was not yet firmly established (see 3.2.2.1). In addition many categories exhibit variation between two competing forms. In some cases, this variation seems to reflect that one set of forms, the innovative forms in (b) in (41), was replacing another, those in (a); compare with the EMJ forms in column (c).

(41)

	a. OJ older	b. OJ innovative	c. EMJ
Exclamatory	*sa*	*kyere*	*kere*
Conditional	*kyeba*	*kupa*	*kupa*
Provisional	*kyeba*	*kyereba*	*kereba*
Concessive	*kyedo*	*kyeredo*	*keredo*

3.2.2 Core forms

The following forms of the adjectival copula were used widely and may be said to constitute the core forms in OJ:

(42)

Conclusive	*si*
Adnominal	*ki*
Exclamatory	*sa*
Infinitive-1	*ku*
Infinitive-2	*mi*
Nominal	*kyeku*

The remaining forms were not frequent: *-kyere* is attested once, but in EMJ this form became frequent. *-kute* was very rare in OJ (only about a dozen wholly or partly phonographically attested examples in the *Man'yōshū*), but common in EMJ. The conditional, provisional and concessive forms are infrequent in the texts, but the innovative set became widely used in EMJ. There are about a dozen attestations altogether of *-kyeba* and *-kyereba*; there is a single example of *-kyeredo* and a handful of *-kyedo*.

The adjectival copula did not generally combine directly with auxiliaries. The only exception was the conjectural, *-(a)m-*: *-kye-m-*. There are approximately twenty or so attestations in the *Man'yōshū*, comprising different inflected forms of *-(a)m-* (conclusive, adnominal *-kye-mu*, exclamatory *-kye-me*, nominal *-kye-maku*). In addition, there are a few examples of the nominal form of the negative *-(a)naku*: *-kye-naku*, indicating that the adjectival copula was earlier used more freely with negative *-(a)n-*: *-kye-n-. These forms disappeared altogether in the transition to EMJ and were instead formed analytically, see 3.2.3 below.

3.2.2.1 Conclusive and adnominal

The conclusive and adnominal are generally associated with those two functions, (43)–(44). Like the verbal adnominal, the adjectival copula adnominal was also used as a nominalized form in headless nominalizations, (49). However, the syntactic specialization between conclusive, *-si*, and adnominal, *-ki*, of the adjectival copula was not complete by OJ and there was some overlap in usage. There are many examples of a *shiku* adjective in what appears to be the conclusive form modifying a noun, e.g. (45), but the traditional interpretation of such examples is as a compound of adjective stem and noun. There is, however, a small number of examples with *ku*-adjectives where a conclusive form of the adjectival copula is used to modify nouns, e.g. *ara-**si** wo* 'tough man' (*ara-* 'wild, violent, tough') or fully lexicalized *yo-**si** nwo* 'good moor; Yoshino (place name)' (*yo-* 'good'). (46) is an illustrative pair of examples with *kagurwo-* 'black (of hair)' (< *ka* 'hair', *kurwo-* 'black'). Examples such as these demonstrate that what can unambiguously be identified as a conclusive form of the adjectival copula could be used to modify a noun. On the other hand, such examples are rare in the OJ texts. As in later stages of the language, and as is the case with verbs, the adnominal form may be used predicatively, but in those cases it either correlates with a focus particle as in later stages of the language, (47) (note that *koso* in OJ correlates with the adnominal of the adjectival copula, not the exclamatory; see 8.9.2), or is associated with interrogative or exclamative function, (48) (cf. 8.9.4).

(43) *a ga* *mune* ***ita-si***
I GEN heart painful-ACOP.CONCL
'my heart aches' (*MYS* 15.3767)

(44) ***kurwo-ki*** *mi-kyesi*
black.ACOP.ADN HON-clothes
'black clothes' (*KK* 4)

(45) ***topotoposi*** *kosi* *no* *kuni*
far.far.away Koshi GEN land
'the far far away land of Koshi' (*KK* 2)

(46) *kagurwo-**ki** kami* 'black hair; black-ACOP.ADN hair' (*MYS* 5.804)
*ka*gurwo-***si*** kami 'black hair; black-ACOP.CONCL hair' (*MYS* 16.3791)

(47) a. aki *to* *ipeba* *kokoro* ***so*** *ita-**ki***
autumn COMP say.PROV heart FOC painful-ACOP.ADN
'when talk is about the autumn, my heart aches!' (*MYS* 20.4307)

b. nwo *wo* *piro-mi* *kusa* ***koso*** *sige-**ki***
moor ACC wide.INF grass KOSO abundant-ACOP.ADN
'with the moor being wide, it is the grass which is abundant' (*MYS* 17.4011)

(48) wa ga koromode *no* puru toki *mo*
I GEN sleeve GEN dry.ADN time ETOP
*na-**ki***
not.exist-ACOP.ADN
'there is no (not enough) time for my sleeves to dry!' (*MYS* 10.1994)

(49) [*ipu* *subye* *no* *taduki* *mo* *na-ki*] *pa*
say.ADN way GEN means ETOP not.exist-ACOP.ADN TOP
a *ga* *mwi* *nari-kyeri*
I GEN body COP-MPST.CONCL
'The one with no means of expressing himself was me!' (*MYS* 18.4078)

3.2.2.2 Nominal

The use of the nominal form does not differ much from the use of the verbal nominal form, (50)–(51). Like it, it was an important and frequent form in OJ, but disappeared in the transition to EMJ.

(50) yononaka *no* *u-kyeku* *tura-kyeku*
this.world GEN sad-ACOP.NMNL hard-ACOP.NMNL
'the sadness and hardness of this world' (*MYS* 5.897)

(51) *kwopwisi-kyeku* *ke* *naga-ki* *mono-wo*
long.for-ACOP.NMNL day long-ACOP.ADN although
'although it has been a long time that I have longed for you' (*MYS* 17.3957)

3.2.2.3 Exclamatory

-Sa, which in later stages of the language functioned as a nominalizer, was in OJ a predicative, exclamatory form, usually occurring in the frame: *NOUN ga/no ADJ-sa* or, more rarely, *VERB/ADJ ga ADJ-sa* (52a, b). As opposed to the verbal exclamatory (and the rare exclamatory-2, *-kyere*), *-sa* could not be used as a stem for forming other inflected forms (cf. 3.4.4.3).

(52) a. *wakayu* *turu* *imwora* *wo* *mi* *ramu*
young.trout fish.ADN girls ACC see(BASE) PCONJ.ADN

pito ***no*** *tomosi-**sa***
person GEN enviable-ACOP.EXCL
'(how) enviable are those who will see the girls fishing young trout!' (*MYS* 5.863)

b. *ware* *yuwe* *ni* *omopiwabu* *ramu*
I reason COP.INF worry.CONCL PCONJ.ADN
imo ***ga*** *kanasi-**sa***
my.beloved GEN dear-ACOP.EXCL
'(how) dear is my beloved who will be worrying for my sake!' (*MYS* 15.3727)

c. *kogu* *punabito* *wo* mi*ru* ***ga*** *tomosi-**sa***
row.ADN boatsman ACC see.ADN GEN enviable-ACOP.EXCL
'(how) enviable it is to see the rowing boatsman!' (*MYS* 15.3658)

d. *apu* *be-ki* *yosi* *no* *na-ki*
meet.CONCL NEC-ACOP.ADN means GEN not.exist-ACOP.ADN
ga *sabusi-**sa***
GEN lonely-ACOP.EXCL
'(how) lonely it is that there is no means of meeting her' (*MYS* 15.3734)

3.2.2.4 The infinitives

3.2.2.4.1 Infinitive-1 The infinitive has two main functions: (a) adverbial, modifying a verbal, (53)–(55); (b) non-finite (56)–(58). In (57), *taputwo-ku* is the predicate of a nonfinal coordinate clause, whereas *kanasi-ku* is the first half of a complex predicate. In (58), *tadasi-ku* is the first half of a complex attribute. The infinitive is one of the most important of the OJ adjective forms. It is also found in the modern standard language with largely unchanged functions (although it has been subject to phonological change in EMJ, see 8.2).

(53) *minatwokaze* *samu-**ku*** *puku* *rasi*
rivermouth.wind cold-ACOP.INF blow.CONCL PRES
'the wind from the rivermouth seems to blow coldly' (*MYS* 17.4018)

(54) *topo* *no* *kuni* *ni* *imada* *mo* *tuka-zu*
distant COP.ADN country DAT yet ETOP arrive-NEG.INF
yamato *wo* *mo* *topo-**ku*** *sakarite*
Yamato ACC ETOP distant-ACOP.INF separated.from.GER
'not yet having reached that distant land and also far separated from Yamato' (*MYS* 15.3688)

(55) *kimi* *ga* *yuki* *ke* ***naga-ku*** *nari-nu*
my.lord GEN go.INF day long-ACOP.INF become-PERF.CONCL
'many days have passed since you, my lord, left' (*KK* 88)

(56) yama-kapa *wo* *naka* *ni* *penarite* ***topo-ku***
mountain-river ACC middle DAT obstruct.GER distant-ACOP.INF
tomo *kokoro* *wo* *tika-ku* *omopose*
although heart ACC close-ACOP.INF think.RESP.IMP
wagimo
my.beloved
'even if we are far apart, with mountains and rivers between us, think our hearts close to one another, my love' (*MYS* 15.3764)

(57) titi-papa *wo* mire*ba* ***taputwo-ku***
father-mother ACC see.PROV awesome-ACOP.INF
mye-kwo mire*ba* ***kanasi-ku*** *megu-si*
woman-child see.PROV dear-ACOP.INF lovely-ACOP.CONCL
'when one sees one's mother and father, they are awesome; when one sees one's wife and child(ren), they are dear and lovely' (*MYS* 18.4106)

(58) tadasi-***ku*** kiywo-*ki* kokoro *wo* moti*te*
true-ACOP.INF pure-ACOP.ADN heart ACC hold.GER
'with a true and pure heart' (*SM* 29)

The infinitive is also used as the stem upon which were built a number of forms: the gerund (*taka-**kute***) and an extended gerund (*taka-**kusite***, usually thought to involve the gerund of *se-* 'do'). The infinitive also formed the basis for the formation of the innovative conditional (*taka-**kupa***) and it combined with the concessive conjunctional particle (*topo-**ku tomo***, cf. (56)). Note that particles such as *pa* followed a finite form of verbs; this is also the case with the concessive particle *tomo* which followed the conclusive of verbs. Also the evidential final particle *miyu* which followed the conclusive of verbs is said to have followed the infinitive of adjectives, but the few examples of this rely on reading tradition and are not phonographically attested.

Finally, there are a very few examples of a nominal use of the infinitive, e.g. *tokizi-ku* adnominalized with *no*: *tokizi-ku no* 'perennial'; and adverbialized with *ni*: *tokizi-ku ni* 'perennially'.

3.2.2.4.2 Infinitive-2 *-mi* is not usually termed infinitive. While it is functionally more limited than the verbal infinitive, those functions which it has are similar to those exhibited by the verbal infinitive or gerund. Functionally it

often corresponds to the infinitive-1 in *-ku*. In later stages of the language, *-mi* came to be used as a concrete nominalizer, but this was very rare in OJ, a singular example being *sige-mi* 'thicket' (*sige-* 'dense, thickly growing'). In OJ, infinitive-2 was a subordinate, predicative form, that is, an adverbial, non-finite form. Its main use was in free adverbial clauses, usually of the form NOUN *(wo)* ADJ-*mi*. It is often interpreted to mean 'as, because', but the basic meaning is simply 'being'.

(59) yama *wo* taka-***mi*** (*MYS* 1.44)
yama taka-***mi*** (*MYS* 3.324)
*yama-daka-**mi*** (*KK* 78)
'the mountain being high; as, because the mountain is high'

(60) *opokimi no kokoro wo yura-**mi** omi no kwo*
great.lord GEN heart ACC slack-ACOP.INF Omi GEN child
no yapye no siba-kaki iritata-zu ari
GEN manyfold COP.ADN twig-fence enter-NEG.INF exist.CONCL
'the heart of the great lord being slack, he does not enter the manyfold twig fence of the young Omi' (*KK* 107)

Infinitive-2 was also used as a complement, i.e. as the predicate in governed adverbial clauses. This use is frequent in *Senmyō* (and in *kanbun-kundoku*), but is found also in the poetry. Used with *omop-* (or a synonym) this means 'find, deem NOUN to be ADJ'. Used with *se-* the meaning is 'treat NOUN as ADJ; find NOUN ADJ'. This pattern was grammaticalized and has over time yielded a number of lexicalized forms which survive into the modern language, e.g. *omonze-/omonzi-* 'value, give weight to' (< *omo-mi se-*; *omo-* 'heavy'), *karonze-/karonzi-* 'make light of' (< *karo-mi se-*; EMJ *karo-* 'light').

(61) *ne-siku wo . . . urupasi-**mi** omopu*
sleep-SPST.NMNL ACC wonderful-ACOP.INF think.CONCL
'I think it wonderful that she slept (with me)' (*KK* 46)

(62) *ima no masaka mo urupasi-**mi** sure*
now COP.ADN this.moment ETOP wonderful-ACOP.INF do.EXCL
'I find also this very moment lovely!' (*MYS* 18.4088)

The formant *-mi* is often regarded as a fossilized infinitive of a verbal derivational morpheme, *-m-. Support for this may be found in singular examples such as *kasikwo-mi-te* (*NSK* 102) 'reverently, with reverence' (*kasikwo-* 'be awe-inspired') which has *-mi* with the verbal gerund formant *-te*, or *yorokobwi*

yorosimi 'rejoicing and being glad' (*EN* 26) where the adjectival infinitive *yorosi-mi* seems to be parallel with the verb infinitive *yorokobwi*. It should be noted, however, that *-mi* also occurs in morpho-syntactic contexts which are unusual for a verbal infinitive. Interestingly, the morpho-syntax of *-mi* is to a large extent shared by desiderative *pori-*, the infinitive of a defective verb *por-* 'wish, want, love', and by *-ni*, the infinitive of the negative auxiliary *-(a)n-*. *por-* is found mainly in a few fixed expressions. It takes a nominal complement, either a noun (usually *me* 'eye' in the set phrase *N ga me wo por-* 'I want to see N') or a verb in the nominal form (usually *V-maku (wo) por-*, i.e. the nominal of the conjectural *-(a)m-* 'I want to V'). *-ni* and *pori* may both be used as free adverbials and to complement *omopu*.

(63) mi-*maku* *pori* *nisi* *no* *mimaya* *no* *two*
see-CONJ.NMNL want.INF west COP.INF stable GEN door
ni *tat-era-masi*
DAT stand-STAT-SUBJ.CONCL
'wanting to see my beloved I would be standing outside the western stables (Umaryō)' (*MYS* 15.3776)

(64) wa ga kokoda *sinwopaku* sira-*ni* pototogisu
I GEN this.much long.NMNL know-NEG.INF cuckoo
idu-pye *no* yama *wo* naki *ka*
which-direction COP.ADN mountain ACC cry.INF Q
kwoyu ramu
cross.over.CONCL PCONJ.ADN
'over which mountain will the cuckoo fly crying, unaware that I long so much' (*MYS* 19.4195)

(65) mi-*maku* *pori* omopu
see-CONJ.NMNL want.INF think.CONCL
'I am thinking that I want to meet/see you' (*MYS* 17.3957)

(66) inabinwo *mo* yuki-sugwi-kate-*ni* omop-yereba
Inabino ETOP go-pass-be.possible-NEG.INF think-STAT.PROV
'as I was thinking that it is not possible to leave Inabino' (*MYS* 3.253)

-mi, *pori* and *-(a)ni* may be further adverbialized with *to*, i.e. they form a slightly irregular gerund in *-to* instead of *-te*.

(67) tukapi *no* *kyereba* *uresi-mito*
messenger GEN come.STAT.PROV happy-ACOP.GER
'(being) happy because a messenger had come' (*MYS* 17.3957)

(68) *kurapasi-yama* *wo* *sagasi-mito*
Kurahashi-mountain ACC steep.ACOP.GER
'Mt. Kurahashi being steep' (*KK* 69)

(69) *naku* *kowe* *wo* *kika-maku* *porito*
sing.ADN voice ACC hear-CONJ.NMNL want.GER
'wanting to hear the crying voice (of the cuckoo)' (*MYS* 19.4209)

(70) *aka-nito*
be.satisfied-NEG.GER
'without being satisfied' (MYS 17.3991)

-mi and *pori* are generally not used in other inflected forms. In addition to *kasikwo-mite* mentioned above, the exceptions, which are very few, include singular instances of adnominal *poru* (*a ga poru tama* (*NSK* 92) 'the pearl that I want/love') and *pori* with *-si* the simple past adnominal (*wa ga pori-si ame pa puri-ki-nu* (*MYS* 18.4124) 'the rain I wished/longed for has started to fall'). Like *-mi*, both *-ni* and *pori* could be predicated by *se-* 'do'. For *pori* this construction was in EMJ phonologically reduced to give the desiderative verb *posse-* < *pori-se-*.[9] Note also the coradical *shiku* adjective OJ *posi* 'be desirous of' which may also be from *pori-si*.

(71) ono ga inoti *wo* naga-ku pori *sure*
own GEN life ACC long-ACOP.INF want.INF do.EXCL
'I wish for my life to be long' (*MYS* 12.2868)

(72) kimi *pa* mi*redo* *aka-ni* *se-mu*
my.lord see.CONC be.satisfied-NEG.INF do-CONJ.CONCL
'although you look at it, my lord, you will not be content' (*MYS* 17.3902)

Finally, *-mi* and *pori* are used with the particle *kamo* (otherwise used with the adnominal of verbs), expressing doubt or exclamation (*-(a)n-*, however, uses the adnominal with *kamo*, cf. (74)).

(73) *pototogisu* *naku* *oto* *paruke-si*
cuckoo cry.ADN sound distant-ACOP.CONCL
satwo-dopo-mi *kamo*
village-distant-ACOP.INF Q
'the sound of the cuckoo crying is distant, maybe because the village is far away' (*MYS* 17.3988)

[9] The only example with *pori* written phonographically before *se-* is in *MYS* 14.3383.

(74) *kokoro so ita-ki . . .* mi-*maku* *pori* *kamo*
heart FOC painful-ACOP.ADN see-CONJ.NMNL want.INF Q
'my heart aches, maybe because I want to see you' (*MYS* 20.4307)

3.2.3 Analytic forms

In addition to its basic inflected forms, the adjectival copula formed extended analytic forms by the infinitive *-ku* followed by *ar-* (or another existential verb): *-ku ar-*. This allowed the formation of forms which were either not or only rarely formed directly in the primary paradigm, including combination with auxiliaries which did not combine directly with the adjectival copula, e.g. (75) with the modal past *-(i)kyer-*. The analytic formations provided full morphophonological versatility for the adjectival copula. The analytic forms sometimes fused phonologically, *-ku ar-* => *-kar-*, which in EMJ gave rise to a secondary conjugation of the adjectival copula; see further 3.4.2.1.1 and 3.4.2.1.2.

(75) *kimi* *ga* *yosopi* *si* *taputwo-**ku***
my.lord GEN attire EMPH admirable-ACOP.INF
ari-kyeri
exist-MPST.CONCL
'your attire, it is admirable, my lord!' (*KK* 7, *NSK* 6)

3.2.4 Ku *versus* shiku *adjectives*

Shiku and *ku* adjectives cannot be formally distinguished on the basis of the conclusive form, which does not make clear whether the final /si/ is part of the stem or is the conclusive form of the adjectival copula, cf. conclusive *yosi* 'good' and *asi* 'bad'. However, adnominal *yoki* and *asiki* show that the stems are *yo-* and *asi-* respectively. Unger and Tomita (1983) point out that on the criterion of phonographic attestation of a form other than the conclusive/stem, it is actually only possible to determine positively the conjugational class of approximately two thirds of all OJ adjectives.

There are a number of points concerning the *shiku* adjectives, especially their origin and original properties, which remain unclear, but two tendencies in the semantic and morphological properties of the members of each class are notable: First, semantically, most *shiku* adjectives are 'psych' adjectives, referring to subjective emotional states, whereas the *ku* adjectives typically express more objective qualities. This was originally observed by Yamamoto (1955). On his calculations there are approximately 20 per cent exceptions to this tendency in OJ (12% for *ku*, 26% for *shiku*) and more exceptions in later

stages. (76) gives for each class a larger group conforming to the tendency and a smaller group of exceptions.

Second, morphologically, the *shiku* adjectives may generally be thought to involve a formant *-si* (or possibly *-Vsi*). It should be noted, however, that in not a few cases it is not possible to isolate a stem for the derivation which may be identified with an otherwise known morpheme. It is thus possible that some *shiku* adjectives have simple stems which happen to end in *si*; for example, *asi-* 'bad' or *wosi-* 'dear' could well be simple stems. On the other hand, there are no *ku* adjectives with stem final /-i/, and it remains likely that all adjective stems ending in /si/ are in fact derived.

It has been noted that a large group of *shiku* adjectives are transparent deverbal derivatives (Yamazaki (1992)); this is shown in (76c) which also gives the verbs that Yamazaki posits as derivational stems for such adjectives. This relation is quite different from that holding between some *ku* adjectives and coradical verbs, e.g. *aka-* 'bright, red', *ake-* 'become bright', *akas-* 'make bright'; *ara-* 'rough', *are-* 'rage; get ruined', *aras-* 'damage, ruin', where the verbs and the adjective originate in a common (nominal) root. It is on the other hand clear that not all *shiku* adjectives are deverbal.

There is no consensus about the origin or morphological status or function of the formant *-si*. Some scholars (including Yamamoto 1955, Yamazaki 1992) believe that it reflects a derivational formant *-si* (or *-Vsi*) etymologically different from the conclusive formant *-si*; varying degrees of functional specificity are posited (adjective formant, psych adjective formant, deverbal psych adjective formant). The other main point of view, represented by e.g. Kawabata (1976) is that *-si* is etymologically identical with the conclusive formant of the *ku*-adjectives and that this *-si* in the course of the formation of the adjective inflection was resegmented as a part of some adjectives (*shiku*), but not of others (*ku*). On that view the semantic specialization between *ku* and *shiku* adjectives and the function of *-si* as a derivational formant are secondary and would be a result of a further reanalysis of the *shiku* adjectives as consisting of a stem and a derivational suffix.

Regardless of its etymology, it seems that the synchronic function of OJ *-si* was simply to derive adjectives from other parts of speech, without any semantic specification. It may be viewed as having arisen in the course of the establishment of adjectives as an independent part of speech.

Identifying *-(V)si* as the carrier of the psychological meaning of the *shiku* adjectives is problematic. First, there is the not insignificant number of exceptions (26%). Second, the psychological meaning may in many cases be seen to reside in the lexical semantics of the verbal stem. Third, there may well have been a tendency for adjectives derived by means of *-si* to specialize semantically without that being a feature of the meaning of *-si*: adnominalizing

or adverbalizing verbs or nouns was morphologically unproblematical and a derived adjective would therefore only be derived if some special aspect of or perspective on the semantics of the stem was required.

(76) *Ku* adjectives

a. ***aka***- 'bright, red'; ***ara***- 'rough'; ***asa***- 'shallow'; ***kata***- 'hard, firm'; ***kurwo***- 'black'; ***mane***- 'frequent'; ***na***- 'non-existent, no'; ***naga***- 'long'; ***nuru***- 'tepid'; ***opo***- 'many, plentiful'; ***paya***- 'fast'; ***puru***- 'old'; ***putwo***- 'thick, sturdy'; ***siru***- 'obvious, as may be expected'; ***sirwo***- 'white'; ***tika***- 'near'; ***topo***- 'distant'; ***usu***- 'thin, weak'; ***waka***- 'young'; ***yasu***- 'peaceful, easy'; ***yo***- 'good'

b. ***ita***- 'painful'; ***kasikwo***- 'fearsome, awesome'; ***niku***- 'disagreeable'; ***tayu***- 'exhausted'; ***u***- 'sad'

Shiku adjectives

c. ***asi***- 'bad, evil'; ***atarasi***- 'precious, regrettable, dear' (*atar*- 'touch'); ***kanasi***- 'dear, sad' (*kane*- 'be unable to'); ***kokidasi***- 'grave, serious'; ***kuyasi***- 'regrettable, vexing' (*ku(y)i*- 'regret'); ***kwopwisi-/ kwoposi-*** 'dear, beloved' (*kwopwi*- 'love'); ***kyesi***- 'strange, unusual'; ***natukasi***- 'dear, yearned for' (*natuk*- 'become familiar with, be fond of'); ***opoposi***- 'dim, gloomy'; ***pasi***- 'beloved'; ***posi***- 'desirable' (cf. *por*- 'want', see 3.2.2.4.2); ***sabusi-/sabisi-*** 'sad, lonely' (*sabwi*- 'get desolate'); ***tagitagisi-*** 'uneven'; ***tomosi***- 'sparse, enviable, poor'; ***uramyesi***- 'regrettable' (*urami-[10] 'regret, resent'); ***uresi***- 'joyous'; ***urupasi***- 'beautiful' (*urup*- 'get wet'); ***wabwisi***- 'lonely' (*wabwi*- 'be embarrassed'); ***wemapasi***- 'likeable, smile-provoking' (*wemap*- 'keep smiling' <= *wem*- 'smile'); ***wosi***- 'dear'; ***yorosi-*** 'suitable' (*yor*- 'approach'); ***yuyusi***- 'awesome'

d. ***ikasi***- 'plentiful, vigorous'; ***iyasi***- 'lowly'; ***munasi-*** 'empty'; ***pisasi***- 'long-lasting'; ***sagasi***- 'steep'

3.2.4.1 Reduplication

A final point about *shiku* adjectives: traditionally, reduplicated adjectives are said to belong to *shiku*, cf. *tagitagisi*- (*tagi- '?') 'uneven' and perhaps *yuyusi*- (*yu- '?') 'awesome'; this is certainly the case from EMJ onwards. While this was in all likelihood also the case in OJ, it should be noted that OJ reduplicated

[10] *urami*- is not phonographically attested in OJ. This verb was UB in EMJ, but may have been *ura-mi*- (UM) 'heart-see' in OJ.

adjectives with identifiable stems are only attested in the conclusive, making it impossible to determine whether they really are *shiku* adjectives, e.g. *naga-nagasi* 'very long' (*naga-* 'long'; EMJ *naganagasi-*), *topotoposi* 'very far, far far away' (*topo-* 'distant'; EMJ *topodoposi-*), *wowosi* 'gallant' (*wo* 'male'; EMJ *wowosi-*).

3.2.4.2 Jiku *adjectives*

The *jiku* adjectives are usually regarded as a subclass of *shiku* adjectives. It is a very small closed class. (77) is a full list of *jiku* adjectives. With the exception of *masizi-*, the OJ *jiku* adjectives are fairly transparent denominal derivatives with a formant *-zi* and have the meaning 'like NOUN, typical of NOUN'. It may be noted that *kono ipyeziku mo* 'also like this house' (*SM* 25, the only attestation of *ipyezi-*) has the noun phrase *kono ipye* as the derivational stem for *-zi*.

(77) ***ipyezi-*** 'like (this) house' (*ipye* 'house'); ***masizi-*** 'negative potential'; ***omozi-*** 'close (as if from same mother)' (*omo* 'mother'); ***onazi-*** ~ ***oyazi-*** 'same, similar' (cf. *ono* 'self'); ***tokizi-*** 'timeless, eternal, constant; untimely' (cf. *toki* 'time'); ***warezi-*** 'like me' (*ware* 'I').
From EMJ: ***imizi-*** and ***susamazi-*** (~ ***susamasi-***)

Note in this connection the OJ construction NOUN-*zi mono* 'like (a) NOUN, typical of (a) NOUN, as befits (a) NOUN' (e.g. *uma-zi mono* 'like a horse'), built on animate or concrete nouns (in addition to a single example with the stem of a *ku*-adjective: *kasikwo-zi mono* (*SM* 14) 'reverently' (*kasikwo-* 'fearsome, awesome')). The construction is attested with the following nouns: *inu* 'dog', *i* 'cormorant', *uma* 'horse', *kakwo* 'fawn', *kamwo* 'duck', *sisi* 'wild animal', *toko* 'bed', *tori* 'bird', *wotokwo* 'man', *yuki* 'snow'.

The *jiku* adjectives and this construction involve a derivational suffix *-zi* < *-nVsi, see 3.5.2. Further, *masizi* and *tokizi* 'untimely, timeless' are usually said to involve a different derivational *-zi* which is negative in meaning and possibly related to the negative auxiliary *-(a)zu*, or the negative conjectural flective *-(a)zi*.

3.3 Copula

OJ had two simple copulas with highly defective inflection. The two copulas are usually not thought of as verbs, but functionally they exhibit the forms in Table 3.9. Morphologically, infinitive *ni*, gerund *nite* and adnominal *tu* look like regular verb forms, whereas *no* and *to* appear to be roots used in adnominal and adverbial function.

Table 3.9 *OJ copula forms*

	OJ		EMJ	
Adnominal	*no*	*(tu)*	*no*	
Infinitive	*ni ~ no*	*to*	*ni*	*to*
Gerund	*nite*		*nite > de*	

The basic function of the copula in Japanese is to predicate, adnominalize or adverbialize nouns. As we saw above (3.2) that also, to a limited extent, included adjective stems in OJ. As the simple inflected copulas were highly defective, analytic constructions with a copula infinitive or gerund followed by an existential verb were used in many functions (3.3.2). All the copula forms in Table 3.9 were also used as case and other particles, see 3.7.8.2.1. With the exception of the loss of *tu*, and the change of *nite* > *de*, they are all still in use today in NJ both as basic copula forms and as case particles, constituting a remarkably stable pervasive feature of Japanese through its attested history, with little change in form or function.

3.3.1 Use of the simple inflected copula forms

Of the two adnominal forms, *tu* was already in OJ restricted and used only to adnominalize nouns or adjective stems directly, and it disappeared in EMJ outside lexicalized expressions; it is usually described as a genitive, cf. 3.7.1. As in NJ, *no* was highly frequent in OJ and was in addition to the marginal use with adjectives (see (38c) above) used in adnominal clauses as in (78). As distinct from the adnominal forms of verbs and auxiliaries, *no* could not be used as a nominalized form and it was not used as the predicate in *kakari-musubi* constructions (see 8.9); for those purposes, analytic forms were used (see 3.3.2).

(78) purupye *no* satwo ***no*** asuka
old.house GEN village COP.ADN Asuka
'Asuka which is the village of my old home' (*MYS* 3.268)

The non-finite copula forms were used in subordinate function 'being, as, like', (79); infinitive *no* is rare and only used in this sense. Infinitive *ni* and *to* could also be used simply to adverbialize, see (80), and in small clause constructions,

see (81), similar to the accusative-with-infinitive construction in some other languages.

(79) a. namida kwosame ***ni*** purite
tear small.rain COP.INF fall.GER
'tears fall like fine rain' (*MYS* 2.230)

b. pana no . . . yuki ***to*** *puri-kye-mu*
flower GEN snow COP.INF fall-SPST-CONJ.ADN
'flowers would fall like snow' (*MYS* 17.3906)

c. *asa-pi* ***no*** *wemi-sakaye-kite*
morning-sun COP.INF smile-prosper-come.GER
'come smiling radiantly as the morning sun' (*KK* 3)

(80) wotomye*ra* no kazasi *no* *tame* ***ni***
ladies GEN hairpin GEN sake COP.INF
miyabwiwo no kadura no *tame* ***to***
elegant.fellow GEN headdress GEN sake COP.INF
'for the sake of the hairpins of the ladies,
for the sake of the headdresses of the elegant fellows' (*MYS* 8.1429)

(81) a. terasu pi ***wo*** yami ***ni*** minasite
shine sun ACC darkness COP.INF see/regard.GER
'seeing the shining sun as darkness' (*MYS* 4.690)

b. kagami no yama ***wo*** miya ***to*** sadamu*ru*
Kagami GEN mountain ACC residence COP.INF decide.ADN
'decide on Mt. Kagami as his residence' (*MYS* 3.417)

3.3.2 Analytic forms

Supplementing the few simple inflected copula forms were analytic forms consisting of an infinitive (*ni*, *to*) followed by *ar*-, e.g. (82a–b), or another existential verb, e.g. (82c), sometimes separated by a particle, e.g. (82d). Analytic forms were the only way of forming most inflected forms for the copula and for combination with tense and mood auxiliaries and with extensions, and the analytic formations were thus an indispensable part of the morphology and syntax of the copula. *Ni ar*- sometimes phonologically fused to give *nar*-, see (82e), which in EMJ became a fully inflected copula (see further 3.4.2.1.1–2):

(82) a. *ima* *koso* *pa* *wa-dori* ***ni*** ***ara-me***
now FOC TOP me-bird COP.INF exist-CONJ.EXCL
'Now, I am my own' (*KK* 3)

b. yononaka pa munasi-ki mono ***to***
this.world TOP empty-ACOP.ADN thing COP.INF
ara-mu
exist-CONJ.CONCL
'this world is empty' (*MYS* 3.442)

c. *na* *koso* *pa* *wo* ***ni*** ***imaseba***
you FOC TOP man COP.INF exist.RESP.PROV
'as you are a man' (*KK* 5)

d. inisipye no nana no sakasi-ki pitodomo
past GEN seven GEN wise-ACOP.ADN people
mo pori-se-si mono pa
FOC want-do-SPST.ADN thing TOP
sake ***ni*** *si* **aru** *rasi*
saké COP EMPH exist.ADN PRES
'It seems that what the seven wise men of old, too, wanted is saké' (*MYS* 3.340)

e. *ume* *no* *pana* *ima* *sakari* ***nari***
plum GEN blossom now in.bloom COP.CONCL
'The plum blossom is in bloom now' (*MYS* 5.850)

3.4 Conjugation classes and morphophonology

OJ has eight verbal conjugation classes. Although not usually included as verbal conjugation classes, the simple past auxiliary (3.1.4.2), the negative auxiliary (3.1.4.2), and the adjectival copula (3.2) have irregular conjugations with forms for most of the categories which verbs inflect for. Further, as mentioned above, the copulas (3.3) may also be considered verbs with highly defective and irregular conjugations, and finally, the infrequent auxiliary verb *-kose-* (3.4.3.3) has a defective and slightly irregular conjugation.

From a synchronic morphophonological point of view, the eight OJ verbal conjugation classes fall in two major groups, *consonant* base and *vowel* base verbs, each with regular and irregular subclasses. The names for the verb classes are explained in 3.4.6.2. There are two major lexical verbal classes: *quadrigrade* (abbreviated QD), which holds approximately 75 per cent of all OJ verbs, and *lower bigrade* (LB), which accounts for approximately 20 per cent of the verbs. The remaining classes have small sets of members. There are around thirty *upper bigrade* (UB) verbs and fewer than a dozen *upper mono-*

grade (UM) verbs. *n*-irregular (*n*-irr) holds two lexical verbs and an auxiliary. *r*-irregular (*r*-irr), *k*-irregular (*k*-irr), and *s*-irregular (*s*-irr) are each defined by a single idiosyncratic and highly frequent verb with some amount of grammatical functions: *ar-* 'be, exist', *ko-* 'come', and *se-* 'do', respectively.

(83)

Consonant base			
Regular	Quadrigrade	***75%*** of OJ verbs	*kak-*, *or-*, . . .
Irregular	*r*-irregular		*ar-*
	n-irregular		*sin-*, *in-*, *-(i)n-*
Vowel base			
Regular	Lower bigrade	***20%*** of OJ verbs	*ake-*, *ware-*, . . .
	Upper bigrade	*c.* 30 verbs	*okwi-*, *ori-*, . . .
Irregular	Upper monograde	*c.* 10 verbs	*mi-*, *ni-*, . . .
	k-irregular		*ko-*
	s-irregular		*se-*

3.4.1 Regular verb classes

3.4.1.1 Quadrigrade verbs

QD is the class of regular consonant base verbs and the largest verb class. The base ends in a consonant and most have the shape CVC-. Those which are longer, CVCVC-, usually incorporate more or less transparent derivational matter. The following base final consonants are found: *-p*, *-t*, *-k*, *-b*, *-g*, *-m*, *-s*, *-r*. Thus, there are no regular consonant base verbs which end in *-d*, *-n*, *-z*, *-y*, *-w*.

3.4.1.2 Bigrade verbs

LB is the large regular class of vowel base verbs, with bases ending in *-e* (that is, $-e_2$, as opposed to $-e_1$ after consonants where the distinction existed). Most LB verb bases are disyllabic, *CVCe-*, but there are a few monosyllabic, *Ce-* (e.g. *e-* 'get', *pe-* 'pass', *ne-* 'sleep'), and some trisyllabic *CVCVCe-* bases as well.

UB verbs end in *-wi* (neutralized as *-i* after alveolars). They number only about thirty. Most are disyllabic, *CVCwi-*, but a few monosyllabic, *Cwi-*, and trisyllabic, *CVCVCwi-*, bases exist. The few *CVi-* bases (*koi-* 'lie down', *kui-* 'regret', *oi-* 'age, get old') are underlyingly //CVyi-// as shown both by derived forms such as *kuyasi-* 'regrettable' and by their EMJ conclusive forms (they are only attested in their base in OJ). (84) is a reasonably full list of UB verbs in OJ and MJ. Note that some verbs are attested only in forms that do not allow an unambiguous identification of their conjugational class in OJ. For example, for *osori-* 'fear' only the base (/infinitive) is attested in OJ, *osori*, and conversely *porobwi-* 'go to ruin' is only attested in the conclusive, *porobu*;

both of these verbs could therefore possibly belong to QD in OJ. Where the attested OJ forms of a verb do not allow us positively to identify the conjugational class, we here project the EMJ class back to OJ, but note relevant facts of attestation.

(84) ***koi-*** 'lie down' [= //koyi//; only base attested in OJ]
komwi- 'enclose, put in' [only attested once in the compound *tumagomwi* 'wife-enclosing'; usually LB *kome-*]
kozi- 'uproot' [only base attested in OJ]
kui- 'regret' [= //kuyi//]
kwopwi- 'love'
nagwi- 'become calm'
negwi- 'appease, solace, pray, be sympathetic with' [only base attested in OJ]
odi- 'fear' [only conclusive attested in OJ]
oi- 'age, get old' [= //oyi//; only base attested in OJ]
okwi- 'arise'
opwi- 'grow/get bigger'
osori- 'fear, dread, be apprehensive' [only base attested in OJ]
oti- 'fall'
pwi- 'winnow' [only one OJ example, in the compound *pwi-gapa* (place name)]
porobwi- 'go to ruin' [only conclusive attested in OJ]
puri- 'get old' [only base attested in OJ]
pwi- 'dry (intr.)'
pwi- 'sneeze'
sabwi- 'get desolate, old, rusty, faded'
sakwi- '(?)be torn' [rare; only base attested in OJ; usually LB *sake-* 'be torn']
sinobwi- 'endure, suffer, hide'
sipwi- 'insist, force'
sugwi- 'pass by, exceed, elapse, pass away'
susabwi- 'become strong' [from EMJ also *susab-* and *susam-*, both QD]
tamwi- 'go round, turn'
todomwi- 'stop (tr.)' [also LB *todome-* which is more common]
tukwi- 'get exhausted'
wabwi- 'be embarrassed, disappointed, apologize for'
wi- 'sit down, settle down' [*wi-*, which is the ancestor of NJ *i-* 'exist', is often treated as an existential verb also in OJ, but it was not used as a simple existential verb until late LMJ, see 12.4].
woti- 'be rejuvenated' [only base attested in OJ]

yodi- 'grab and pull' [only base attested in OJ]
yokwi- 'avoid, go around'.

. . . ***-bwi-*** 'be like . . .'; derives verbs from nouns or adjectives, e.g.: *arabwi-* 'be wild' (*ara-* 'wild'), *manabwi-* 'imitate; learn' ~ *manebwi-* 'imitate' (cf. *mane* 'imitation').
. . . ***-sabwi-*** 'act as/display characteristics appropriate for . . .'; derives verbs from nouns, e.g.: *kamusabwi-* 'act godly, be old' (*kamu-/kamwi* 'spirit, god'), *yamasabwi-* 'be mountain-like' (*yama* 'mountain').

Additional UB verbs (attested in EMJ unless noted):

(85) ***abi-/ami-*** 'bathe oneself in'
iki- 'live' [from LMJ; QD in OJ and EMJ]
isati- 'weep' [attested as UM in OJ]
kabi- 'turn mouldy'
kobi- 'flatter'
koi- 'freeze' [= //koyi//]
kori- 'feel regret/remorse for, learn by experience'
kuti- 'rot'
miti- 'become full' [from LMJ; QD in OJ and EMJ]
mukui- 'requite' [= //mukuyi//]
nebi- 'get, seem old'
nedi- 'twist, screw'
nobi- 'extend, lengthen, spread, grow'
obi- 'tie around the waist' [QD in OJ]
ori- 'descend, alight'
padi- 'feel shame'
potobi- 'swell with water'
simi- 'freeze'
tibi- 'get worn out' [not attested until LMJ]
todi- 'shut'
urami- 'resent, regret'

3.4.1.3 Formation of basic inflected forms

There are five basic inflected verb forms (see Table 3.10). Their formation on the regular verbs can synchronically be described in terms of combination of the verb base with a flective, with reference to simple phonological rules, in particular the regular rules of vowel deletion which apply when two vowels are brought together in composition or morphological derivation, see 2.6.1. We will briefly discuss alternative analyses which have been proposed.

Table 3.10 *Formation of OJ inflected verb forms*

QD

Base	***kak-***	***war-*** 'break (tr.)'
Infinitive	*kaki* <= *kak* + *i*	*wari* <= *war* + *i*
Imperative	*kakye* <= *kak* + *ye*	*ware* <= *war* + *ye*
Conclusive	*kaku* <= *kak* + *u*	*waru* <= *war* + *u*
Adnominal	*kaku* <= *kak* + *ru*	*waru* <= *war* + *ru*
Exclamatory	*kake* <= *kak* + *re*	*ware* <= *war* + *re*

LB

Base	***ake-***	***ware-*** 'break (intr.)	***e-*** 'get'
Infinitive	*ake* + Ø	*ware* + Ø	*e* + Ø
Imperative	*ake* (+ *yo*)	*ware* (+ *yo*)	*e* + *yo*
Conclusive	*aku* <= *ake* + *u*	*waru* <= *ware* + *u*	*u* <= *e* + *u*
Adnominal	*akuru* <= *aku* + *ru*	*waruru* <= *waru* + *ru*	*uru* <= *u* + *ru*
Exclamatory	*akure* <= *aku* + *re*	*warure* <= *waru* + *re*	*ure* <= *u* + *re*

UB

Base	***okwi-***	***ori-*** 'descend'	***pwi-*** 'dry'
Infinitive	*okwi* + Ø	*ori* + Ø	*pwi* + Ø
Imperative	*okwi* + *yo*	*ori* + *yo*	*pwi* + *yo*
Conclusive	*oku* <= **okwu* <= *okwi* + *u*	*oru* <= *ori* + *u*	*pu* <= **pwu* <= *pwi* + *u*
Adnominal	*okuru* <= *oku* + *ru*	*oruru* <= *oru* + *ru*	*puru* <= *pu* + *ru*
Exclamatory	*okure* <= *oku* + *re*	*orure* <= *oru* + *re*	*pure* <= *pu* + *re*

The conclusive, adnominal and exclamatory endings are common to the two conjugational classes. The following regular phonological rules apply in the formation of these three forms. The important point about the vowel deletion rules is that when two vowels come together, the second is only deleted in the context -(C)V-VCV-, i.e. when a monosyllabic morpheme is followed by a disyllabic vowel initial morpheme; in all other cases, it is the first vowel which is deleted (cf. 2.6.1):

(86) Vowel deletion

ake + *u* => *aku*

okwi + *u* => *okwu => *oku*

e + *u* => *u*

Consonant deletion

kak + *ru* => *kaku*

kak + *re* => *kake*

The vowel base verbs attach the adnominal and exclamatory endings to the conclusive, rather than to the base. The conclusive is thus used as a stem in addition to its use as an inflected form. An alternative analysis would posit the adnominal and exclamatory formants as *-uru* and *-ure*, respectively, for the vowel base verbs. This would mean, first, that the flectives are different for the consonant and vowel stem verbs. Second, and more importantly, specific morphophonemic rules would be required to generate the correct forms, for as shown in (87b), the application of the regular phonological rules would generate incorrect forms for the monosyllabic vowel bases. The proposed analysis gives the correct forms in all cases, (87a):

(87)		a. Conclusive + *-ru (/-re)*	b. Base + *-uru (/-ure)*
	UB *CVCwi*	*oku* + *ru* => *okuru* 'rise'	*okwi* + *uru* => (*okwuru* =>) *okuru*
	LB *CVCe*	*aku* + *ru* => *akuru* 'open'	*ake* + *uru* => *akuru*
	UB *Cwi*	*pu* + *ru* => *puru* 'dry (intr.)'	*pwi* + *uru* => *pwiru
	LB *Ce*	*pu* + *ru* => *puru* 'pass, elapse'	*pe* + *uru* => *peru

The consonant base verbs form their infinitive and the imperative by means of flectives, *-i* and *-ye*, respectively, whereas the vowel base verbs simply use the base as infinitive and, originally, imperative (see further 3.4.5). The infinitive formant is sometimes said to be *-i* also for the vowel base verbs. However, that, again, would mean that otherwise unmotivated morphophonemic rules would have to be posited, for the regular phonological rules give incorrect forms for the LB verbs, (88b):

(88)		a. Base = infinitive	b. Base + *-i*
	UB *CVCwi*	*okwi*	*okwi* + *i* => *okwi*
	UB *Cwi*	*pwi*	*pwi* + *i* => *pwi*
	LB *Ce*	*pe*	*pe* + *i* => *pi
	LB *CVCe*	*ake*	*ake* + *i* => *aki

3.4.2 Irregular consonant base verbs

3.4.2.1 r-*irregular*

r-irr is the class of irregular *-r* base verbs. It is defined by the highly frequent existential verb *ar-* and in addition also holds a number of frequent and important grammatical morphemes which are all thought to incorporate *ar-* etymologically. *r*-irr also includes the verb *wor-* 'be sitting'. *Wor-* is the ancestor of the NJ existential verb *or-* and usually *wor-* is said to have been an existential verb also in OJ. However, it has been shown by Kinsui (2006) that *wor-* was not used as a simple existential verb until LMJ. In fact,

wor- was used as a lexicalized stative form of (an ancestor of) *wi-* 'sit/settle down', in OJ and EMJ, meaning 'be sitting' and for example as the antonym of *tat-er-* 'be standing' (the morphological stative of *tat-* 'stand up'). Note also that while *wor-* had grammatical uses from OJ – as an auxiliary verb to form progressives (cf. 3.1.4.7.4) – it was not used in the grammatical functions exhibited by truly existential verbs until late LMJ. See further 12.4 about *wor-*. *r*-irr also comprises two existential verbs *pab(y)er-* and *imasugar-* which are both thought to involve developments of something + *ar-*, although it is not clear what that something might be; neither is phonographically attested in OJ.

(89) *r*-irr verbs
ar- 'be; exist'
wor- 'be sitting'

From EMJ:
pab(y)er- ~ ***panb(y)er-*** 'exist.humble; exist.polite' (no phonographic OJ attestation, but is found in the reading tradition of 侍 which otherwise represents words meaning 'serve').
imasugar- ~ ***imasukar-*** 'exist.RESP' (variant shapes: *imasogar-*, *imasokar-*, *imasigar-*, *imasikar-*, *mimasugar-*, *mimasukar-*).

Other members of the *r*-irr class are thought to incorporate *ar-*, in most cases conspicuously so. The forms in (90) are fully lexicalized grammatical morphemes, etymologically incorporating *ar-*. The forms in (91) are phonologically fused, but not yet lexicalized forms (some of which are not attested until EMJ in their fused form); the forms in (91a) and (91b) became important grammatical morphemes in EMJ, see 3.4.2.1.2 immediately below.

(90) Lexicalized grammatical morphemes
Auxiliaries
-kyer- modal past < *ki-ar- 'come-exist' or 'SPST-exist'
-yer- stative < *-i-ar- 'INF-exist'

Evidential verb extensions
nar- 'sounds' < *na-ar- 'sound-exist'
From EMJ:
mer- [not robustly attested until EMJ] 'looks' < *me-ar- 'eye-exist' or *mi-ar- 'see-exist'

(91) Phonologically fused forms
a. Periphrastic stative (gerund + *ar-*)
-tar- <= *-te ar-*

b. Infinitive + *ar-*
nar- <= *ni ar-* 'COP.INF + exist'
tar- [not until EMJ] <= *to ar-* 'COP.INF + exist'
kar- <= *ku ar-* 'ACOP.INF + exist'
be-kar- <= *be-ku ar-* 'NEC-ACOP.INF + exist'
zar- <= *zu ar-* 'NEG.INF + exist'

c. Adverb + *ar-*
kakar- 'be this way' <= *kaku ar-* 'this.way + exist'
sikar- 'be that way' <= *sika ar-* 'that.way + exist'
sar- 'be that way' [not until EMJ] <= *sa ar-* 'that.way + exist'

r-irr verbs can be said to be syntactically rather than morphologically irregular: they use the infinitive as conclusive form. In other respects they are like the QD verbs. Note that the verb extensions *be-*, *rasi-*, and *ram-*, which regularly attach to the conclusive, attach to the adnominal of *r*-irr verbs: *aru be-*, *aru rasi-*, *aru ram-*. Another way of describing this would be to say that *r*-irr verbs have two conclusive forms: regularly formed *aru* (<= *ar+u*) used only with verb extensions, and *ari* (= infinitive) used elsewhere.

(92)

	ar-
Infinitive	*ari* <= *ar* + *i*
Imperative	*are* <= *ar* + *ye*
Conclusive	*ari* <= *ar* + *i*
Adnominal	*aru* <= *ar* + *ru*
Exclamatory	*are* <= *ar* + *re*

3.4.2.1.1 Grammatical uses of *ar-* and other existential verbs In addition to its lexical meaning, the existential verb *ar-* has been used extensively in grammatical functions throughout the history of Japanese. Thus, some important grammatical morphemes which were lexicalized already in OJ etymologically incorporate *ar-*, see (90) above. However, productively *ar-* – and other existential verbs, in particular the exalted synonyms of *ar-* (e.g. in OJ respectful *mas-*, *imas-*) – has been used to form extended, analytic or periphrastic forms through all attested stages of Japanese (reflected in (91) above), and these uses of existential verbs are a pervasive feature of Japanese. There are two overall uses: First, reflecting its semantics as a stative verb, *ar-* formed stative verb forms, both in the morphological stative (*-yer-* < *-i ar-) and in the periphrastic stative construction (*-te ar-*), see 3.1.4.7.3. Second, used with infinitives of the adjective auxiliary (*-ku*), negative auxiliary (*-zu*) or the defective copula verbs (*ni*, *to*), *ar-* functioned to provide analytic, morphologically versatile forms; such extension was in itself semantically vacuous (and seems

to have been applied playfully in the poetic texts on occasion), but for adjectives and the negative it provided freer combination with auxiliaries (see 3.1.4.8 and 3.2.2.4.1), and for the copula it was the only way of forming most inflected forms for use in nominal predications (see 3.3.2 above). Such constructions also made it possible to focus an adjectival, negative, or nominal predicate with a focus particle after the infinitive (cf. 8.9). (93) exemplifies a focused negative (*mi-zu so aru*) and also has the combination of an adjective (here the extension *be-*) with a tense auxiliary (*be-ku ari-kyeru*):

(93)	*imo*	*wo-ba*	*mi-zu*	*so*	***aru***	*be-ku*
	beloved	ACC-TOP	see-NEG.INF	FOC	exist	NEC-ACOP.INF
	ari-kyeru					
	exist-MPST.ADN					

'I should never have met my beloved' (*MYS* 15.3739)

3.4.2.1.2 Fused forms; secondary conjugations Periphrastic and analytic forms with *ar-* sometimes fused as summarized in (94):

(94) a.	Periphrastic stative	*-te ar-* =>	*-tar-*
b.	Copula	*ni ar-* =>	*nar-*
	Negative	*-zu ar-* =>	*-zar-*
	Adjective	*-ku ar-* =>	*-kar-*

In EMJ these phonologically fused forms were lexicalized to form a new stative auxiliary, *-(i)tar-* (see 8.4.2), and secondary conjugations of the copula, adjective, and negative auxiliary (see 8.2.1), but in OJ they were simply fused forms produced according to the rules of vowel deletion (cf. 2.6.1). The fused forms are concentrated in the younger sections of the *Man'yōshū* (and in the *Bussokuseki-ka*), but are not found in the early OJ materials,[11] and the unabbreviated forms were much more frequent than the fused forms. The distribution of fused and full forms in the OJ texts is sometimes taken to indicate that the phonological fusion itself was of fairly recent origin. However, this seems rather to be a matter of writing: contrast the following examples, with *taputwo-* 'awe-inspiring' from the *Kojiki* and *Nihon shoki*, (95) below, and from the *Bussokuseki-ka*, (96) below. Although the *Kojiki* and *Nihon shoki* form is unabbreviated, the line is metrically irregular in having eight rather than seven feet, indicating that *ta.pu.two.ku.a.ri.kye.ri* actually is a morphophonemic writing of *taputwokarikyeri*, the form phonemically written in the *Bussokuseki*

[11] Thus, *-zar-* and *-kar-* are not found in the *Kojiki*, *Nihon shoki*, *Norito*, or *Senmyō*, and *-tar-*, which is attested less than fifty times in the entire OJ corpus, not in the *Kojiki* or *Nihon shoki*, only once in the *Norito*, and three times in the *Senmyō*.

-ka. Most, although far from all, examples of unabbreviated *-ku ar-* occur in lines with an extra written foot (so-called *ji-amari* 'surplus letters'), which we expect to represent a fused pronunciation. This suggests that phonological fusion was more usual than the simple ratio between written fused and full forms would otherwise lead us to believe. More importantly, unabbreviated writing of a fused pronunciation, as in (95), testifies clearly to awareness of the morphemic constituency and to the lack of lexicalization.

(95) *kimi ga yosopi si taputwo-**ku** **ari-kyeri***
my.lord GEN attire EMPH admirable-ACOP.INF exist-MPST.CONCL
'your attire, it is admirable, my lord!' (*KK* 7, *NSK* 6)

(96) *ima no kusurisi taputwo-**kari**-kyeri*
now COP.ADN master.of.medicine admirable-ACOP-MPST.CONCL
'the present master of medicine is worthy of praise!' (*Bussoku* 15)

3.4.2.2 n-*irregular*

n-irr is the class of irregular *n*-base verbs; it has only three members, the perfective auxiliary *-(i)n-* and two lexical verbs *sin-* 'die' and *in-* 'depart, pass'. It is likely that this conjugation class is constituted and maintained as an independent class by the very frequent auxiliary, rather than by the two lexical verbs. Note that there are no regular *n*-base verbs, that is to say, no *n*-bases among the QD verbs and this, too, may have contributed to maintaining the morphological independence of this class. Traditional etymology posits a relation between the two lexical verbs (< *(s)in(V)-) and also derives the perfective auxiliary from *in-*. It is more likely, however, that the perfective *-n-* derives from a copula *nV, see 3.5.2. *n*-irr verbs have the infinitive and imperative of the consonant base verbs, but the adnominal and exclamatory are formed using the conclusive as stem, like the regular vowel base verbs. For that reason *n*-irr is often referred to as a 'hybrid' conjugation. Perfective *-n-* has no imperative. No imperative or nominal form is attested for *in-*; for *sin-*, no nominal form is found and the adnominal and exclamatory are not attested phonographically. Neither *sin-* nor *in-* forms a perfective or stative in OJ.

(97)

	sin-	***-n-*** perfective auxiliary
Infinitive	*sini* <= *sin* + *i*	*-ni* <= *-n* + *i*
Imperative	*sine* <= *sin* + *ye*	–
Conclusive	*sinu* <= *sin* + *u*	*-nu* <= *-n* + *u*
Adnominal	*(sinure* <= *sinu* + *re)*	*-nure* <= *-nu* + *re*
Exclamatory	*(sinuru* <= *sinu* + *ru)*	*-nuru* <= *-nu* + *ru*

3.4.3 Irregular vowel base verbs

There are three classes of irregular vowel base verbs, all of which are monosyllabic, with bases of the shape CV-.

3.4.3.1 Upper monograde

UM is the class of short (one-mora) verb bases of the shape *Ci*, which is thought originally to have been Ci_1. However, there are only two UM verbs which are identifiable as Ci_1 as opposed to neutral *Ci* (*mi-* 'see, look' and *ki-* 'wear'), and there is a single verb with Ci_2, *mwi-* 'turn', which is thought to have been UB originally, partly because it has Ci_2, partly on the basis of the existence of the UB verb *tamwi-* 'go round, turn'. Also *wi-* 'sit down, settle down' was UB originally, but had all but completed the shift to UM by the beginning of OJ (there are a few examples of conclusive *u* = //wu//). Thus already in OJ a migration of one-mora UB verbs to UM was under way; this was completed in EMJ. UM comprises only the following simple verbs (in addition there are a few compounds in *-mi-* 'see', *-wi-* 'bring along').[12]

(98) ***i-*** 'cast (metal)'
i- 'shoot'
ki- 'put on'
mi- 'see, look'
mwi- 'turn' (originally UB)
ni- 'resemble'
ni- 'boil'
wi- 'lead, bring along'
wi- 'sit down, settle down' (originally UB)

Additional UM verbs from EMJ
i- 'pour' [no OJ attestation]
moti-wi- 'use' [no OJ attestation]
pi- 'dry out' [UB in OJ: *pwi-*]
pi- 'sneeze' [UB in OJ: *pwi-*]
pi- 'winnow' [UB in OJ: *pwi-*]

UM is usually not considered an irregular class, mainly because of its position in later stages of Japanese when eventually all *-i-* base verbs became UM, but

[12] UB *arabwi-* 'be wild' is in a few instances found with the adnominal *arabwiru* of the UM paradigm (*SM* 62, *EN* 12) as opposed to the regular UB adnominal *araburu* (e.g. *EN* 27); EMJ UB *isati-* 'weep' is in the *Kojiki* given as *isatiru*. Both may be taken as very early, idiosyncratic instances, which did not gain currency, of the later general shift of UB verbs to UM (see 15.1.3). *Arabwi-* is *ara-* 'wild' + *-bwi-* and may have shifted for some speakers to UM in analogy with the shift of UB *wi-* 'sit down' to UM.

it is irregular vis-à-vis the regular vowel bases in the selection and attachment of inflectional endings and in the syntactic use of conjugational forms. There is no attestation in OJ of a UM verb in simple conclusive function. Furthermore, as opposed to other vowel base verbs, this class uses the base as the stem for all formations, including the adnominal and exclamatory forms, and the modal verb extensions *be-*, *rasi-*, *ram-* and the conjunctional particle *tomo*, which otherwise regularly attach to the conclusive, attach to the base of UM verbs, *mi-be-*, *mi-ram-*, *ni-rasi-*, *mi-tomo*. All this has been taken to mean that at an earlier stage the base of UM verbs was used as the conclusive form, but use of the stem in conclusive function is not attested in OJ either. From EMJ onwards, UM verbs use the adnominal (*miru*) also as conclusive. We include the stem in brackets as conclusive in the paradigms for this class to reflect that this form was used with extensions, but not (attested) in conclusive function. In the imperative, *-yo* is generally used, but *mi-* 'see' occasionally uses the base as imperative.

(99)

	mi-	***ni-*** 'boil'	***mwi-*** 'turn'
Infinitive	*mi*	*ni*	*mwi*
Imperative	*mi* (+ *yo*)	*ni* + *yo*	*mwi* + *yo*
Conclusive	*(mi)*	*(ni)*	*(mwi)*
Adnominal	*mi* + *ru*	*ni* + *ru*	*mwi* + *ru*
Exclamatory	*mi* + *re*	*ni* + *re*	*mwi* + *re*

3.4.3.2 k-*irregular and* s-*irregular*

These two classes are defined by and consist of a single verb each, *ko-* 'come' and *se-* 'do'.

(100)

Base	***ko-***	***se-***
Infinitive	*ki* <= *ko* + *i*	*si* <= *se* + *i*
Imperative	*ko*	*se* (+ *yo*)
Conclusive	*ku* <= *ko* + *u*	*su* <= *se* + *u*
Adnominal	*kuru* <= *ku* + *ru*	*suru* <= *su* + *ru*
Exclamatory	*kure* <= *ku* + *re*	*sure* <= *su* + *re*

From EMJ, two additional verbs which were originally QD and for which some fluctuation was found throughout MJ migrated to *s*-irr: *imase-* (OJ QD *imas-*) 'be, come, go.RESP; respect auxiliary verb', *opase-* (no OJ attestation) 'be, come, go.RESP; respect auxiliary verb'. *Ko-* and *se-* are very similar in conjugation. They are like other vowel base verbs in using the base as imperative and, as with the bigrade and *n*-irr verbs, their adnominal and exclamatory

are formed on the conclusive. They are set off from other vowel stem verbs by having a morphological infinitive which is distinct from the base: *ko* + *i* => *ki*, *se* + *i* => *si*. However, in the prohibitive form, *se-* uses the base in the prohibitive: *na-se-so* 'don't do', and in some contexts where other verbs use the infinitive as a stem, they use the base, see 3.4.4.1.

3.4.3.2.1 Grammatical uses of *ko-* and *se-* *Se-* 'do' is not mainly used as a lexical verb. As in NJ, it functions as a *pro-verb*, standing in for other verbs, and following a verb infinitive *se-* functions as a supporting *light verb*, carrying finite predicational markers, as in (101) where the infinitive *omopi* 'think of with deep emotion' is focused by *so* (cf. 3.7.2) and *se-* completes the predication. This function of forming extended predicate structures is shared with *ar-*, which is used in a similar way with the regular and adjectival copulas and with the negative auxiliary (see 3.4.2.1.1).

(101) apa-nu pi *mane-mi* **omopi** *so* a ga
meet-NEG.ADN day many-ACOP.INF think.INF FOC I GEN
su*ru*
do.ADN
'not having been able to see you for many days, I do think deeply and fondly of you' (*MYS* 19.4198)

Ko- is used as an important and frequent *auxiliary verb* 'come to . . .' (3.1.2). Etymologically, it is likely that a number of grammatical morphemes are cognate with (the sources of) these two verbs, see 3.5.2.

3.4.3.3 -Kose-

The auxiliary verb *-kose-* is an irregularly and incompletely inflected vowel base: *kose-*. It was already at the OJ stage infrequent and limited in use, and in EMJ went out of use. The meaning of *-kose-* itself seems to have been 'do for me', but it is only found in fixed optative expressions, either in combination with modals or in the imperative, meaning 'please let there, I wish there'd (/you'd), won't you (please)'. It has the forms in (102). The conclusive is only used as in (103) followed by prohibitive *na yume* (prohibitive particle *na* and emphatic adverb *yume*); the base is used as a combinatory stem to form the optative (*uti-yame-kosene* 'I wish they'd stop') and with the negative *-kose-nu* (always followed by *kamo*, as in e.g. *arikosenu kamo* 'couldn't there be?; I wish there were'). What is interesting about this verb and what sets it off from a rare and incompletely inflected LB verb is its imperative *-koso* which seems to use the etymological root in that function. In EMJ a regular, analogically formed, imperative *-kose* is found in a few examples. Furthermore, the root of this verb may be directly reflected in the focus particle *koso*.

(102)

Base	*-kose-*
Conclusive	*kosu*
Imperative	*koso*
Optative	*kosene*
Negative	*kose-n-*

(103) a. *ware yukite kapyeri-kuru made tiri-kosu*
I go.GER return-come.ADN until fall-do.for.me.CONCL
na yume
PROHIBITIVE
'I wish they wouldn't fall until I have gone and come back' (*MYS* 15.3702)

b. *ume ga pana tira-zu ari-koso*
plum GEN blossom fall-NEG.INF exist-do.for.me.IMP
'I wish the plum blossom would remain without scattering' (*MYS* 5.845)

3.4.4 Extended inflectional forms; combinatory stems

In addition to the five basic inflected forms, the formation of almost all other inflectional forms can be described in terms of combination of a verb stem and an inflectional suffix (comprising both flectives and auxiliaries). Table 3.11 shows combinations of stems and inflectional suffixes, combining the forms in 3.1.3 and 3.1.4 above, sorted into three sets by verb stem. The majority of inflectional forms are formed directly on the base of regular vowel base verbs, but have either *-i-* or *-a-* after the base in the consonant base classes. That is to say, they are either built on the infinitive (3.4.4.1) or on a derived stem in *-a* (3.4.4.2). Two forms are built on the exclamatory (3.4.4.3). From EMJ, a number of forms were built on a new stem, the *onbin* stem (8.1.4).

3.4.4.1 The infinitive as stem

The forms under (a) in Table 3.11 are readily segmentable into infinitive plus an invariant (consonant initial) formant: *-te*, *-tutu*; *-te-*, *-n-*, *-ki*, *-kyer-*. This is shown most clearly by those verbs which have a morphological infinitive distinct from the base: the consonant base verbs and *s*-irr and *k*-irr. The only exception is the use of the base of *s*-irr and *k*-irr verbs with the *s-* initial forms of the simple past: *se-si* (adnominal), *se-sika* (exclamatory), *ko-si*, *ko-sika*. *Ko-* does not combine with the *k-* initial forms of the simple past, but *se-* has *si-ki* (conclusive). The suffixes which attach to the infinitive will be noted as in (104). This is a morphophonemic notation to show that these suffixes select the infinitive: /i/ is not part of the phonemic shape of the suffixes and

Table 3.11 *OJ forms ordered according to formation on* katsuyōkei

	QD	*r*-irr	*n*-irr	LB	UB	UM	*s*-irr	*k*-irr
Base	*kak-*	*ar-*	*sin-*	*ake-*	*okwi-*	*mi-*	*se-*	*ko-*
(a) Infinitive/base	*kaki*	*ari*	*sini*	*ake*	*okwi*	*mi*	*si*	*ki*
Gerund	*kakite*	*arite*	*sinite*	*akete*	*okwite*	*mite*	*site*	*kite*
Continuative	*kakitutu*	*aritutu*	*sinitutu*	*aketutu*	*okwitutu*	*mitutu*	*situtu*	*kitutu*
Perfective	*kakite-*	*arite-*	–	*akete-*	*(yodite-)*	*mite-*	*site-*	*kite-*
Perfective	*(sakin-)*	*arin-*	–	*(token-)*	*okwin-*	*min-*	*sin-*	*kin-*
Simple Past	*kakiki*	*ariki*	*siniki*	*akeki*	*okwiki*	*miki*	*siki/sesi*	*-/kosi*
Modal Past	*kakikyer-*	*arikyer-*	*sinikyer-*	*akekyer-*	*okwikyer-*	*mikyer-*	*sikyer-*	*kikyer-*
(b) *a*- stem/base	*kaka-*	*ara-*	*sina-*	*ake-*	*okwi-*	*mi-*	*se-*	*ko-*
Neg.Conj.	*kakazi*	*arazi*	*sinazi*	*akezi*	*okwizi*	*mizi*	*sezi*	*kozi*
Optative	*kakana*	*arana*	*sinana*	*akena*	*okwina*	*mina*	*sena*	*kozi*
Conditional	*kakaba*	*araba*	*sinaba*	*akeba*	*okwiba*	*miba*	*seba*	*koba*
Respect	*kakas-*	–	*sinas-*	–	–	*myes-*	*ses-*	–
Causative	*kakasime-*	*arasime-*	*sinasime-*	*akesime-*	*okwisime-*	*misime-*	*sesime-*	*kosime-*
Passive	*kakaye-*	*araye-*	*sinaye-*	–	–	*miye-*	–	–
Passive	*kakare-*	*arare-*	*sinare-*	–	–	–	–	–
Negative	*kakan-*	*aran-*	*sinan-*	*aken-*	*okwin-*	*min-*	*sen-*	*kon-*
Negative	*kakazu*	*arazu*	*sinazu*	*akezu*	*okwizu*	*mizu*	*sezu*	*kozu*
Conjectural	*kakam-*	*aram-*	*sinam-*	*akem-*	*okwim-*	*mim-*	*sem-*	*kom-*
Subjunctive	*kakamasi*	*aramasi*	*sinamasi*	*akemasi*	*okwimasi*	*mimasi*	*semasi*	*komasi*
(c) Exclamatory	*kake*	*are*	*sinure*	*akure*	*okure*	*mire*	*sure*	*kure*
Provisional	*kakeba*	*areba*	*sinureba*	*akureba*	*okureba*	*mireba*	*sureba*	*kureba*
Concessive	*kakedo*	*aredo*	*sinuredo*	*akuredo*	*okuredo*	*miredo*	*suredo*	*kuredo*

will not be included when we give inflectional paradigms of the auxiliaries, as for example in 3.1.4.2 above, and we will segment *saki-nu*: 'bloom-PERF. CONCL'.

(104)

Gerund	*-(i)te*
Continuative	*-(i)tutu*
Perfective	*-(i)te-*
Perfective	*-(i)n-*
Simple Past	*-(i)ki*
Modal Past	*-(i)kyer-*

An alternative analysis, however, does propose that the basic phonemic shapes of these formants have initial *-i-* (e.g. *-ite* or *-ikyer-*) and attach directly to the base. However, this would require otherwise unmotivated morphophonemic rules to derive the correct surface forms, for the regular phonological rules of vowel deletion (cf. 2.6.1) would give incorrect surface forms for the *s*-irr, *k*-irr, and polysyllabic LB verbs, see (105b), whereas (105a) gives the correct surface forms:

(105)

	a. Infinitive + *-te*	b. Base + *-ite*
LB *CVCe*	*ake* + *te* => *akete*	*ake* + *ite* => *akite
LB *Ce*	*pe* + *te* => *pete*	*pe* + *ite* => *pete*
s-irr	*si* + *te* => *site*	*se* + *ite* => *sete
k-irr	*ki* + *te* => *kite*	*ko* + *ite* => *kote

It is not surprising that the infinitive is a stem upon which a number of forms are built, for the infinitive is also used as the first member of compounds and may be thought a general compositional form in addition to its independent use, and it is also a nominalized form (the substantive). Thus, the infinitive functions both as a word form and as a combinatory stem.

3.4.4.2 The a- *stem*

The forms under (b) in Table 3.11 are formed by attaching an invariant formant (*-zi*, *-na*, *-ba*; *-s-*, *-sime-*, *-ye-*, *-re-*, *-n-*, *-zu*, *-m-*, *-masi*) to the base of the vowel base verbs, but to a stem in *-a-* for the consonant base verbs. The only exception is the combination of UM verb and respect auxiliary which is lexicalized with those verbs on which it is formed, see 3.1.4.3. The stem of the consonant base verbs used in these combinations is a secondary, bound stem which is derived by adding *-a-* to the basic stem. We refer to it as the '*a*- stem'; it corresponds to the *mizenkei* in Japanese traditional grammar, see 3.4.6. The suffixes which select the base of vowel base verbs and the *a*- stem of consonant base verbs will be noted as in (106) to show this. As

above, this is an abstract morphophonemic notation: /a-/ is not part of the phonemic shape of the suffixes, and we segment: *saka-nu* 'bloom-NEG.ADN'. Note, however, that the /a/ of the *a-* stem diachronically reflects resegmentation of suffixes in initial *a- and that in some cases, /a/ formed part of an earlier shape of the suffix, for example the negative which goes back to *an-; see 3.5.2.

(106)

Negative conjectural	*-(a)zi*
Optative	*-(a)na*
Conditional	*-(a)ba*
Respect	*-(a)s-*
Causative	*-(a)sime-*
Passive	*-(a)ye-*
Passive	*-(a)re-*
Negative	*-(a)n-*
Negative	*-(a)zu*
Conjectural	*-(a)m-*
Subjunctive	*-(a)masi*

Again, however, an alternative analysis proposes that these suffixes synchronically have initial *-a-* (e.g. *-aba*, *-asime-*) and attach directly to the verb base. This, too, would require otherwise unmotivated morphophonemic rules to derive the correct surface forms, as the regular phonological rules (cf. 2.6.1) would give incorrect forms with the polysyllabic vowel base verbs, (107b), whereas (107a) gives the correct forms.

(107)

	a. Bigrade base + *-ba*	b. Bigrade base + *-aba*
UB	*okwi* + *ba* => *okwiba*	*okwi* + *aba* => **okwaba/*okaba
LB	*ake* + *ba* => *akeba*	*ake* + *aba* => *akaba

3.4.4.3 The exclamatory as stem

Two forms, the provisional and the concessive, are formed by attaching the flectives *-ba* and *-do* to the exclamatory. This is quite clear if considering also the exclamatory, provisional and concessive of the simple past, as in (108). The two suffixes are noted in (109) to show that they select the exclamatory, but as before, /e/ is not part of the phonemic shape of the suffix.

(108)

	QD	LB	Simple past
Exclamatory	*kake*	*akure*	*-sika*
Provisional	*kakeba*	*akureba*	*-sikaba*
Concessive	*kakedo*	*akuredo*	*-sikado*

(109) Provisional *-(e)ba*
Concessive *-(e)do*

3.4.4.4 The stative and the nominal

The stative may be analysed as having the shape *-yer-* attached to the base of QD verbs: *sak-* + *-yer-* => *sakyer-*; *mot-* + *-yer-* => **motyer-* => *moter-*. Diachronically the stative reflects contraction of a construction with the infinitive followed by the existential verb *ar-*: **saki-ar-* 'bloom-is; is blooming' > *sakyer-*. With *s*-irr, *k*-irr and the few UM verbs on which it is formed, a fairly high degree of lexicalization must be envisaged, reflecting the etymological source: *si-ar- > *ser-*, *ki-ar- > *kyer-*.

(110)

	QD	QD	*s*-irr	*k*-irr	UM
Base	*sak-*	*mot-*	*se-*	*ko-*	*ki-*
Infinitive	*saki*	*moti*	*si*	*ki*	*ki*
Stative	*sakyer-*	*moter-*	*seri*	*kyer-*	*kyer-*

The *nominal* form, too, is thought to reflect an originally analytic construction, consisting of the adnominal form and a following (formal) noun/nominalizer *aku 'thing, place' which is not independently attested. At some point *aku must have been reinterpreted as a flective and the combination morphologized. *-kyeku*, which is the shape of the adjectival copula as well as a very rare simple past variant, must be thought to be an older lexicalized form, which does not reflect regular synchronic vowel deletion, but diachronic contraction (*ia > *ye*) at a stage where the adnominal was juxtaposed with *aku.

(111)

QD	*kakaku*	<= *kaku-aku*
r-irr	*araku*	<= *aru-aku*
-(a)n- Negative	*-naku*	<= *-nu-aku*
LB	*akuraku*	<= *akuru-aku*
UB	*okuraku*	<= *okuru-aku*
UM	*miraku*	<= *miru-aku*
k-irr	*kuraku*	<= *kuru-aku*
s-irr	*suraku*	<= *suru-aku*
n-irr (perfective)	*-nuraku*	<= *-nuru-aku*
Simple past	*-siku*	<= *-si-aku*
	~ -kyeku	< **-ki-aku*
Adjectival copula	*kyeku*	< **ki-aku*

3.4.5 Consonant versus vowel base morphophonology

The overall morphophonological difference between consonant and vowel base verbs lies in the extensive use of the base of vowel base verbs. Whereas

Table 3.12 Katsuyōkei *paradigms for OJ verbs*

	QD	*r*-irr	UM	*n*-irr	LB	UB	*s*-irr	*k*-irr
mizenkei	*kaka*	*ara*	*mi*	*sina*	*ake*	*okwi*	*se*	*ko*
ren'yōkei	*kaki*	*ari*	*mi*	*sini*	*ake*	*okwi*	*si*	*ki*
shūshikei	*kaku*	*ari*	*miru*	*sinu*	*aku*	*oku*	*su*	*ku*
rentaikei	*kaku*	*aru*	*miru*	*sinuru*	*akuru*	*okuru*	*suru*	*kuru*
izenkei	*kake*	*are*	*mire*	*sinure*	*akure*	*okure*	*sure*	*kure*
meireikei	*kakye*	*are*	*mi(yo)*	*sine*	*ake(yo)*	*okwiyo*	*se(yo)*	*ko*

the consonant base verbs use a stem-forming suffix to form the *a*- stem and flectives to form the infinitive and the imperative, the vowel base verbs simply use the bare base as the combinatory stem used for forming extended forms, as infinitive and, originally, imperative. In OJ some LB verbs use the bare base as imperative, but mostly the imperative use of the base is reinforced by the exclamatory particle *yo*. UB verbs generally use *yo* in the imperative. Of the irregular vowel base verbs, *k*-irr always uses and both UM and *s*-irr sometimes use the basic stem as imperative. In OJ *yo* seems to have been in the course of changing from an exclamatory particle reinforcing the imperative use of the base to an inflectional ending marking the imperative, but it is not yet a flective proper at the OJ stage. This in itself is an important difference between the vowel and consonant base verbs: despite the superficial resemblance between the imperative formants, the synthetically integrated flective *-ye* of the consonant base verbs is not directly related to the fairly independent exclamatory particle *-yo*.

(112)

	QD	LB	UB
Base	*kak-*	*ake-*	*okwi-*
a- stem	*kak.a-*	*ake-*	*okwi-*
Infinitive	*kak.i(-)*	*ake(-)*	*okwi(-)*
Imperative	*kak.ye*	*ake (yo)*	*okwi (yo)*

3.4.6 The katsuyōkei *system*

The description given above is in many respects similar to the presentation and analysis of verb forms in traditional Japanese grammar where verb forms are said to inflect for six basic '*katsuyōkei*' (活用形, inflected forms or stems), which function as *word forms* and/or *stems* used for forming extended forms. Table 3.12 gives the *katsuyōkei* of the eight verb classes (note that only one class has distinct shapes for all six *kei*, viz. *n*-irr).

3.4.6.1 Names and uses of the six katsuyōkei

Mizenkei (未然形 'irrealis') is a stem used for combination with some flectives and auxiliaries. ***Ren'yōkei*** (連用形 'adverbial form; the form followed by inflected words (用言 *yōgen*)') is the infinitive, also used as a stem with some flectives and auxiliaries. ***Shūshikei*** (終止形 'concluding form') is the conclusive. ***Rentaikei*** (連体形 'adnominal form; the form followed by uninflected words (体言 *taigen*)') is the adnominal. ***Izenkei*** (已然形 'realis') is the exclamatory, also used as a stem with some flectives. ***Meireikei*** (命令形 'imperative') is the imperative. *Izenkei* and *mizenkei* are named in contrast after the meaning of the forms ending in *-ba*: realis: *sinure-ba* 'as, when one dies'; irrealis *sina-ba* 'if one dies'.

3.4.6.2 Traditional names for the verb classes

The analysis in the *katsuyōkei* system is intimately linked to the *kana* script (and is therefore not able to segment forms into units smaller than a mora) and the traditional nomenclature for the verb classes is based on the *katsuyōkei* system in combination with the so-called *gojūonzu* ('fifty sound table') arrangement of the *hiragana* and *katakana* letters in ten columns (*gyō* 行) by five rows (*dan* 段), see 6.1.4. The names of the verb classes refer to the positions in the *kana* arrangement of the final *kana* used in writing the *kei*. This terminology pays no heed to the *kō-otsu* syllable distinctions, as it was devised for the classical language of the late Heian period (where those distinctions had merged) and particularly as it was established before the final discovery of the *kō-otsu* distinctions; also, transparent endings written in a full *kana* are ignored for the purposes of these names, namely the *-ru*, *-re* and *-yo* of UM, UB, LB, *k*-irr, *s*-irr and *n*-irr.

Quadrigrade (*yodan* 四段) verbs are thus named because they are written with final *kana* from 'four rows' (*yo-dan*), i.e. with four different vowels, e.g. the *ta*, *ti*, *tu*, *te* of the forms of *tat-* 'stand':

(113)	*mizenkei*	ta*ta*	たた
	ren'yōkei	ta*ti*	たち
	shūshikei	ta*tu*	たつ
	rentaikei	ta*tu*	たつ
	izenkei	ta*te*	たて
	meireikei	ta*te*	たて

As shown in Table 3.12, the *izenkei* and *meireikei* were in fact different (although the difference was neutralized after coronal consonants); thus *izenkei* (exclamatory) *kake* was distinct from *meireikei* (imperative) *kakye*. However, once /Cye/ and /Ce/ had merged at the beginning of the EMJ period (7.3.2.1), *izenkei* and *meireikei* became homophonous. **Upper monograde** (*kami ichidan*

上一段) verbs are written with *kana* from 'one row' (*ichi-dan*), the *-i* row (cf. invariant ***mi*** in Table 3.12), which is in the 'upper' (*kami*) half in the vertical sequence in the columns in the *kana* tables. **Upper bigrade** (*kami nidan* 上二段) verbs are written with final *kana* from 'two rows' (*ni-dan*), the *-u* and *-i* rows (cf. ***oku***, (*okwi* > EMJ ***oki***)), of which the *-u* row is in the middle and *-i* in the upper half. **Lower bigrade** (*shimo nidan* 下二段) verbs are written with final *kana* from two rows, the *-u* and *-e* rows (cf. ***aku***, ***ake***), of which the *-e* row is in the 'lower' (*shimo*) half. The final four conjugation classes are termed 'irregular' (変格 *henkaku*). ***k*-irregular** (*kagyō-henkaku* カ行変格; *ka-hen* カ変) is written with *kana* from the '*k*-column' (*ka-gyō*); ***s*-irregular** (*sagyō-henkaku* サ行変格; *sa-hen* サ変) is written with *kana* from the '*s*-column' (*sa-gyō*); ***r*-irregular** (*ragyō-henkaku* ラ行変格; *ra-hen* ラ変) verbs are written with *kana* from the '*r*-column' (*ra-gyō*); ***n*-irregular** (*nagyō-henkaku* ナ行変格; *na-hen* ナ変) with *kana* from the '*n*-column' (*na-gyō*).

3.4.6.3 Discussion; basic paradigms

The *katsuyōkei* system is the standard way of thinking or talking about Japanese verb morphology within the Japanese tradition, but it has not been widely adopted, and is sometimes even denounced, in English language descriptions of Classical Japanese.

The *katsuyōkei* system was devised to account for the verb forms of Classical Japanese. To some extent it has the same functions as the four principal parts of Latin verbs. If you know your *amo* 'I love', *amare* 'to love', *amavi* 'I have loved', *amatum* 'loved', or *video* 'I see', *videre* 'to see', *vidi* 'I have seen', *visum* 'seen', you can generate and identify any form of those verbs, as well as identify their conjugational class. And likewise if you know the six *katsuyōkei* of a Classical Japanese verb. The *katsuyōkei* framework evolved in the work of Japanese philologists through the Edo period and only found its current form early in the nineteenth century. It is therefore possible that the conception of it was inspired in some way by Latin grammar, which was made available in Japan in Jesuit publications as early as the 1580s. If so, however, this influence remains unacknowledged. All expositions of the history of the *katsuyōkei* framework simply describe it as an indigenous development.

First of all, five of the *katsuyōkei* correspond to five basic inflected forms: *ren'yōkei* = infinitive, *shūshikei* = conclusive, *rentaikei* = adnominal, *izenkei* = exclamatory, *meireikei* = imperative. Second, in its account of the formation, or segmentation, of most other inflected forms, the *katsuyōkei* system shows that some inflected forms (infinitive = *ren'yōkei*, exclamatory = *izenkei*) also function as stems for forming other forms, and as we saw above, the segmentation of the *katsuyōkei* analysis – for example *saka-ba* rather than *sak-aba* – gives a simpler and more regular account of the formation of the forms,

Table 3.13 Katsuyōkei *paradigms for negative, simple past, adjectival copula and subjunctive*

	Negative	Simple past	Adjectival copula	Subjunctive
Mizenkei	*zu*	*kye ~ -se*	*kye*	*mase*
Ren'yōkei	*zu ~ ni*	–	*ku*	–
Shūshikei	*zu*	*ki*	*si*	*masi*
Rentaikei	*nu*	*si*	*ki*	*masi*
Izenkei	*ne*	*sika*	*kyere ~ kye*	–

without recourse to ad hoc rules. Thus, the *katsuyōkei* system captures important facts of the language.

It is, on the other hand, also easy to agree that the *katsuyōkei* system has some shortcomings. The major conceptual shortcoming is that it presents the six *katsuyōkei* as equivalent morphological primitives. However, (a) they are not equivalent: some are exclusively word forms (*shūshikei* = conclusive, *rentaikei* = adnominal, *meireikei* = imperative) and some are both word forms and stems (*ren'yōkei* = infinitive, *izenkei* = exclamatory); in particular, the *mizenkei* is not a word form at all, but simply a derived stem, but in applications of the *katsuyōkei* system, the *mizenkei* is generally treated as a word form on a par with the basic inflected forms. And (b), nor are they morphological primitives. As we saw above, inflected forms, including those presented in the *katsuyōkei* system, are fairly easily segmentable, but it is not possible to capture this in the *katsuyōkei* system which is constrained by the moraic *kana* writing, precluding a proper analysis of the forms in it. Some presentations of the *katsuyōkei* do provide an analysis of the forms, but it is so bound up with the *kana* writing that it for example presents the six *katsuyōkei* of *kak-* as: stem (*gokan* 語幹) *ka-* + ending (*gobi* 語尾) *-ka*, *-ki*, *-ku*, *-ku*, *-ke*, *-ke*, which is of course misleading.

Katsuyōkei paradigms are also presented for those auxiliaries which do not belong to one of the main verbal conjugation classes (the negative, simple past, adjective and subjunctive auxiliaries), see Table 3.13. The descriptive value of these paradigms is, however, far more limited than for the verbs and they contribute to the confusion about the status of the *mizenkei*.

Finally, some inflected forms, the morphological stative and the nominal form (see 3.4.4.4), are not consistently describable by the *katsuyōkei* system, which also does not include the frequent and important adjectival copula exclamatory-1, *-sa*, and infinitive-2, *-mi*.

However, bearing in mind its limitations, or perhaps better, its proper application, the *katsuyōkei* presentation is useful in at once capturing basic inflected

Table 3.14 *Basic paradigm for OJ verbs*

	QD	*r*-irr	*n*-irr
Base	*kak-*	*ar-*	*sin-*
a- stem	*kaka-*	*ara*	*sina*
Infinitive	*kaki*	*ari*	*sini*
Conclusive	*kaku*	*ari*	*sinu*
Adnominal	*kaku*	*aru*	*sinuru*
Exclamatory	*kake*	*are*	*sinure*
Imperative	*kakye*	*are*	*sine*

	LB	UB	LM	*s*-irr	*k*-irr
Base	*ake-*	*okwi-*	*mi-*	*se-*	*ko-*
Infinitive	*ake*	*okwi*	*mi*	*si*	*ki*
Conclusive	*aku*	*oku*	*miru*	*su*	*ku*
Adnominal	*akuru*	*okuru*	*miru*	*suru*	*kuru*
Exclamatory	*akure*	*okure*	*mire*	*sure*	*kure*
Imperative	*ake(yo)*	*okwiyo*	*mi(yo)*	*se(yo)*	*ko*

forms *and* providing a basis for understanding the formation of extended inflectional forms. Especially in the latter function it gives a simple overview of the morphophonological differences between the verbal conjugation classes. We can adjust the presentation a little, as shown in Table 3.14, so that it includes the basic stem, shows that the *mizenkei* is a derived stem and not a word form, and does not include a *mizenkei* for the vowel base verbs. We will use this mode of presentation and refer to it as the 'basic paradigm' for verbs.

3.5 Proto-Japanese and pre-Old Japanese morphology

3.5.1 Verb classes; bigrade verbs

There are two facts about bigrade verbs which contribute to an understanding of their pre-history. First, many bigrade verbs take part in transitivity alternations as in (114), where QD (a) and bigrade (b) verbs have opposite transitivity values:

(114)

a.	b.
Intransitive	Transitive
tuk- 'stick to'	*tuke-* 'attach'
ap- 'meet'	*ape-* 'join'
yam- 'pause'	*yame-* 'stop'
tum- 'pile up'	*tume-* 'pile up'

Transitive	Intransitive
tok- 'untie'	*toke-* 'be untied'
war- 'break'	*ware-* 'be broken'
yak- 'burn'	*yake-* 'be burned'

Second, a number of bigrade verbs take part in apophonic alternations similar to those exhibited by some nouns (see 2.7.2.2), e.g. (115). In particular, a number of these verbs are clearly related to adjectives.

(115)

		Bigrade verb base (exposed form)	Adjective stem (covert form)
LB	-e- ~ -a-	*ake-* 'redden, lighten'	*aka*(-) 'red'
UB	-wi- ~ -u-	*sabwi-* 'get desolate, fade'	*sabu-* 'lonely'
	-wi- ~ -o-	*opwi-* 'get big; grow'	*opo-* 'big'

In these alternations it is the bigrade verb base which corresponds to the 'exposed form', whereas the 'covert form' is reflected in adjective or noun stems or in other verb stems. It is thought that the bigrade verb bases, like the exposed form of the apophonic nouns, phonologically reflect contraction of *Vy diphthongs, so that for example *ake-* 'to redden' derives from *akay. For the bigrade verbs, the second part of the diphthong is believed to reflect a separate morpheme, thus for example *ake-* < *akay < *aka-y. We refer to this morpheme *-y as the bigrade 'theme' and to verbs whose basic stem incorporates the theme as 'thematic'. Ohno's groundbreaking study (1953) proposed that the bigrade formant *-y is identical with the OJ QD infinitive formant *-i* (cf. 3.4.1.3), but since Unger (1977/1993) it has come to be widely accepted that the theme is an originally derivational morpheme, although there is no consensus about its earlier shape or function; the function reconstructed for this formant will have to cover both of the main functional characteristics of bigrade verbs, the derivation from adjectives (115) and the transitivity alternations (114). See Whitman (2008) and Frellesvig (2008) for some recent proposals.

3.5.1.1 Diachronic classification of verbs

The verb classes may be classified diachronically into *secondary* (*thematic*) verbs whose basic stem incorporates the bigrade theme and *primary* (or *a-thematic*) verbs whose basic stem does not incorporate the theme, see (116). There is no one-to-one correlation between the two major diachronic classes (primary and secondary) and the two major synchronic OJ classes (consonant and vowel base verbs), although all consonant base verbs are athematic; note, however, that it is possible that the *n*-irr infinitive reflects an older thematic

stem: *sini* ??< *sinu-y or *sini-y. *k*-irr *ko-* is primary, directly reflecting its root *kə- or *kɨ. We classify *s*-irr *se-* 'to do' as secondary on the hypothesis that it represents a regular contraction from *sə-y and thus incorporates the bigrade theme; this is based on forms such as the *-so* in the prohibitive form (3.1.3.1) which suggests a root *so* < *sə which is also reflected in the basic stem: *se-* < *sə-y. Other reconstructions divorce the *so* in the prohibitive from the verb *se-* 'do' and posit *se-* < *sye < *se through (partial) mid vowel raising; if so, *s*-irr should also be classified as primary.

(116)	Primary (athematic)	QD, *r*-irr, *n*-irr, UM, *k*-irr
	Secondary (thematic)	LB, UB, *s*-irr

I have elsewhere argued that the secondary verb classes constitute a *younger* morphological layer in the Japanese language than the primary verb classes and that this means that no account should be taken of the forms of the secondary verbs in reconstructing simple pJ verb paradigms. We will not discuss this any further here, but see Frellesvig (2008).

3.5.2 Pre-history of verb suffixes

Apart from basic inflected forms, most auxiliaries and flectives in OJ attach either to the *a*-stem or to the infinitive of the consonant base verbs. The *a*-stem reflects a diachronic reanalysis of the combination of verb stems and suffixes in initial /a/. This reanalysis is shown in (117), exemplified by the combination of *anV, the ancestor of the negative auxiliary *-(a)n-*, and *yak- 'burn' and *yəsə- 'draw close'.[13]

(117)

Original pre-OJ formation	Reanalysis	Later (pre-)OJ formation
{(*yak- =>) *yak-anV => *yakanV}	>	{*yakanV* <= *yak.a-nV* (<= *yak-*)}
{(*yəsə- =>) *yəsə-anV => *yosanV}	>	{*yosanV* <= *yos.a-nV* (<= *yos-*)}

Note that in the original formation, a base final vowel was deleted in accordance with the vowel deletion rules set out in 2.6.1 above, as in *yəsə-anV => *yosanV. There are three parts to the reanalysis: (a) resegmentation of the shape of the suffix from /*-anV-/ to /-nV-/ and conversely (b) resegmentation of the verb stem to which the suffix attached as a derived stem ending in *-a-* (*yaka-*, *yosa-*); (c) reinterpretation of the basic stems of all ancestors of QD verbs as having the basic stem shape CVC-, regardless of their etymological root shape (OJ *yak-* < pre-OJ *yak-, *yos-* < *yəsə-).

[13] OJ *yos-* 'draw close$_{TR}$' is used as an example of a QD verb originating in an open (vowel final) root; OJ *yosor-* 'approach' and *yose-* 'draw close$_{TR}$' point to a common open root for all three, *yəsə.

In addition to the negative auxiliary, also the conjectural auxiliary seems to reflect old suffixes in initial /a/, *amV, as both have likely MK cognates in initial /a/. They are also both involved in the formation of other suffixes. Respect *-(a)s-* also has a possible MK cognate, but with a different initial vowel.

(118)	Negative	*-(a)n-* < *-anV (cf. MK *a ni* negative adverb)
		-(a)zu < *-ani-su (see 3.1.4.8.3)
		-(a)zi negative conjectural < *-ani-si
	Conjectural	*-(a)m-* < *-amV (cf. MK *¨amwo* 'any, anyone')
		-(a)ba conditional < *-amu-pa
		-(a)masi subjunctive < *-amV-(a)si
	Respect	*-(a)s-* < *-VsV (cf. MK *-usi-* RESP)

Other suffixes which attach to the *a*-stem (optative *-(a)na* (etc.), causative *-(a)sime-*, passive *-(a)ye-* and *-(a)re-*), which have no obvious external links, may also reflect old suffixes in initial /a/, or they may have been analogically formed.

In any case, at least some suffixes attaching to the *a*-stem reflect fairly old /a/ initial suffixes. The suffixes which attach to the infinitive are different. They are more transparently agglutinating and their use as suffixes seems to be younger. It is possible to view all of them as being derived from a few elements which are also reflected in a number of other grammatical morphemes in OJ. They fall in two groups: (a) forms in $k \sim s$, and (b) forms in $t \sim n$.[14] It is likely that both sets reflect earlier copulas which were morphologized.

The ***k ~ s*** forms involve both of the past tense auxiliaries, *-(i)ki* and *-(i)kyer-* (of which the modal past seems to be derived from the simple past: *kyer-* < *ki-ar-). The forms of the adjectival copula overlap to a large extent with the past tense auxiliaries, see Table 3.15. This suggests that these are different morphologizations of the same (copula) material, and lends further support to the analysis of the adjectival endings as a restricted copula. Note also that the $k \sim s$ alternation is exhibited by the adjectival copula infinitive *-ku* and the *-su which takes part in formation of the innovative negative forms (3.1.4.8.3) and which may also form part of the semblative (see immediately below); the morpho-syntactic similarities between the adjectival infinitive and the negative infinitive are easier to explain if, as suggested here, they originate in variant copula forms. It is further possible that the verb *se-* 'do', the focus particles *so* and *ka* (see 3.7.2), and the demonstratives *ko* and *so* (see 3.8.3) are root-related to these forms.

[14] I have suggested elsewhere (Frellesvig 2008) that the *t- ~ n-* forms have Korean cognates.

Table 3.15 *OJ adjectival copula and possibly related grammatical forms*

	Adjectival copula forms				
Conclusive	*si*				
Adnominal		*ki*			
Exclamatory	*sa*		*kyere* < *ki-are		
Infinitive				*ku*	*ku*
Gerund				*kute*	
Conditional		*kyeba* < *ki-amu-pa		*kupa*	
Concessive		*kyedo*	*kyeredo*		
Provisional		*kyeba*	*kyereba*		
Nominal		*kyeku* < *ki-aku			
Conjectural		*kyem-* < *ki-am-			
Negative (nominal)		*kyenaku* < *ki-anu-aku			

	Simple Past		Modal Past	Negative	Semblative
Conclusive		***ki***	*kyeri*	*zu* < *ani-**su***	
Adnominal	***si***		*kyeru*		*nasu* < *na-**su***
Exclamatory	*sika*		***kyere***		
Infinitive				*zu* < *ani-**su***	*nasu* < *na-**su***
Gerund				*zute* < *ani-**su**-te*	
Conditional	*seba*	***kyeba***		*zupa* < *ani-**su**-pa*	
Concessive	*sikado*		***kyeredo***		
Provisional	*sikaba*		***kyereba***		
Nominal	*siku*	***kyeku***	*kyeraku*		
Conjectural		***kyem-***			

The ***t*** ~ ***n*** forms are reflected among the verbal suffixes in the perfectives, *-(i)te-* and *-(i)n-*,[15] and in gerund *-(i)te* and continuative *-(i)tutu*. They are, however, also reflected directly in the simple non-finite copula forms (3.3), *ni*, *no*, *to*, *tu* and the particles derived from the copulas. Other related grammatical forms possibly include *nasu* which is used after nouns and the conclusive verb form to mean 'like N, as V-ing', e.g. *asapi nasu* 'like the morning sun'. This construction can be used adverbially or adnominally and in EOJ a variant *nosu* is found alongside *nasu*. *Nasu* ~ *nosu* are most likely derived from a copula root, *nV-*, by means of a formant *-su* which, as mentioned above, may be related to the *-su involved in the reformation of the negative auxiliary and further to *-ku*, the infinitive of the adjectival copula. Also the formant *-zi* which derives denominal *jiku* adjectives 'like N' (see 3.2.4.2) may be thought to be *-zi* < *nVsi < *nV- copula + *-(V)si* adjective formant (3.2.4).

(119)

	t-	***n-***
Copula	*to, tu*	*ni, no*
Case and conjunctional particles	*to*	*ni, nite*
Genitive particle	*tu*	*no*
Gerund	*-te*	
Perfective	*-te-*	*-n-*
Continuative	*-tutu*	
Semblative		*nasu* ~ *nosu* ~ *zi*

3.6 Verb extensions

Verb extensions are inflecting clitics. They follow a finite verb form to form an extended verb syntagm. They are all *modal* in meaning: ***be-*** necessitive, ***masizi*** negative potential, ***ram-*** present conjectural, ***rasi-*** presumptive, ***nar-*** evidential.

(120)

	QD	LB	*r*-irr	UM
Pres. conject.	*kaku ram-*	*aku ram-*	*aru ram-*	*mi ram-*
Presumptive	*kaku rasi*	*aku rasi*	*aru rasi* ~ *arasi*	*mi rasi*
Necessitive	*kaku be-*	*aku be-*	*aru be-*	*mi be-*
Neg. potential	*kaku masizi*	*aku masizi*	–	–
Evidential	*kaku nar-*	*aku nar-*	*ari nar-*	*miru nar-*

[15] This is contrary to the traditional account, which diachronically derives *-(i)te-* from the verb *(s)ute-* 'discard, throw away' and *-(i)n-* from *in-* 'disappear, pass, elapse' and accordingly believes that the function of assertion is a secondary development (e.g. Ohno 1990: 1473). See Frellesvig (2001) about this.

Table 3.16 *Inflected forms of OJ verb extensions*

	be-	*masizi-*	*rasi-*	*ram-*	*nar-*
Conclusive	*be-si*	*masizi*	*rasi*	*ramu*	*nari*
Adnominal	*be-ki*	*masizi-ki*	*rasi-ki*	*ramu*	*naru*
Exclamatory	–	–	–	*rame*	*nare*
Infinitive-1	*be-ku*	–	–	–	–
Infinitive-2	*be-mi*	*masizi-mi*	–	–	–

Be-, *rasi-*, *ram-*, *masizi-*, *nar-* (evidential) follow the conclusive form of verbs from all other conjugation classes than *r*-irr and UM (*masizi* is, however, not attested with UM or *r*-irr verbs); note that when *rasi* is used with a *r*-irr verb, a reduction often takes place: *aru-rasi* => *arasi*, *-kyeru-rasi* => *-kyerasi*. *Nar-* evidential follows the conclusive of all verb classes (but from EMJ the adnominal of *r*-irr). The verb extensions exhibit the inflected forms in Table 3.16. *Rasi-* and *masizi-* belong to the adjectival *shiku* (*jiku*) conjugation; *be-* belongs to the adjectival *ku*-conjugation. *Ram-* belongs to QD and *nar-* to the *r*-irr conjugation.

Necessitive *be-*: expresses necessity, obligation and strong probability. This is the morphologically most versatile verb extension, mainly due to its infinitive which forms the basis for extensions with *ar-*. **Negative potential *masizi-***: is most commonly found with verb forms involving some expression of ability, giving the meaning 'probably/surely cannot'. *Masizi-* is rare and restricted in use already in OJ; it is not found in EMJ, but is thought to correspond to (to have become) EMJ *mazi*. **Present conjectural *ram-***: 'probably, apparently, seemingly' usually relates to the present. **Presumptive *rasi-***: 'presumably'. The adnominal is very rare (one example in the *Man'yōshū*); instead the conclusive was used in noun modifying function. **Evidential *nar-***: has two main functions: (a) 'it sounds as if somebody does, one can hear somebody do'; (b) 'it is said/reported that somebody does'. (a) was somewhat more frequent than (b) in OJ. Note that evidential *nar-* is distinct from the assertive use of the homophonous EMJ copula *nar-*. The assertive copula follows the adnominal form, unlike evidential *nar-*. Evidential *nar-* may be thought to be a lexicalized contraction from *na 'sound' + *ar-* 'exist'.

3.7 Particles

Particles are bound postpositional grammatical words which attach to a host, minimally a word. We do not know how closely particles attached to their host in OJ, but sound changes taking place in the first half of EMJ show that particles then were phonologically integrated with the host word (8.7). Particles contribute to the syntax, semantics and/or pragmatics of an utterance. As the best-studied OJ texts are poetry it is difficult to gain a complete picture

of the use and functions of the particles. The literary or rhetorical style employed in the OJ poetry means that it is full of exclamations, invocations, lamentations, etc. These are, of course, features of language use in any culture at any time, but this feature of the available OJ materials seems to have skewed the view of the grammar of OJ particles, many of which traditionally are glossed 'emphatic'.

In Japanese school grammar, *joshi* (助詞) 'auxiliary word' includes particles, but also a number of verbal inflectional endings, for example the conditional formant *-(a)ba*. This is because Japanese part of speech classification traditionally is a classification into morpheme types, not into word types. However, the two are also different morpheme types: inflectional endings take part in forming a word whereas particles attach to a full word, phrase, or clause. Note, however, that there is some *functional* overlap between inflectional verbal endings on the one hand and conjunctional, final, and interjectional particles on the other, in that both contribute to the expression of modality and interclausal syntax. But note also that not all such inflectional endings, e.g. the imperative formant *-yo* used with vowel base verbs, are included among *joshi* in the traditional classification.

The following classification of particles into six types is traditional (although other classifications are found). It is based first of all on functional criteria and several particles belong in more than one class.

(121) a. case particles (*kaku-joshi* 格助詞; 3.7.1)
b. topic and focus particles (*kakari-joshi* 係助詞; 3.7.2)
c. restrictive particles (*fuku-joshi* 副助詞; 3.7.3)
d. conjunctional particles (*setsuzoku-joshi* 接続助詞; 3.7.4)
e. final particles (*shū-joshi* 終助詞; 3.7.5)
f. interjectional particles (*kantō-joshi* 間投助詞; 3.7.6)

Limiting the classification to particles proper (i.e. excluding inflectional endings) and with the addition of a class not provided in the school grammar, namely that of *complementizer* (3.7.7), this grouping is by and large valid for OJ and the following stages of Japanese.

3.7.1 Case particles

Case particles attach to nouns and nominalized forms of verbs and adjectives, specifying grammatical relations within a clause. Although we speak of these OJ particles as case particles, they do not form a fully developed case system comparable to that of NJ – or of languages with case inflection – nor do they reflect an inherited case system (see 3.7.8). Case marking of core arguments, subject and object will be discussed below in 3.7.1.2. Using familiar names for cases, the OJ case particles are as follows, divided into three groups on the basis of their use and productivity:

(122)

Main	
Accusative	*wo*
Genitive	*no*; *ga*
Dative	*ni*
Ablative	*ywori* (~ *ywo* ~ *yuri* ~ *yu*)
Comitative	*to*
Obsolete	
Nominative	*i*
Genitive	*tu*; *na* (~ *da*)
Emerging	
Ablative	*kara*
Allative	*pye*

3.7.1.1 Main Old Japanese case particles

Accusative ***wo*** is mainly used to mark direct and traversal objects, as in NJ, but was in OJ used more widely also to mark durational adverbials. It is also used as a conjunctional particle and as an interjectional particle. When followed by the topic particle *pa* the resulting form is *wo-ba*.

Dative ***ni*** is the general oblique case, marking both argument and non-argument oblique nominals. The main uses are indirect object, allative, purposive,[16] agent, instrumental, locative, temporal. A variant *nite* is used in some of the peripheral functions, especially instrumental, locative, temporal.

Ablative ***ywori ~ ywo ~ yuri ~ yu*** are used about source of movement, comparison, material, and means: 'from, than, with'. There is no discernible difference in function between the four variants; in EMJ only the shape *yori* survived.

Comitative ***to*** is used as coordinative, comitative, and comparative: 'with, and, than'.

3.7.1.1.1 Genitives Both *no* and *ga* were productive genitive particles. In addition to being a genitive, *no* was in OJ, and still is in NJ, an adnominal form of the copula (3.3, 15.2). The genitive function of *no* seems to derive in pre-OJ from the function as adnominal copula (see 3.7.8.2.1), but at the OJ stage both functions of *no* were firmly established and fully independent. From the common genitive use, *no* and *ga* developed differently and have quite different functions in NJ. The major changes are as follows: *Ga* acquired an additional use as a conjunctional particle ('and, but, as') in EMJ (8.7.2), and

[16] In this function *ni* is also used in purpose-of-motion constructions after verb infinitives, e.g. *tumi ni ku* 'come to pick (flowers)'.

changed from a genitive to a nominative case particle in late LMJ (see 12.6.2). *No* remains a copula and a genitive case particle, but in early NJ acquired the additional function of a nominalizer (see 12.6.1.3); however, in between, in EMJ and early LMJ, *no* had more nominative-like functions than *ga* (see 12.6.2). The different developments of *no* and *ga* will be summarized in 12.6.2, but (123) shows the main OJ and NJ functions; it is clear that *ga* functionally has changed more than *no*, but it must be emphasized that of the five main functions exhibited by *ga* and *no* in NJ shown in (123), genitive, nominalizer, and nominative are in a number of dialects distributed differently over the two particles; in some dialects, for example, *no* is nominative, whereas *ga* is genitive and nominalizer.

(123)

	OJ	NJ
no	Copula Genitive	Copula Genitive Nominalizer
ga	Genitive	Nominative Conjunctional particle

In OJ, genitive *no* and *ga* had two main functions (124). Both *ga* and *no* were used in both functions.

(124) a. Adnominalization (NP-*ga/no* NP)
b. Subject marking (NP-*ga/no* VP)

Adnominalization is the primary function associated with the genitive in Japanese, as seen in many of the examples cited in other contexts throughout the book. Furthermore, both *ga* and *no* were used to mark subjects from OJ onwards. In OJ, this was limited to marking subjects of adnominal clauses and other subordinate clauses, and more rarely non-declarative main clauses (cf. 8.9).

(125) a. *kimi* ***no*** *imasi-seba*
my lord GEN exist.RESP-SPST.PROV
'If my lord had still been here' (*MYS* 3.454)

b. *sawarabi* ***no*** *moye*-duru paru *ni*
fern GEN sprout-emerge.ADN spring DAT
'In the spring when the fern sprouts' (*MYS* 8.1418)

c. *kimi* ***ga*** *imasaba*
my lord GEN exist.RESP.COND
'If my lord is (still there)' (*MYS* 19.4280)

Genitive subject marking is sometimes thought to have originated in reanalysis of relative constructions like (126a) as (b), but whether or not that is the case, genitive marking of subjects in subordinate clauses was firmly established already in OJ and not limited to adnominal clauses.

(126) a. wa ga [yuku miti]
I GEN go road
'my road to go, that I'm going'

b. [wa ga yuku] miti
'the road I go'

3.7.1.1.2 Differences between *no* and *ga* Both *no* and *ga* were used in both of the functions in (124), but there were significant differences in the constituents they could mark. Generally *ga* was far more restricted in use than *no*: *ga* was only used to mark noun phrases referring to humans (or personified animals or things), whereas *no* could mark all nouns, including those referring to humans. This is also reflected in the fact that personal pronouns take *ga* (*wa-ga* 'mine', *na-ga* 'yours', *si-ga* 'his', *ta-ga* 'whose'), whereas demonstratives take *no* (*ko-no* 'this', *so-no* 'that') (cf. 3.8.1). The systematic restrictions on the use of *ga* with nouns continued in EMJ and early LMJ where *ga* was increasingly curtailed; see further 12.6.2. Importantly, the restrictions on the use of *ga* applied both when the genitive was used to adnominalize and when it marked subjects. In both cases *ga* was severely restricted. This shows that it was properties of the host noun which determined whether *ga* could be used.

There was, however, one context which favoured *ga* over *no*, namely marking clauses with the predicate in the adnominal form. In the examples in (127), *ga* functions as a kind of complementizer to adjoin a clause to a head: In (127a) *ga* connects a modifying clause to a head noun; also *no* was used in this way (127b), but in OJ and EMJ less so than *ga*. This use was restricted to ungapped modifying clauses, see further 12.6.1.1.1. Only *ga*, never *no* was used to adjoin clauses to a predicate, as in (127c–d). In the parallel construction in (127c), *ga* adjoins a clause to the first *gotoku*, whereas the clause attaching to the second *gotoku* does not have *ga*, showing that the use of this construction was optional in this case, although it seems to have been obligatory when connecting a complement clause to adjectives in the exclamatory form (cf. 3.2.2.3), as in (127d).

(127) a. *wagimokwo* *ni* *mise-mu* ***ga*** *tame* *ni*
my.beloved DAT show-CONJ.ADN GA sake COP.INF
'In order to show (them) to my beloved' (*MYS* 19.4222)

b. taye-*mu* ***no*** kokoro
end-CONJ.ADN NO heart
'The intention to end it (our relationship)' (*MYS* 12.3071)

c. puku kaze *no* mi-*ye-nu* ***ga***
blow.ADN wind GEN see-PASS-NEG.ADN GA
goto-ku,
like-ACOP.INF
yuku midu *no* *tomara-nu* *goto-ku*
go.ADN water GEN cease-NEG.ADN like-ACOP.INF
'like the blowing wind is not visible, like flowing water does not cease' (*MYS* 19.4160)

d. *kogu* *punabito* *wo* mi*ru* ***ga*** *tomosi-sa*
row.ADN boatsman ACC see.ADN GA enviable-ACOP.EXCL
'How enviable it is to see the rowing boatsmen!' (*MYS* 15.3658)

Finally, it is often claimed that the use of *no* and *ga* was subject to social differentiation, for example so that *ga* is said to have been used with 'sentient nouns whose referent is someone close to the speaker or the person who dominates the narrative viewpoint', whereas *no* is used with 'exalted or indefinite animate nouns' (see Takeuchi 1999: 159–60 who offers (128) as an illustrative example); cf. also 12.6.2.1 about this.

(128) *titi-papa* ***ga*** *tame* *ni* *moropito*
father-mother GEN sake COP.INF all.people
no *tame ni*
GEN sake COP.INF
'for the sake of father and mother, and for the sake of all people' (*Bussoku* 1)

3.7.1.2 Case marking of subject and object

The case marking of subjects and objects in OJ has recently attracted a great deal of interest from scholars working in different theoretical frameworks, but from a neutral descriptive point of view, the case marking of subjects and objects in OJ can be summarized as in (129):

(129)

	Subject	Object
Declarative main clauses	Ø	*wo*, Ø
Subordinate/non-declarative main clauses	Ø, *ga*, *no*	*wo*, Ø

(130) gives examples of these possibilities. As in EMJ, LMJ and NJ, marking of subject and object was optional. It is first of all an important fact of OJ

(holding also for EMJ) that it did not have means of explicitly marking the subject of a declarative main clause and the development of a nominative case particle is one of the few major syntactic changes to have taken place in Japanese. This will be discussed in 12.6.2. As mentioned above (3.7.1.1.2), *ga* was more restricted than *no* in the subjects it could mark. Marking of objects overall seems not to have changed greatly since OJ: as today, objects could in OJ be marked by *wo* or left unmarked. While clear-cut conditions which force or bar *wo*-marking of objects have yet to be established, it seems clear that adjacency between object and verb tends to disfavour *wo*-marking of objects, although there are plenty of examples of *wo*-marked objects adjacent to their verb, such as (130e). However, one clear regularity is that if subject and *wo*-marked object co-occur in a clause, the *wo*-marked object precedes the subject, as in (130f). Finally, it should be noted that *wo*, like accusative markers in many languages, in addition to marking objects can mark a range of durational, temporal and locative adjuncts, e.g. (130g).

(130) a. *ume* *no* *pana* ***Ø*** *tiri*
plum GEN blossom **Ø** scatter.INF
'the plum blossoms scatter . . .' (*MYS* 5.838)

b. kimi ***ga*** *yuku* miti
my.lord GEN go.ADN way
'the way my lord goes' (*MYS* 15.3724)

c. *ugupisu* ***no*** *ki-naku* *yamabuki*
warbler GEN come-cry.ADN kerria
'the kerria for which the warbler comes and cries' (*MYS* 17.3968)

d. *ume* *no* *pana* ***Ø*** *wori*
plum GEN blossom Ø break.off.INF
'breaking off the plum blossoms' (*MYS* 5.843)

e. *awo-yanagi* *ume* *to* *no* *pana* ***wo*** *wori*
green-willow plum COM GEN blossom ACC break.off.INF
'breaking off the blossoms of the green willow and the plum' (*MYS* 5.821)

f. *ware* ***wo*** *yami* *ni* *ya* *imo* ***ga***
I ACC dark DAT Q beloved GEN
kwopwitutu *aru* *ramu*
long.for.CONT exist.ADN PCONJ.ADN
'Would my beloved be longing for me in the dark?' (*MYS* 15.3669)

g. akikaze *no* samu-ki asake ***wo*** sanu
autumn.wind GEN cold-ACOP.ADN dawn ACC Sanu
no woka kwoyu ramu kimi
GEN hill cross.CONCL PCONJ.ADN my.lord
'my lord, who will cross the hills of Sanu during dawn when the autumn wind is cold' (*MYS* 3.361)

3.7.1.3 Obsolete and peripheral case particles

Genitives (*attributives*): ***Tu*** which only survived into EMJ in lexicalized collocations was somewhat fossilized already in OJ, in expressions like *nipa-tu-tori* 'garden-GEN-bird; chicken'. It is sometimes termed 'locative genitive' as it is often found after nouns denoting some kind of place; this is, however, not likely to be an original feature of this particle which derives from a copula and which is also used to adnominalize adjectives and other words (see 3.2, 3.3). A genitive marker ***na*** is usually included in grammars, based on lexicalized forms such as *ma-na-kwo* 'eye-*na*-?child; eye(ball)' and *ta-na-soko* 'hand-*na*-bottom; palm of the hand'.[17] Based on a few words such as *keda-mono* 'beast' (*ke-da-mono* 'hair-*da*-being'), ***da*** is often said to be an obsolete variant of *na*.

Nominative ***i*** is rare in the OJ texts – about twenty examples in the entire text corpus – and it went out of use after OJ (except for extensive use in EMJ *kunten* glosses to Chinese texts, see 9.1.1). Most OJ examples of *i* are found in the *Senmyō* which are thought in several respects to preserve archaic language usage (and which also are the OJ text type most heavily influenced by *kanbun-kundoku*). There is no consensus about the main or basic function of OJ *i*. It has for example been studied in Miller (1989) and Vovin (1997) who arrive at strikingly different conclusions, namely that *i* is an old *accusative* marker (Miller) or an *active* marker in a vestigial active/passive alignment system (Vovin). What is clear is that *i* is used to mark certain subjects, mainly in subordinate clauses. It is also traditionally said to be used for emphasis, and there are some occurrences where it appears to nominalize. This particle is at best marginal already in OJ, but is of some diachronic interest, as it may be related to the Korean nominative particle *i*, either as a cognate or as a loan into Japanese.

[17] Vovin suggests (1994: 253) that *na* is an old dual or plural marker, based on the occurrence of *na* mostly with paired body parts (*ma-na* 'eyes', *ta-na* 'hands') or uncountable nouns (*mi-na* 'water'). There seems to be only a single compelling counterexample to Vovin's suggestion: *momo* ***na*** *pito* (*NSK* 11) '100-*na*-person; 100/many people'. However, another possibility is that the forms in *-na* in fact reflect pre-OJ root final consonants supported by an epenthetic vowel, so that *tana* diachronically should be segmented *tan-a*. This hypothesis is based on the observation that this *na* almost only occurs with nouns that take part in apophonic alternations and would entail that the root final consonant in some cases yodized and contracted with the preceding vowel (see 2.7.2.2) and in others was reflected as *n* + epenthetic vowel *a*, so that for example *taC 'hand' showed two developments: (a) *taC > *tay > *te*, and (b) *taC > *tan > *tana*.

(131) ipye naru imo ***i*** obobosi-mi
house be.at.ADN beloved NOM be.depressed-ACOP.INF
se-mu
do-CONJ.CONCL
'my beloved who is at home will be worrying about me' (*MYS* 12.3161)

(132) naka*maro* ***i*** itupari kadam-ye*ru* kokoro
Nakamaro NOM lie.INF be.deceitful-STAT.ADN heart
wo moti*te* ikusa *wo* okosi
ACC hold.GER army ACC raise.INF
'Nakamaro, having a lying and deceitful heart, raised an army' (*SM* 28)

3.7.1.4 Emerging case particles

Kara, which in later periods replaced *ywori* to become the general ablative marker, was in OJ coming into use, but was not yet fully grammaticalized. It is thought to derive from a noun 'will, way, extent' which was all but obsolete in OJ; it is often followed by *ni* and often expresses *reason* rather than *source of motion*. It is also found in the conjunctional particle *mono kara* and had other semi-grammatical uses as well. The noun *pye* 'side, direction' was being grammaticalized as an *allative* case particle ***pye***, but in the OJ period had not yet acquired that status.

3.7.2 Topic and focus particles

Pragmatic particles single out a nominal constituent for comment or as focus for emphasis or question. They express no particular syntactic relationship between the marked constituent and its predicate and may be divided into two groups, topic and focus particles.

The *topic* particles ***pa*** and ***mo*** are used with little difference from later stages of Japanese, although OJ *pa* (> EMJ *wa*) mainly marks a contrastive topic; *mo* marks an emphatic topic, 'also, even'. When combining with accusative *wo*, *pa* becomes *ba*: *wo-ba*.

The *focus* particles mark the focus of emphasis, ***namo***, ***so***, ***koso***, or question, ***ya***, ***ka***. *Namo* is found only once in *Man'yōshū*, but is relatively frequent in *Senmyō*; it has the rarer variant *namu* which replaces *namo* in EMJ. *So* has the variant *zo* which replaces *so* in EMJ. The focus particles take part in the important construction *kakari-musubi* which will be discussed in detail in 8.9.

3.7.3 Restrictive particles

Restrictive particles form adverbial phrases of extent or degree and include: ***bakari*** 'about, approximately', ***dani*** 'at least; even', ***made(ni)*** 'until; so much

that', ***nomwi*** 'only, solely, alone; entirely, fully, completely', ***sape*** 'also, further, besides, even, at least', ***si*** 'also, even', ***simo*** 'even', ***sura*** 'even, at least'. Note that *made(ni)* which in EMJ came to be used as a conjunctional particle already in OJ could follow a nominalized clause.

3.7.4 Conjunctional particles

Conjunctional particles follow finite verb forms to form subordinate clauses: ***gane*** (only in postposed purposive clauses to volitional main clauses) 'so that'; ***gani*** 'as if'; ***mono-wo*** 'although'; ***mono-kara (ni)*** 'as, while'; ***mono-yuwe (ni)*** 'as, while'; ***nape (ni)*** 'at the same time as, together with'; ***ni*** 'as, when, because'; ***to*** (a) concessive '(even) if, although' (in this use, often followed by *mo*: ***tomo***; cf. here the etymologically related concessive formant *-do(mo)*), (b) purposive 'in order to, (so) that'; ***wo*** 'as, because'; ***yuwe (ni)*** 'because' (also after nouns: 'because of'). Most conjunctional particles follow the adnominal, but *gane*, *gani* and concessive *to(mo)* follow the conclusive; however, *to(mo)* follows the infinitive of the adjectival copula (*na-ku tomo*, 'although there isn't') and the base of UM verbs (*mi tomo*, 'although he sees').

As mentioned, in Japanese school grammar this class of particle includes a number of verbal inflectional endings, namely those forming non-finite verbal forms such as the gerund formant *-te* or the concessive formant *-do*.

3.7.5 Final particles

Final particles occur in utterance final position, after finite verb forms or predicate nominals, contributing to the specification of the modality to the entire utterance: ***kamo*** (a) interrogative, (b) exclamatory (after a negative adnominal functioning as an optative: *konu kamo* 'won't she come?; I wish she'd come'); ***moga(mo)*** optative (after nominals and adverbials, including infinitive-1 of adjectives); ***miyu*** evidential ('it seems, appears'), see 3.7.8.2; ***na*** (a) prohibitive ('don't!'), (b) exclamatory.

3.7.6 Interjectional particles

Interjectional particles form interjections or invocations: ***ro*** (rare in the Central dialect, but frequent in Eastern), ***we***, ***wo***, ***ya***, ***yo***. It is worth noting that the imperative verb endings for the secondary verb classes conspicuously originate in interjectional particles (*-yo* for Central, *-ro* for Eastern).

3.7.7 Complementizer

The complementizer ***to*** is used after directly or indirectly quoted utterances, see (133), sometimes with the verb of utterance omitted, and in naming constructions, see (134). The latter use is closely related to and is difficult to distinguish from use of *to* as a copula infinitive in some constructions, such as in (81) in 3.3.1, and the use of *to* as a complementizer likely derives from the copula used in such constructions. In Japanese school grammar the complementizer is grouped with the case particle *to*, but the two have different functions.

(133) ametuti pa *piro-si* ***to*** *ipedo*
world TOP wide-ACOP.CONCL COMP say.CONC
'although you/they say that the world is wide' (*MYS* 5.892)

(134) *yamato no kuni* ***wo*** *akidusima* ***to*** *ipu*
Yamato GEN country ACC Akizushima COMP call.CONCL
'call the country of Yamato Akizushima' (*NSK* 75)

3.7.8 Etymology

Suggestions of more or less plausible internal etymologies for the particles abound in dictionaries, but are in many cases not persuasive. However, a good number of particles do have fairly obvious internal etymologies, indicating that they either were in the process of being, or fairly recently had been, grammaticalized.

3.7.8.1 Nominal sources

The directional case particles have conspicuous nominal sources: the ablatives *ywori ~ywo ~yuri ~yu* derive from the source of the OJ noun *yuri* 'after(wards)' < pre-OJ *yori,[18] and the emerging allative and ablative particles are based on the nouns *pye* 'side, direction' and *kara* 'will, way, extent', respectively. The restrictive particle *bakari* is from the noun *pakari* 'estimate, limit' (from *pakar-* 'to measure, plan'). The conjunctional particles *yuwe (ni)*, *mono-wo*, *mono-kara (ni)*, *mono-yuwe (ni)* incorporate the nouns *yuwe* 'reason', *mono* 'being, thing', *kara* 'will, way, extent', in most cases optionally adverbialized by the copula infinitive *ni*. As in later stages of Japanese, conjunctional particles are frequently grammaticalized from full or dependent nouns, usually optionally adverbialized by the copula infinitive *ni*.

[18] In order to account for the four OJ variants, a shorter pre-OJ variant, *yo, must be assumed, reflecting either shortening or that *yori itself originally is bi-morphemic ***yo-ri***. Mid vowel raising gave *yori > OJ *yuri*, *yo > OJ *ywo*. The remaining two OJ forms, *ywori* and *yu*, would then be analogical, based on *yuri ~ ywo*.

It is most likely that other particles have nominal sources which we can no longer recover with the same amount of certainty, e.g. *made (ni)* which is optionally adverbialized by the copula infinitive and already in OJ could follow clauses with the predicate in the adnominal form; in EMJ, it came to function as both terminative case particle and conjunctional particle ('until'). As a nominal source, *matwo* 'target' may be suggested, possibly from pre-OJ form *mato(C).

3.7.8.2 Verbal sources

The restrictive particle *si* most likely derives from the ancestor of *se-* 'to do'. A single particle, the final particle *miyu*, derives from a finite verb form, namely the conclusive of *miye-*, the passive of *mi-* 'to see'. Traditionally *miyu* is not classified synchronically as a particle, but it should be, as (a) it follows the conclusive of verbs, and (b) it is only found in the form *miyu*, not in any other inflected forms, in this function. It was presumably grammaticalized from biclausal or bisentential constructions.

3.7.8.2.1 Copula A number of particles derive from copula forms: *no*, *ni*, *nite*, *tu*, and *to* all functioned as particles *and* as forms of a copula at the same time; see 3.3.1. This double function has continued into NJ where *no*, *ni*, *to*, and the new form *de* still are productive forms of the copula, as well as particles. As mentioned in 3.5.2, the focus particles *so* and *ka* may have originated in an earlier copula in *k-* ~ *s-*, also reflected in the adjectival copula.

3.7.8.2.2 Roots of other verbs *Wo*, which functions as an accusative case particle, a conjunctional particle, and an interjectional particle, is sometimes said to have been grammaticalized from the root underlying the existential verb *wor-*; however, it has been shown (by Kinsui 2006) that *wor-* originates in a lexicalized stative form of the verb *wi-* 'sit down' (cf. 3.4.2.1) which is a less likely source of an accusative particle. The focus particle *koso* is most likely from the root of the defective optative auxiliary verb *-kose-* (< *kəsə-i). Note that *-koso* is used as the imperative of *-kose-* (see 3.4.3.3), distinct from its use as a focus particle.

3.7.8.3 External etymology

The *external* etymology of OJ particles has mainly been studied within an Altaic, or Japanese–Korean, comparative framework. However, particles are short words and the risk of chance resemblances is therefore not small. More importantly, it is a fact which is not usually addressed explicitly in comparative studies that no significant subset of particles within OJ represents an inherited *system*. This is well illustrated by the case particles of which only *ga*, *i*, *wo* and *rwo* lack a transparent internal source, as opposed to *no*, *tu*, *ni*, *ywori*, *to*,

Table 3.17 *Pronominal forms of OJ*

	Short	Long	Locational
Personal			
1st	***wa, a***	***ware, are***	
2nd	***na***	***nare***	
3rd	***si***	–	
interrogative	***ta***	***tare***	
reflexive	***ono***	–	
Demonstrative			
proximal	***ko***	***kore***	***koko***
non-proximal	***so***	–	***soko***
interrogative	***idu-***	***idure***	***iduku***

na, *kara*, *pye* whose grammmaticalization from other parts of speech is either documented or easily recoverable. This suggests that comparison of individual particles with forms in other languages has to be interpreted with caution.

3.8 Pronouns

The main pronominal forms of OJ are shown in Table 3.17. In addition to these forms, there is an interrogative noun *nani* 'what' which does not form part of any morphological system. There are also a number of alternative terms of address, see 3.8.2.1, and there is a trace of an earlier proximal demonstrative *i*, see 3.8.5. Some descriptions include the noun *woto-~woti* 'distant place or time' as a 'distal' demonstrative, but that is a lexical noun.

3.8.1 Short versus long forms

A conspicuous feature of the pronominal system is the existence of short and long forms. The original function of the *-re* of the long forms is not known; it has been hypothesized to reflect the plural marker *-ra* + case particle *-i*, *ra-i > *-re* (Vovin 1997). Whatever the original grammatical distinction between short and long forms, it is not systematically reflected in their use and distribution in OJ. In OJ, the two sets of forms are formally in free variation in many environments, but in complementary distribution with respect to some criteria:

(135)

	short	**long**
Use in isolation	–	+
Use with genitive particles	+	–
Use in nominal compounds	+	–

The long forms are used in isolation, i.e. without a following particle, as subject/topic and as emphatic and exclamatory forms, e.g. (136a, b). The short forms are not used in this way, but must be followed by a particle. Conversely, the long forms are not used with genitive particles (apart from a small number of examples of *idure-no* (136b), which has no corresponding independent short form, and of *kore-no*) nor as a modifying first element in a nominal compound. The overwhelming majority of examples of the short forms are with genitive particles, used both attributively and as subjects in subordinate clauses: *wa ga kokoro* 'my heart'; *wa ga mure-inaba* 'when I go away' (*KK* 4). The personal pronouns (*a*, *wa*, *na*, *si*, *ta*, *ono*) take *ga* and the demonstratives (*ko*, *so*) take *no*.[19] The contraction involved in forms such as *wagipye* 'my house' suggests that in some cases short form + genitive particle had been univerbated into a possesive pronoun, here *waga* 'my', so that *wagipye* < *waga ipye* (see 2.6.1). In addition, the short forms (except for *si* and *ta*) are used to form compounds, e.g. *a-duma* 'my wife', *wa-dori* 'my (own) bird', *na-dori* 'your bird', *ko-yopi* 'tonight'; usually the second member of the compound undergoes *rendaku* (2.6.2). Note that interrogative *idu-* is found only as a constituent of derived forms and compounds (except for a single example in the meaning 'where' in EOJ, *MYS* 14.3549), never with a particle, genitive or otherwise.

Thus, use of the short forms is quite restricted, most examples being in compounds or with a genitive particle; this was the only use of the short forms which survived productively into EMJ. In OJ they are, however, also found with other case particles and with focus particles, which seems to reflect an earlier wider use as free pronouns, e.g. (136c). There are even parallel examples with full equivalence between short and long forms, such as *KK* 4 which in a repetitive sequence has . . . ***kore pa pusapazu*** 'this will not do' (*pusap-* 'befit, be suitable') in the first instance, but . . . ***ko mo pusapazu*** 'this too will not do' in the second. In contrast with the short forms, the long forms are used more freely, but it must also be noted that some long forms were not well established in OJ. Non-proximal *sore* is not attested in any of the core sources and has only two marginal attestations (in the *Kakyō hyōshiki* and in handwritten material). Thus *sore*, which became very frequent from EMJ onwards, was not part of the productive OJ system and is not included in Table 3.17 above. Likewise, reflexive *onore* is only attested twice (*MYS* 12.3098, 16.3883) and 2nd person *nare* is textually rare. Both are frequent in EMJ. Also proximal *kore* is not as highly frequent as in following periods. Thus, rather than presenting a system with a stable grammatical distinction, such as combining (=short) versus free (=long) forms, the OJ pronominal system was in a state

[19] There is a small number of examples of *ga* used with the demonstrative *so* when it functions as an inanimate 3rd person pronoun, e.g. *so ga pa* 'its leaves' (*KK* 101), which however has *so no pana* 'its flowers' in the next line.

of morphological transition with long forms replacing older short forms whose use was gradually being circumscribed, regardless of what the original grammatical distinction between them may have been.

(136) a. *wegusi* *ni* ***ware*** *wepi-ni-kyeri*
saké.of.smiles DAT I get.drunk-PERF-MPST.CONCL
'**I** have become drunk on the saké of smiles' (*KK* 49)

b ***idure*** ***no*** *sima* *ni* *ipori* *se-mu* ***ware***
which COP.ADN island DAT hut do-CONJ.ADN I
'**I**! on **which** island shall I make my hut (for the night)' (*MYS* 15.3593)

c. ***a pa*** *mo* *yo* *mye* *ni* *si* *areba*
I TOP ETOP EMPH woman COP.INF EMPH exist.PROV
na *wokite* [= *wo okite*] *wo* *pa* *na-si*
you ACC.put.GER man TOP exist.not-ACOP.CONCL
'**Me**, because I am a woman, apart from **you** I have no man' (*KK* 5)

3.8.2 Personal pronouns

The personal pronouns form a morphological class by taking *ga*, never *no*, as genitive marker. Furthermore, the personal pronouns, as opposed to non-pronominal terms of address, were not used with plural markers in OJ. This may be taken to support Vovin's etymology for the long forms as involving a plural marker, but note that 1st person *are* seems not to have been used with plural reference. 1st and 2nd person pronouns were used frequently in OJ, much more than in later stages of the language. Among the 1st person forms, the *a-* forms are only used for exclusive, singular reference ('I, me (alone)'), whereas the *wa-* forms also are used for inclusive and plural reference ('me/us (including you)') and also reflexively ('myself, oneself'), suggesting that the *wa-* forms reflect an earlier indefinite personal pronoun 'one'. The *wa-*forms were used more than the *a-* forms and there is also some overlap in usage, indicating that a shift from *a-* to *wa-* as the 1st person pronoun was in the course of being completed; the *a-* forms went out of use and are not used in EMJ. An EOJ form *wanu*, corresponding to central *ware*, is attested in a few cases. *Maro*, which in EMJ was used as a 1st person pronoun with some frequency, is attested in a song found in both the *Kojiki* (*KK* 48) and *Nihon shoki* (*NSK* 39) in the phrase *maro ga ti* 'my father'.

The 3rd person *si* is used both with animate and inanimate reference. *Si* is not used much, however; 3rd person reference is mostly expressed by the

non-proximal demonstrative *so*, from which *si* is diachronically derived (explaining why *si* alone among the short forms does not form compounds): *si-i > *si*. *Si* is sometimes said to be used for 2nd person reference, but the few examples which may be cited in support of this are not persuasive. *Ono* is sometimes believed to have alternated with an obsolete word *ana* 'self'; *na* is sometimes said originally to have been used for the 1st person and to be a reduced form of *ana*, but, again, with little good evidence.

3.8.2.1 Other terms of address

In addition to 2nd person pronouns, OJ had a number of terms of address of which the following are the most prominent. Whereas the personal pronouns proper do not combine with plural markers, some of the alternative terms of address do.

(137) *imasi*, *masi*, *mimasi* 'you, respectful'; cf. *imas-* 'be, exist, respect'; *mimasi* is thought to be more respectful than *imasi* and *masi* and to be from the honorific prefix *mi-* + *imasi*; another possibility is that *mi* represents a heavily nasalized initial [ĩ]*masi*

kimi 'you (my lord), respectful'

namuti 'you' (originally respectful, but neutral at the OJ stage; < *na* 'you' + *-muti* 'esteemed person; honorific suffix in names and titles', cf. *mutu-* honorific prefix)

namutati 'you, plural' (thought to be from *namuti-tati*)

wake 'you, pejorative' (also 'I, humble') (< 'lowly person')

ore 'you, pejorative'

i 'you, pejorative' (only *i-ga* 'you-GEN')

3.8.3 Demonstratives

Most accounts of OJ demonstratives posit a three-term 'proximal – mesial – distal' system, built on *ko – so – ka*. However, there is no evidence within OJ of *ka* being a productive member of the demonstrative system. Two forms are attested in OJ: long *kare* is found once, see (138d) below; what may be taken to be short *ka*, as distinct from the adverb *ka* 'this way', is attested at most twice, both in EOJ poems (*MYS* 14.3565, 20.4384). While these forms most likely represent the budding of the distal demonstrative which is so frequent in EMJ, they did not form a central part of the OJ system of demonstratives. Other *ka*-based forms often cited are in fact attested only from EMJ.

The description of the semantics of the OJ demonstrative system is due to Hashimoto Shirō (1966) whose study is the first to consider the OJ system on its own merits, rather than in terms of the EMJ system. The *ko-* versus *so-* system is entirely speaker based, with no primary reference to the hearer.

Proximal, *ko-*, refers to what is within the speaker's domain of direct sensory perception, or experience. **Non-proximal**, *so-*, refers to what is outside of the speaker's domain of direct experience. The facts of the use of the main *ko-* and *so-* forms in OJ are as follows:

- The *ko-* forms are almost entirely used deictically, referring to what may be directly experienced by the speaker; the only form used anaphorically is *koko*.
- The *so-* forms are mostly used anaphorically, with some examples of reference to something which is implied, but has not been mentioned explicitly; temporal deictic reference to past events is not infrequent. This anaphoric, or conceptual, and temporal deictic reference follows from the definition as being outside the speaker's domain of direct experience. There are no clear examples of spatial deictic use of *so*; also *soko* is mostly used anaphorically, but there are a few examples of spatial deictic use with reference to the hearer.

The ternary 'proximal – mesial – distal' system of EMJ and later arose, Hashimoto argues, through a subdivision of the direct domain into 'close vs. far', with the form *ka* being drawn in as an alternant of *ko*. Note that the reference of the single OJ example of *kare*, (138d), is within the field of direct visual perception of the speaker. The few examples of *soko* with spatial deictic use form part of the development of the three-term system.

In addition to the three main forms built on the proximal – non-proximal – interrogative/indefinite bases *ko – so – i* included in Table 3.17, there are a number of other forms, see Table 3.18. While these forms show the pervasiveness of the *ko – so – i* system, and no trace of a distal *ka*, it is also clear that the system is not as well developed as in EMJ and later stages. Some of what later become derivational morphemes expressing syntactic and/or fairly definite semantic categories, were at this stage apparently semantically vaguer. The designations for the derived categories are those used for later stages of the language; it is not clear that they are entirely appropriate for OJ. Locational *-ko/-ku* is thought to be from an obsolete noun *-ka/-ko/-ku* 'place', cf. e.g. *miyako* 'capital; palace-place'; note, however, the widespread use of short *ko* to mean 'here', as well as the use of *koko* and *soko* without locational meaning. Likewise, *koti* is not particularly directional, but rather meant 'this way, this side'. Recall that *sore* was not yet a productive part of the system (3.8.1).

The secondary *-du-* in the interrogatives (of unknown meaning and origin) was reinterpreted as part of the base in what emerged as the productive system; *idu* later changed to the *do* of the well known *ko – so – a – do* system of NJ. Other OJ forms built on *i-du-*: *idura* 'where(abouts)', *idupye* 'which direction'

Table 3.18 *OJ demonstrative forms*

	Proximal ***ko(-)***	Non-prox. ***so(-)***	Interr. ***i-/idu-***
Short	*ko*	*so*	*idu-*
Long, *-re*	*kore*	–	*idure*
Locational, *-ko/-ku*	*koko*	*soko*	*iduku*
Directional, *-ti*	*koti*	–	*iduti*
Degree, quantity, *-kV/-ku*	*kokV-*	*sokV-*	*iku-*
Manner	*ka*	*sate*	–
Manner, (*-ku/*)*-ka*	*kaku*	*sika*	*ika*
Time, *-tu*	–	–	*itu*

(*pye* 'side'), *idusi* 'which direction' (*-si* 'side'). Attested OJ forms built on *i-ku-* 'how much' *-ko-kV-* 'this much' *-so-kV-* 'that much' with the derivational elements *-ra*, *-da*, *-ba* and the adverbializer *-ku*: *ikura*, *ikuda*; *kokoda*, *kok(w)ida*, *kokodaku*, *kok(w)idaku*, *kokoba*, *kokobaku*, *kok(w)ibaku*; *sokoraku*, *sokidaku*, *sokoba*.

(138) a. *are* *pa* *wasurezi* ***ko*** *no* *tatibana* *wo*
I TOP forget.NCONJ this GEN orange.blossom ACC
'I will never forget it, **this** orange-blossom' (*MYS* 18.4058)

b. wakare-ko-*si* ***so*** *no* pi
part-come-SPST.ADN that GEN day
'**that** day when I left' (*MYS* 17.3978)

c. *uwe-si* ta *mo* *maki-si* *patake*
plant-SPST.ADN paddyfield ETOP sow-SPST.ADN paddyfield
mo *asa-goto* *ni* *sibomi-kare-yuku.* ***So***
ETOP morning-each DAT wither-dry-do.more.and.more That
wo mi*reba* . . .
ACC see.PROV
'For each morning the planted rice fields and the sown fields wither and dry out more and more. When I see **that** . . .' (*MYS* 18.4122)

d. *a* *gamopu [= ga omopu]* *kimi* *ga* *mi-pune* *kamo* ***kare***
I GEN.think.of.ADN my.lord GEN HON-boat Q that
'is it the boat of my beloved lord, **that**' (*MYS* 18.4045)

3.8.4 Basic pre-Old Japanese pronominal system

Based on the OJ forms, a simple and basic pronominal system may be posited for pre-OJ, see (139), with a 'speaker – non-speaker – interrogative' system for both the personal and the demonstrative pronouns, and including *wa* as an indefinite:

(139) *Personal*

speaker	1st	***a***
non-speaker	2nd	***na***
indefinite		***wa***
interrogative	'who'	***ta***

Demonstrative

speaker	proximal	***ko***
non-speaker	non-proximal	***so***
interrogative	'which'	***i-***

Reflexive ***ono***

3.8.5 Proto-Japanese demonstratives

Although the OJ system of demonstratives clearly is a two-way 'speaker – nonspeaker' plus interrogative system, it seems in fact that it goes back to a three-way 'proximal – mesial – distal' plus interrogative system, like that found in EMJ and later stages of the language, and like that found in Korean, as proposed by Frellesvig and Whitman (2008), see (140) and (141).

(140)

pJ	Proximal ***i**	Mesial ***kɨ**	Distal ***sɨ**	Interrogative ***e**
pre-OJ.a	Proximal ***i**	Mesial ***kɨ**	Distal ***sɨ**	Interrogative ***i**
pre-OJ.b		Participant ***kɨ**	Non-participant ***sɨ**	Interrogative ***i**
OJ	Speaker ***ko (~ i)***		Non-speaker ***so***	Interrogative ***i- ~ idu-***
EMJ	Proximal ***ko***	Mesial ***so***	Distal ***ka***	Interrogative ***i- ~ idu-***

The system posited for pJ includes interrogative **e* and proximal **i*, lost by the time of OJ, but vestigially attested in a lexicalization such as *ima* 'now' < **i* + *ma* 'space'. The change between pJ and pre-OJ.a of interrogative **e* > **i* is a result of mid vowel raising (see 2.7.2.3), resulting in homonymy

between proximal and interrogative. This homonymy (between two paradigmatically opposed terms) was resolved by eliminating proximal **i* and reinterpreting **kɨ* and **sɨ* as speech event participant and non-participant, respectively, in pre-OJ.b and then in OJ as speaker and non-speaker. Typological pressure then resulted in the system being augmented with a distal *ka*, but not until EMJ.

It has long been observed that the MK mesial and distal demonstratives, *ku* and *tye*, respectively, present a good form fit with Japanese proximal and mesial *ko* and *so*, but that the semantics do not fit. The changes between pJ and OJ explain that; and the pJ system is a good phonological *and* semantic fit with the MK system.

(141)

	Proximal	Mesial	Distal	Interrogative
MK	***i***	***ku***	***tye***	***e***
pJ	***i**	***kɨ**	***sɨ**	***e**

REFERENCES

General: Iwai 1970, Vovin 2005, 2008, Yamada 1913/1954. Verb morphology: Frellesvig 2007. Derivatives: Martin 1987: 790ff., Unger 1977/1993: 127ff. Uses of the adnominal and nominal forms: Wrona 2008. Tense and aspect: Takeuchi 1987, 1999: 101ff., Washio 2002. Adjectives: Kawabata 1976, Martin 1968, 1987: 801ff., Unger and Tomita 1983, Yamazaki 1992. Pre-OJ verb morphology: the seminal article is Ohno 1953; further: Frellesvig 2001, 2008, Hattori 1979, Martin 1987, Ohno 1978a, Unger 1977/1993, Whitman 2008. Particles: *Ga*, *no*, *wo* and OJ syntax: Kuroda 2007, Shibatani 1990: 333ff., Yanagida and Whitman 2009a, Wrona and Frellesvig 2010. Focus particles: Ohno 1993, Takeuchi 1999: 169ff., Whitman 1997, Wrona 2008: 195ff. Etymologies: Frellesvig 2001, Martin 1990, 1991, Miller 1971, Murayama 1957. Pronouns: Takeuchi 1999: 180ff., Whitman 1999, Yanagida and Whitman 2009b. Demonstratives: Frellesvig and Whitman 2008b: 27ff., Hashimoto 1966, 1982.

4 Loanwords

The etymological study of the lexicon of OJ, the question of its provenance, and the sorting of inherited from borrowed wordstock are part of the issue of the genetic affiliation of Japanese: OJ words which are similar to, or take part in sets of sound correspondences with, words from another language may be cognate with these words; they may be borrowings from the other language; they may themselves be the source of borrowings into the other language; or the similarities may simply be due to chance. We are more fortunate for the later historical stages of Japanese, but for the OJ language as it presents itself to us, we simply very often do not know and many attempts to identify old loanwords within OJ are highly speculative.

It is beyond doubt that OJ includes old loanwords from the languages around Japan – especially words relating to agriculture, seafaring, warfare, spiritual and religious life, government, and administration – but that we will not be able to identify many of them as loanwords on other than extra-linguistic grounds. It is for example a strong hypothesis that OJ *iraka* 'roof, roof tile' is a loanword, but we do not know from where. In other cases we believe that a word must be borrowed and can come up with several likely sources but cannot choose between them. It is, for example, very likely that the Japanese word for 'horse', OJ *uma*, is borrowed and there are indeed words in surrounding languages which mean 'horse' and which are similar to *uma*, e.g. EMC *maɨ', MK *mol*, Mongolian *morin*. Mongolian and Korean may be genetically related to each other and/or to Japanese, so the MK and Mongolian forms may be cognate, but one language may also have borrowed the word from the other. And is the Chinese word the source of borrowing, or, perhaps more likely, itself borrowed?

On the other hand, note should be taken of the recent work of Unger (2001, 2003) which provides a linguistic basis for identifying candidates for loanwords. Unger brings attention to a number of cases in which a word in Korean (e.g. MK *pal* 'leg, foot') corresponds phonologically to a Japanese word with a narrower meaning (*pagi* 'shin'), whereas the Japanese word (*asi* 'leg, foot') which is synonymous with the Korean word is unrelated to it, that is:

(1)

OJ	MK
pagi 'shin'	*pal* 'leg, foot'
asi 'leg, foot'	

Unger argues that in such cases it is likely that the *pagi* type word is inherited, but was displaced, and became semantically specialized, by the *asi* type word which may either be from a substratum or a super- or ad-stratum, much like the displacement and specialization in English of Anglo-Saxon *tell* and *loft* by *count* and *air* borrowed from French. On the other hand, it is of course quite possible that displacement or replacement of the *asi* type word was entirely internal and not motivated by borrowing.

Linguistic borrowing presupposes some degree of bilingualism. It can be difficult to draw a line between loanwords and foreign words. Foreign words are not infrequently used by bilingual speakers. This is known as *code mixing* and takes place for various reasons including prestige, novelty, or clarity. A foreign word used in this way can gain currency and eventually be accepted by members of a speech community as part of their language. In the transitional phase the word will be foreign to some speakers but native to others. It may be pronounced with the phonology of the foreign language, with an adapted native phonology, or totally phonologically (and morphologically) assimilated to the native language. It is only in the latter situation we shall talk about loanwords as an established part of OJ. This point is particularly relevant when considering loanwords from Chinese.

It is traditional to distinguish between 'loanwords' and 'Sino-Japanese' vocabulary, the former usually designating loanwords from other languages than Chinese. Here the distinction will be made between fully assimilated nativized loanwords, including those of Chinese origin, and the Sino-Japanese vocabulary which has been used through the history of Japanese. The two are different and only the former will be discussed here. Sino-Japanese will be discussed in Chapter 9, but it should be said here that Chinese served not only as a source of Chinese vocabulary but also as a medium for the introduction of much Buddhist vocabulary originally from Sanskrit. Sacred names and terms were usually not translated into Chinese, but adapted phonologically and in writing transcribed phonographically (1.1.2.4).

4.1 Ainu

It is difficult to identify loanwords in OJ from Ainu, perhaps with one notable and remarkable exception: OJ *kamwi* ~ *kamu-* 'spirit, deity' may well be borrowed from an ancestor of Ainu *kamuy* 'bear; deity' (although the opposite direction of borrowing has also been proposed). Recently it has also been

suggested by Vovin (2009) that some OJ words (including place names) attested only in EOJ (see Chapter 5), e.g. *sida* 'time', in fact reflect prehistoric borrowings from Ainu. Though not lexical borrowing as such, it should be mentioned that Ainu words are also preserved in some place names in Japan. Best known are those ending in *-betu*, Ainu *pet* 'river', or *-nai*, Ainu *nay* 'stream, valley, river'. It is likely that many other place names reflect now unrecognizable Ainu words.

4.2 Continental loanwords: Korean, Chinese, Sanskrit

We do not know what languages other than the ancestor of Japanese were spoken in the Japanese archipelago in pre- and proto-historic times. The time depths involved make it hazardous to attempt to identify borrowed (or sub-stratum) vocabulary from Austronesian or Austro-asiatic, although some languages belonging to those families may have been spoken in Japan at the same time as Japanese. For example, OJ has a word *tape* 'bark-cloth; cloth made from the bark of the mulberry tree' which must somehow be related to the Austronesian word *tapa* which means exactly the same, but we can say nothing about the direction or route of transmission. However, it is possible to point out a number of OJ words which in all likelihood are loans and for which we can plausibly identify a source, from Korean, Chinese or Sanskrit.

In proto-historic times, many Japanese speakers would have had some facility in one or more foreign languages, ranging from fishermen and traders communicating with their continental colleagues to highly learned clergy. Different continental languages, especially Korean languages and varieties of Chinese, would have been used at different times and particularly within limited social or professional circles. Most continental culture was in pre- and proto-historic times transmitted to Japan via the Korean peninsula. Today Korean is one language with dialects which all descend from Middle Korean, which in turn continues the linguistic tradition of the Kingdom of Shilla. Until unification under Shilla in 668 there were, however, three main kingdoms on the Korean peninsula: Koguryo, Paekche, and Shilla. From the early fifth century contact with and immigrants from the kingdom of Paekche seem to have played a particularly significant part in the transmission of continental material and intellectual culture, including Chinese language, writing, learning, and also later, Buddhism which seems to have been introduced from around the middle of the sixth century. With this came new vocabulary. In addition to loanwords from Korean languages, much borrowed vocabulary of ultimately further origins must be thought to have entered Japanese through some Korean language, or through Korean speakers, but in most cases details remain obscure.

4.2.1 Korean

The following words have been thought to have been borrowed from a Korean language. Usually we do not know which, but must simply assume an earlier cognate form of an attested Middle Korean word as the source, (2a) below, but in a few cases we can tentatively identify Paekche as the source language (reconstruction of Paekche forms follows Bentley 2000), see (2b). If Japanese and Korean are genetically related, some of these words may in fact be cognate rather than borrowings, for example *kudira*, *para*, or *uri*.

(2) a. ***karamusi*** 'ramie (fabric, cloth), Chinese silk plant', MK *mwosi* 'ramie fabric, cloth'; cf. OJ *kara-* 'China, Korea, foreign'
kasa 'bamboo hat, umbrella', MK *kas* 'id.'
kudira 'whale', MK *kworoy* 'id.'
mori 'woods', MK *ˈmwoyh* 'mountain', pK **mwoˈlih*
para 'field, plain', MK *pel* 'id.'
patake 'field', MK *path* 'id.'
pyera 'spatula, pallet', MK *pyet* 'moldboard'
sarapi 'spade' (EMJ), MK *salp* 'id.', possibly further from OC *tshrap; cf. OJ *sapi*, *sapye* 'spade' (see (3a) below)
sitogi (EMJ) 'rice cake for ceremonial purposes', MK *ˈstek* 'rice cake'
uri 'melon', MK *ˈwoy* 'cucumber', pK *woˈli*

b. ***kopori*** 'district', Paekche *kəpəri, MK *ˈkwo wolh*
kuti 'hawk', Paekche *kutI 'falcon'
kwi 'fortress, walled city', Paekche *kɨ 'id.'
sasi 'walled city', Paekche *casI, MK *ˈcas* 'id.'

4.2.2 Chinese

The following may be thought to be early loans from Chinese. At least some of the words in (3a) are quite old and represent direct borrowings. Those in (3b) are transparent and probably not very old borrowings from Chinese, but they are not usually thought of as SJ vocabulary; two (*gakwi* and *zeni*) have initial media, which were not allowed in the native vocabulary (2.7.1.2). In Chinese historical phonology EMC and LMC are fairly securely reconstructed and more recently our understanding of OC phonology, too, is improving. We are therefore sometimes able to determine which stage of Chinese a borrowing originates in. (Reconstructions of EMC and LMC are from Pulleyblank (1991); the reconstructed OC forms follow Miyake (1997 and p.c.).)

(3) a. ***kama*** 'pot', 坩 OC *khaam
kama 'sickle', 鎌 OC *gryam
ke 'spirit' 気 EMC *k^{h}ɨj^{h}, OC *khɨys
kinu 'silk', 絹 OC *kwyans (EMC *kjwian^{h})
kuni 'country', 郡 OC *guns (EMC *gun^{h})
saga 'characteristic; good omen', 性 OC *saŋ (EMC *siajŋ^{h}) 祥 EMC *ziaŋ
sapi, ***sapye*** 'spade', 鍤 EMC *tʂ^{h}əip/tʂɛ:p, OC *tshrap. Cf. *sarapi* (see (2a) above). It is possible that OJ *sapi/sapye* was borrowed from EMC while the ancestor of MK *salp* was borrowed from OC, further being borrowed into EMJ in the shape *sarapi*
sugu-roku no saye 'pair-six-Gen-game' (name of a game of dice), written 双六乃佐叡 in *MYS* 16.3827. *Sugu* is the traditional reading of 双 OC *sroŋ, EMC*ʂaɨwŋ/ʂœɨwŋ; *saye* which is spelt out phonographically reflects 賽 LMC *saj ` 'game (of dice)'. From EMJ this word became *sai*
ume 'plum', 梅 OC *hmay

b. ***gakwi*** 餓鬼 EMC *ŋa^{h} kuj' 'glutton, hungry ghost'
pakase 博士 EMC *pak dʑɨ' 'expert, authority'. This word is not phonographically attested in OJ, but was surely used as it was an important official title in the *ritsuryō* system. The regular SJ character readings were used in coining the SJ word cNJ *hakushi* '(academic) doctor, PhD'
puse 布施 EMC *pɔ^{h} ɕi^{h} 'temple offering, charity' (Chinese loan translation of Skt. *dāna* 'offering, alms')
saka 尺 EMC *tɕiajk 'unit of measure; *shaku*'
we 画 EMC *ɣwaɨj^{h} 'picture, drawing'
zeni (EMJ) 銭 EMC *dzian 'money'

4.2.3 Sanskrit

The words in (4) below can be traced back to Sanskrit (or, in one case, Pali). Almost all ultimately derive from Buddhist contexts but became everyday words and most remain in use today. A few words are included which are not attested until EMJ but which were probably in use in OJ. In a few famous examples we can trace the route of transmission and find both Chinese and Korean related forms (4b), but for several of the oldest and most naturalized loanwords that is not possible (4a). It is worth noting that in addition to Buddhist inspired vocabulary, a word for 'rice', which surely must be old, is

of Sanskrit origin. It is also worth noting that most of these words survive into the modern language.

(4) a. ***ama*** (EMJ) 'nun', Pali *ammā* 'mother'
kapara (EMJ) 'ceramic roof tile', Skt. *kapāla* 'cup, jar, dish; cover'
kasa 'scab, the pox', Skt. *khasa* 'itch, scab'
mara 'penis', Skt. *māra* 'death; the evil one, the tempter; god/passion of love'
pata 'banner, standard', Skt. *patākā* 'id.'
sara 'plate', Skt. *śarāva* 'shallow cup, dish, plate'
uru- (EMJ) in e.g. *urusine* 'nonglutinous rice', Skt. *vrīhi* 'rice' (OJ *-sine* (~ *ine~ina-*) 'riceplant')

b. ***potoke*** 'Buddha, Buddha image', MK *pwuthye* 'Buddha', 仏陀 EMC *but t^{h}a 浮屠 OC *buu daa, Skt. *buddha* 'Buddha'
pati 'bowl', MK *pali*, 鉢 OC *pat, Skt. *pātra* 'vessel'
tera 'temple', MK *tyel* 'temple', 刹 EMC *tʂ^{h}aɨt < OC *tshraat/ksraat, Skt. *kṣetra* 'place'
kyesa 'priest's robe' 袈裟, EMC *kaɨ ʂaɨ; LMC *kja: ʂa:, Skt. *kaṣāya* 'the (yellow) robe of Buddhist clergy'. In addition to preserving *kesa* in the original meaning, NJ also reflects this word in *oogesa-na* 'pompous' (*oo-* 'big').

The words in (5) are some examples from the EMJ period of common Buddhist names and terms ultimately deriving from Sanskrit, but which were taken in via Chinese renditions.

(5) ***aka*** 'Buddhist water offering; container for this; wine (priests' secret language)' 閼伽, EMC *ʔat gɨa, Skt. *argha*, 'value', *arghya* 'water offered to a guest at a respectful reception'
amida 'Amitabha Buddha' 阿弥陀, EMC *ʔa mji t^{h}a, Skt. *amitābha*
bosatu 'Boddhisatva' 菩薩, EMC *bɔ sat, Skt. *bodhisattva*
butu 'Buddha, Buddhism' 仏(陀), EMC *but (t^{h}a), Skt. *buddha*
daruma 'Bodhidharma' 達磨, EMC *dat ma, Skt. *bodhidharma*
naraku 'hell' 奈落, EMC *na^{h} lak, Skt. *naraka* 'id.'
setuna 'instant' 刹那, EMC *tʂ^{h}aɨt na^{h}, Skt. *kṣaṇa* 'instant, moment'

4.3 Phonological adaptation

In most cases we do not know the exact shape of the proximate source of loanwords in OJ and it is therefore difficult to say much about the phonological adaptations that took place in the course of the borrowing. Some forms appear shortened or otherwise simplified in comparison with the distal source, e.g. OJ *uru-* from Skt. *vrīhi*, but it is not possible to know where in the course of transmission of this word the abbreviation took place. In particular, if a word passed through Chinese it would have been made to conform to its largely monosyllabic morpheme structure, cf. Skt. *pātra-* giving OC *pat.

It is easy, however, to see that some adaptation to the simple CV syllable structure of OJ took place, most conspicuously in the insertion of epenthetic vowels to avoid syllable final consonants. The addition of *-i* or *-u* to loanwords is well known from NJ, and OJ ***kinu*** and ***kuni*** are examples of that, as is perhaps also ***pati*** (although the MK shape ***pali*** makes us suspect that the *-i* may predate the arrival of the word in Japan). However, this seems to have become more common from EMJ onwards. A more widely used strategy in the older loanwords was the insertion of a post-consonantal echo vowel to echo the preceding vowel, e.g. OJ ***kasa*** (cf. MK *kas*), ***kama*** (OC *khaam), ***pakase*** (EMC *pak dʑɨ'), ***para*** (MK *pel*) are straightforward examples of this, but others include ***pyera*** < pre-OJ *pyara (MK *pyet*), ***tera*** < *tyara (MK *tyel*).

REFERENCES

General: Miller 1967, K. Satō 1982, T. Satō 1982, Umegaki 1978, Unger 2001, 2003. Korean: Bentley 2000, Kanno 1978. Chinese: Miyake 1997. Sanskrit: Suzuki 1978.

5 Eastern Old Japanese

The language of the overwhelming majority of OJ sources is that of the capital Nara, or the area around it. However, some sources reflect dialects from eastern Japan, usually collectively referred to as Eastern Old Japanese. Volume 14 of the *Man'yōshū* consists of 230 *azuma uta* 'eastern songs/poems', out of which 89 have the place of origin identified in the commentary. Most of these were recorded by people from the capital travelling to the east in various official capacities. Volume 20 includes 93 *sakimori uta* 'borderguard songs', composed by soldiers from the east, mostly serving on guard duty in Kyushu, again mostly recorded by scribes from the capital area. Finally, the *Hitachi fudoki* includes 9 songs in EOJ. This material does not reflect a single dialect, but at least three main areas which may be labelled: (a) northern EOJ, comprising poems from (or dialects spoken in) the provinces of Kazusa, Mutsu, and Shimotsuke; (b) central EOJ: Hitachi, Kōzuke, Musashi, Sagami and Shimōsa; and (c) southern EOJ: Shinano, Suruga, and Tōtōmi. These dialect areas differ appreciably between each other, with southern being least and northern most different from the central OJ dialect of the capital area. However, the majority of the poems and songs which are identified as EOJ have no information about their provenance.

The EOJ material reflects features of EOJ to varying degrees, presumably due to the fact that most of it was written down by speakers of the central OJ dialect who were not equipped – or inclined – to record all linguistic features in the EOJ poems they were writing down. It is thought that the *azuma uta* of volume 14 of the *Man'yōshū* among the EOJ material reflect least systematically the language of the eastern dialects. For example, there are more examples of phonological differences between central OJ and EOJ in the *sakimori uta* than in the *azuma uta*. The quality and especially small quantity of this material mean that what we can learn about EOJ dialects from it is limited. However, it is the only material which to any significant extent reflects dialectal material among the pre-modern Japanese text corpus, and at least one identifiably EOJ feature is reflected in standard cNJ (the vowel base imperative, see 16.4). In the following we outline some main features of EOJ, but do not attempt a full description or a characterization of individual dialect areas.

5.1 Phonology

In phonology, the EOJ material does not exhibit the *kō-otsu* syllable distinction in syllables with front vowels. That is to say, although the *man'yōgana* used to write individual words in the EOJ texts can be interpreted as having the sound values they had when writing OJ, there is no systematic distinction between Ce_1 (/Cye/) and Ce_2 (/Ce/), or between Ci_1 (/Ci/) and Ci_2 (/Cwi/). However, Co_1 (/Cwo/) and Co_2 (/Co/) appear to have been kept distinct. At present no phonological distinctions in EOJ which are not found in central OJ have been discovered. If they existed, the small quantity of the material would make them very difficult to identify.

Where central OJ contracted juxtaposed *-i-a to /-ye/ in some grammatical morphemes, EOJ gave /a/, e.g. the morphological stative *i-ar- > central OJ *-yer-* :: EOJ *-ar-*, or with the adjectival copula conditional: *-ki-aba > central OJ *-kyeba* :: EOJ *-kaba*.

5.2 Morphology

In morphology there are several well-known features of EOJ which differ from central OJ. First, EOJ is usually thought to have had a distinction between conclusive and adnominal among the QD verbs, see the basic paradigm in (1) of *yuk-* 'go', which also shows that there was no distinction between exclamatory and imperative:

(1)

	Central OJ	EOJ
a- stem	*yuka-*	*yuka-*
Infinitive	*yuki*	*yuki*
Conclusive	*yuku*	*yuku*
Adnominal	*yuku*	*yukwo*
Exclamatory	*yuke*	*yuke*
Imperative	*yukye*	*yuke*

This is based on examples such as (2a) below. However, the picture is not as clear-cut as it is often presented. First, the number of examples of the distinct EOJ adnominal form is very small, with many more examples of verbs in adnominal function, and the other functions associated with the adnominal form, ending in *-u* as in central OJ. Second, it is noteworthy that a few instances of the EOJ forms ending in *-wo* occur in contexts where usually a conclusive form would be used. Third, even in some songs which also have the distinct adnominal form in *-wo*, a form ending in *-u* is used in adnominal function, e.g. *ikiduku* in (2b), which also has the distinct EOJ adnominal forms *sumo* and *kinwo*.

(2) a. *ikapo no nero ni* ***purwo*** *ywoki*
Ikao GEN top DAT fall.ADN snow
'snow falling on the top of Ikao' (*MYS* 14.3423)

b. *oki ni* ***sumo*** *wokamo no mokoro*
offing DAT live.ADN small.duck GEN like
yasakadori ***ikiduku*** *imo wo*
deeply sigh.ADN beloved ACC
*okite ki-****nwo*** *kamo*
leave.GER come-PERF.ADN EMPH

'I have come, leaving behind my beloved who sighs deeply like the little duck that lives in the offing' (*MYS* 14.3527)

The distinct EOJ adnominal is an important form for reconstruction of pJ verb morphology. It may be thought to reflect a pre-OJ ending *-o, which had raised fully to *-u* in central OJ (by mid vowel raising, cf. 2.7.2.3), but only partially to *-wo* in EOJ. It is therefore usual to reconstruct pre-OJ conclusive/adnominal formants on the basis of the EOJ forms, as shown in (3). However, it is equally possible to reconstruct, as in (4), a single finite verb formant, pre-OJ *-o, which in central OJ exhibits full mid vowel raising (*-o > *-u*), but in EOJ has full (*-o > *-u*) and partial (*-o > *-wo*) mid vowel raising in different environments and is morphologized differently in different positions.[1]

(3)		Central OJ		pJ		EOJ
	Conclusive	*-u*	<	*-u	>	*-u*
	Adnominal	*-u*	<	*-o	>	*-wo*

(4)	Central OJ/NJ		p(re-O)J		EOJ
					-u
	-u	<	*-o	>	
					-wo

The adnominal of the adjectival copula appears in some cases as *-ke*, e.g. *naga-ke* 'long' :: central OJ *naga-ki*. This is thought to reflect an older shape, which is also attested in a single song in the *Kojiki*, which raised (again by mid vowel raising) *-ke > (pre-OJ *-kye* >) central OJ *-ki* :: EOJ *-ke* (~ *-ki*).

[1] Support for this hypothesis may be found in the fact that most occurrences of conclusive verb forms are found not in absolute sentence final position (which is one environment where partial mid vowel raising is expected), but before extensions and final particles which probably combined with the preceding verb to form a single phonological word, landing the verb ending in nonfinal position which is an environment which favours full mid vowel raising (cf. 2.7.2.3); this would explain why it is the conclusive which has the fully raised reflex of *-o. See further Frellesvig (forthcoming) about this.

EOJ has a negative auxiliary, *-(a)nap-*, which is not found in central OJ. It has main forms as shown in (5), which appears to be a hybrid, defective conjugation (note that the exclamatory only is attested as a stem), but which clearly is verbal as opposed to the composite central OJ negative auxiliary (cf. 3.1.4.5.1). An example is given in (6a). Other EOJ negative forms not found in central OJ are shown in (6b–c), labelled gerund and infinitive, respectively. (6b) also contains one of the few words that are found in EOJ, but not in central OJ, namely *sida* 'time' (which may be a loan from Ainu, cf. 4.1). In NJ, negative formation is a major isogloss which separates western Japanese dialects (including Kyoto) from eastern dialects, including standard cNJ (cf. 16.3), but it is not likely that the eastern and standard NJ forms reflect the EOJ negative.

(5)

Base	*-(a)nap-*
a- stem	*-napa-*
Conclusive	*-napu*
Adnominal	*-nape*
Exclamatory	*-nape-*

(6) a. tusima *no* *ne* *pa* *sita-gumo* ara-***napu***
Tsushima GEN top TOP under-cloud exist-NEG.CONCL
'There are no low clouds on the peak of Tsushima.' (*MYS* 14.3516)

b. *ywopi* *na* *pa* ko-***nani*** *ake-nu*
night you TOP come-NEG.INF dawn-PERF.CONCL
sida *kuru*
time come.ADN
'you won't come at night, but you will come when it has dawned' (*MYS* 14.3461)

c. *ura-gare* *se-**nana***
treetop-wither.INF do-NEG.GER
'When the (leaves on the) treetops do not wither …' (*MYS* 14.3436)

The imperatives of vowel base verbs attach *-ro*, as opposed to central OJ *yo*. The EOJ imperative is the only feature of distinct EOJ grammar which is directly reflected in standard cNJ (cf. 16.4).

(7)

Central OJ	EOJ
mi-yo	*mi-ro*

REFERENCES

Fukuda 1965; Mizushima 1984, 2003; Russell 2006, forthcoming; Vovin 2005, 2008.

Part II

Early Middle Japanese

6.1 Writing

From the late Nara period a gradual process of simplification and abbreviation of the shapes of *man'yōgana* took place, slowly at first but gaining speed through the early part of the Heian period, and eventually resulting in the development of *hiragana* and *katakana* scripts. As the reduced shapes were dissociated from the *kanji* from which they derived, they unambiguously became phonograms and it became possible to write Japanese phonographically in a simple and straightforward manner. From the beginning of the tenth century until the end of the eleventh century we have a large body of text which was written almost entirely phonographically in *hiragana*, with only a few logographically used *kanji*. This style of writing is a continuation of the tradition of extensive phonographic writing which developed in the context of writing or recording poetry in the OJ period. Most of the EMJ *hiragana* texts were written by women who traditionally were excluded from the world of learning and Classical Chinese. The first line of the *Tosa nikki* from 935 where its author, Ki no Tsurayuki, somewhat tongue-in-cheek pretends to be a woman trying her hand at writing diaries, is famous: *wotoko mo su naru nikki to ipu mono wo wonna mo site mimu tote suru nari* 'I as a woman will also try keeping this thing called diary, which men are said to keep.' It also shows, however, that it is a stereotypical oversimplification to view prose writing in *hiragana*, or in Japanese, during the Heian period as the exclusive domain of women. Also both men and women wrote poetry in *hiragana*.

6.1.1 Kanji-kana majiribun

From the middle of the Heian period the style of writing known as *kanji-kana majiribun* ('mixed *kanji* and *kana* writing') started being used more widely. *Kanji-kana majiribun* has a heavy preponderance of *kanji*, used logographically for most content and some grammatical words, and with *kana* (almost always *katakana*), used phonographically for some grammatical words and endings. The earliest extant text identified as being written in *kanji-kana*

majiribun is the *Tōdaiji fujumonkō* from the early ninth century, but it was not until after the middle of the EMJ period that *kanji-kana majiribun* became more widespread and was used in particular in the *setsuwa* literature (6.2.1). From the LMJ period it became widely used and largely replaced *hiragana* writing in prose. Thus the writing of extensive text passages in *hiragana* (or *katakana*) did not survive the early LMJ period and, regrettably perhaps, the elegant and economic tradition of simple *hiragana* writing, which is a fully sufficient means of representing Japanese, was lost. It also is worth noting that the establishment and spread of *kanji-kana majiribun* from the late LMJ period coincides with the establishment of the written norm of Classical Japanese. The way Japanese is written today is a direct descendant not of the *hiragana* writing of the EMJ period, but of the *kanji-kana majiribun* of EMJ and LMJ, although the proportions of logographic and phonographic writing differ between then and now, with more words spelled out phonographically today and with *hiragana* replacing *katakana*.

All logographic writing of Japanese derives from reversing *kanbun-kundoku* (the rendition in Japanese of a Chinese text) from reading to writing (see 9.1.2), but the link is particularly close for *kanji-kana majiribun* which arose as a transfer of the techniques of *kunten* (annotations to a Chinese text) to the writing of Japanese. Especially the practice of writing grammatical elements of a Japanese rendition in *kana* glosses next to Chinese text served as the model for the *kanji-kana majiribun* style of writing. In 9.1.1 there is a short example of an annotated Chinese text and it is conspicuous that its appearance is quite similar to Japanese texts written in *kanji-kana majiribun*. Also the Japanese language used in *kanji-kana majiribun* was heavily influenced in style, vocabulary and grammar by the Japanese language used in *kanbun-kundoku*, the so-called *kuntengo* (9.1.6).

6.1.2 Kana (hiragana, katakana, hentaigana)

The *hiragana* and *katakana* developed as reduced shapes of *man'yōgana*. Table 6.1 shows the standard sets of *kana* and their *kanji* origins in the usual 'fifty-sound table' (*gojūonzu* 五十音図, see further 6.1.4) grid arrangement, in ten columns by five rows plus an extra slot for the moraic nasal. The columns (*gyō* 行) are referred to by the head *kana*, e.g. *a-gyō* ア行 the '*a*-column' or *ka-gyō* カ行 the '*ka*-column', and the rows (*dan* 段 (or *retsu* 列) by the single vowel *kana*, e.g., *a-dan* ア段 the '*a*-row' or *i-dan* イ段 the '*i*-row'. Of these, the following *kana* from the *wa*-column were abolished in the writing of modern Japanese in the script reform of 1946 and are today only used in the writing and presentation of Classical Japanese: ゐ/ヰ, ゑ/ヱ, を/ヲ (except that the latter pair is used for the accusative particle *o*). As may be seen, in some cases equivalent *hiragana* and *katakana* originate in the same

Table 6.1 Kanji *origins of* kana

a	ka	sa	ta	na	ha	ma	ya	ra	wa	
あ 安	か 加	さ 佐	た 太	な 奈	は 波	ま 末	や 也	ら 良	わ 和	
ア 阿	カ 加	サ 散	タ 多	ナ 奈	ハ 八	マ 末	ヤ 也	ラ 良	ワ 和	
i	ki	si	ti	ni	hi	mi		ri	wi	
い 以	き 幾	し 之	ち 知	に 仁	ひ 比	み 美		り 利	ゐ 為	
イ 伊	キ 幾	シ 之	チ 千	ニ 仁	ヒ 比	ミ 三		リ 利	ヰ 井	
u	ku	su	tu	nu	hu	mu	yu	ru		
う 宇	く 久	す 寸	つ 川	ぬ 奴	ふ 不	む 武	ゆ 由	る 留		
ウ 宇	ク 久	ス 須	ツ 州	ヌ 奴	フ 不	ム 牟	ユ 由	ル 留		
e	ke	se	te	ne	he	me		re	we	
え 衣	け 計	せ 世	て 天	ね 祢	へ 部	め 女		れ 礼	ゑ 恵	
エ 江	ケ 介	セ 世	テ 天	ネ 祢	ヘ 部	メ 女		レ 礼	ヱ 恵	
o	ko	so	to	no	ho	mo	yo	ro	wo	N
お 於	こ 己	そ 曾	と 止	の 乃	ほ 保	も 毛	よ 与	ろ 呂	を 遠	ん 无
オ 於	コ 己	ソ 曾	ト 止	ノ 乃	ホ 保	モ 毛	ヨ 与	ロ 呂	ヲ 乎	ン

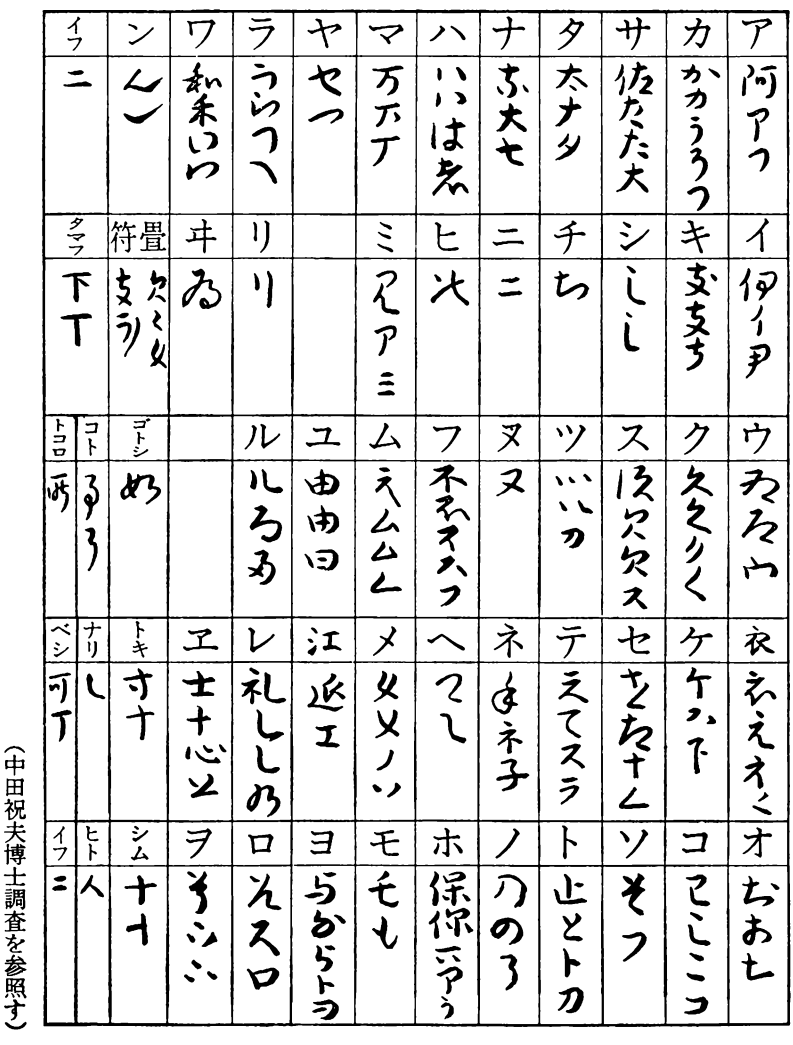
kanji (e.g. *ka*, *na*, *ma*, *ya*, *ra*, *wa*), in others not (e.g. *a*, *sa*, *ta*, *ha*). Overall, the reduction took place in two ways: as a general rule, *hiragana* resulted from a *cursive writing* of full *man'yōgana*, whereas *katakana* originate in *parts* of *man'yōgana*. It must be emphasized, however, that *man'yōgana* in their full shape continued to be used through the MJ period as phonograms. It is not the case that *hiragana* or *katakana* simply replaced *man'yōgana*: they coexisted for a long time.

Published editions of texts today use the standard sets of *kana*, but that inventory of graphs was only settled upon in the Elementary School Order (*Shōgakkōrei*) of 1900. Through the MJ period a large number of competing and variant, but phonographically equivalent, shapes of *kana* was used; these are today known as *hentaigana*. Table 6.2 below gives an impression of the variability among early *kana* shapes.

Legend ascribes the invention of *hiragana* to the priest Kūkai (774–835), but it is clear that *kana* were not invented by any individual or group. Many *man'yōgana* are graphically complex and cumbersome to write, and it is only natural that they would tend to be simplified and abbreviated for ease of writing. With the discovery of more primary material from the OJ period, mostly in the form of *mokkan* (wooden tablets, see 1.2.2), examples of graphically reduced *man'yōgana* are attested in increasing numbers also from the OJ period. It is important to note that many of the reduced shapes which became *hiragana* in Japan are found in Chinese cursive writing styles, and that many of the reduced shapes which became *katakana* are found in Korean *kugyŏl* (cf. 9.1.1). This suggests that in fact the *hiragana* and *katakana* letter shapes were not the result of independent developments in Japan, but followed continental models.

It is only possible to follow the gradual emergence of the *hiragana* and *katakana* letter shapes through authentic, contemporary writing, not later copies of manuscripts. In addition to *mokkan* and fragments of various sorts, many of which seem to be incidental writing, or even writing practice, the most important authentic, primary text materials from the EMJ period are annotations on texts, the so-called '*kunten*' materials (see 6.2.2). The study of the development of *kana*, especially *katakana*, and of *dakuten* (below) is intimately linked to the study of *kunten* materials. There exists a large amount of such materials in the possession of Buddhist temples in Japan, exhibiting great diversity in traditions of annotation, including letter shapes. Tsukishima (1986) reproduces in several hundred tables examples of *kana* shapes and other writing conventions drawn from such *kunten* materials. In Table 6.2 (taken from Tsukishima 1977: 98) are shown *kana* shapes (and a few other annotation conventions) from annotations (dating from 883) of the *Dìzàng shílún jīng* (地蔵十輪経, Japanese *Jizō-jūrin-gyō*), a Chinese translation of the 'Sutra of the Ten Cakras of Kṣitigarbha'. Even a cursory

Table 6.2 *Early* kana *shapes (from Tsukishima 1977: 98)*



inspection gives a clear impression of the diversity and also reveals that both *hiragana*- and *katakana*-like letter shapes are found in a single set of annotation. It is important to emphasize that all the letter shapes in Table 6.2 are from a single set of annotations and do not represent different steps in the development of letter shapes.

As such materials are further studied and published, the details of our knowledge about the development of the use of *hiragana* and *katakana* will change, but the general picture is as explained in this section.

The different strategies of reduction eventually resulted in two distinct sets of *kana*, *hiragana* and *katakana*, which were associated with functional differences: Parts of *man'yōgana*, which yielded *katakana*, were practical to use in annotations on texts where space is limited; in addition to that use, *katakana* also later became used in *kanji-kana majiribun*. Throughout the premodern period *katakana* maintained some connection with *kanbun* or writing incorporating *kanji*. Cursively written whole *man'yōgana*, eventually giving the modern *hiragana*, were thought aesthetically pleasing and were associated with literature and calligraphy. The *hiragana* were also practical for writing running text, as strings of letters could be written without lifting the brush from the paper. When talking about writing in the Heian period, *hiragana* has become known, and was also at the time occasionally referred to, as '*onna-de*' ('woman's hand'). Whereas academic, intellectual or public writing was mostly done in Classical Chinese (or in *hentai kanbun*, see 9.1.2.1), personal and private writing in Japanese in *hiragana* was associated with the leisure of the (female) aristocracy and as is well known, much of the prose literature in Japanese from this period was written in *hiragana* by women. One exception was composition and writing of poetry in *hiragana*, which to some extent was public and in which men certainly took part. As noted above it is wrong to view writing in *hiragana*, or in Japanese, during the Heian period as the exclusive domain of women.

It must further be noted that just as there was a gradual transition and long period of coexistence between *kana* and *man'yōgana*, so the letters which we today identify as *hiragana* and *katakana* were not functionally differentiated from the outset, or conceived of as belonging to different sets of letters; thus among the *hentaigana* we find *katakana*(-like) letter shapes, and conversely the *kunten* materials also hold many *hiragana*(-like) letter shapes mixed in with *katakana*, see Table 6.2. The *hiragana* and *katakana* for *e* illustrate this well: they originate in different *man'yōgana* which represented distinct syllables in OJ and early EMJ, え< 衣 (OJ /.e/) and エ < 江 (OJ /.ye/), and they can both be found within single bodies of text. Overall, however, the later functional differentiation between *hiragana* and *katakana* has its origin and basis in the two different strategies of simplification and tendencies in use.

6.1.2.1 Sei'on *and* daku'on

In OJ, different *man'yōgana* were generally used for *sei'on* and *daku'on* (i.e. syllables with initial tenuis (/p, t, k, s/) or *media* (/b, d, g, z/), respectively; see 2.2.2). Thus, for example, the syllables *ka* and *ga* could be written with different characters (e.g. *ka* 加; *ga* 我). However, even in OJ this phonological distinction was sometimes ignored, and for example a *man'yōgana* normally

used for *ka* could be used for *ga* (e.g. 加 sometimes used for the genitive particle *ga*). In early EMJ the *sei–daku* distinction came to be consistently ignored in general writing and the same *man'yōgana* were used for *pa/ba*, *ta/da*, *ka/ga*, *sa/za*, etc. Accordingly, as the abbreviated *hiragana* and *katakana* forms developed, separate letters for *sei'on* and *daku'on* did not evolve and this phonological distinction remained unnoted through most of the MJ period in general writing. Thus, although the tenues and mediae clearly were phonologically distinct in MJ, as in OJ and in NJ, the *kana* letters from the *pa*, *ta*, *ka*, and *sa*-columns were used to represent /pV, bV/, /tV, dV/, /kV, gV/, and /sV, zV/, respectively, such that for example は was used for both *pa* and *ba*, た for *ta* and *da*, か for *ka* and *ga*, and さ for *sa* and *za*. This is reflected even today in the absence of separate *kana* letters for *sei'on* and *daku'on*, with *daku'on* being indicated by adding a diacritic, the *dakuten* ゛, to the top right corner of *kana* for *sei'on*, e.g. か *ka*, が *ga*.

Such a development towards a more underspecifying script is not unique, but paralleled, for example, in the transition from the older (Germanic) to the younger (Scandinavian) runes (completed by the middle of the ninth century). As opposed to the older runes (with twenty-four letters), the younger system of sixteen letters did not have separate letters for tense (unvoiced) and lax (voiced) stops, in addition to other instances of equivalence. Thus there was in the Scandinavian runes only one letter for /p, b/, /t, d/, and /k, g/, respectively, although these sounds were and remained distinctive in the Scandinavian languages and had had separate orthographic representation in the Germanic runes. It is not difficult to understand that native readers have few problems with a phonologically underspecifying script in general writing, as they know the words in the language and can guess from context which words are meant, even in the absence of an absolute phonological identification, in the same way that phonologically or phonetically underspecified speech is usually readily understandable. Thus, for example, most scripts have no expression of suprasegmental features such as accent or tone. Note finally that although tenues and mediae have been phonologically distinct through the history of Japanese, a sizeable proportion of mediae derives from tenues either in morphophonemic alternations as a result of *rendaku* (see 2.6.2), e.g. *sakura* 'cherry' + *pana* 'flower' => *sakurabana* 'cherry flower', or by automatic phonological neutralization of tenues as mediae after nasals (cf. 7.1.2.2), e.g. (*yom-* 'read' =>) *yoN-* + *-te* => *yonde* 'read.GER'. In such cases no ambiguity in morpheme identification arises, even if the phonemic shape is not unambiguously noted.

6.1.2.2 Dakuten

In specialized writing, however, various means were used, when it was thought necessary, to give a more precise indication of the phonological shape of a word, noting whether a syllable was *sei* or *daku*, or its phonological pitch.

This was mostly done by diacritics, but there are also cases of, for example, inverted *kana* being used to specify a *daku'on*. The earliest attested use of diacritics to mark *sei* or *daku* on *man'yōgana* is from the late ninth century, and on *kana* from the eleventh century. Diacritics include for example 〻, short for 濁(音) *daku(on)* (or the verb *nigoru* 濁 'pronounce as a *daku'on*'), for example 婆〻 to specify *ba* (and not *pa*). Usually, however, diacritics were dots, circles, lines, even triangles, or combinations of these. They could be added to *man'yōgana* or *kana*, but they were also added to logographically used *kanji* to give a hint to their reading. EMJ texts in which *sei-daku* (or pitch) was noted were almost exclusively annotations (*kunten* materials) or dictionaries, and *sei-daku* thus remained un-noted in general writing in the EMJ period. During the early LMJ period (Kamakura), annotation of EMJ texts in Japanese to specify features of pronunciation along those lines became current, but it was not until the NJ period, from the beginning of the Edo period, that the *dakuten* we know today became established and widespread in general writing.

The *dakuten* eventually settled upon appears to have had two sources: (Chinese) tone dots and the letter *anusvāra* from the *Siddham* script. In some documents, including the dictionary *Ruiju-myōgi-shō* from 1081 (6.2.3), *sei-daku* was noted in conjunction with pitch by means of tone dots which mark tone or pitch by their position next to a *kanji*; see 7.4.1. When extended to note *sei-daku*, the tone dots indicate pitch by *position*, while their *shape* indicate *sei-daku*, usually using single dots or circles for *sei'on*, but double dots or circles for *daku'on*.

However, separate from the notation of pitch was a diacritic of the shape ◦ which in early annotations was placed next to or below a letter, but later customarily on the right hand top corner, the position in which the present-day *dakuten* is placed. This diacritic is thought to have its origin in the letter *anusvāra* from the Indic script *Siddham*. The *Siddham* script was used to write Sanskrit and was brought to Japan in the early ninth century, most likely by the priest Kūkai when he returned in 806 from his study tour to China where he also studied Sanskrit. *Siddham* was, in India and other places, later generally replaced by the *Devanagari* script to write Sanskrit, but *Siddham* has remained in use in Shingon Buddhism in Japan, mainly for copying out sutras and mantras. The letter *anusvāra* represents a nasality feature of Sanskrit. It is usually romanized as *-ṃ*, but in *Siddham* it appears as a dot above a syllable, e.g. *sa* 𑖯 versus *saṃ* 𑖯𑖾. The diacritic ◦ was used with *kana*, *man'yōgana*, or with *kanji* used logographically for Chinese words. In addition to marking *daku'on*, it is also in some annotations used to mark syllables with an initial nasal (*n-*, *m-*); there are also cases of it being added to *kana* for *.u* or *.i* to note a nasalized vowel [ũ] or [ĩ] which were renditions of Chinese /-ŋ/, e.g., 𛀀◦ [ũ] or 𛀁◦ [ĩ]. This diacritic was in other words used mainly as a nasality mark;

this provides further evidence that the mediae phonetically were pre-nasalized in EMJ (see 2.2, 7.1.4.3, and 11.1) and were associated with the phonetic feature of nasality also in metalinguistic consciousness.

6.1.2.3 Handakuten

OJ /p/ was lost in most contexts in the course of sound changes which occurred through the MJ period, merging with /-w-/ in medial position after vowels (but preserved after /Q/) in EMJ (7.3.1.2), and changing to /f-/ in initial position in LMJ; see 11.3 for details and exceptions. The limited contexts in which /p/ was preserved were readily recognizable and a specific way of writing /p/ as distinct from /f/ did not develop in the MJ period. However, towards the end of the LMJ period Portuguese Jesuit missionaries instituted the use of a circle on the top right corner of a *kana* for contemporary *fV* (present-day *hV*) in order to write unambiguously *pV*, e.g. は (*fa/ha*), ぱ (*pa*). This was first used in the *Rakuyōshū*, a *kanji* dictionary published by the Jesuit press in Amakusa in 1598. Since then this practice gradually spread and is, of course, today a fully integrated feature of Japanese writing.

6.1.2.4 Orthographic categories; the Iroha-uta, *the Japanese 'alphabet'*

Although there was a large inventory, including variant shapes, of *man'yōgana* and early *kana* letters with widespread phonographic equivalence (i.e., many different letters for the same sound), the establishment and awareness of distinct orthographic categories is evident from three mnemonic word lists or poems from the first half of the Heian period in which each distinct letter category occurs only once: the *Ame-tsuchi no kotoba* (or *Ame-tsuchi no uta*), *Taini-uta*, and *Iroha-uta* (see (1)–(3) on pp. 166–7). These lists/poems functioned as a kind of ABC to remember the distinct letter categories and were also used for writing practice. The *Ame-tsuchi no kotoba* is basically a list of words, but the *Taini-uta* and *Iroha-uta* are organized into phrases or clauses conforming to some extent to the native poetic metre of alternating seven- and five-mora-long verse lines. The *Ame-tsuchi no kotoba* has forty-eight letters, but the latter two have the forty-seven *kana* letter categories which, with the later addition of ん/ン, are in use today.

The three lists concur in not having distinct letters or representation for *sei'on* and *daku'on*. Chronologically, they all reflect a stage of the language *after* the merger of the *kō–otsu* distinctions (as there is no representation of these distinctions), but *before* the merger of non-initial /-p-/ with /-w-/ (950–1000, see 7.3.1.1), as some *pV kana* are illustrated in positions in which /p/ later merged with /w/, e.g. かは *kapa* 'river' (> *kawa*) distinct from ゆわ *yuwa* 'sulphur', as well as before the merger of /.wo/ and /.o/ (*c.* 1000, see 7.3.2.3), as both を *wo* and お *o* find distinct representation. The *Ame-tsuchi no kotoba* alone reflects a stage of the language prior to the merger

(1) ***Ame-tsuchi no kotoba***

あ め つ ち ほ し そ ら や ま か は み ね た に く
も き り む ろ こ け ひ と い ぬ う へ す ゑ ゆ わ
さ る お ふ せ よ え の え を な れ ゐ て

あめ	つち	ほし	そら	やま	かは	みね
ame	*tuti*	*posi*	*sora*	*yama*	*kapa*	*mine*
heaven	earth	star	sky	mountain	river	peak

たに	くも	きり	むろ	こけ	ひと	いぬ
tani	*kumo*	*kiri*	*muro*	*koke*	*pito*	*inu*
valley	cloud	mist	room	moss	man	dog

うへ	すゑ	ゆわ	さる	おふ	せよ
upe	*suwe*	*yuwa*	*saru*	*opu*	*seyo*
above	end	sulphur	monkey	grow	do!

え	の	え	を	なれ	ゐて
e	*no*	*ye*	*wo*	*nare*	*wite*
hackberry tree	GEN	branch	ACC	get used to	sitting

(2) ***Taini-uta***

た ゐ に い て な つ む わ れ を そ き み め す と
あ さ り お ひ ゆ く や ま し ろ の う ち ゑ へ る
こ ら も は ほ せ よ え ふ ね か け ぬ

たゐ	に	いて	な	つむ	われ	を	そ
tawi	*ni*	*ide*	*na*	*tumu*	*ware*	*wo*	*zo*
paddy	DAT	go out	greens	pick	I	ACC	FOC

きみ	めす	と	あさり	お ひ ゆ く
kimi	*mesu*	*to*	*asari*	*opi-yuku*
you	see.RESP	when	hunt	chase-after

やましろ	の	うちゑへる	こら
yamasiro	*no*	*utiweperu*	*kora*
Yamashiro	GEN	is very drunk	dear.girl

もは	ほせよ	え	ふね	かけぬ
mopa	*poseyo*	*e-*	*pune*	*kakenu*
seaweed	dry!	(un)able-	boat	doesn't anchor

(3) ***Iroha-uta***

い ろ は に ほ へ と ち り ぬ る を わ か よ た れ
そ つ ね な ら む う ゐ の お く や ま け ふ こ え
て あ さ き ゆ め み し ゑ ひ も せ す

いろ	は	にほへと	ちりぬる	を
iro	*pa*	*nipopedo*	*tirinuru*	*wo*
colour	TOP	although shines beautifully	will scatter	EXCL

わ	か	よ	たれ	そ	つね	ならむ
wa	*ga*	*yo*	*tare*	*zo*	*tune*	*naramu*
I	GEN	world	who	FOC	always	will be

うゐ	の	おくやま	けふ	こえて
uwi	*no*	*okuyama*	*kepu*	*koete*
material.world	GEN	deep.mountain	today	crossing

あさき	ゆめ	みし	ゑひ	も	せす
asaki	*yume*	*mizi*	*wepi*	*mo*	*sezu*
shallow	dream	won't see	get drunk	ETOP	doesn't

'Colors are fragrant, but they fade away. In this world of ours none lasts forever. Today cross the high mountain of life's illusion [i.e. rise above this physical world] and there will be no more shallow dreaming, no more drunkenness [i.e. there will be no more uneasiness, no more temptations].'
(Translated by Andrew N. Nelson (1974: 1014))

of the syllables /.ye/ and /.e/ (before 950, see 7.3.2.2). The *Ame-tsuchi no kotoba* is not attested until the *Minamoto Shitagō-shū*, compiled 967, in *hiragana*, but it has two occurrences of *hiragana* え *e*, interpreted as representing /.e/ and /.ye/, respectively, as there is no other good explanation for the double occurrence of え. The current *kana* inventory of course does not include representation of that earlier distinction, but in addition to *man'yōgana*, *kana* annotations from the ninth and early tenth century have distinct letters for /.e/ and /.ye/, shown under 衣 in the *a*-column and under 江 in the *ya*-column in Table 6.2 above, and as mentioned the current *hiragana* and *katakana* in fact derive from early *kana* for /.e/ and /.ye/, respectively: え< 衣 /.e/, while エ < 江 /ye/.

The *Taini-uta* is attested first in 970 (in the *Kuchizusami*, a first primer for the education of boys of the nobility, compiled by Minamoto no Tamenori, ?–1011), which is probably fairly close to the time it was made, but the first attestation of the *Iroha-uta* is in the *Konkōmyō saishōō kyō ongi* from 1079, although it must have been made earlier. Nothing is known about the

authorship of the *Iroha-uta*. It has, like the *hiragana* script itself, been attributed to Kūkai, but the phonological system reflected in the *Iroha* postdates Kūkai, who died in 835, by a century and a half.[1] The *Iroha-uta* soon superseded the *Ame-tsuchi no kotoba* and *Taini-uta*, which today are not generally known, and from the late Heian period became widely and commonly used to remember the *kana* categories and for writing practice.

It is difficult to exaggerate the importance of the establishment and dissemination of the *Iroha-uta*. It is a defining event in the history of Japanese script and writing, and more generally in Japanese culture, linguistic and otherwise, and for example played an essential role in the spread of literacy in Japan through the LMJ and NJ periods. The *Iroha-uta* finally defined the forty-seven *kana* letter categories which, with the sole addition of ん/ン for the moraic nasal, still are in use today, enshrined the principle of not having separate letters for *sei'on* and *daku'on*, and fixed the sequence of the letters in a list. Phonographically, the *Iroha* inventory may be seen to have some shortcomings, viz. the systematic underspecification of the *sei'on–daku* distinction, and a lack of distinct, single-letter representation of 'new' sounds in the language (see 6.1.2.5), so that as the *Iroha* letter categories became set in stone, the relation between sound and written representation over time was obscured by sound changes (see 6.1.3). However, the *Iroha* represents a good phonological analysis of the number of (free) moras in the language at the time it came into being, and it successfully established and fixed a small manageable inventory of letter categories, thus providing a writing system which was learnable for larger groups of people. The fact that the letter categories of the *Iroha* by and large are those used to write Japanese today is testimony to its success.

In addition to establishing the letter categories and being used for writing practice, the *Iroha-uta* also in more general cultural terms assumed the functions of an *alphabet* or *ABC*. It came to be used as a sequence for enumeration, and is still today sometimes used in lists, or in subclassification within dictionary entries. One enormous impact of the *Iroha* was in its use for ordering ('alphabetizing') dictionaries from as early as the end of the Heian period (used first in the *Iroha jiruishō* compiled between 1144 and 1181), making possible the compilation of dictionaries with easy, phonologically based look-up. The *Iroha* continued to function as the Japanese alphabet, as a linguistic and meta-linguistic organizing principle until it eventually was replaced by the *gojūonzu* in the middle of the Meiji period (6.1.4). The word *Iroha* (written variously as いろは, 伊呂波, 色葉) also came to be used as a name for the

[1] It is interesting to note that *iropa* is a word meaning 'natural mother' which could be taken in the sense of 'origin, beginning', here of writing. The similarity between the words ***alpha**bet* and ***iropa*** is surely fortuitous.

kana inventory and by extension to mean 'ABC' in the sense of 'rudiments, basics (of something)', cf. living expressions such as *iroha no 'i' no ji mo shiranai* (いろはの「伊」の字も知らない) 'be utterly illiterate, or ignorant; lit.: not even know the 伊 [first] letter of the *Iroha*'.

6.1.2.5 New sounds

In the transition between OJ and EMJ and during the EMJ and LMJ periods, a number of sound changes occurred. Some of these were mergers, or loss of phonological distinctions, for example the loss of the *kō–otsu* distinction, resulting in fewer orthographic distinctions being necessary. Others, however, introduced entirely new phonemes into the language, all related to syllable structure: *bound moras* (6.1.2.5.1), *complex syllable onsets* (6.1.2.5.2), *and SJ syllable final* -t (6.1.2.5.3). As the *kana* letters were reduced forms of *man'yōgana*, which reflected the phonology of OJ, they did not readily provide for the representation of sounds which were new to EMJ, so although some of these new sounds were present in the language at the time the *Iroha* was established, the letter categories of the *Iroha* do not include distinct representation of these new sounds.

6.1.2.5.1 Bound moras Through the *onbin* sound changes (7.1.4), Japanese became quantity sensitive and acquired long syllables which included bound moras (7.1.1): *moraic consonants* /C, Q, N/ and *moraic vowels*, both *oral* /I, U/, and *nasal* /Ĩ, Ũ/. Of these only the moraic nasal /N/ eventually got its own letter, the only addition to the letter categories of the *Iroha*. Various means were employed to write the *moraic consonants* /C, Q, N/ which appeared in the language in the transition between OJ and EMJ: Apart from not being represented (e.g. *moQ-te* (< *motite*) 'holding; hold-GER' written as モテ *mo.te* (in annotations from the late ninth century) or *siN-zi* (< *sini-si*) 'dead; die-SPST.ADN', written (in the *Tosa nikki*, 935) as しし *si.si*), they have been written with a number of different *kana* for CV syllables. There was some overlap – within individual texts – in representation, and for example *man'yōgana* 牟 (*mu*) or its reduced *katakana* shape ム have both been used to represent both *N* and *Q*; this reflects the phonemic identity between *N* and *Q* (as /C/ unspecified for nasality) in morpheme-internal position (cf. 7.1.2.1). The current use of *kana* for *tu* (つ) to write the moraic obstruent /Q/ was originally inspired by the *kana* transcription of syllable final *-t* as つ (*tu*) (cf. 6.1.2.5.3) and was settled upon relatively late. The convention of writing this letter in small size dates from the script reform of 1946. The *hiragana* letter ん for the nasal moraic obstruent /N/ is thought to derive from the *man'yōgana* 无 (*mu*), though some scholars believe it to be a further development from (a precursor of) *katakana* ム (itself < 牟). Table 6.2 above shows two letter shapes for /N/, one like present-day *hiragana* ん and the other

seemingly a precursor of *katakana* ン. However, the *katakana* letter ン seems to be first attested as such in the eleventh century. It appears to have no recognizable source in a *man'yōgana*, but to have been invented specifically for the purpose of representing /N/. The dot in ン may well reflect the nasality diacritic ゛ which was also a source of the *dakuten* (6.1.2.2).

The bound moraic vowels, /I, U, Ĩ, Ũ/, emerged through the *onbin* sound changes, and also came to be used in SJ loanwords. These sounds were written with *kana* for the free moras /i, u/, i.e. い, う, with no differentiation of oral and nasal, apart from the use in some annotations of the nasality diacritic ゛ with *kana* for *.i, .u* to specify /Ĩ, Ũ/ in renditions of Chinese /-ŋ/.

6.1.2.5.2 Syllables with complex onsets Syllables with complex onsets, consisting of consonant + (palatal or labial) glide, represented in the moras *Cya*, *Cyo*, *Cyu* and *kwa*, *gwa*, *kwe*, *gwe* (7.2), entered the language in the EMJ period through the large-scale adoption and adaptation of SJ vocabulary. Somewhat later *Cya*, *Cyo*, *Cyu* also arose outside SJ vocabulary through sound changes.

Also these sounds were not written with distinct single letters, but instead as combinations of *kana* for *Ci* or *Cu* with *yV* and *wa*, i.e., *Ciya*, *Ciyo*, *Ciyu* (e.g. きや *kya*, きよ *kyo*, きゆ *kyu*) and *kuwa*, *guwa*, *kuwe*, *guwe* (くわ *kwa*, ぐわ *gwa*, くゑ *kwe*, ぐゑ *gwe*), respectively. Since the script reform in 1946 the glide initial letter has been written in smaller size: きゃ, きょ, きゅ; くわ, ぐわ.

6.1.2.5.3 Syllable final /-t/ A sound found only in SJ loanwords was syllable final /-t/ (cf. 7.1, 7.2). This remained pronounced as /-t/, i.e. without a following vowel, until well into the NJ period (see further 11.4). In annotations it was noted in a number of different ways, but in general writing it was transcribed by *kana* つ or sometimes ち. It has recently been shown, however, that some LMJ sources made an orthographic distinction between /-tu/ and /-t/ by means of variant *kana* (*hentaigana*) which were originally used as equivalents for /tu/. In NJ final /-t/ acquired an epenthetic vowel to give *tu* or *ti*.

6.1.2.5.4 Recent Modern Japanese loanwords Finally, we should note 'new sounds', and especially new combinations of sounds, which have come into the Japanese language in the contemporary period, since the Meiji restoration, through the intake of large numbers of western loanwords (see 14.6). Like moras with complex onsets, the moras containing these sounds have been written with combinations of *kana*, e.g., *fan* ファン, *paatii* パーティー 'party', *sherii* シェリー 'sherry', *tsaitogaisuto* ツァイトガイスト 'zeitgeist', *jerii* ジェリー 'jelly', *vandaru-jin* ヴァンダル人 'Vandals (Germanic

tribe)'. The moras /wi, we, wo/ have reappeared in the language, but they are today written as *wirusu* ウィルス 'virus', *wesutan* ウェスタン '(a) western', *wokka* ウォッカ 'vodka', rather than using the old *kana* for these moras (*キルス, *エスタン, *ヲッカ).

6.1.3 Orthographic norms

After the establishment of the orthographic categories in the *Iroha*, a number of sound changes took place through the MJ period which rendered several originally phonographically distinct *kana* categories equivalent, in some cases unconditionally and in others only in word non-initial position: /-p-/ merged in intervocalic position with /-w-/ as /-w-/, but remained /-p-/ after /Q/ (7.3.1), and /w/ was lost in both word initial and non-initial position, first before /o/, and later before /i, e/ (7.3.2.3). These changes are often stated as syllable mergers, as shown in (1) which lists the merged syllables, the outcome of the merger, and the previously phonographically distinct *kana* categories which could now be used to represent the outcome of the merger.

(4) Sound changes resulting in *kana* equivalence

	In initial position			
a.	/.o/, /.wo/	>	/.o/	お を
b.	/.i/, /.wi/	>	/.i/	い ゐ
c.	/.e/, /.we/	>	/.e/	え ゑ
	In non-initial position			
d.	/-.o/, /-.wo/, /-po/	>	/-.o/	お を ほ
e.	/-.i/, /-.wi/, /-pi/	>	/-.i/	い ゐ ひ
f.	/-.e/, /-.we/, /-pe/	>	/-.e/	え ゑ へ
g.	/-.wa/, /-pa/	>	/-.wa/	わ は
h.	/-.u/, /-pu/	>	/-.u/	う ふ

These sound changes had two consequences for the representational value of *kana*: First, the *kana* categories おを, いゐ, and えゑ, respectively, became phonographically equivalent in all contexts. Second, the representational value of the *p*-column *kana* changed in non-initial position, but the fact that there was no orthographic distinction between *sei'on* and *daku'on* and that the *kana* from the *p*-, *t*-, *k*-, *s*- columns were used to represent both *sei'on* and *daku'on* (/pV, bV/, /tV, dV/, /kV, gV/, and /sV, zV/, respectively) complicated the situation further. In initial position, as before, は could represent /pa/ or /ba/ (mainly in SJ loanwords which had no restrictions on the occurrence of word initial media, see 7.2), and likewise ひ could stand for /pi/ or /bi/. However,

in non-initial position は could now represent /wa/ or /ba/, or, after /Q/, /pa/. Likewise, ひ came to represent /i, bi/ (or /pi/), へ could represent /e, be/ (or /pe/), ほ /o, bo/ (or /po/), and ふ /u, bu/ (or /pu/). The result of this was partial phonographic overlap, but not full equivalence, of non-initial ほ, ひ, へ, は, and ふ with お/を, い/ゐ, え/ゑ, わ, and う, respectively. See (2) which shows the phonographic value of はひふへほ, and the partial phonographic equivalence with other *kana* in non-initial position:

(5)

	Initial position		Non-initial position		/Q/ __
は	/pa/	/ba/	/ba/	/wa/ (= わ)	/pa/
ひ	/pi/	/bi/	/bi/	/i/ (= い, ゐ)	/pi/
ふ	/pu/	/bu/	/bu/	/u/ (= う)	/pu/
へ	/pe/	/be/	/be/	/e/ (= え, ゑ)	/pe/
ほ	/po/	/bo/	/bo/	/o/ (= お, を)	/po/

Thus, for example, *kawa* 'river' could now be written かわ or かは, whereas *kaba* 'Japanese white birch' would only be written かは. Likewise, *kai* 'shellfish' could be written かい, かゐ, or かひ, whereas *kabi* 'mildew' would only be written as かひ. Conversely, from the reader's point of view, かわ spelled *kawa*, whereas かは spelled *kawa* and *kaba*; and かい and かゐ unambiguously spelled *kai*, whereas かひ spelled *kai* and *kabi*. Taking account of the spelling of words with /-Qp-/ complicates the situation further. For example, *kappa(-to)* 'with a thud' could also be spelled かは (depending on how /Q/ was represented). In (6), the spellings of some words which illustrate the polyvalence are listed, showing the potential difficulties for both readers and writers:

(6)

kawa 'river'	かは	かわ		
kaba 'Japanese white birch'	かは			
kappa(-to) 'with a thud'	かは			
kabi 'mildew'	かひ			
kai 'shellfish'	かひ	かゐ	かい	
kae 'change (tr.); inf.'	かへ	かゑ	かえ	
kabe 'wall'	かへ			
io '500'	いほ	いを	いお	
ibo(ziri) 'praying mantis'	いほ			
kau 'change (tr.); concl.'	かふ	かう		
kabu 'head'	かふ			

Clearly, for the *kana* categories おをほ, いゐひ, えゑへ, うふ, わは, the mapping between sound and writing became much less straightforward than it had been at the time the *kana* letter categories were established and they more closely reflected the phonology of the language, and as may be expected, this led to efforts to regularize the use of these *kana* letters. Rather than reforming the inventory and use of *kana* letters along simple phonographic lines (for example, by dropping the letter categories をゐゑ altogether and specializing non-initial はひふへほ to represent /bV/ and /pV/ after /Q/), an etymological spelling principle gained ground from around the beginning of the thirteenth century, eventually resulting in the so-called *rekishi-teki kana-zukai* (歴史的仮名遣い) 'historical *kana*-usage, historical spelling', which is still in use today in the presentation of premodern Japanese texts and in writing Classical Japanese. The simple principle is that any word is to be spelled the way it was, or would have been, spelled within the *kana* categories of the *Iroha*, before the sound changes outlined in (4) took place. Thus *kai* 'shellfish' (< OJ *kapi*) is spelled かひ, *ai* 'indigo' (< *awi*) is あゐ, and *kai* 'rudder' (< *ka.i*) is かい. In addition to the notion of faithfulness to the shape of earlier texts, the adoption of the etymological spelling principle seems to show that the orthographic categories in the *Iroha-uta* were taken as given and not to be tinkered with. The widespread adoption of the etymological spelling principle means that many text copies as well as much writing preserves orthographic distinctions long after the phonological distinctions they originally represented had been lost, making such materials useless for the dating of sound changes.

At the beginning of the NJ period further sound changes took place which resulted in phonographic equivalence between more *kana* categories: /z/ and /d/ merged (as /z/) before the high vowels /i, u/ (see 14.1), rendering the *kana* pairs じ/ぢ and ず/づ phonographically equivalent. By then the etymological principle had long been established and so the principle for which *kana* to use was fairly clear, although usages which can be thought of as 'spelling mistakes' (for example writing *mizu* 'water' < *midu* as みず rather than the etymologically correct みづ) are frequent in Edo period NJ texts.

The *rekishi-teki kana-zukai* was only abolished as the norm with the orthographic reforms in 1946 when the *gendai kana-zukai* ('present-day *kana* usage, present-day spelling') was adopted, in which the etymological principle largely has been abandoned, with a few well-known exceptions. For example, the second half of the long vowel /oo/ is mostly written う (e.g. *gakkoo* 'school' がっこう, *oogi* 'fan' おうぎ, *toota* 'asked' とうた), but お when having earlier been spelled ほ or を (e.g. *too-* 'far' とお (< とほ) or *too* 'ten' とお (< とを)); and the grammatical particles *wa* 'topic', *e* 'allative', and *o* 'accusative' retain their earlier spelling は, へ, and を; finally, /zu/ and /zi/ (/ji/) are usually spelled ず and じ, except where they derive from *rendaku* of /tu/ (/tsu/) and /ti/ (/chi/) in transparent compounds (e.g., *mi-zika* 'close, familiar'

みぢか, cf. *mi* 'body' and *tika-* 'close' or *kana-zukai* '*kana*-usage' かなづかい, cf. *tsukai* 'usage').

The etymological spelling principle was first explicitly proposed by the poet and scholar Fujiwara no Teika (1162–1241) as part of his recommendations of the spelling of individual words set forth in works such as *Gekanshū* (?1241, a book on poetics) or *Hekianshō* (1226, annotations on poems from earlier poetry anthologies). His primary concern was the production of faithful editions of earlier texts, preserving their original shape, and he based his spelling proposals on inspection of earlier manuscripts and text versions in order to establish original, 'correct' spellings. It seems that there was some consensus about the spelling of individual words before Teika, at least to some extent on an etymological basis, as shown for example in the surviving handwritten letters of the wife of Fujiwara no Tamefusa (1049–1115) in which the distinctions between /-p-/ and /-w-/ and between /.wo/ and /.o/ are maintained orthographically although the sound changes in (4) had already taken place, but it was Teika who explicitly established the etymological spelling principle which, together with his specific spellings of individual words, came to be known as *Teika kana-zukai* 'Teika *kana*-usage, Teika spelling' and soon acquired the status of a spelling norm, not just for copying, but also for composing text. Thus, from Kamakura to early Edo most literary writing was spelled according to the *Teika kana-zukai*.

There are inevitably not a few mistakes in Teika's etymological spellings (for example *too-* 'far' < mid EMJ *towo* < OJ *topo-* spelled とを 'to.wo' instead of etymologically correct とほ 'to.po'), first of all because the copies of texts he had access to and worked from already contained errors. Furthermore, his recommendations concern only a relatively small number of words and only address the use of えゑへいゐひおを.[2] Therefore later scholars elaborated on and amplified Teika's work. The *Kanamoji-zukai* (completed after 1363) by the monk Gyōa (fourteenth century, dates unknown, born Minamoto no Tomoyuki) is the first large-scale application of Teika's principles. It lists spellings for more than a thousand words and for a long time served as the major normative spelling reference. In addition to unsuccessfully attempting to credit his own grandfather, rather than Teika, with the establishment of the etymological spelling principle, Gyōa takes the important step of incorporating also the letter categories ほわはうふ (as well as む which was

[2] However, with regard to お and を, Teika partly employed a phonographic, rather than etymological, spelling principle, following the example of earlier dictionaries like *Iroha jiruishō* (late Heian) in assigning letters in word initial position according to phonetic pitch, such that /.o-/ would be written with を if it was [high], e.g. をく *oku* (< OJ *oku*) 'fall, descend (of dew, snow, frost); put', but with お if it was [low], e.g. おる *oru* (< OJ *woru*) 'break'. This caused much confusion later, as pitch changed, obscuring the basis for the spellings which had become fixed and normative.

problematic for entirely different reasons). Mention must also be made of the *kokugaku* scholar Keichū (1640–1701) who provided the basis for the *rekishiteki kana-zukai* in use today. In his *Waji shōranshō* (1695) he set the etymological spelling principle on a firm philological footing with systematic citation for specific spellings of sources which even by today's standards mostly are reliable.

6.1.3.1 Undoing the etymological kana*-spelling*

In printed editions premodern texts are usually presented in the etymological *kana*-spelling, so that for example the word for 'front' is written まへ regardless of the date of the text. Table 6.3 is a simple guide to phonemic transcription of EMJ and LMJ texts, giving the phonemic shapes represented by the *kana* letters えゐゑをはひふへほ in different periods, reflecting the sound changes which took place during the EMJ and LMJ periods (see 7.3, 11.2, 11.3). As these sound changes generally are not systematically reflected in the copies available to us today of most of the literary texts (due to the application of the etymological spelling principle in copying the texts), the assignment of texts to definite stages within these sound changes is to some extent arbitrary. This holds particularly for the texts dating just after 1000, including important literary texts such as the *Genji monogatari*, *c.* 1001–1010, and the *Makura no sōshi*, *c.* 1000. It is clear that these texts postdate the change of /-p-/ to /-w-/ (cf. 7.3.1) and that what is written かは in editions of these texts represents /kawa/. However, it is difficult to say for certain that the change of /wo/ to /o/ (cf. 7.3.2.3), which is thought to have been complete *c.* 1000, was in fact entirely complete by the time these texts were written, and it is therefore less certain that what is written かほ in editions of these texts represents /kao/ and not /kawo/; it is quite certain, though, that it does not represent /kapo/. Note also that the difference between /.e/ and /.ye/, which pertains only to morpheme initial position, is not represented in the standard inventory of *kana* letter categories (cf. 6.1.2.4), but it is relevant to the earliest *kana* materials, such as *Tosa nikki* or *Kokinwakashū*, it can easily be verified in good dictionaries whether words written with initial え had /.ye/ (e.g. *yeda* 'branch') or /.e/ (e.g. *e-* 'to get'); in morpheme non-initial position え always represents /.ye/ before *c.* 950 (e.g. *nuye* 'thrush').

Through the EMJ and LMJ periods, し, ち, つ, じ, ず, ぢ, づ straightforwardly represent /si, ti, tu, zi, zu, di, du/ and should be transcribed *si*, *ti*, *tu*, *zi*, *zu*, *di*, *du*, e.g. つち *tuti* 'earth', しづか *siduka* 'quiet', みず *mizu* 'doesn't see', みじ *mizi* 'won't see', and やまぢ *yamadi* 'mountain path'. Only in texts dating from the early NJ period do じ and ぢ both come to represent /zi/ and should be transcribed *zi*, and ず and づ both come to represent /zu/ and should be transcribed *zu*. In terms of the etymological *kana* spelling (which in text editions supplies *dakuten* although these were not used in general writing, cf.

Table 6.3 *Phonemic transcription of the historical* kana*-spelling*

Kana	Examples	< 950 *Tosa nikki, Kokinwakashū, Ise monogatari*	*c.* 950 *Taketori monogatari**	950 > *Kagerō nikki, Utsuho monogatari, Ochikubo monogatari*	1000 > *Genji monogatari, Makura no sōshi*	1100 > *Konjaku monogatari, Heike monogatari, Hōjōki*	1300 > *Tsurezure-gusa, Soga monogatari*
え	え	e					
	えだ	yeda	eda				
	ぬえ	nuye	nue				
ゐ	ゐ	wi					i
	まゐる	mawiru				mairu	
ゑ	ゑみ	wemi					emi
	こゑ	kowe				koe	
を	をか	woka			oka		
	あを	awo			ao		
は	はら	para					fara
	かは	kapa		kawa			
ひ	ひら	pira					fira
	かひ	kapi		kawi		kai	
ふ	ふね	pune					fune
	たふ	tapu		tau			
へ	へら	pera					fera
	まへ	mape		mawe		mae	
ほ	ほか	poka					foka
	かほ	kapo		kawo	kao		

Glosses: え 'to get', えだ 'branch', ぬえ 'thrush', ゐ 'well', まゐる 'goes humbly' ゑみ 'smile', こゑ 'voice', をか 'hill', あを 'blue', はら 'stomach', かは 'river', ひら 'plain', かひ 'shell', ふね 'boat', たふ 'stupa', へら 'shovel', まへ 'front', ほか 'other place', かほ 'face'.

*Note that the dating of the *Taketori monogatari* is very tentative.

6.1.2.2), other *kana* letters than those discussed above have not changed in representational value since the beginning of the EMJ period and can be transcribed for premodern texts as they are for modern texts.

6.1.4 Sound tables

Today the *kana* letters are not presented in the *Iroha* sequence, but arranged in the *gojūonzu*, as shown in Table 6.1 above. This arrangement of the *kana* letters is now used universally in school education and for dictionary entries; it also influenced the *katsuyōkei* system for describing verbal inflection (see 3.4.6). *Kana* lists (*onzu* 音図 'sound tables') which formed the basis of the current grid arrangement began appearing from the beginning of the eleventh century. The first known such *onzu*, which is incomplete consisting of only eight columns, appears at the back of the *Kujakukyō-ongi* (from around 1000) and the first complete *onzu* is found in Meigaku's *Han'on-sahō* (1093).

As opposed to the *Iroha*, the *onzu* went beyond simply enumerating the distinct orthographic categories: the *onzu* involve a substantial phonographic and phonological analysis and systematization of the letter categories defined in the *Iroha*. They provide a phonological analysis into consonants and vowels of the moras represented by the *kana* letters, such that *kana* in a column share the same initial consonant and *kana* in a row share the same vowel, i.e. an analysis of a syllable into subsyllabic parts, or a mora into submoraic parts. In the early *onzu* the *kana* were not actually arranged in grids with columns next to each other, but as lists with the columns following each other, but the principle is the same, of columns sharing an initial consonant and the vowels appearing in a fixed order within the columns. The grid system, incidentally, also makes it possible to refer in Japanese to consonants, which have no separate *kana* representation, with reference to this arrangement, e.g. the 'consonant of the *ka*-column' (*ka-gyō no shi'in* カ行の子音), viz. *k*. Ingenious as they were, the *onzu* remained in the realm of scholastic and academic writing and annotations until late in the Edo period; in the Meiji period the *gojūonzu* was adopted for educational purposes and eventually replaced the *Iroha* as the presentation of the *kana* categories and as an organizing principle for dictionaries. The first Japanese dictionary ordered in the *gojūonzu* sequence was the *Genkai* (言海) published in 1891 by Ōtsuki Fumihiko, who was also an important figure in the development of a 'standard' Japanese, see 13.2.2.

The phonological analysis of the *onzu* was principally inspired by Chinese rhyme tables and in particular by the *fǎnqiè* (反切) spelling system, in which the 'reading' of a *kanji* is described in terms of two other *kanji*, in a form like '*x yz*' to be understood as '*x* has the initial of *y* and the rhyme of *z*', for example, 東 德紅 which means that 東 has the initial of 德 (EMC *tək, LMC

*təšk) and the rhyme of 紅 (EMC *ɣəwŋ, LMC *xɦəwŋ), and thus that 東 has the readings EMC *təwŋ, LMC *təwŋ.

However, knowledge of Indic script, again in particular the *Siddham* script (cf. 6.1.2.2), exerted an important influence on the development of the sound tables. The sequence of vowels (あ *a*, い *i*, う *u*, え *e*, お *o*) as well as consonants (か *k-*, さ *s-*, た *t-*, な *n-*, は *p-* (> *f-* > *h-*), ま *m-*, や *y-*, ら *r-*, わ *w-*) in the *gojūonzu* corresponds to that of the letters in the *Siddham* script. The letters representing true consonants come first, followed by approximants and the liquid: *y*, *r*, *w*. For the true consonants, this arrangement moves from velar to labial place of articulation, with nasals following oral consonants. The position of *s-* in the arrangement has been taken to support the view that early EMJ /s/ phonetically was not a simple sibilant (which in *Siddham* are placed at the end together with approximants and liquids), but rather an affricate (2.2.3); the arrangement also clearly shows that the reflex of OJ /p/ was an obstruent, most likely a stop, /p/, and certainly not an approximant such as bilabial [ɸ] (7.3.1.3). Between the earliest *onzu* from the beginning of the eleventh through the twelfth century, there was considerable variation in the order of vowels or consonants. The first *onzu*, from the *Kujakukyō-ongi*, has the vowels in the order *i, o, a, e, u* and the consonants *k, s, t, y, m, p, w, r*. The *onzu* in Meigaku's *Han'on-sahō* has the vowels in the current order and the consonants ordered strictly by place but not manner of articulation: *k*, *y*, *s*, *t*, *n*, *r*, *p*, *m*, *w*. Eventually the order we know today, which is first found in the *Shittan-yōjū-ki* (1075) by Kanchi (1045–1111), was settled upon.

The sound changes mentioned in (4) above also affected the phonological analysis and arrangement of the *kana* in the *onzu*. In particular, when the *kana* pairs おを, いゐ, and えゑ became phonographically equivalent, standing for /.o, .i, .e/, respectively, the historically correct assignment of these letters to the *a-gyō* or *wa-gyō* within the *onzu* became far from obvious, and until quite late there are many examples of *onzu* with incorrect assignments. In his 1682 *Shittansanmitsushō*, the monk Jōgon finally established い and ゐ in the *a-gyō* and *wa-gyō*, respectively, and Keichū (1640–1701) did the same for え and ゑ in 1693 in the *Wajinshōranshō*. It was not until 1776 that お and を were finally fixed in the *a-gyō* and *wa-gyō*, respectively, in Moto'ori Norinaga's *Jion kana-zukai*.

6.2 Sources

The sources available to study EMJ are far more extensive and varied than the sources for OJ. In particular, we have from the middle of the period a large set of prose texts, written in *kana* in a language form which is thought to be close to the vernacular of the nobility and officials at court and which gives a more comprehensive picture of the contemporary language than we have for any other time in premodern Japanese. Sources are scarce until the early tenth

century, and there is thus a substantial gap of more than one hundred years in the solid attestation of Japanese. This is usually attributed to the fact that written Chinese almost succeeded in establishing itself as *the* written language in Japan at the expense of Japanese, particularly outside poetry, but thankfully, from a linguistic and cultural point of view, that did not happen. Apart from the first one hundred years, the Heian period is characterized by a lively culture of writing in Japanese which produced some of the masterpieces of Japanese literature, such as the *Genji monogatari* and the *Makura no sōshi*. After around 1100 the language in the written sources increasingly fossilized and towards the end of the Heian period the written norm known as 'Classical Japanese' (*bungo* 文語 'written, literary language') became relatively fixed and served as the dominant base for writing in Japanese until the beginning of the twentieth century. The sources for this period, in addition to the prose and poetry texts (6.2.1), include a body of annotated Chinese texts, the so-called *kunten shiryō* (6.2.2), as well as dictionaries (6.2.3) and Sanskrit-studies (6.2.4). Table 6.4 is a chronological list of some of the important sources.

6.2.1 Prose and poetry

Prose texts are by far the most valuable material for studying the language of this period. There exist a number of poetry anthologies and many poems within prose texts from the Heian period, but while the literary merit of some of these poems is considered high, their value for linguistic studies is more limited: a norm concerning grammar and lexis soon became established in poetry, and most of the poetic texts are therefore conservative in these regards, for example avoiding both SJ vocabulary and native forms with bound moras (in particular *onbin* forms). In this respect the Heian period poetry is in stark contrast with the poetry from the Nara period, which constituted the major source of OJ. Among the poetry anthologies from the period, the first one, *Kokin wakashū* (古今和歌集 'Collection of Japanese poetry'), stands out, but in particular for its Japanese preface (*kanajo* 仮名序 '*kana*-preface') on poetics, written in *hiragana* by the editor Ki no Tsurayuki (?868–?945). This is the first piece of prose writing in *hiragana* and it contrasts with most later *hiragana* writing from the period by being expository and non-fictional.

The prose material includes *monogatari* ('stories') of various sorts, ranging from collections of short stories, or even what would today be called 'short-short stories', e.g. *Ise monogatari* (伊勢物語 'Tales from Ise'), to Murasaki Shikibu's very long novel *Genji monogatari* (源氏物語 'Tale of Genji'); other well-known works are *Taketori monogatari* (竹取物語 'The tale of the bamboocutter'),[3] *Utsuho monogatari* (宇津保物語 'The tale of the hollow tree'),

[3] *Taketori monogatari* is thought to date from the middle of the tenth century, but the earliest extant copies date from the Edo period, making it less reliable as linguistic evidence than the other EMJ texts.

Table 6.4 *Important EMJ sources*

Ninth century
Konkōmyō saishōō kyō (annotations *c.* 830)
Tōdaiji fujumonkō (before 834)
Zaitōki (*c.* 842)
Shinsen jikyō (898–901)
Tenth century
Kokin wakashū (914)
Ise monogatari (early tenth century)
Wamyō-ruiju-shō (*c.* 934)
Tosa nikki (935)
Taketori monogatari (mid tenth century)
Kagerō nikki (second half of tenth century)
Utsuho monogatari (970s)
Ochikubo monogatari (late tenth century)
Sanbō ekotoba (984)
Eleventh century
Hokke mongu (annotations *c.* 1000)
Makura no sōshi (*c.* 1000)
Genji monogatari (1001–10)
Tsutsumi chūnagon monogatari (*c.* 1055)
Sarashina nikki (1059–60)
Konkōmyō saishōō kyō ongi (*c.* 1079)
Han'on sahō (1093)
Ruiju-myōgi-shō (*c.* 1100)
Twelfth century
Daijionji sanzōhōshi-den (annotations 1099–1116)
Konjaku monogatari-shū (*c.* 1120)
Shittan yōketsu (1101)
Ōkagami (*c.* 1119)
Iroha jiruishō (compiled between 1144 and 1181)
Kohon setsuwa-shū ((late?) twelfth century)
Shittan kuden (*c.* 1180)

Ochikubo monogatari (落窪物語 'The tale of Ochikubo'), *Tsutsumi chūnagon monogatari* (堤中納言物語 'The tales of the Tsutsumi middle counsellor'), and the historical novel *Ōkagami* (大鏡 'The great mirror'). Other genres include *nikki* ('diaries'), which mostly were not private, but written for circulation, such as Ki no Tsurayuki's *Tosa nikki* (土佐日記 'The Tosa diary'), the *Kagerō nikki* (蜻蛉日記 'Gossamer diaries') or the *Sarashina nikki* (更級日記 'The Sarashina diary'); and *zuihitsu* (随筆 'jottings'), especially Sei Shōnagon's famous *Makura no sōshi* (枕草子 'The pillow book'). Much of this literature was written in *hiragana* and by women in a free-flowing, lively, elaborate and sometimes playful language.

The so-called *setsuwa* (説話 'tales, legends') literature comprises a variety of short tales ranging from didactic Buddhist stories set in India or China to local folktales and ghost stories. The *setsuwa* literature is generally written in *kanji-kana majiribun* (6.1.1), i.e. with a large proportion of *kanji*. The language in the *setsuwa* and other writing in *kanji-kana majiribun* is relatively simple and straightforward compared to the *hiragana* literature, but it is also more heavily and obviously influenced by Chinese, because of the influence of *kanbun-kundoku* (see 9.1) on *kanji-kana majiribun*. The most famous anthology is the *Konjaku monogatari-shū* (今昔物語集 'A collection of tales of times now past'); others from this period are the *Kohon setsuwa-shū* (古本説話集 'A collection of old tales') and *Sanbō ekotoba* (三宝絵詞 'Illustrated stories about the three jewels'). As the written language fossilized, the narrative was kept in the classical written language, but dialogue was in a form close to the vernacular. From the end of the period, and especially in the LMJ period, the dialogue passages in the *setsuwa* literature are important sources of changes in the language, as most other writing was done in the by then fossilized classical written language.

These types of prose texts together present a comprehensive view of the contemporary language and its use. As with materials from the Old Japanese period, the text portions written logographically are less useful for studying the lexicon and grammar of the language than those written in *kana*. However, because of the volume of text this is in practice not a severe problem for studying the grammar. The texts from this period have come down to us in later copies, subject to scribal error and to conservative and normative redaction, especially in applying the etymological spelling principle (6.1.3), making them unreliable as guides to the dating of sound changes. However, much effort has been invested in producing reliable critical editions of the texts. Most editions present the texts in etymological spelling and, for the texts written in *kana*, with many interpolated *kanji* (for the convenience of modern readers), but good editions will also indicate the written form of the base text.

6.2.2 Annotated texts

Kunten shiryō (訓点資料 'materials with reading marks') are texts in Classical Chinese with annotations which provide a guide to the pronunciation of Chinese or SJ words, or to understanding or rendering the texts into Japanese. See below (9.1) for details about *kunten*, *kuntengo* (the language used in *kunten* texts), and *kanbun-kundoku*, the practice of rendering Chinese into Japanese. A great amount of *kunten* texts exist, mostly unpublished and in the possession of Buddhist temples. A great effort of both fieldwork and philology has been and is being invested in gaining access to, charting, cataloguing and studying these materials. Impressive advances have been made in

recent years, but there is still enormous scope for further research, in addition to the sheer volume of material also because the material is complex and difficult to interpret. The study of *kunten* texts and *kuntengo* has become a major focus for research for Japanese linguists and philologists since the 1950s, more recently attracting overseas scholars as well. *Kunten* texts are important, primary sources. Often the date of annotation and the identity of the annotator are noted in the text. It is in particular thought that early *kunten* texts from the late eighth and ninth centuries reflect some form of contemporary vernacular language and that they thus can contribute to filling the gap left by the lack of other types of sources between late OJ and 900. Glosses provide valuable information about vocabulary not attested elsewhere and often provide an accurate indication of pronunciation and therefore of contemporary phonology, making them important material for dating sound changes, as opposed to the versions of the prose texts which we have today. It must be kept in mind, however, that *kunten* texts are annotations and therefore orthographically severely underspecified and on some points very difficult to interpret, so their evidential value is not straightforward. In addition, strict, dogmatic norms of rendition and annotation arose, so that *kunten* texts from after the middle of the EMJ period generally cannot be taken to reflect contemporary language, but especially in grammar and vocabulary had become conservative and even archaic. As mentioned above (6.1.2), the *kunten* material is also very important to the study of the history of the development of the *kana* scripts. The amount of published studies and material increases steadily, but well-publicized *kunten* texts include *Konkōmyō saishōō kyō* (金光明最勝王経 'Sutra of the golden light and most victorious king', annotated *c.* 830), the *Hokke mongu* (法華文句 'The sentences and phrases in the Lotus sutra', *c.* 1000, see 9.1.1), and the *Daijionji sanzōhōshi-den* (大慈恩寺三蔵法師伝, *Kōfukuji* manuscript, annotated 1099 and 1116).

6.2.3 Glossaries and dictionaries

Closely related to the *kunten* materials are so-called *ongi* ('glossaries', 音義 'sound-meaning') which are lists of words extracted from individual texts in Classical Chinese, explaining the pronunciation or meaning of words or characters and aiding the reading of individual texts. There are a number of *ongi* from the period. An important example is the *Konkōmyō saishōō kyō ongi* (Glossary for the 'Sutra of the golden light and most victorious king', 1079) which also contains the first attestation of the *gojūonzu* (6.1.4) and of the *iroha-uta* (6.1.2.4). Originating as combinations of glossaries for single texts, the period also saw the compilation of the first *dictionaries*, which were independent of individual texts. These include the *Shinsen jikyō* (新撰字鏡), *Wamyō-ruiju-shō* (倭名類聚抄 or *Wamyōshō* 倭名抄), *Ruiju-myōgi-shō* (類

聚名義抄), and the *Iroha jiruishō* (色葉字類抄 or 伊呂波字類抄), compiled in the second half of the twelfth century and the first dictionary ordered according to the sequence of letters in the *Iroha-uta*. Meigaku's *Han'on sahō* (反音作法) explains the pronunciation of *kanji* by the *fǎnqiè* spelling principle (and by *kana* glosses).

6.2.4 Sanskrit studies

The term *shittangaku* (悉曇学), which literally means the study of the Indic script *Siddham*, is used to refer to the study of Sanskrit and materials in Sanskrit related to Buddhism. It comprises a number of materials which are important to the study of Japanese, especially its phonetics, because they explain or exemplify features of Sanskrit by comparison with Japanese, although the interpretation of the material in many cases is far from straightforward. They include the *Zaitōki* (在唐記 'Record of a sojourn in China'), the priest Ennin's notes on the pronunciation of Sanskrit (which he studied in China) illustrated by means of sinograms used as *phonetic* (rather than phonemic) symbols, based on their sound values as *man'yōgana* or their Japano-Chinese readings (see 9.2.1), supplemented with comments; Meigaku's *Shittan yōketsu* (悉曇要訣 'Essentials of Sanskrit learning') which has many valuable observations on Japanese in order to illustrate points about Sanskrit pronunciation; and Shinren's *Shittan kuden* (悉曇口伝 'The secrets of Sanskrit learning') which exemplifies Sanskrit sounds by notes and by Japanese sounds.

REFERENCES

Writing, general: Kōno 1969, Mabuchi 1993, Ohno 1977, Seeley 1991, Tsukishima 1977: 52–120, Tsukishima 1986. Runes: Elliott 1959: 14ff., 21ff.. *Hentaigana* for /-t/: Sugahara 2000: 23ff. Sources: Tsukishima 1987: 12ff. *Kunten* materials: Kasuga 1985, Nakada 1969, Nishizaki 1992, Tsukishima 1986, Yoshida *et al.* 2001. Dictionaries: Nishizaki 1995.

The major phonological changes which distinguish OJ from NJ took place through this period, leaving Japanese at the end of the period phonologically looking much like NJ. The main structural change is the introduction of the distinction between *short* and *long* syllables and the emergence of bound moraic phonemes which occurred after a free mora to form a long syllable. At the end of the period, EMJ had the sound inventory shown in Table 7.1. The phoneme inventory in free moras was the same as in OJ, but a number of changes had taken place concerning their distribution. Of the free moras in Table 7.1, /wi, we/ were almost exclusively found only in word initial position (cf. 7.3.2.3), and /pa, pi, pu, pe, po/ in word initial position and after /Q/ (cf. 7.3.1; also 11.3), whereas /b, d, g, z, r/ as in OJ were not used in word initial position, except in SJ loanwords. The intake of SJ loanwords also introduced syllables with complex onsets /Cy-/ and /Cw-/. The traditional term for these moras is *yō'on* 'twisted sounds' 拗音 which is opposed to *choku'on* 'straight sounds' 直音, i.e. moras with simple onsets. Of the bound moraic phonemes, /-t/ was used only in SJ vocabulary, which is also often said to have had a distinction between syllable final *-n* and *-m*. See further 7.2 about some of the special phonological features of SJ vocabulary.

Whereas OJ writing seems to have represented the contemporary phonology well, EMJ writing was, as explained above, underspecifying in several respects, no longer noting the distinction between tenues and mediae in general writing (6.1.2.2), and having no distinct ready means of writing new phonemes in the language (6.1.2.5). For that reason, the phonemics and phonetics of the new moraic phonemes are not directly or systematically reflected in the written sources. This holds in particular for the feature of nasality. However, interpreting the sources in the light of what we know or can reconstruct of OJ phonology and the sound changes involved in the introduction of the new phonemes (7.1.4), combined with later reflexes and writings of the forms in the texts, we are able to obtain a fairly detailed picture of the phonological system of the moraic phonemes also during EMJ. The late LMJ Christian materials provide an invaluable reference point for consonantal nasality, but the distinction we posit between /I, U/ and /Ĩ, Ũ/ had been lost by late LMJ. As we saw in 6.1.2.2,

Table 7.1 *EMJ sound inventory*

Free moras									
a	ka	sa	ta	na	pa	ma	ya	ra	wa
i	ki	si	ti	ni	pi	mi		ri	wi
u	ku	su	tu	nu	pu	mu	yu	ru	
e	ke	se	te	ne	pe	me		re	we
o	ko	so	to	no	po	mo	yo	ro	
	ga	za	da		ba				
	gi	zi	di		bi				
	gu	zu	du		bu				
	ge	ze	de		be				
	go	zo	do		bo				
	kya	sya	tya	nya	pya	mya		rya	
	(kyu)	syu	(tyu)	(nyu)	(pyu)	(myu)		(ryu)	
	(kyo)	syo	(tyo)	(nyo)	(pyo)	(myo)		(ryo)	
	gya	zya	dya		bya				
	(gyu)	zyu	(dyu)		(byu)				
	(gyo)	zyo	(dyo)		(byo)				
	kwa								
	kwe								
	gwa								
	gwe								
Bound moraic phonemes									
Vowels	/-I, -U, -Ĩ, -Ũ/								
Consonants	/-N, -Q, -C, -t/								

some pronunciation glosses to Chinese texts note [ĩ, ũ] as distinct from [i, u], but the main reason we reconstruct this distinction for Japanese in EMJ is the observable traces in morphophonological rules or in lexicalizations of the postnasal neutralization rule, whereby tenues became mediae after nasals (7.1.2.2). Concretely, we posit /Ĩ, Ũ/ in *onbin* (and related) forms (7.1.4.3) where the source syllable had initial /b, g, m/, and in SJ vocabulary which reflects EMC *-ŋ.

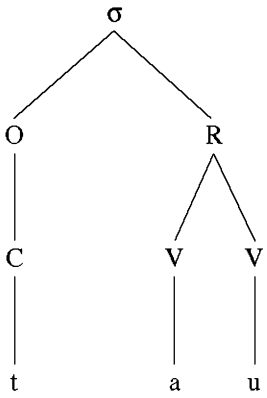
7.1 Syllable structure

EMJ and later stages of Japanese are set off from OJ by a major change in phonological structure: the introduction of the distinction between metrically *short* (or light) and *long* (or heavy) syllables. Through this change Japanese became *quantity sensitive* and the *mora* became a relevant unit in the language. Moras are measures of syllable length (sometimes more concretely thought of

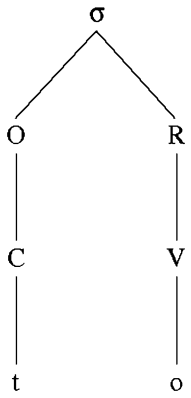
as syllable constituents) of which a short syllable consists of one and a long syllable of two. Whereas OJ only had syllables with a single vowel as peak and nucleus (2.5), EMJ acquired long syllables of the form CVV and CVC, which have an additional position or constituent in the syllable *after* the peak vowel; see 7.2 about complex *onsets* (e.g. Cy- or Cw-) which do not contribute to syllable length. The structure of the syllables of EMJ may informally be illustrated as in (1), which simply has long syllables with branching *rhymes*, with a new syllable position within the rhyme. (Depending on the theoretical framework of phonological representation, long syllables are usually more formally represented with a complex (or branching) syllable nucleus or a syllable coda.)

(1) a. *tauto-* 'exalted'

Long CVV syllable

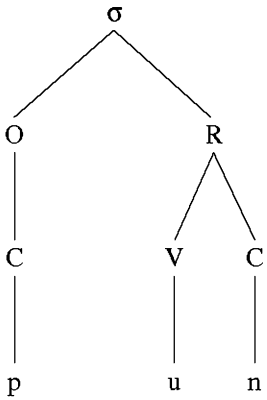


Short syllable

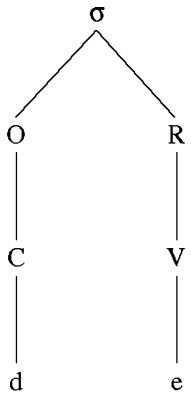


b. *punde* 'brush'

Long CVC syllable



Short syllable



7.1.1 Bound moraic segments

This general syllable structure, with a position after the peak, has remained part of the phonological make-up of Japanese since EMJ. The phonemes which occur in this 'post-peak' position in the syllable are *moraic*, because they measure, or constitute, one mora, and they are *bound* (or dependent) because they do not form a syllable peak, but occur after the syllable peak. The EMJ system of bound moraic phonemes is somewhat different from NJ which has /I, U, V, Q, N/. The overall EMJ inventory was as shown in (2); these phonemes are customarily noted in capital letters.

(2)

	Vowel				Consonant			
	Front		Back					
Nasal	–	+	–	+	–	+	α	
	/ I	Ĩ	U	Ũ	Q	N	C	/

/I, Ĩ, U, Ũ/ are high vowels, distinguishing front versus back, and oral versus nasal; /Q, N/ are consonants whose only distinctive feature is nasality and which are not otherwise specified for place and manner of articulation, but copy these features from the following consonant, as in NJ; and /C/ is a consonant which is not even specified for nasality, but copies all features, including nasality, from the following consonant, see (4) in 7.1.2.1. /Q/ was only used before /p, t, k, s/, and /C/ was not used before /r, w, y/. Other than that, the distribution of these segments depended on position in the word and morpheme, (3). It is sometimes said that two moraic nasal consonants were distinguished, /N/ and /M/, but see below (7.2).

(3)

morpheme non-final	/I, U, C/
word final	/I, Ĩ, U, Ũ, N/
morpheme final, word internal	/I, Ĩ, U, Ũ, Q, N/

It should be noted that few descriptions recognize the nasality distinction among the vowels, and most assume that nasality was phonemic for the consonants in all positions. That is to say, most descriptions posit a system for EMJ which is identical to that for NJ, and we will therefore in the following discuss the system we here posit for EMJ. Recall that the written representation of the new bound moraic phonemes was incomplete and inconsistent (6.1.2.5.1) and that the difference between tenues and mediae was not noted in general writing (6.1.2.1). This means that the phonemics and phonetics of the new moraic phonemes are not directly or systematically reflected in the written sources. What we can say is based on *kunten* glosses (the interpretation of which is not always straightforward) and what we know or reconstruct from OJ phonology and from the sound changes involved in the introduction of the

new phonemes (7.1.4), combined with later reflexes and writings of the forms in question (see 11.1).

7.1.2 Nasality

The major difference between the EMJ system and NJ lies in the distribution and distinctiveness of the feature of *nasality* in the bound moraic segments. In EMJ nasality was distinctive for both vowels and consonants (/I, U, Q/ versus /Ĩ, Ũ, N/) in morpheme final position, but not distinctive for either consonants or vowels in morpheme non-final position. In the inventory of bound moraic phonemes found from LMJ onwards (/I, U, V, Q, N/ cf. 11.1.2), on the other hand, bound moraic *consonants* have distinctive nasality, whereas moraic *vowels* do not, in both cases regardless of position in the morpheme. This difference is closely linked to the phonetic prenasalization of the mediae (2.2.2) and the overall role of the feature of nasality in OJ and MJ.

7.1.2.1 Morpheme internal position

In morpheme internal position, moraic consonants in EMJ only occurred before /p, t, k, s, b, d, g, z, m, n/. What we have reflected as NJ /Q/ and /N/ in morpheme internal position were in EMJ in complementary distribution, with phonetically *oral* moraic consonants occurring only before /p, t, k, s/, e.g. *tatto-* 'precious', and phonetically *nasal* moraic consonants occurring only before /b, d, g, z, m, n/, e.g. *kanbasi* 'fragrant', *punde* 'brush' (> NJ *hude*), *pingasi* 'east' (> NJ *higashi*), *wonna* 'woman' (> late EMJ *onna*).

There were in other words no minimal pairs such as NJ *sittai* (/siQtai/) 'blunder' versus *sintai* (/siNtai/) 'body'. Consequently, there was no distinction between /N/ and /Q/ in morpheme internal position and nasality was not distinctive in the moraic segments in that position. In other words, in morpheme internal position there was only one moraic consonant, /C/, which was phonemically unspecified for nasality, but derived its phonetic nasality value from the following consonant. /C/ was [–nasal] before tenues (/p, t, k, s/), but [+nasal] before nasals and mediae (/m, n, b, d, g, z/). This makes it easy to understand the overlap in the writing of oral and nasal moraic consonants in early materials mentioned in (6.1.2.5.1), as they in morpheme internal position were not distinct phonemes, but conditioned automatic variants of one phoneme /C/. The forms just cited were therefore phonemically /taC.ta, kaC.ba.si, puC.de, piC.ga.si, woC.na/, with '.' showing a syllable boundary. The phonetic nasality value of /C/ was copied from the following consonant (of which /b, d, g, z/ were phonetically prenasalized: [^{m}b, ^{n}d, ^{ŋ}g, ^{n}z]), see (4), alongside other features of manner and place of articulation.

(4) /C/ => [αnasal] / __ [αnasal]

This is a simple phonetic rule, anticipatory assimilation of the phonetic nasality of the following nasal or media, or, leftwards *spreading* of nasality, in the same way that vowels preceding nasals and mediae were nasalized (2.3). This is shown in (5) in a phonetic notation which shows first prenasalization (and medial voicing of tenues) and then nasality spreading:

(5)

	Prenasalization		Nasality spreading	
/kaC ba si/	=>	[kaC ^{m}ba ʒi]	=>	[kãmbaʒi]
/puC de/	=>	[puC ^{n}dʲe]	=>	[pũndʲe]
/piC ga si/	=>	[piC ^{ŋ}ga ʒi]	=>	[pĩŋgaʒi]
/woC na/			=>	[wõn:ã]

7.1.2.2 Morpheme final position; postnasal neutralization

On the other hand, nasality seems to have been distinctive both for moraic consonants and vowels in morpheme final position. The main reason for positing this distinction is the effect it had on a following tenuis. This is mainly relevant in verb inflection and in SJ loanwords. For example in the innovative EMJ gerund forms in (6a) (see 8.1.4) the most straightforward interpretation is that they derive from underlying phonemic forms as in (6b). As opposed to monomorphemic forms such as *punde* /puC.de/ mentioned above, where /d/ was prenasalized [^{n}d] and provided the phonetic nasality, the nasality in these forms must have derived from the moraic segment.

(6)

a.		b.
sin-de 'dying'	<=	/siN + te/ (<= sin-)
yoũ-de 'calling'	<=	/yoŨ + te/ (<= yob-)
sit-te 'knowing'	<=	/siQ + te/ (<= sir-)
ou-te 'pursuing'	<=	/oU + te/ (<= op-)

The rule which derives the (a) forms in (a) from their underlying (b) forms is a neutralization rule such as (7) which neutralizes the distinction between tenues and mediae after nasal segments. This rule was not limited to the formation of verb forms, but applied to all combinations of underlying morpheme final /Ĩ, Ũ, N/ + /p, t, k, s/, including in SJ vocabulary, where we find many words which have lexicalized the effects of postnasal neutralization, for example *saũzi(n)-mono* 'vegetarian food' (< *saũ+sin* 精進 'devotion to Buddhism; abstention' + *mono* 'thing, stuff'), *saũzok-* 'dress up' < *saũzoku* 'dress' < *saũ+soku*; see further (11.1.1). Interestingly, postnasal neutralization was, at least in LMJ, the subject of an explicit reading rule, *umu no sita nigoru* (see 11.1.1.1).

(7) /p, t, k, s/ => /b, d, g, z/ / /Ĩ, Ũ, N/ __

Postnasal neutralization suggests strongly that medial voicing of the tenues took place also after /N/, which, combined with prenasalization of the mediae, made it close to impossible to maintain a distinction between tenues and mediae after nasals: /siN/ + /te/ => [sin] + [dʲe] versus /siN/ + /de/ => [sin] + [ⁿdʲe].

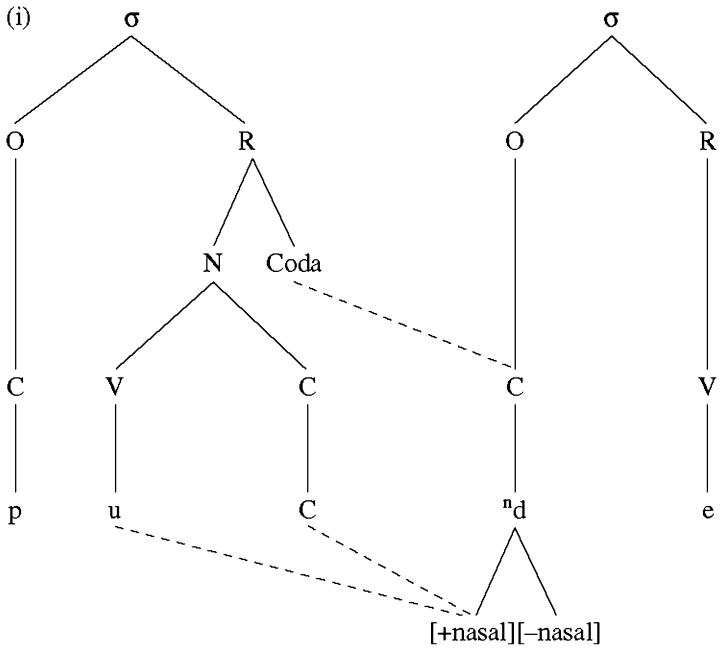
7.1.2.3 Word final position

In word final position, /Q, C/ were not used. /I, U/ were used in both native and SJ vocabulary. /Ĩ, Ũ, N/ were mostly used in SJ vocabulary where they were frequent, but at least /N/ and /Ũ/ were also found in some native forms, such as *-N* or *-Ũ* (reduced variants of *-mu*, the conclusive/adnominal of the conjectural *-(a)m-*, cf. 8.5). In LMJ, word final /N/ had the sound value [n] (see 11.4.2). We have little positive basis for determining the EMJ sound value, but it seems likely that it was [n].

7.1.2.4 Nasality harmony

Although the two rules of nasality spreading (4) and neutralization after nasals, (7), are different types of rules, their effect was much the same. In both cases the rules result in a certain amount of nasality harmony.[1] Compare two sets of

[1] In Frellesvig (1995), I suggested that OJ and MJ had a process of *ambisyllabication* which consists in linking up a non-final syllable with a following syllable, through an underlyingly empty coda slot. On that analysis, the domain of nasality harmony is the syllable rhyme, whose nasality could derive from the onset consonant of the following syllable, incorporated as coda through ambisyllabication, as in [pũndʲe] <= /puCde/:



homophonous words: [pũnd$_j$e]: (a) a monomorphemic noun 'brush' and (b) a verbal gerund 'stepping'; and [tũĩ.^{n}d$_j$e]: (a) a monomorphemic noun 'order, sequence' and (b) a verbal gerund 'following'. Although phonetically identical, we posit different underlying phonemic forms and different derivations, see (8). Note that in the (b) forms, prenasalization had little effect and that nasality spreading took place from the nasal moraic segment to the preceding peak vowel.

(8)

a.	Prenasalization		Nasality spreading	
/puCde/ 'brush'	=>	[puC ^{n}d$_j$e]	=>	[pũnd$_j$e]
/tuIde/ 'sequence'	=>	[tui ^{n}d$_j$e]	=>	[tũĩ.^{n}d$_j$e]
b.	Postnasal neutralization		Nasality spreading	
/puN + te/ 'stepping'	=>	/puN de/	=>	[pũnd$_j$e]
/tuĨ + te/ 'following'	=>	/tuĨ de/	=>	[tũĩ.^{n}d$_j$e]

Nasality harmony was an important feature of OJ and EMJ phonology, together with prenasalization of the mediae making nasality a prominent feature in the phonetic make-up of OJ and EMJ. Nasality, postnasal neutralization, and medial voicing made OJ and EMJ sound very different from late LMJ and NJ, in which prenasalization, nasality harmony, postnasal neutralization, and medial voicing were all lost, cf. 11.1.

7.1.3 Transcription of moraic segments

The notation of moraic segments by capital letters looks clumsy. Furthermore, the morphological constituency of a form is not always obvious, making the notation or not of phonemic nasality arbitrary, so unless it is relevant in the context and we wish to show this in phonemic transcription, we shall use a semi-allophonic notation and transcribe forms such as /taCto-/ 'precious' as *tatto-*, /siQ-te/ as *sitte*, /puCde/ and /puN-te/ as *punde*, and /tuIde/ and /tuĨ-te/ as *tuĩde*.

7.1.4 Sources of long syllables: the onbin *sound changes*

The phonological changes that triggered the introduction of long syllables were the so-called '*onbin*' sound changes, which occurred some time between the late eighth and the early tenth century. The term *onbin* (音便) 'euphony' is used in Japanese philology to refer to a variety of phenomena, including both sound changes and morphophonological processes, which have in common that they are conceived of as being motivated by ease of articulation

and involve the change of a CV syllable to a bound moraic segment. (Earlier, *onbin* was used in a wider sense to include also *rendaku* (2.6.2) or *renjō* (7.2.1, 11.4.21.)

It is no exaggeration to say that the *onbin* changes comprise the defining set of linguistic changes of the EMJ period, which left the language looking at the end of the period much like NJ in phonology and morphophonology. First of all, the *onbin* changes led to the introduction of the new syllable structure and system of bound moraic phonemes described above, resulting in a major typological change in the phonological structure of Japanese. Second, they resulted in changes in the morphophonology of verb and adjective inflection, as the sound change affected the shape of inflected forms and eventually was regularized to give consonant base verbs a new stem derived from the infinitive (e.g. OJ *ipi-* 'say' > EMJ *ipi-* ~ *iU-/iQ-* or *yomi-* 'read' > *yomi-* ~ *yoŨ-/yoN-*, see 8.1.4) and the adjectives new shapes of some forms (e.g. *taka-ki* 'tall-ACOP.ADN' > *taka-I* and *taka-ku* 'tall-ACOP.INF' > *taka-U*); see 8.2).

Table 7.2 gives some examples. Those listed under (a) exemplify the core sound changes involved in *onbin* (see 7.1.4.1); those under (b) are morphophonemic changes analogically modelled on the sound changes, and (c) are sporadic word individual changes which fall outside the core sound change, but which yielded forms with one of the new phonemes. The outcomes are listed in two columns, depending on whether the resulting bound moraic phoneme was a consonant or a vowel. In addition to their far-reaching influence on the phonology of Japanese, there are several noteworthy features which set the *onbin* changes off from regular sound changes. First, the *onbin* changes give rise to variant shapes of individual words or morphemes which coexisted for a considerable period. Second, they were non-automatic sound changes; that is to say, there were certain phonological conditions which had to be met for *onbin* to take place, but no identifiable phonological conditions under which the changes must occur. Third, as shown by the examples in Table 7.2, some source syllables gave variable outcome, either a moraic consonant or vowel.

Onbin forms appear in the written sources from the beginning of the EMJ period. These forms are first of all found in Buddhist *kunten* glosses, which to a large extent reflect the contemporary vernacular (see 6.2.2). A relative dating based on phonological criteria shows first that the *onbin* changes predate the merger of intervocalic /-p-/ with /-w-/ which took place in the second half of the tenth century (7.3.1.1): (a) /-pV/, but not /-wV/ syllables gave *onbin*; (b) *onbin* took place before /-p/, but not before /-w/. Furthermore, OJ /Ci/ syllables took part in the *onbin* changes, whereas OJ /Cwi/ syllables did not. OJ /Cwi/ was much less frequent in the lexicon than /Ci/ and this correlation could therefore be fortuitous, but it is significant that UB verbs whose base

Table 7.2 *Examples of* onbin

	Vocalic	Consonantal
(a) Core sound changes		
nipi + *ta* 'new + field; proper name' >	*niuta*	*nitta*
omopite 'think.GER'	*omoute*	*omotte*
taputwo- 'exalted' >	*tauto-*	*tatto-*
aki + *bito* 'trade + man' >	*akiūdo*	*akindo*
ywobite 'call.GER'	*yoūde*	*yonde*
*sabusabu-si 'lonely' >	*saūzaū(-si)-*	
kwo + *miti* 'small + road' >	*koūdi*	*kondi*
tumite 'pluck.GER'	*tuūde*	*tunde*
pi + *muka(si)* 'sun-facing, east' >	*piūga* 'proper name'	*pingasi* 'east'
saki-dat- 'ahead-set.out, precede' >	*saidat-*	
okite 'put.GER' >	*oite*	
yo-ki 'good-ACOP.ADN' >	*yo-i*	
kani-pata 'crab-loom, silk brocade' >		*kanbata*
sinite 'die.GER' >		*sinde*
kinu + *kaki* 'silk + fence; silk curtain'		*kingai*
takume 'wholly' >	*taume*	
yo-ku 'good-ACOP.INF'	*yo-u*	
tugite 'order, sequence' >	*tuīde*	
kogite 'row.GER'	*koīde*	
kagupasi- 'fragrant' >	*kaūbasi-*	*kanbasi-*
(b) Purely morphological *onbin*		
motite 'hold.GER'		*motte*
torite 'take.GER'		*totte*
idasite 'take.out.GER'	*idaite*	
aru meri 'exist EVID'		*anmeri*
(c) Sporadic word individual changes		
mawos- 'say.HUM'	*maus-*	
mawide 'visit.HUM'	*maude* 'shrine/temple visit'	
*kari-ta 'harvest-field; proper name'		*katta*
*kari-da 'harvest-field; proper name'		*kanda*
kagapuri 'headwear (on the head)'	*kaūburi*	
kaditori 'oarsman'		*kandori*

and infinitive in OJ ended in /-wi/ did not acquire *onbin* stems (8.1.4), see (9). If the *onbin* changes had taken place after the merger of /Ci/ and /Cwi/ there is no good phonological reason that the UB verbs should not have acquired *onbin* variants. This means that the *onbin* changes in the main predate the merger of /Ci/ and /Cwi/ (7.3.2.1), i.e. they must originate in the OJ period.

(9)

		OJ	EMJ
QD	'put'	*oki-*	*oki-* ~ *oi-*
UB	'arise'	*okwi-*	*oki-*

Thus although both the *kunten* materials and a relative dating of *onbin* show that *onbin* forms were an established part of the language by the time of the appearance of the first mainstream EMJ written sources at the beginning of the tenth century, *onbin* forms are underrepresented in the literary sources and remained so through the entire MJ period. In particular, *onbin* forms are virtually never found in poetry, because of the conservative norm in poetic language (6.2.1).[2] On the whole, we are not able to say much in detail about the course of use and spread of *onbin* forms through MJ on the basis of the written sources, for *onbin* forms never became part of the written norm which emerged in the course of the second half of the EMJ period. However, as may be expected we find *onbin* forms far more widely represented in prose writing, but also there the impression is that *onbin* forms are often used to impart a flavour of spoken language. This is confirmed by remarks by the priest and Sanskrit philologist Meigaku in his *Shittan yōketsu* (1101) where he characterizes *onbin* forms as being characteristic of spoken language, but the source forms as being used in writing (cited in Mabuchi 1971: 92). His examples are mainly various forms of morphological *onbin*, e.g. *be-i* NEC-ACOP.ADN being used in speech instead of *be-ki* which was used in writing and therefore considered more correct, or *tot-te* 'take-GER' being used instead of *tori-te*, but he also gives lexical examples, saying for example that *komiti* (< OJ *kwo-miti*) is the correct form, but that people popularly say *koūdi*. These remarks, some centuries after the *onbin* sound changes took place, reflect both a tension between written conservative norms and spoken language, as well as the fact that the *onbin* changes first of all led to the establishment of variant forms of words or morphemes, in many cases not the replacement of an older form by a newer. The variation that was established was in some cases stylistic (like the variation in NJ between *watakushi*, *watashi*, *atashi* for 'I'), and in others it became grammatically conditioned, as with the variation between infinitive and *onbin* stem of consonant base verbs, e.g. OJ *sini-* 'die.INF' > EMJ *sini-* (INF) ~ *sin-* (*onbin*-stem) subject to selection by different contexts (8.1.4). In some cases the *onbin* form eventually replaced the source form, e.g. *kaki-mami-* ('peek at through a fence; lit. fence-space-look') > *kaimami-* 'have a glimpse of', or the adjectival adnominal formant *-ki* > *-i*. Conversely, in other cases it was the source form which survived into NJ, e.g. *saki-dat-* 'ahead-

[2] The first EMJ poetry anthology *Kokinshū* (914) stands out by having a single *onbin* form, *tika-u* ('close-ACOP.INF'), which is however used in a peculiar context, as an allusion to the flower *kitikau*.

set.out; precede' and not the attested *onbin* form *sai-dat-*. A number of *onbin* forms are attested only in *kunten* glosses or dictionaries, but were not used in general writing. Finally, in some cases both *onbin* and source forms survived. This includes the establishment of grammatical alternation, such as between infinitive and *onbin* stem of the consonant base verbs, but also lexical items with some semantic or other specialization between the source and *onbin* form. A lexical example which illustrates the different possibilities is OJ *kwo-miti* 'small-road', which gave the attested *onbin* variants EMJ *koũdi* and *kondi*. Of these, *kondi* did not survive into NJ, whereas both *koũdi* > NJ *koozi* 'narrow backstreet' and *kwo-miti* > *ko-miti* 'small road' are used today, but with semantic specialization.

7.1.4.1 Onbin *as sound changes*

Of the various forms which in Japanese philology are identified as *onbin*, it is possible to identify a set of core changes which can be said to constitute the '*onbin* sound changes', i.e. a core of changes of certain CV syllables to bound moraic phonemes. This was preliminarily shown in Table 7.2 (a) above and is summarized here in (10). These changes account for the great majority of early examples of changes of lexical words. It was these changes which formed the basis for the accompanying change in phonological structure and whose outcomes served as a model for the morphophonological changes involved in *onbin*. Changes which are usually classified as *onbin* but are not included here are either sporadic changes which yielded forms with one of the bound moraic phonemes, but generally are late and singular, listed in Table 7.2 under (c), or morphophonological developments such as those in (b) in Table 7.2 which are only found in the verb paradigms; such developments are analogical formations (see 8.1.4). It is traditional to distinguish four types of *onbin*: I-, U-, N-, and Q-*onbin*, depending on the bound moraic phoneme which resulted from the change, but we expand that to allow for the distinction between /I, U/ and /Ĩ, Ũ/. As mentioned above, some source syllables gave variant outcomes, either a consonant or a vowel.

(10)

Source syllable	Moraic vowel		Moraic consonant	
	Oral	Nasal	Oral	Nasal
-pi, -pu >	U		Q	
-bi, -mi, -mu, -gu >		Ũ		N
-ni, -nu >				N
-bu >		Ũ		
-ki >	I			
-ku >	U			
-gi >		Ĩ		

The source syllables thus consist of one of the consonants /p, k b, g m, n/ and one of the high vowels /i, u/. As mentioned above, (7.1.4), syllables with OJ /Cwi/ (i.e., /pwi, kwi, bwi, gwi, mwi/) do not appear to have given *onbin*, with one or two exceptions.[3] In other words, it was only OJ /Cu/ and /Ci/, not /Cwi/, which gave *onbin*. A phoneme following the syllable that gave *onbin* was always a true consonant, /p, t, k, s, b, d, g, z, m, n/; that is to say, *onbin* did not take place before /r, w, y/.

7.1.4.2 Syllable reduction

The *onbin* changes are sometimes described as, or illustrated by, segment loss and consonantal assimilation, so that for example *takume* is said to have become *taume* through simple loss of /k/; or *pimuka(si)* > *piũga* through loss of /m/, and > *pingasi* through loss of /u/, in which case the alternative outcomes are explained by the loss of different segments. It is possible to state the correspondence between source form and *onbin* form in this way in some cases, but not for example /pi/ > /U/, or /bi, mi/ > /Ũ/. A more comprehensive and realistic account describes these changes as phonemic reinterpretations of a phonetically reduced or weakened realization of a CV syllable as the realization of a single segment. This is illustrated in (11) which shows three stages of realization of the syllables involved. Column (a) shows the phonemic shape of the source syllable. (b–c) show phonetic variation which was regular in OJ: in both of them are shown intervocalic voicing of tenues and prenasalization of mediae (cf. 2.2.2), and tonality adjustment (2.3), and (c) alone shows continuousness variation among /p, k, b, g/ (2.2.3). In column (d) is shown a phonetically further reduced realization in which the CV syllable is realized as one unit with little contrast. The *onbin* sound changes thus consist in the reinterpretation of the realization in column (d) as deriving from a different phonemic representation than that in (b) and (c), shown in the final column, (e).

The kind of phonetic reduction shown in (d) is typical of languages, like OJ, whose main or only syllable structure is CV; in particular, neutral or high vowels are usually reduced except in slow, distinct or elaborate diction, as is the case in NJ, especially as long as they leave a trace in the preceding consonant, for example in the form of palatalization or labialization. The more familiar or frequent in use a word is, the more likely it is to be phonetically reduced (this general principle is known as *Zipf's Law*). Many of the words which gave *onbin* were complex forms which at the same time were

[3] An apparent exception is *tuitati* 'the first day of the month' which is usually said to be from *tukwi-tati 'moon-rise' (which is itself not attested in OJ). It is possible, however, that this is a folk etymology and that *tuitati* instead derives from a compound of *tuki* 'arrive' + *tati* 'depart' (notionally paralleling the name of the month January, named after the double-faced god Janus in Roman mythology). If so, *tukitati > *tuitati* is no exception.

univerbated, e.g. *kwo-miti* 'small road' > *koũdi* 'narrow backstreet', and this fits the same pattern. Whether individual words gave *onbin* will have depended on their individual use, and this helps account for the fact that the *onbin* changes were non-automatic, yet pervasive. The role played by familiarity, and speech tempo and style, also makes it easy to understand the stylistic variation and distribution among unchanged and *onbin* forms of the same words.

(11)	(a)		(b)		(c)	(d)		(e)
	/-pu/	=>	[b^wu	~	β^wu]	[β^w:]	>	/U, Q/
	/-pi/	=>	[b_ji	~	β_ji]	[β_j:]	>	/U, Q/
	/-ki/	=>	[g_ji	~	γ_ji]	[γ_j:]	>	/I/
	/-ku/	=>	[g^wu	~	γ^wu]	[γ^w:]	>	/U/
	/-gi/	=>	[$^{ŋ}g_j$i	~	$^{ŋ}\gamma_j$i]	[$^{ŋ}\gamma_j$:]	>	/Ĩ/
	/-gu/	=>	[$^{ŋ}g^w$u	~	$^{ŋ}\gamma^w$u]	[$^{ŋ}\gamma^w$:]	>	/Ũ, N/
	/-bi/	=>	[$^m b_j$i	~	$^m\beta_j$i]	[$^m\beta_j$:]	>	/Ũ, N/
	/-bu/	=>	[$^m b^w$u	~	$^m\beta^w$u]	[$^m\beta^w$:]	>	/Ũ/
	/-mi/	=>	[m_ji]			[m_j:]	>	/Ũ, N/
	/-mu/	=>	[m^wu]			[m^w:]	>	/Ũ, N/
	/-ni/	=>	[n_ji]			[n_j:]	>	/N/

7.1.4.3 Nasality

As shown in (10) and (11), syllables with initial tenuis, /-pi, -pu, -ki, -ku/, gave oral *onbin*, /I, U, Q/, whereas syllables with nasal or media, /-bi, bu, -gi, -gu, -mi, -mu, -ni, -nu/, gave a nasal moraic phoneme, /Ĩ, Ũ, N/, reflecting the phonetic nasality of the reduced realization which in the case of the mediae derived from prenasalization. As shown in the examples in Table 7.2, a tenuis in the source form changed to a media after a moraic phoneme deriving from /-bi, bu, -gi, -gu, -mi, -mu, -ni, -nu/, e.g., *pumite* 'brush' > *punde* and *pumite* 'stepping' > *punde*. However, there was a difference in the phonemic shape of such *onbin* forms, depending on the morphemic constituency. As discussed above (7.1.2), nasality was not phonemic for moraic phonemes in morpheme internal position; thus, in the course of the *onbin* sound changes, the phonetic nasality of the reduced realization of the source forms (11d) was phonemicized differently depending on whether it was in morpheme final or morpheme non-final position, see (12), which shows both the phonemic shape and phonetic realization of (a) the source forms and (b) the resulting *onbin* forms. For the forms in (b), which have a morpheme boundary after the moraic phoneme, the phonetic nasality was assigned phonemically to the morpheme final moraic phoneme (/puN- + -te/) and the phonetic realization was derived as described in (8b) in 7.1.2.4. However, for the forms in (c), which have the moraic

phoneme in morpheme internal position, the phonetic nasality in the source form in (a) was in the course of the *onbin* sound change re-interpreted as deriving from a (prenasalized) media, but the moraic phoneme itself as being underspecified for nasality, i.e. /puCde/ whose phonetic realization was derived by the rules described in (8a) in 7.1.2.4.

(12)	a.		b.
(i)	{/pumi + te/ 'step + GER'	=> [pũm$_j$:d$_j$e]} >	{[pũnd$_j$e] <= /puN + te/}
	{/tugi + te/ 'follow + GER'	=> [tũ^{ŋ}ɣ$_j$:d$_j$e]} >	{[tũĩ.^{n}d$_j$e] <= /tuĨ+te/}
(ii)	{/pumite/ 'brush'	=> [pũm$_j$:d$_j$e]} >	{[pũnd$_j$e] <= /puCde/}
	{/tugite/ 'sequence'	=> [tũ^{ŋ}ɣ$_j$:d$_j$e]} >	{[tũĩ.^{n}d$_j$e] <= /tuIde/}

Thus, phonemically the diachronic correspondences involved in *onbin* differ a great deal depending on whether the resulting moraic phoneme occurred in morpheme initial or non-initial position: /pu**mi** + te/ > /pu**N** + te/, but /pu**mi**te/ > /pu**Cd**e/. However, in both of these cases we speak of /mi/ as having given N-*onbin*, in the one case (morpheme final position: /pu**mi** + te/ > /pu**N** + te/) reflected as a nasal moraic phoneme, in the other (morpheme internal position: /pu**mi**te/ > /pu**Cd**e/) reflected as a moraic phoneme underspecified for nasality but followed by a media which contributed the phonetic nasality to the surface form, through prenasalization and nasality spreading. And in both cases we use *punde* as a convenient transcription when we are not specifically concerned with the phonemic shape of the form (cf. 7.1.3). An important point about (12) is that it shows that the *onbin* sound changes involve the change of one set of phonemic representation and its phonetic representation, shown in column (a), into another set of phonemic representation and its phonetic realization, column (b). Thus, stating the change as {/pumite/ 'brush' => [pũm$_j$:d$_j$e]} > {[pũnd$_j$e] <= /puCde/} says a great deal more than does a simple statement of the phonemic diachronic correspondence, such as /pumite/ > /puCde/.

Some examples in OJ of syllable loss seem to involve the same kind of phonetic reduction as was involved in *onbin*, see (13), cf. 2.7.1.1. The main difference between the developments in (13) and in (12) is that no moraic phoneme arose in the examples from OJ in (13). This suggests that the phonetics which in the transition between OJ and EMJ gave rise to *onbin* already were a feature of OJ. Similar reductions took place through EMJ, e.g. the copula gerund *nite* > *de* (8.2) or *nani-to* > *nado* (8.7.3). The morphophonological process of *rendaku* is also thought to originate in similar reduction of particles such as genitive *no* and dative *ni* (2.6.2).

(13) {/*ami-piki/ 'net-pulling' => [ãm$_j$:b$_j$ig$_j$i]} > {[ã^{m}b$_j$ig$_j$i] <= /abiki/}
{/kigisi/ 'pheasant' => [k$_j$i^{ŋ}ɣ$_j$:^{d}ʒi]} > {[k$_j$indʒi] <= /kizi/}

7.1.4.4 Major class; consonant or vowel?

An interesting aspect of the *onbin* changes is that some source syllables gave both vocalic and consonantal outcomes. The basis for this is that the reduced realization in (11d) in some cases lent itself to interpretation as both a vowel and a consonant. It is not possible based on the written sources to determine any systematic distribution of the consonantal and vocalic outcomes. The consonantal outcome is often assumed to have been characteristic of eastern dialects. This is quite possible, and fits in well with an overall characterization of western Japanese as more vocalic and eastern Japanese as more consonantal, as reflected today for example in western NJ having less vowel devoicing than eastern NJ; however, for *onbin* this characterization is to some extent based on the present-day distribution of verbal *onbin* forms, which up until the end of the LMJ period was not as clear-cut as it is today (see 12.3.2).

7.1.4.5 Other sources of bound moraic phonemes

Apart from *onbin*, /I, U/ result from the regular loss of /p/ and /w/ before /i/ and /u/ which took place between the end of the tenth century and 1100, see below (7.3.1.1, 7.3.2.3). As mentioned below, bound moraic phonemes, both consonants and vowels, were frequent in SJ vocabulary (7.2). Both the *onbin* sound changes and the introduction of syllable length are often ascribed to influence from SJ. This is, however, doubtful, because of the fact, amongst others, that *onbin* forms and SJ vocabulary occupied different ends of the stylistic spectrum. It is rather the case that incorporation of SJ loanwords was facilitated by the emergence of the new syllable structure and phonemes. Thus, early loans from Chinese in OJ reflect EMC *-ŋ by /gV/ (e.g. EMC *ʂa̍iwŋ/ ʂœ̍iwŋ (双) > OJ *sugu* 'double'), but later by /Ũ, Ĩ/ (cf. the SJ readings of 双: *soũ, saĩ*; see 9.2.2.3).

7.2 The sound shape of Sino-Japanese vocabulary

The intake of SJ loanwords resulted in the introduction of some sound combinations and distributions which were initially only used in the SJ vocabulary. First of all, there were no restrictions on the occurrence of word initial mediae or liquid, that is to say, SJ loanwords freely had initial *b-*, *d-*, *g-*, *z-*, *r-* (see e.g. examples in 9.2.3.2 below). Second, as shown in Table 7.1 above, SJ loanwords (re)introduced complex syllable onsets, /Cy-/ or /Cw-/: /Cya, Cyo, Cyu; Cwa, Cwe/. Note that on the phonemic interpretation of the *kō–otsu* distinctions adopted in this book, also OJ had complex

syllable onsets, but with different combinations of glides and vowels: /Cye, Cwi, Cwo/ (cf. 2.1.4). Of the new complex syllable onsets introduced by SJ loanwords, /Cya-/ was the most frequent. With the exception of a number of words with /syu, syo, zyu, zyo/, other /Cyu, Cyo/ were rare in EMJ, although they are frequent in NJ; they are included in Table 7.1 in brackets because they are found in a few SJ words taken in towards the very end of the EMJ period. SJ /Cyo/ occurs in /Cyoku/ and a few words with /Cyoũ/ (:: EMC *-iŋ) in words mostly taken in during LMJ. Most other instances of NJ SJ /Cyu, Cyo/ are found in the outcome of LMJ contractions of EMJ /Cyaũ, Ceu, Ciu, Cipu/ (see 11.5). Through sound changes in LMJ, /Cya, Cyo, Cyu/ also came to occur in the native vocabulary. /Cw-/ is only found in /kwa, gwa, kwe, gwe/; /kwe/ is found in a single native word, the verb *kwe-* 'kick' (see 8.1.2).

In general, SJ loanwords are characterized by a high frequency of long syllables. In addition to the bound moraic phonemes /I, U, Ĩ, Ũ, N, Q/, some SJ words had syllable final /-t/ (reflecting EMC *-t, see 9.2.2.3). There was no ready way in the *kana* script of noting syllable final /-t/ and the main evidence is from forms exhibiting *renjō* (see 7.2.1 immediately below) and in late LMJ transcriptions, such as *fotnet suru* 'get fever'. Syllable final /-t/ will be further discussed in 11.4. Furthermore, SJ words are often said to have had a distinction between syllable final /-n/ and /-m/ (reflecting EMC *-n and *-m). The sole evidence invoked for positing syllable final /-m/ is from forms exhibiting *renjō*, but it seems likely that the distinction between /-m/ and /-n/ was in fact never a part of Japanese, see below (7.2.1).

7.2.1 Renjō

Renjō (連声 'liaison') is a phonological process of *onset creation* which originates in SJ. When a syllable final consonant was followed by a syllable with initial vowel or glide, the consonant was carried over to create the onset in the next syllable. An early example from (some versions of) the *Genji monogatari* is *onmyaũzi* 'fortuneteller' (> cNJ *onmyōji*) which reflects Chinese 陰陽師 whose individual SJ *kanji* readings are *on*, *yaũ* (> cNJ *yō*), and *si*. The form *onmyaũzi* thus both shows creation of an onset in the second syllable, as well as suggesting that the reading of the first *kanji* ended in *-m*, as in EMC: *ʔim. *Sanmi* 'third rank' (*san* 'three' 三 :: EMC *sam + *wi* 'rank' 位) is a similar example. However, rather than positing distinct syllable final /-m/ within the SJ sound system on the basis of a few examples like these, it seems more likely that forms such as *onmyaũzi* or *sanmi* were taken in as whole words from the Chinese used in Japan ('Japano-Chinese') and that there never was a distinction between /-m/ and /-n/ in SJ. See 9.2 about the distinction between Japano-Chinese and Sino-Japanese.

Renjō is usually thought to have been a regular, automatic feature of pronunciation which was therefore not noted much in writing. It ceased to be productive early in NJ where it only survives in a small number of lexicalized forms, see (14) which gives the cNJ form and the EMJ form if the two are different, as well as the cNJ and EMJ readings of the individual *kanji*. Note that /C.wV/ => /C.CV/, with loss of the labial glide.

(14) *tennō* 'emperor' < *tennaũ* (天 *ten* + 皇 *ō* < *waũ*)
ninnaji < *ninnazi* 'Ninnaji' (仁 *nin* + 和 *wa* + 寺 *zi*)
rinne 'transmigration (of souls)' (輪 *rin* + 廻 *e* < *we*)
hannō < *pannoũ* 'reaction' (反 *han* < *pan* + 応 *ō* < *oũ*)
kannon < *kwannon* 'Kannon' (観 *kan* < *kwan* + 音 *on*)
innen 'destiny' (因 *in* + 縁 *en*)
annon 'peace and quiet' (安 *an* + 穏 *on*)
ginnan 'gingko nut' (銀 *gin* + 杏 *an*)
kannō < *kannoũ* 'sympathetic response' (感 *kan* + 応 *ō* < *oũ*)
unnun 'such and such' (云 *un* + 云 *un*)

setchin < *settin* 'toilet; "tight spot"' (雪 *seti/setu* < *set* + 隠 *in*)
kuttaku 'worry' (屈 *kuti/kutu* < *kut* + 惑 *waku*).

7.3 Regular segmental sound changes

Two overall sets of sound changes took place during EMJ: the merger of intervocalic /p/ with /w/ (7.3.1.), and the loss of (phonemic) glides before /e, i, o/, including the loss of the *kō-otsu* distinction (7.3.2). These changes are all mergers and are textually manifested as loss of orthographic distinctions. An overall summary of sound changes which took place during this and following periods is given in the appendix.

7.3.1 Changes affecting OJ /p/

One of the main issues in Japanese historical phonology is the complex of changes which have affected OJ /p/, from EMJ to NJ. They are summarized in (15):

(15) a. /-p-/ > /-w-/ > Ø before all vowels except /a/.
b. /p/ > /f/ > /h/ (realized as [ɸ] before /u/ and as [ç] before /i/)

The first change, of intervocalic /-p-/ to /-w-/ took part in EMJ and will be discussed in this chapter. After that merger /-w-/ was lost before /i, e, o/, see 7.3.2.3. The second change, of /p/ to /f/, appears to have taken place in LMJ

and will be addressed in 11.3, as the first positive evidence we have for the stage /f/ is in the Christian materials from the end of LMJ, but we will in this chapter briefly review evidence for the sound value /p/ in EMJ of the reflex of OJ /p/ (7.3.1.3). The change of /f/ > /h/ took place during NJ, see 14.3.

7.3.1.1 Merger of intervocalic /-p-/ with /-w-/

Some time in the second half of the tenth century, intervocalic /-p-/ changed to /-w-/ (merging with earlier /-w-/), except before /u/ where /-p-/ was lost, see (16). This change resulted in the merger of previously distinct forms, exemplified in (17). Although this sound change properly dates from the second half of the tenth century, early word-individual examples are seen already at the beginning of EMJ. A well-known early example is the adjective OJ *urupasi* 'splendid' which already in some early EMJ materials is spelled *uruwasi*.

(16)

p	>	w /	V__	{i, e, a, o}
		Ø /	V__	u

(17)

OJ		EMJ
apa 'millet' ≠ *awa* 'foam'	>	*awa*
api 'meeting' ≠ *awi* 'indigo'	>	*awi*
kapo 'face' ≠ *awo* 'blue'	>	*(k)awo*
upe 'top' ≠ *uwe* 'planting'	>	*uwe*
apu 'meets'	>	*au*

OJ did not have the syllable /wu/. The fact that /-p-/ in the course of this sound change was lost before /u/, and did not give /wu/, indicates that the absence of the syllable /wu/ was no accident, but the result of an active phonotactic restriction. The /u/ which resulted from /-pu/ was incorporated into the preceding syllable as the second mora of a long syllable. /-pu/ > /-u/ was frequent in SJ vocabulary, where early EMJ /-pu/ reflected EMC syllable final *-p (see 9.2.2.3), e.g. early EMJ *sipu* (集) 'anthology' > late EMJ *siu*, *tapu* (塔) 'stupa' > *tau*.

The sound change /-p-/ > /-w-/, merging intervocalic /-p-/ with the already existing phoneme /-w-/, is a straightforward sound change, which is observable and dateable through the written sources as orthographic confusion of *kana* for *pV* and *wV* syllables in medial position. Its background is the phonetic realization of tenues (/p, t, k, s/) mentioned in 2.2.3, so that intervocalic /-p-/ already in OJ was phonetically non-distinctively *voiced* and had both *occlusive* and *fricative* variants: /-p-/ => [b ~ β]. The phonemic change consisted in the reinterpretation and phonemic identification of these phonetic realizations of /-p-/ with /-w-/, due to the phonetic similarity of [β] and [w], likely related to

the fricative realization of /-p-/ gradually becoming more prominent or frequent.

(18)
/-p-/ =>	[b ~ β]	>	/-w-/	
/-w-/ =>	[w]			
/apa/ =>	[aba ~ aβa]	>	/awa/	
/awa/ =>	[awa]			

This scenario makes it easy to understand the appearance of early word-individual instances of this change, such as *urupasi* > *uruwasi*: as long as /-p-/ and /-w-/ coexisted, individual words could be reinterpreted as having /-w-/ rather than /-p-/, with no effect on the phonological system. Although all intervocalic tenues were phonetically voiced, and it is likely that /k/, like /p/, had fricative variants, /k/ or /t/ did not undergo similar changes, as there was no scope for reinterpreting their phonetic realizations as realizations of other already existing phonemes; thus, [-g- ~ -ɣ-] and [-d-] remained uniquely identifiable as realizations of /-k-/ and /-t-/. As the mediae, /b, d, g, z/, were prenasalized, [^{m}b, ^{n}d, ^{ŋ}g, ^{n}z] (cf. 2.2.2), they were readily distinguishable from the intervocalic phonetically voiced realizations of /-p-, -t-, -k-, -s-/.

It was earlier thought that the change /-p-/ > /-w-/ proceeded through the EMJ period as something like (/-p-/ [p] >) /-ɸ-/ [ɸ] > /-w-/. On that view, the change mainly consisted in the introduction of intervocalic voicing of /-ɸ-/ to /-w-/, but not of other tenues. However, it has, since Günther Wenck proposed this in the early 1950s, become increasingly accepted that OJ tenues were phonetically voiced intervocalically, and following on from that that /-p-/ > /-w-/ did not have the intermediary stage /-ɸ-/ [ɸ], but that the change proceeded as in (18).

7.3.1.2 Retention of /-p-/

/-p-/ is retained after the moraic oral obstruent /Q/. A few emphatic words have preserved /-p-/ after /Q/ into NJ, e.g. *appare* 'splendid, brilliant; bravo!' (OJ *apare* 'wow!; being evocative of deep feeling' which also has the reflex *aware*, from Heian times a central element in Japanese aesthetics), *moppara* 'solely, entirely' (OJ *mopara*). It is likely that these words existed as emphatic forms with a forcefully articulated /p/ already in OJ and were phonemicized with /Qp/ after the emergence of /Q/. Among SJ vocabulary the retention of /p/ after /Q/ has been regularized so that all NJ SJ morphs with initial /h-/ (< /p/) have automatic allomorphs in /-p-/ for use after /Q/. Similar alternations are preserved in the native vocabulary, so that for example NJ verbs in initial /h-/ have /p-/ when combining with the prefix *hiQ-* (< EMJ *piC-* < *piki-* 'pull.INF'), e.g. *hip-par-* 'pull, drag' (cf. *har-* 'stretch, spread'). Such forms are now

lexicalized and some may have been formed analogically, but the point remains that they show preservation of /p-/ after /Q/.

When /p/ was initial in the second component of a transparent compound it did not change to /w/, but remained /p/. All such remaining cases of /p/ later became LMJ /f/ > NJ /h/ (see further 11.3, 14.3; for example OJ *asa-pi* 'morning-sun' > LMJ *asa-fi* > NJ *asa-hi*. Proper names present an interesting case. Some, such as early EMJ *Pudi(-)para* > late EMJ *Pudiwara* > NJ *Fujiwara*, have /-p-/ > /-w-/ and therefore seem to have been univerbated, whereas others such as EMJ *Naka-para* > NJ *Nakahara*, or *Kiyo-para* > *Kiyohara* behave like compounds and have /-p-/ > /f/ > NJ /h/, although all involve a reflex of OJ *para* 'plain, field' as their second element. The present-day name of the supposed main character of the *Ise monogatari*, *Ariwara no Narihira* (825–80), which in his day would have been *Aripara no Naripira*, exhibits both types of development. Apart from a negligible number of exceptions, such as *bipa* (琵琶) 'flute' > *biwa*, SJ vocabulary with a second component in initial /p/ were treated as transparent compounds in which /-p-/ did not change to /-w-/, but was left as /p/ which eventually gave LMJ /f/ > NJ /h/, for example OJ/EMJ *mupon* (謀反, EMC *muw puan'; only logographic OJ attestation) 'rebellion, treason' > LMJ *mufon* > NJ *muhon*.

7.3.1.3 The Early Middle Japanese sound value of the reflexes of initial /p-/

Until fairly recently the widely accepted consensus view was that pre-OJ /*p/ already by, or in, OJ had changed to become /ɸ/. There is, however, little to support this view.

With some regular exceptions (see 11.3), the reflex in NJ of pre-OJ initial /*p-/ is /h/, realized as [ɸ] before /u/ and as [ç] before /i/. The Christian materials from the end of the LMJ period where the reflex of OJ /p-/ is transcribed 'f', e.g. *fito* 'person' (OJ *pito*, NJ *hito*), *fana* 'nose' or 'flower' (OJ *pana*, NJ *hana*), show that an intermediary stage /f/ (possibly realized as [ɸ]) existed between pre-OJ /*p/ and NJ /h/: /*p/ > /f/ > /h/. Apart from the *terminus ante quem* provided by the Christian materials, the change of /p-/ > /f-/ is very difficult to date for it resulted in no mergers and found no orthographic expression. The only materials available are evidence from the sound values of *man'yōgana*, internal alternations, some notes on pronunciation from the EMJ period, considerations of the *kana* orthography, and foreign transcriptions, primarily the Christian materials.

First of all, it is no longer controversial to regard the reflex of pre-OJ /*p/ as having been /p/ in OJ. That is amply shown by the reconstructed EMC sound values for the *ongana* used for those syllables, as well as by phonological alternations between OJ /p/ and /b/. The earliest indication which is taken

as positive evidence for a sound value [ɸ] or [f] for the reflex of OJ /p-/ are articulatory descriptions or instructions in the priest Shinren's *Shittan kuden* and *Shittan sōden*, both from the very end of the Heian period. This material has traditionally been taken to show that the reflex of OJ /p-/ at that time was pronounced as a bilabial fricative or approximant, but the fact is that Shinren's descriptions are obscure and difficult to interpret.[4] Furthermore, as Kiyose (1985) points out, there are strong internal arguments in favour of viewing the EMJ reflex of OJ initial /p-/ as having been /p/ throughout that period and probably well into the LMJ period. The main facts are (a) that the reflex of OJ /p-/ was paired with /b/ in exactly the same way /t, k/ were paired with /d, g/ in sound changes and phonological rules through the EMJ period; and (b) that single *kana* letters were used to represent both /pV/ and /bV/ syllables, for example は used to represent both /pa/ and /ba/, in exactly the same way that か represented both /ka/ and /ga/ and た both /ta/ and /da/ (6.1.2.1); this shows that /p/ and /b/ were phonologically paired as tenuis and media at the time the letter categories of the *kana* writing became fixed. Finally, when the change /p/ > /f/ took place, /p/ was preserved after /N, Q/, providing further support for /p/ having remained /p/ into the LMJ period, see 11.3.

7.3.1.4 Summary

Summarizing, OJ /p/ phonemically remained /-p-/ in intervocalic position (with voiced and both occlusive and fricative variants [b ~ β]) in EMJ until its merger with /-w-/ in the tenth century. Furthermore, we follow Kiyose that initial /p-/ phonemically remained /p/ (with both occlusive and fricative variants [p ~ ɸ]) through the entire EMJ period and (well) into the LMJ period. Sometime before the end of the LMJ period /p/ > /f/ (either a bilabial fricative [ɸ] or a labio-dental [f]); this change will be discussed in 11.3.

7.3.2 Loss of labial and palatal glides

In the course of a set of changes which took place between late OJ and around 1300, the palatal glide /y/ was lost before /e/ and the labial glide /w/ was lost before /o, i, e/.

[4] The relevant passages are (cited from Kobayashi 1981): *Shittan kuden* (悉曇口伝): in the articulation of "pa", use 脣内分 'the inner part of the lips', but in the articulation of "ma", use 脣外分 'the outer part of the lips'. *Shittan sōden* (悉曇相伝): to articulate "pa": 上下脣合竦呼a而成音 'Join the upper and lower lips and say out "a" softly/gently. Then this sound will result'; and for "ma": 閉脣外極強呼a(而)成音 'Close the outside of the lips and say "a" very forcefully. (Then) this sound will result.' Much speculation has gone into interpreting these instructions and they are conventionally taken to mean that "p" was a bilabial fricative [ɸ], but this is far from clear.

7.3.2.1 Merger of kō-rui *and* otsu-rui *syllables; loss of post consonantal /y/ before /e/ and /w/ before /i, o/*

The first of these changes is a loss of the distinction between the *kō-* and *otsu-rui* syllables (cf. 2.1). These mergers set in towards the end of OJ and are in the main thought to have taken place in the transition between OJ and EMJ, but the distinction between *Cwo* and *Co* was retained longer in some contexts, notably for *kwo* ≠ *ko* into the tenth century (in texts written in *man'yōgana*). On the phonemic interpretation introduced in 2.1.4 above, and reflected in the transcription used in this book, these changes may simply be understood as the loss of a glide between the onset consonant and nuclear vowel in complex syllables, see (19). The consonant in /Ce, Ci/ which resulted from these mergers was palatalized ([C_je, C_ji]); palatalization before /e, i/ was a feature already of OJ (cf. 2.3) and it was maintained into the NJ period before /e/ and is still found today before /i/ (cf. 2.3; 11.6.1, 14.4.1).

(19) *me* 'eye' ≠ *mye* 'woman' > *me*
pi 'sun' ≠ *pwi* 'fire' > *pi*
ko 'this' ≠ *kwo* 'child' > *ko*

7.3.2.2 Loss of syllable initial /y/ before /e/

Next, syllable initial /y/ was lost before the front vowel /e/ during the first half of the tenth century, e.g. (20). This change is thought to have been complete by 950 after which *man'yōgana* and *kana* for /.ye/ and /.e/ were no longer kept distinct. Recall that the *Ame-tsuchi no kotoba* reflects a language stage when /.ye/ and /.e/ were distinct, whereas the language reflected in the *Taini-uta* and the *Iroha-uta* no longer had that distinction (6.1.2.4). This change is usually thought of as a merger of the syllables /.ye/ and /.e/, but that only holds in word initial position, as OJ had no word internal /V.e/ sequences (cf. 2.5). Note, importantly, that /.e/ resulting from the loss of phonemic /.y/ before /e/ was pronounced with an automatic phonetic onglide, as shown in (20); see further 7.3.2.4–5 below.

(20) *e-* 'to get' ≠ *ye* 'inlet' > *e* ([ʲe])
nuye 'mountain thrush' > *nue* ([nuʲe])

7.3.2.3 Loss of syllable initial /w/ before /o, i, e/

Finally, syllable initial /w/ was phonemically lost before /o, i, e/. This happened in the following stages over several centuries and was not concluded until the LMJ period. These changes took place after intervocalic /-p-/ > /-w-/ (7.3.1.1) and they therefore also affected /-w-/ < /-p-/. As opposed to the loss of post-consonantal /y, w/ and of syllable initial /y/ before /e/, the loss of syllable initial /w/ took place after the establishment of the letter categories of the *Iroha* and it is therefore ignored in the etymological *kana* spelling which retains several separate letters for each of /.o, .i, .e/ (6.1.3).

Syllable initial /w/ was lost before /o/, completed around 1000. This change is usually thought of as a merger between the syllables /.wo/ and /.o/, but as with the loss of /.y/ before /e/, that only holds in word initial position as OJ had no /V.o/. Also the outcome of this change was pronounced with an automatic phonetic onglide (see 7.3.2.4–5).

(21)

oto 'sound' ≠ *wotoko* 'man'	>	*oto(ko)* ([ʷodogo])
awo 'blue'	>	*ao* ([aʷo])
(*kapo* >) *kawo* 'face'	>	*kao* ([kaʷo])

The loss of /w/ before /i, e/ proceeded in two phases: in word medial position, concluded around 1100, e.g. (22); in word initial position, concluded around 1300, i.e. in the LMJ period, e.g. (23).

(22)

mawiru 'comes, goes (HUM)' >	*mairu*
(*mapi* >) *mawi* 'dance' >	*mai*
(OJ *opwi* > EMJ *opi* >) *owi* 'grow.INF' >	*oi*
uwe 'plant.INF' >	*ue*
(*upe* >) *uwe* 'top' >	*ue*
(OJ *mapye* > EMJ *mape* >) *mawe* 'front' >	*mae*

(23)

winaka 'countryside' >	*inaka*
wemi 'smile' >	*emi*

7.3.2.4 /-i, -u/ versus /-.e, -.o/

The changes outlined above resulted in a large number of instances of vowel sequences (/-Vi, -Vu, -Ve, -Vo/) from different sources, see (24). Recall that vowel sequences were not allowed by OJ phonotactics (cf. 2.5).

(24)

		-Vwi				
-Vi	<	-Vwi	<	-Vpi		
		-Vwi	<	-Vpi	<	-Vpwi
-Vu	<	-Vpu				
		-Vye				
-Ve	<	-Vwe				
		-Vwe	<	-Vpe		
		-Vwe	<	-Vpe	<	-Vpye
-Vo	<	-Vwo				
		-Vwo	<	-Vpo		

In EMJ, /i, u/ were incorporated into the preceding syllable to form the second mora of a long syllable, just like the /I, U/ resulting from the *onbin* changes (7.1.4) and were later subject to various contractions (11.5). On the other hand, /e, o/ maintained their independence as syllable initials and were pronounced with an automatic onglide which created a phonetic onset to the syllable: [^{j}e] and [^{w}o], respectively; this was also the case in word initial position, (25).

(25)

e- 'to get'	[^{j}e]	(< OJ *e-*)
eda 'branch'	[^{j}ẽ^{n}da]	(< OJ *yeda*)
emi 'smile'	[^{j}ẽm$_{j}$ĩ]	(< OJ *wemi*)
nue 'thrush'	[nu^{j}e]	(< OJ *nuye*)
koe 'voice'	[ko^{j}e]	(< OJ *kowe*)
ue 'top'	[u^{j}e]	(< OJ *upe*)
oto 'sound'	[^{w}odo]	(< OJ *oto*)
otoko 'man'	[^{w}odogo]	(< OJ *wotoko*)
ao 'blue'	[a^{w}o]	(< OJ *awo*)
kao 'face'	[ka^{w}o]	(< OJ *kapo*)

This feature of pronunciation is still clearly reflected in the Jesuit texts from the end of the sixteenth century (11.8) and was maintained into the NJ period (14.4), for /.e/ late enough to be reflected in for example the English name, *yen*, for the Japanese currency (NJ *en*). The palatal onglide preceding /.e/ was phonetically, if not functionally, related to the palatalization of consonants before /e, i/ (cf. 7.3.2.1 above).

7.3.2.5 Phonemicization: /-.e, -.o/ or /-.ye, -.wo/

The question arises why the glide in forms such as [ta^{j}e] and [a^{w}o] is interpreted as being automatically inserted and not as being present in phonemic representation. This is a particularly pertinent question because there were no words in OJ of the structure /CVe/ and /CVo/, so in word internal position no merger took place between /.ye/ and /.e/, or between /.wo/ and /.o/. In other words, why do we interpret these forms as /tae/ (=> [ta^{j}e]) and /ao/ (=> [a^{w}o]), and not simply as /taye/ and /awo/, which is the shape they had in OJ? This question can be viewed as a theoretical one which will receive different answers in different phonological frameworks, but from a more practical point of view several factors suggest that word internal, syllable initial /y/ was lost before /e/, and /w/ before /o/, at the same time as in word initial position, i.e. (26), even if these changes in non-initial position resulted in no mergers and are not manifested as a loss of orthographic distinctions.

(26)

/Vye/ >	/Ve/
/Vwo/ >	/Vo/

First of all, the sound changes that were discussed above form a clear pattern, with loss of /y, w/ before certain vowels. The loss of post-consonantal /y/ before /e/ and /w/ before /o/ (7.3.2.1), and of word initial /y/ before /e/ and /w/ before /o/ (7.3.2.2–3), is then paralleled by the loss of word internal, syllable initial /y/ before /e/, and /w/ before /o/. They may all be captured simply as (27), with no reference to position in the word or syllable:

(27) /y/ > Ø / __ /e/
/w/ > Ø / __ /o/

Second, /w/ has gradually been lost before other vowels than /a/ (11.2) and /w/ is thus only preserved in the environment of maximal contrast; it would be difficult to explain the retention of phonemic /w/ before /o/ later than before /i, e/ where the contrast is greater than before /o/. Conversely, the phonetic onglide is functionally well motivated, providing a phonetic syllable onset to a vowel initial syllable. Third, as described above, also /w/ changed before /e/, merging [ta^{j}e] 'cease' < OJ *taye* and [ta^{j}e] 'bark-cloth, tapa' < *tawe* < OJ *tape*. If the outcome of this merger, [ta^{j}e], was interpreted as /taye/, that change would then be /we/ > /ye/, which while not impossible is less straightforward than either of /ye/ > /e/ and /we/ > /e/. Fourth, observe the verb forms in (28) (see further 8.1.2). The infinitives are identical (in their last syllable), whereas the conclusive of (a) is different from (b) and (c). As shown, verbs in (a) reflect OJ stems in *CVye* whose regular conclusive is OJ and EMJ *CVyu*. This shows that the (a) verbs have underlying stems in //CVye// whose palatal glide surfaced when followed by /u/. On the other hand, the stems in (b) reflect OJ *CVpe* and those in (c) reflect OJ *CVwe*, which changed in the course of the sound changes above, merging the shape of the infinitive with the reflex of OJ *CVye*, as part of the overall changes which merged /Vye, Vwe, Vpe/. However, the conclusives are different, with the (b) and (c) verbs showing no trace of a glide. That suggests that they underlyingly had no glide, as opposed to the verbs in (a), and that the glide in the infinitive is the result of automatic glide generation. (Much later, in late LMJ, all three verb base types merged, see 12.3.1.)

(28)		Infinitive	Conclusive	EMJ base < OJ base	
a.	'cease	[ta^{j}e]	*tayu*	//taye//	*taye-*
b.	'exchange'	[ka^{j}e]	*kau*	//kae//	*kape-*
c.	'starve'	[u^{j}e]	*uu*	//ue//	*uwe-*

Finally, in LMJ /au/ was contracted to a long vowel [ɔɔ:]. When /.au/ was syllable initial, the outcome of the contraction was [^{w}ɔ:], as shown by the

transcription in the Jesuit texts where the contracted form for example of *au* 'meets' is written *vǒ* (see 11.5). In this case the glide is clearly automatically generated.

7.4 Prosody; 'accent'

NJ is usually characterized as a 'pitch accent' language. Certainly, the great majority of Japanese dialects employ pitch phonologically, and Tokyo (Standard) NJ has a prototypical pitch accent system, with distinctive lexically specified accent, which is realized by pitch, as well as accentuation rules assigning accent to inflected and other complex forms. This is also the case for modern Kyoto, whose system, however, is more complicated, involving both accent and tone, more narrowly defined. And for some dialects, for example Kagoshima, the phonological system of pitch seems to be less, if at all, accentual than tonal. In Japanese, the phonetics and phonology of pitch is customarily referred to as *akusento*, and this usage has to a large extent carried over into English language scholarship as 'accent', but in order to avoid potential misunderstandings, we refer to the phonology of pitch as 'prosody', distinguishing 'accent' and 'tone' as appropriate. Some dialects, usually referred to as 'accentless', do not employ pitch distinctively and generally words in those dialects have flat pitch contours.

Prosody plays a big part in Japanese dialectology. Three main types of systems using pitch distinctively have been identified, with Tokyo, Kyoto and Kagoshima being typical representatives, but with much variation within the systems. The basic descriptive facts of the three types of systems are as follows.

Kyoto distinguishes four basic word melodies, [high], [falling], [rising], and [rising-falling]. Phonologically this manifests combinations of lexical *word-tone* (defined by the initial pitch of a form) and *accent* (defined by a drop in pitch). Words with [high] or [falling] melodies, which begin on a high pitch, have a lexical High wordtone (H-words), those with [rising] or [rising-falling] melodies, which begin on a low pitch, have a lexical Low wordtone (L-words). In Samuel E. Martin's terminology, wordtone is referred to as 'register'; see further about lexical wordtone 7.4.3. Words with a drop in pitch, [falling] and [rising-falling], have an accent on the last mora before the drop. The location of an accent is distinctive, but initial accent is allowed only in H-words and final accent only in L-words of two moras (or three-mora words of the shape /CVQ.CV/, e.g. *makka* LLH(L) 'deep red'). Lexical words of the shape HHL have recently changed to HLL, so in words of less than four moras the location of an accent is no longer distinctive (but cf. *nadesiko* 'pink (name of flower)' LHLL versus *irogami* 'coloured paper' LLHL). The accent bearing unit is the mora, but of the bound moras, /Q/ cannot be accented. Unaccented L-words have an automatic final rise in pitch. Noting High wordtone by a preposed

acute, Low wordtone by a grave, and accent by an acute over the accented mora, we can note these forms as in (29), using "o" for mora:

(29)

High		Low	
Unaccented	Accented	Unaccented	Accented
HH	HL	LH	LF
ˊoo	ˊóo	ˋoo	ˋoó
ame 'candy'	*hana* 'flower'	*kasa* 'umbrella'	*ame* 'rain'
HHH	HLL	LLH	LHL
ˊooo	ˊóoo	ˋooo	ˋoóo
katati 'shape'	*awabi* 'abalone'	*usagi* 'rabbit'	*tokage* 'lizard'

Tokyo distinguishes three basic melodies, [rising], [falling], and [rising-falling], manifesting unaccented versus initial or non-initial accented; thus wordtone is not relevant in Tokyo. The location of an accent is distinctive. The accent-bearing unit is the syllable; long accented syllables have a drop in pitch from the first to the second mora. Lexical final accent is allowed, but only realized when followed by a neutral particle, as shown in (30). Unless the first syllable is accented, the first mora in a word is phonetically [low].

(30)

Unaccented	Accented		
	Initial	Non-initial	
LH(-H)	HL(-L)	LH(-L)	
oo	óo	oó	
ame 'candy'	*ame* 'rain'	*hana* 'flower'	
LHH(-H)	HLL(-L)	LHL(-L)	LHH(-L)
ooo	óoo	oóo	ooó
katati 'shape'	*awabi* 'abalone'	*tamago* 'egg'	*atama* 'head'

Kagoshima distinguishes two basic word melodies [(rising-)falling] and [rising], often classified as 'A' and 'B', respectively. The location of a pitch drop or rise is not distinctive, but is always between the penultimate and final mora, moving if a neutral particle is attached. This system is often thought of in terms of wordtone, such that the [falling] melody manifests High or Falling tone and the [rising] melody a Low or Rising tone, but it may be analysed in different ways, for example in terms of accented versus unaccented, or as penultimate versus final accent. Here the forms are noted with preposed circumflex for High-Falling tone and grave for Low(-Rising) tone.

(31)

Falling (A)	Rising (B)
HL ~ LH-L	LH ~ LL-H
^oo	`oo
ame 'candy'	*ame* 'rain'
LHL ~ LLH-L	LLH ~ LLL-H
^ooo	`ooo
katati 'shape'	*awabi* 'abalone'

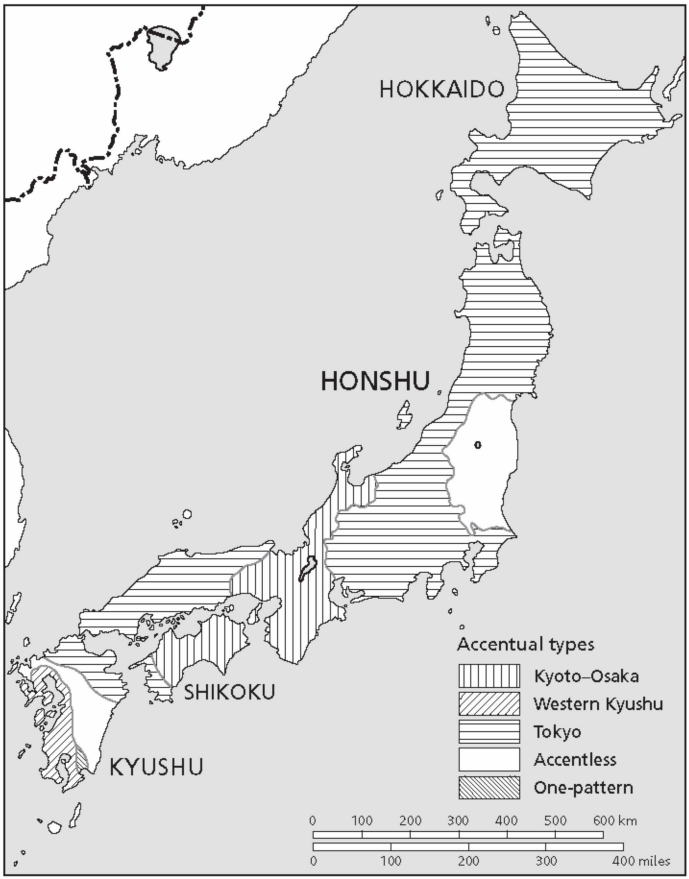
Despite the waves of standardization and dialect eradication and erosion which have taken place since the beginning of the twentieth century, dialectal prosodic systems and differences between dialects are well preserved, and prosody figures prominently in meta-linguistic consciousness about speakers and varieties of Japanese. Dialect groupings based on prosody to some extent yield different divisions than isoglosses based on segmental phonology, vocabulary, or grammatical features. Prosody has for that reason, amongst others, attracted great interest in reconstruction of proto-Japanese. Map 7.1 shows the distribution of the main types of prosodic systems.

Much scholarship on Japanese prosody is couched in terms of phonetic pitch patterns or melodies. Correspondences between dialects or between different language stages are often stated as relations between phonetic pitch patterns, e.g. HL :: LH, or LL > HL, with no regard to underlying phonological shapes. This facilitates comparison between structurally different systems, but sometimes obscures the differences between them.

7.4.1 Prosodic classes

Based on correspondences between Japanese dialects, combined with the written sources noting pitch, Kindaichi Haruhiko (1913–2004), one of the pioneers of diachronic and historical research into Japanese prosody, proposed a reconstruction of the prosodic classes of pJ. The classes, which form the basic framework for work on Japanese prosody, are noted in the form 'x.y', where 'x' represents the number of moras and 'y' the class, such that for example '2.3' is class three among two mora nouns. The classes are shown below in Table 7.3 and representative members of each are given in (35).

The reconstruction overall works well for the mainland dialects of Japanese. Different classes have merged differently in different dialects, but the classes exhibit fairly good correspondences both across time and between dialects, so that it is largely possible to predict the prosodic shape of a word in a dialect from its prosodic class (with the exception of inevitable word individual changes in class membership over time); for example class 2.3 nouns, e.g.



Map 7.1 Distribution of prosodic systems (from Shibatani 1990: 211)

mimi 'ear' or *yama* 'mountain', have the pitch shapes HL in Kyoto, LH(-L) in Tokyo, and LH in Kagoshima.

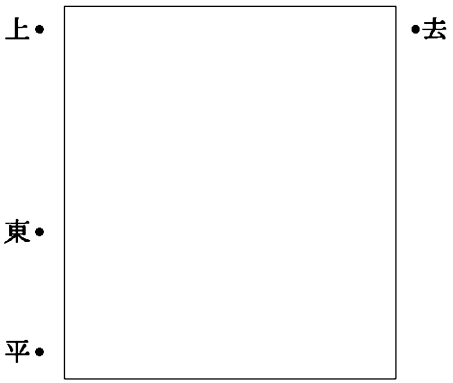
While the classes provide a reconstruction of the distinct pJ classes, they are largely established on the basis of correspondences between prosodic classes of living dialects and the words that populate them include both loanwords or transparent compounds, both of which of course do not belong to pJ. It should also be mentioned that classifications of Japanese dialect do not always make clear whether their criteria for classification are the phonological systems of the dialects, or the patterns of merger of prosodic classes exhibited by the dialects.

Prosody thus plays a central part in the reconstruction of pJ and the charting of how different dialects have branched off, and in general in an overall

understanding of the pre-history of Japanese. Here, however, we focus on the attested changes in prosody, between the earliest sources and the present day, that is, modern Kyoto (or Kansai) Japanese which is the descendant of the language reflected in the OJ and MJ sources. Reference will only be made in passing to reconstructions of pJ or to Tokyo Japanese, whose contemporary prosodic system is well described, but of which no records exist before the NJ period.

The first extensive sources for prosody are from the second half of the EMJ period and consist in materials which note the pitch of words by means of *tone dots*. The main source with tone dots is the dictionary *Ruiju-myōgi-shō*, whose original date is 1081. Tone dots originate in China where they were used to mark the tone of a syllable by their position next to a *kanji*. This practice was first adopted in *kunten* materials to indicate the tone of Chinese words, but later adapted to note the pitch of Japanese words. As mentioned above (6.1.2.2), tone dots are in some materials extended to note the distinction between tenues and mediae, with pitch indicated by the *position*, but tenuis or media by the *shape* of the marks. The tone dots used in Japanese materials are as shown in (32), with four positions: 平声 *hyōshō* 'even tone', 東声 *tōsei* 'east tone', 上声 *jōshō* 'rising tone', and 去声 *kyoshō* 'departing tone', respectively.

(32)



The 平 and 上 dots are overwhelmingly frequent. The 去 dot represents a combination of 平上, whereas the 東 dot is taken to represent the opposite sequence: 上平. The dots are thought to represent the following pitches:[5]

[5] An ingenious alternative, chiefly developed and articulated by S. Robert Ramsey, reverses the reconstruction of the phonetic pitches of the 平 and 上 dots, so that 平 represents H, and 上 L (see Ramsey 1979a, 1979b, and more recently de Boer forthcoming). This proposal entails that EMJ in terms of phonetic pitches was similar to present-day Tokyo and that present-day Kyoto represents a dramatic innovation. Ramsey's ideas have, however, been rejected by most other scholars.

(33) 上 [H]igh pitch
平 [L]ow
東 [F]alling (= [HL])
去 [R]ising (= [LH])

Earlier materials have been shown to have differentiated otherwise segmentally equivalent phonograms (*man'yōgana* or *kana*) to distinguish between high and low pitched syllables or moras, for example in some parts of the *Nihon shoki*. This usage is, however, not independently recognizable, but mainly established in correspondence with the tone dot material.

Tone dots are mostly used to note the pitch of individual words in isolation. While we have a fairly good idea of the basic phonetic and phonological system and the pitch shapes of a large group of words, particularly nouns, what we can say about the morphophonology of pitch, for example in the formation of inflected word forms, is quite limited.

Kindaichi's prosodic classes to large extent coincide with, and were established with reference to, the distinct classes in the *Ruiju-myōgi-shō*. This has led to an unfortunate identification of the prosodic system of eleventh-century Kyoto with that of pJ (which is obviously much older), and further to the idea that the prosodic systems in modern dialects, sometimes even including Ryukyuan, all somehow derive from eleventh-century Kyoto. These views to some extent still inform research on the historical prosody of Japanese. However, as might be expected, subsequent research has shown that the dialectal correspondences are more complicated, especially when material from Ryukyuan is taken into consideration, and that more classes must be reconstructed for pJ than are in Kindaichi's system or in the *Ruiju-myōgi-shō*. Furthermore, philology has made clear that the *Ruiju-myōgi-shō* system has more classes than Kindaichi's system. Kindaichi's system is, however, so ingrained that newly discovered or established classes are designated as subclasses, here noted by letters, e.g. 1.3a.

7.4.1.1 Final falling pitch: classes 2.5, 3.5b, and 1.2

The 東 dot (or 'light' (軽) 平 dot) is used in some copies of the *Ruiju-myōgi-shō* and a few other sources. It is mainly used to mark the final syllable of nouns belonging to class 2.5. It is not, however, used in all copies of the *Ruiju-myōgi-shō*, nor is it used consistently for all 2.5 nouns. Membership of class 2.5 may well have expanded over time, so it is difficult to judge in all cases whether particular nouns originally belong to 2.5 or 2.4. Today the distinctive LF melody of this class is emblematic of Kansai dialects and the class is attracting members which originally belonged to other classes, as well as new words, such as *demo* 'demonstration', *ero* 'erotic(ism)', *sabo* 'sabotage'. The distinction between class 2.4 and 2.5 is only very rarely reflected outside Kansai and vicinity, with almost all dialects on either side having no distinction

between 2.4 and 2.5, and it has been suggested that class 2.5 split off from class 2.4 as an innovation in Kyoto which only spread in Kansai, but that is very unlikely, especially because there are no identifiable conditioning criteria for the split, and because the distinction seems to be reflected for example in Taisha-chō in Shimane prefecture (Uwano Zendō, p.c.).

A small number of 3.5 nouns usually noted as 平平上 are in some sources noted as 平平東. For that reason 3.5 is divided into two subclasses, one with the melody LLH (3.5a) and one with LLF (3.5b), paralleling the final fall of class 2.5. Classes 3.5a and 3.5b merged completely soon after EMJ and the original membership of 3.5b is difficult to establish; the words listed in (32) below seem to be those for which this shape is attested.

It is finally noteworthy that the falling pitch of 1.2 nouns is not noted in isolation by a 東 dot, but by a 上 dot.

7.4.1.2 Initial rising pitch: class 1.3b; classes 2.3–5b?

The 去 dot is primarily used for the pitch of a number of one-mora nouns. They belong to class 1.3b, with rising pitch, and are distinct from 1.3a nouns which have low pitch, noted by 平. These two subclasses merged soon after the EMJ period. In addition, there is a small number, fewer than twenty, dissyllabic nouns noted with initial 去, see (34), where SJ loanwords are in bold:

(34) a. 去平 (RL): *pagi* 'shank' (also *paagi* LHL), ***yasi*** 'palm tree', *madu* 'first' (also HL), *mosi* 'perchance', ***yupi*** 'mating (of livestock)'

b. 去上 (RH): *yuri* 'lily', *yusi* 'name of a spinous evergreen', *sime* 'hawfinch', *para* (no NJ attestation) 'trumpet (horn)', *tisa* 'lettuce', *yaya* 'a bit', *pemi* 'snake' (also *penbi* LHH), ***kisa/kiza*** 'elephant', ***kuko*** 'Chinese matrimony vine'

c. 去平 ~ 去上: *pobo* 'about', *pime* 'hawfinch', ***biwa*** 'loquat', ***goma*** 'sesame' (pitch only noted for the first mora; could be either RH or RL)

These words which have rising pitch in the first syllable are generally treated as variants of LL (2.3) or LH (2.4). They are sometimes interpreted as having had long vowels with a LH contour, e.g. *yuuri [LHH] 'lily' or *paagi* [LHL] 'shank'. The double attestation of 'shank' might be taken to support that, but the question arises whether *paagi* is not rather written with two vowels to accommodate a notation of a pitch contour, rather than the contour showing the presence of a long vowel. None of these forms is written in contemporary sources with long vowels or is reflected in later stages of the language with long vowels, so it is unlikely that the forms in (34) in EMJ had phonemically long vowels. It is more likely that the pitch contour reflects a much earlier,

pJ or pre-OJ, long vowel or diphthong with a LH melody which by EMJ had contracted or shortened but retained the contour. Note, however, that *yupi* is a phonologically simplified SJ loanword (遊牝) which is also found in the shape *iupin* (> NJ *yuuhin*), supporting the possibility of a long vowel in this word. SJ loanwords are in boldface in (34); it is possible that the R pitch in these words reflects length or contour in the source words. It is thus likely that classes 2.3 and 2.4 (and possibly 2.5) should actually each be split into two classes, which had all but merged at the EMJ stage, but were originally distinguished by whether they originally had low pitch in the first syllable (2.3a, 2.4a, 2.5a) or rising pitch (2.3b, 2.4b, 2.5b).

7.4.2 Eleventh-century prosodic classes and later changes

Table 7.3 shows the prosodic classes, including subclasses, represented in the *Ruiju-myōgi-shō* (but not pJ classes not reflected in the *Ruiju-myōgi-shō*) and the reflexes in cNJ Kyoto (based on Martin 1987, especially pp. 600–34), that is to say, the changes between EMJ and cNJ Kyoto. Table 7.3 also shows some basic correspondences with Tokyo and Kagoshima dialects; importantly, the latter show inter-dialectal correspondences, *not* developments from the EMJ system: the split between the Kyoto and Tokyo branches appears to predate by quite some time the (EMJ) written sources noting pitch shapes. Note, however, that Kagoshima class 'A' corresponds well to EMJ (Kyoto) High wordtone and Kagoshima class 'B' to EMJ Low wordtone, although the changes which took place between EMJ and cNJ Kyoto to some extent have upset these correspondences. It must also be noted that these correspondences are general tendencies and that many individual words do not conform to them. Examples of words belonging to the classes are given in (35).

(35)
- 1.1 *ka* 'mosquito', *yo* 'world', *ti* 'blood', *mi* 'fruit'
- 1.2 *pa* 'leaf', *pi* 'sun', *na* 'name', *ya* 'arrow', *ne* 'price'
- 1.3a *na* 'weed', *me* 'eye', *ki* 'tree', *ko* 'basket', *no* 'field'
- 1.3b *pa* 'tooth', *me* 'woman', *su* 'nest', *mo* 'garment'
- 2.1 *pana* 'nose', *ame* 'gluten', *niwa* 'garden', *kuti* 'mouth', *sake* 'saké'
- 2.2 *kami* 'paper', *pito* 'person', *mura* 'village', *piru* 'daytime', *kata* 'direction'
- 2.3 *pana* 'flower', *kami* 'spirit, god', *kami* 'top hair', *yama* 'mountain', *mimi* 'ear'
- 2.4 *wara* 'straw', *ine* 'riceplant', *kami* 'above', *kesa* 'this morning', *ito* 'thread'
- 2.5 *ame* 'rain', *asa* 'morning', *koto* 'harp', *yoru* 'night', *muko* 'bridegroom'

Table 7.3 *Prosodic classes*

	EMJ		cNJ		
			cNJ Kyoto	Tokyo	Kagoshima
1.1	上	H	H (ˊo)	L-H (o)	H-L (A)
1.2	上	F	F (ˊó)	L-H (o)	H-L (A)
1.3a	平	L	L (ˆo)	H-L (ó)	L-H (B)
1.3b	去	R	H (ˊo) L (ˆo) F (ˊó)	H-L (ó) L-H (o)	L-H (B)
2.1	上上	HH	HH (ˊoo)	LH-H (oo)	HL (A)
2.2	上平	HL	HL (ˊóo)	LH-H (oo) LH-L (oó)	HL (A)
2.3	平平	LL	HL (ˊóo)	LH-L (oó)	LH (B)
2.4	平上	LH	LH (ˆoo)	HL-L (óo)	LH (B)
2.5	平東	LF	LF (ˆoó)	HL-L (óo)	LH (B)
3.1	上上上	HHH	HHH (ˊooo)	LHH-H (ooo)	LHL (A)
3.2	上上平	HHL	LHL (ˆoóo) HLL (ˊóoo)	LHH-H (ooo) LHH-L (ooó)	LHL (A)
3.3	上平平	HLL	HLL (ˆóoo)	HLL-L (óoo)	LHL (A)
3.4	平平平	LLL	HLL (ˊóoo) (~ HHL (ˊoóo))	LHH-L (ooó)	LLH (B)
3.5a	平平上	LLH	HLL (ˊóoo)	LHH-H (ooo)	LLH (B)
3.5b	平平東	LLF	HLL (ˊóoo)	HLL-L (óoo)	LLH (B)
3.6	平上上	LHH	LLH (ˆooo)	LHH-H (ooo)	LLH (B)
3.7	平上平	LHL	LHL (ˆoóo)	LHH-H (ooo) HLL-L (óoo)	LLH (B)

(35 *cont.*)

3.1 *katati* ‘shape’, *akari* ‘light’, *minato* ‘port’, *kodomo* ‘child’, *miyako* ‘capital’

3.2 *tokage* ‘lizard’, *aduki* ‘red bean’, *mukade* ‘centipede’, *tubasa* ‘wing’, *enoki* ‘hackberry’

3.3 *patati* ‘twenty’, *ikura* ‘how much?’, *komugi* ‘wheat’, *tikara* ‘strength’, *awabi* ‘abalone’

3.4 *kagami* ‘mirror’, *pakama* ‘man’s skirt’, *inori* ‘prayer’, *kotoba* ‘word’, *otoko* ‘man’

3.5a *iraka* 'roof tile', *sekido* 'barrier', *asuka* 'morning fragrance', *taira* 'flat', *teboko* 'spear'
3.5b *aoto* 'bluegrey grindstone', *pirome* 'broad-leaf seaweed', *pitoe* 'single layer', *tamaki* 'elbow guard', *akidu* 'dragonfly', *pimidu* 'ice water'
3.6 *usagi* 'rabbit', *padaka* 'naked', *kamome* 'seagull', *nezumi* 'rat', *sumomo* 'plum'
3.7 *kabuto* 'helmet', *tamago* 'egg', *kusuri* 'drug', *sukosi* 'a little', *mabuta* 'eyelid'

7.4.3 Phonological interpretation

As mentioned above, work on historical Japanese prosody is often done in terms of phonetic pitch shapes or patterns, but there is no standard or agreed description of the phonological categories of which the pitch patterns are manifestations. Unsurprisingly, EMJ appears similar to modern Kyoto (see Table 7.3 above) which is its direct descendant, but it is not possible to describe all the distinctive pitch shapes exhaustively in terms of wordtone and accent, without recourse to multiple accents within a word, or more than two levels of wordtone, so we need something else, or something more. However, it is clear that wordtone, the initial pitch of a form, is a significant lexical property, also at the EMJ stage. This is reflected in a simple, but important, general rule for the prosodic shape of complex forms which seems to have held through the history of the language, given in (36). See (39) below for a number of examples involving compounds.

(36) A complex form has the wordtone of its initial component.

This rule, which is sometimes referred to as 'Kindaichi's Law' (after Kindaichi Haruhiko), applies to the formation of all complex forms, including prefixation, compounding, and formation of inflected forms. Because of this rule, wordtone is often used as a diagnostic in etymological studies, such that two words are thought not to be derivationally related, and therefore not cognate, if their wordtone, as reflected in the earliest materials, is different. However, although wordtone is a basic lexical property it can, like all other phonological features, change in the course of conditioned regular sound change, as shown by the changes in Table 7.3, where for example classes 2.3, 3.2, 3.4, and 3.5 today have a different wordtone from EMJ. This type of change has resulted in many attested cases of tonal divorce of related forms, as exemplified in (37) where the wordtone of a lexicalized compound has changed in the course of regular sound change and is now different from that of its original first component. This type of change must also be envisaged before the attestation of

prosody in the written sources, showing that it is dangerous to rely too heavily on tonal correlations in etymology. Opinions differ about whether wordtone as such should be reconstructed for pJ, or whether it reflects one or more segmental features of pJ, such as for example initial voiced consonants or vowel length, but it is clear that wordtone is a significant lexical property in Japanese whose history goes much further back than the eleventh-century sources.

(37)

	EMJ	NJ Kyoto
kage (~ *kaga-*) 'light, shade, reflection'	LF	LF
kagami 'mirror' (< *kaga-* + *mi-* 'see')	LLL	HLL

Wordtone alone can account for the shape of those prosodic classes whose overall phonetic melodies are level (with no drop in pitch), either all [high] (1.1, 2.1, 3.1) or all [low] (1.3a, 2.3, 3.4), and interestingly these classes have more members than the other classes. Thus 2.1 and 2.3 together hold 59 per cent of all two-mora nouns, and 3.1 and 3.4 hold 66 per cent of three-mora nouns (Martin 1987: 273). Note also that, with a small group of exceptions, the only lexically relevant prosodic feature for verbs and adjectives is wordtone.

The prosodic classes with [falling] (1.2, 2.2, 3.2, 3.3) or [rising] (1.3b, 2.4, 3.5a, 3.6) melodies can be accounted for by the presence of a single *change* in pitch. The identification of change in pitch (rather than specifically a fall or rise) as a distinctive phonological property is due to Samuel E. Martin (1987), and we here follow his analysis and refer for EMJ to the locus of pitch change as 'accent'. Thus, the difference between HLL (3.3) and LHH (3.6) is the wordtone: both have an accent, which changes the pitch, between the second and third moras. Noting wordtone as above, but accent by a small vertical line, before the change in pitch, we can note these forms phonologically as /ˊo'oo/ and /ˆo'oo/.

Apart from the more abstract notion of accent, this system is like that of Modern Kyoto. However, it cannot account for those classes which have a [rising-falling] melody (2.5, 3.5b, 3.7), i.e., L-words which also have a drop in pitch. One way of accounting for such forms is to posit an additional word internal Low tone. Noting this L-tone by a grave over the mora which carries it, we can show the difference between LHH (3.6) and LHL (3.7) as /ˆo'oo/ versus /ˆo'où/. Table 7.4 shows this interpretation applied to the attested pitch patterns for one- to six-mora nouns. On this analysis, the structural changes between EMJ and NJ Kyoto consisted in (a) the loss of the free L-tone and (b) the change of accent from marking a change in pitch to marking a fall in pitch. These are not great changes and comparing EMJ with NJ Kyoto (cf. Table 7.3 above) shows that the phonetic changes are not great either.

Table 7.4 *Phonological interpretation of the EMJ pitch patterns*

One-mora nouns		
1.1	H	´o
1.2	F	´ˈo (or ´ò)
1.3a	L	`o
1.3b	R	`ˈo
Two-mora nouns		
2.1	HH	´oo
2.2	HL	´o'o (or ´oò)
2.3	LL	`oo
2.4	LH	`o'o
2.5	LF	`o'ò
Three-mora nouns		
3.1	HHH	´ooo
3.2	HHL	´oo'o
3.3	HLL	´o'oo
3.4	LLL	`ooo
3.5a	LLH	`oo'o
3.5b	LLF	`oo'ò
3.6	LHH	`o'oo
3.7	LHL	`o'oò
Four-mora nouns		
4.1	HHHH	´oooo
4.2	HHHL	´ooo'o
4.3	HHLL	´oo'oo
4.4	HLLL	´o'ooo
4.5	LLLL	`oooo
4.6	LLLH	`ooo'o
4.7	LLHH	`oo'oo
4.8	LHHH	`o'ooo
4.9	LHHL	`o'ooò
4.10	LHLL	`o'oòo
4.11	LLHL	`oo'oò
Five-mora nouns		
5.1	HHHHH	´ooooo
5.2	HHHHL	´oooo'o
5.3	HHHLL	´ooo'oo
5.4	HHLLL	´oo'ooo
5.5	HLLLL	´o'oooo
5.6	LLLLL	`ooooo
5.7	LLLLH	`oooo'o
5.8	LLLHH	`ooo'oo
5.9	LLHHH	`oo'ooo
5.10	LHHHH	`o'oooo
5.11	LHHHL	`o'oooò
5.12	LHHLL	`o'ooòo
5.14	LLHHL	`oo'ooò
5.15	LLHLL	`oo'oòo
5.16	LLLHL	`ooo'oò
Six-mora nouns		
6.1	HHHHHH	´oooooo
6.2	HHHHHL	´ooooo'o
6.3	HHHHLL	´oooo'oo
6.7	LLLLLL	`oooooo
6.10	LLLHHH	`ooo'ooo
6.11	LLHHHH	`oo'oooo
6.13	LHHHHL	`o'oooòo
6.14	LHHHLL	`o'oooòo
6.17	LLHHHL	`oo'oooò
6.20	LLLHHL	`ooo'ooò
6.21	LLLHLL	`ooo'oòo
6.22	LLLLHL	`oooo'oò
6.23	LLLLLH	`ooooo'o

Needless to say, other analyses are possible and have been proposed. Regardless of the specifics, it is clear that EMJ had wordtone as a distinctive lexical category, and that at least one more tone is required. On the analysis proposed here, the free L-tone is severely restricted: it occurs only after an accent in L-words and always on the final or penultimate mora. It is quite possible that the pre-EMJ system had more tones in freer combination, cf. the initial rising tone attested in a small group of two-mora nouns (7.4.1.2). It is possible to describe EMJ entirely as a restricted tone language, with more tones, but with no reference to accent. It would also be possible to view H-words with a fall as having the free L-tone, rather than an accent, so that

for example HLL is analysed as /ˊoòo/ instead of /ˊo'oo/, leaving the category of accent relevant only in L-words. However, the EMJ system developed into that of modern Kyoto with wordtone and accent, so the free L-tone, and any other tones posited in alternative analyses, were lost in any case. It is also clear that accent in subsequent periods has been a relevant phonological category for both H- and L-words. The question of when accent was introduced in the language (if it was not always there), could be seen to depend on theoretical definitions of 'accent'. More importantly for the analysis of EMJ, however, is the fact that what little we can say about the prosodic shape of complex forms seems to involve the notion of accent as defined by Martin (as the locus of change in pitch, rather than specifically a rise or fall).

7.4.4 Complex forms

An important part of the prosody of NJ is the morphophonology of wordtone and accent, in composition, in the combination of lexical words and particles, and in the formation of inflected word forms. Unfortunately, material with tone dots mostly gives information about lexical words in isolation, so apart from composition we know little in detail about this for EMJ.

7.4.4.1 Noun + noun compounds

Noun + noun (N_1 + N_2) compounds, like other complex forms, have the wordtone of the initial morpheme, that is of N_1. As a general rule, compounds except under the conditions stated in (38b) have an accent, a change of pitch, usually either between the penultimate and final mora, or between the ante-penultimate and penultimate mora (in which latter case it is usually followed by a free L-tone); (38a) is the general rule, but spells out the circumstances in which the compound has an accent.

(38) a. A compound has an accent (if it has five or more moras **or** if N_2 has an accent)
b. unless it has four moras or fewer **and** N_2 has no accent, in which case the compound has no accent.

The regularities in (38) hold reasonably well, but there are a fair number of exceptions. A few examples are given in (39), ordered by the rule determining whether the form has an accent and secondarily by the wordtone of N_1.

(39) a. *kawa* HL 'river' + *yanagi* HHH 'willow' => *kawayanagi* HHHHL ˊoooo'o 'purple willow'
pito HL 'person' + *kasira* LLL 'head' => *pitogasira* HHHHL ˊoooo'o 'weathered skull'

pana HH 'nose' + *pasira* LLH 'pillar' =>	*panabasira* HHHHL ´oooo'o 'bridge of the nose'
kami HL 'paper' + *zeni* LH 'money' =>	*kamizeni* HHHL ´ooo'o 'paper money'
turi HH 'fishing' + *pune* LH 'boat' =>	*turibune* HHHL ´ooo'o 'fishing boat'
kasira LLL 'head' + *kasa* HH 'sore' =>	*kasiragasa* LLLHL `ooo'oò 'scalp sore(s)'
mimi LL 'ear' + *kusari* HHH 'chain' =>	*mimigusari* LLLHL `ooo'oò 'ear pendant'
yama LL 'mountain' + *ubara* HHL 'rosebush' =>	*yamaubara* LLLHL `ooo'oò 'wild rose'
pama LL 'beach' + *kuri* HL 'chestnut' =>	*pamaguri* LLLH `ooo'o 'clam'
kinu LH 'silk' + *kasa* LH 'umbrella' =>	*kinugasa* LLLH `ooo'o 'silk umbrella'
tate LF 'vertical' + *isi* HL 'stone' =>	*tateisi* LLLH `ooo'o 'upright stone'
b. *tori* HH 'bird' + *ami* LL 'net' =>	*toriami* HHHH ´oooo 'bird net'
kasa HH 'sore' + *puta* HH 'lid' =>	*kasabuta* HHHH ´oooo 'scab'
wara LH 'straw' + *puta* HH 'lid' =>	*warabuta* LLLL `oooo 'straw lid'
mayu LF 'eyebrow' + *sumi* LL 'ink, charcoal' =>	*mayuzumi* LLLL `oooo 'eyebrow paint'
mugi LH 'barley' + *nawa* LL 'rope' =>	*muginawa* LLLL `oooo 'cruller'

7.4.4.2 Particles

Judging from the very limited material which notes the pitch of short phrases, it is usually assumed that many grammatical particles had their own independent prosodic shapes in EMJ, contrasting with later stages of the language where many particles are prosodically neutral. Thus EMJ particles such as *ga* (genitive), *wo* (accusative), *wa* (topic), and *kara* (ablative) seem to have been H or HH. Genitive *no* is thought to have been prosodically neutral, like NJ *ga*. On the other hand, *to* (comitative), *mo* (emphatic apic), and *yori* (ablative) seem to have been 'pre-accenting', that is, to have begun on the opposite pitch of the immediately preceding mora, so that *yori* had the shape 'oò, e.g.

moto-yori LLHL (ˆoo'oò) 'originally' (*moto* LL 'root, original'), but *kuti-yori* HHLL (ˊoo'oo) 'from the mouth' (*kuti* HH 'mouth'). Overall, the material is inconclusive in most matters of detail.

7.4.4.3 Verbs and adjectives

Verbs and adjectives are mainly lexically specified for wordtone. This has held through all attested stages of the language and sets the inflecting, predicative word classes, verbs and adjectives, off from the uninflecting nouns. As in the modern language, it is thought that different inflected forms in EMJ were assigned different prosodic shapes by rule.

Adjectives fall in two prosodic classes distinguished by wordtone. However, the distinction is not found among one-mora stems, all four of which (*yo-* 'good', *na-* 'non-existent, lacking', *su-* 'sour', *to-* 'sharp') are L (*ko-* 'thick' and *u-* 'lamentable, sad' do not have safe attestations of their prosodic class). Adjective stems of two or more moras distinguish H and L wordtone, e.g. H *atu-* 'thick', *aka-* 'red', *atarasi-* 'precious', *kaūbasi-* 'fragrant', but L *atu-* 'hot', *taka-* 'high', *pagesi-* 'severe', *omosiro-* 'amusing'. That distinction largely later merged also for stems of two or more moras. Too little is known about the prosodic shapes of inflected forms to say anything firm.

For verbs, the relevant prosodic classes are H- and L-verbs, with a further subdivision in terms of accent only among three-mora L-verbs. The 'prosodic length' of a verb is a relevant lexical property which for consonant base verbs is measured on the basis of its infinitive or conclusive form, not its basic stem, hence notations like *ok(V)-*. The class of accented L-verbs of three or more moras is not large; they are usually thought to have originated in compounds. The basic verb paradigms in Table 7.5 are based on Martin (1987: 191–8).

In the sources, verbs are generally listed in the conclusive form. Other forms are rarely represented and the evidence is conflicting. The forms posited here are to some extent hypothetical and differ in some respects from other accounts, reflecting that much is still unclear about the prosodic shape of inflected forms. Importantly, for example, for two-mora L-verbs and three-mora class 1 L-verbs Kindaichi and Wada (1980) give the conclusive form as LF and LLF and the adnominal form as LH and LLH. It must also be noted that the prosodic shape of the adnominal form is based on fairly small numbers of examples of phrases in which a noun is modified by an adnominal form, and it is not possible to know whether the adnominal used in other contexts had the same prosodic shape.

The overall regularities governing the paradigms in Table 7.5 are that the conclusive and the infinitive have an accent (change of pitch) between the penultimate and final mora (or, in the case of mono-moraic forms, within that mora), whereas the adnominal does not have an accent. The exception to this rule are the lexically accented 3-mora L-verbs (Class 2), which have a Low

Table 7.5 *Prosodic shapes of EMJ basic inflected verb forms*

	One-mora H-verb		One-mora L-verb		
	ki- 'wear'	*se-* 'do'	*mi-* 'see'	*ko-* 'come'	*pe-* 'pass'
Conclusive	*kiru*	*su*	*miru*	*ku*	*pu*
	HL	F / H	LH	R	R
	´o'o	^o / ´o	`o'o	`'o	`'o
Adnominal	*kiru*	*suru*	*miru*	*kuru*	*puru*
	HH	HH	LL	LL	LL
	´oo	´oo	`oo	`oo	`oo
Infinitive	*ki*	*si*	*mi*	*ki*	*pe*
	F / H	F / H	R	R	R
	^o / ´o	^o / ´o	`'o	`'o	`'o

	Two-mora H-verb		Two-mora L-verb	
	ok(V)- 'put'	*age-* 'raise'	*puk(V)-* 'blow'	*puke* 'deepen'
Conclusive	*oku*	*agu*	*puku*	*puku*
	HL	HL	LH	LH
	´o'o	´o'o	`o'o	`o'o
Adnominal	*oku*	*aguru*	*puku*	*pukuru*
	HH	HHH	LL	LLL
	´oo	´ooo	`oo	`ooo
Infinitive	*oki*	*age*	*puki*	*puke*
	HL	HL	LH	LH
	´o'o	´o'o	`o'o	`o'o

	Three-mora H-verb		Three-mora L-verb			
			Class 1 (unaccented)		Class 2 (accented)	
	sukum- 'crouch'	*padime-* 'begin (vt.)'	*tutum(V)-* 'bundle'	*awase-* 'join (vt.)'	*aruk(V)-* 'walk'	*kegare-* 'get soiled'
Conclusive	*sukumu*	*padimu*	*tutumu*	*awasu*	*aruku*	*kegaru*
	HHL	HHL	LLH	LLH	LHL	LHL
	´oo'o	´oo'o	`oo'o	`oo'o	`o'oò	`o'oò
Adnominal	*sukumu*	*padimuru*	*tutumu*	*awasuru*	*aruku*	*kegaruru*
	HHH	HHHH	LLL	LLLL	LHH	LHHH
	´ooo	´oooo	`ooo	`oooo	`o'oo	`o'ooo
Infinitive	*sukumi*	*padime*	*tutumi*	*awase*	*aruki*	*kegare*
	HHL	HHL	LLH	LLH	LHL	LHL
	´oo'o	´oo'o	`oo'o	`oo'o	`o'oò	`o'oò

tone after the lexical accent in the conclusive and infinitive, but not the adnominal. The substantive noun is segmentally identical with the infinitive, but has no accent: *kobi* HH 'flattery' (*kobi-* (H) 'flatter'), *ami* LL 'knitting' (*am-* (L) 'knit'), *kazari* HHH 'ornament' (*kazar-* (H) 'adorn'), *urami* LLL 'resentment' (*uram-* (L) 'resent'). The gerund is attested in a number of cases and generally seems to have attached the gerund formant *-te* with an invariable H pitch to the infinitive, e.g. *oite* HLH (*ok-* 'put'), *konde* LHH (*kom-* 'enter'), *padimete* HHLH (*padime-* 'begin (tr.)'), *agaite* LLHH (*agak-* 'paw'), but there are already instances of *-te* attaching with the same pitch as the preceding mora, e.g. *oite* HLL (*ok-* 'put').

REFERENCES

EMJ phonology: Frellesvig 1995, Mabuchi 1971, Martin 1987, Wenck 1959. *Onbin*: Frellesvig 1995, Komatsu 1981: 161ff. On the development of OJ /p/: Kiyose 1985. *Renjō*: Vance 1987: 164–7. Accent: Frellesvig 1998, Kindaichi and Wada 1980: 7–10, Martin 1987, Matsumori 2008, Shimabukuro 2007, Uwano 1977.

8.1 Verbs

8.1.1 Morphological categories

The overall range of inflected categories and forms is not very different between OJ (cf. 3.1.3) and EMJ; the EMJ inventory of forms is shown in Table 8.1.

The greatest change was the loss of the nominal form, e.g. *kakaku*, *akuraku*, which was prominent in OJ (see 3.1.3.3). It went out of general use, but was archaically retained in *kanbun-kundoku* and to some extent in poetry; it also survived in some set expressions, cf. 9.1.7. The OJ uses of the nominal form came to be taken over (a) by nominalizations by nouns such as *koto* 'thing' which increased in use both in abstract nominalizations ('the fact that . . .') and in complement clauses, and (b) by the adnominal form which expanded its use in headless nominalizations; see further 12.6.1 about changes in the use of the adnominal form.

The prohibitive retained only the one pattern shown in Table 8.1, and of the several related optative verb forms, only that in *-(a)namu* survived in EMJ, with the general sense 'I wish'. A new optative form in *-(a)baya* came into use, e.g. *kikabaya* 'I wish to hear', originating in the combination of conditional *-(a)ba* and the particle *ya*. The segmental distinction between exclamatory and imperative, e.g. OJ *kake* versus *kakye*, was lost, in the course of the merger of OJ *Ce* and *Cye* (7.3.2.1). For vowel base verbs, the imperative now generally ended in *-yo* (which had been optional with some verb classes in OJ, cf. 3.4.5).

8.1.2 Conjugation classes and basic paradigms

In the first half of EMJ, the so-called *shimo ichidan* (lower monograde, abbreviated as 'LM') conjugation class appeared, morphophonologically patterning with UM, but with a different stem vowel. The basic paradigm of the EMJ vowel base verbs is given in Table 8.2.

Through EMJ and part of LMJ the LM verb class only had a single member, namely 'to kick'. It is thought that this verb had the base shape /kwe/, rather

Table 8.1 *EMJ inflected verb forms*

	QD	LB
Base	*kak-*	*ake-*
Finite		
Conclusive	*kaku*	*aku*
Adnominal	*kaku*	*akuru*
Exclamatory	*kake*	*akure*
Imperative	*kake*	*akeyo*
Negative conjectural	*kakazi*	*akezi*
Optative	*kakabaya*	*akebaya*
Optative	*kakanamu*	*akenamu*
Prohibitive	*na kaki so*	*na ake so*
Non-finite		
Infinitive	*kaki*	*ake*
Gerund	*kaite ~ kakite*	*akete*
Continuative	*kakitutu*	*aketutu*
Conditional	*kakaba*	*akeba*
Provisional	*kakeba*	*akureba*
Concessive	*kakedo(mo)*	*akuredo(mo)*

Table 8.2 *Basic paradigm of EMJ vowel base verbs*

	LM	LB	UM	UB	*s*-irr	*k*-irr
Base	*kwe-* 'kick'	*ake-* 'open, tr.'	*mi-* 'see'	*oki-* 'arise'	*se-* 'do'	*ko-* 'come'
Infinitive	*kwe*	*ake*	*mi*	*oki*	*si*	*ki*
Conclusive	*kweru*	*aku*	*miru*	*oku*	*su*	*ku*
Adnominal	*kweru*	*akuru*	*miru*	*okuru*	*suru*	*kuru*
Exclamatory	*kwere*	*akure*	*mire*	*okure*	*sure*	*kure*
Imperative	*kweyo*	*akeyo*	*miyo*	*okiyo*	*seyo*	*koyo*

than just /ke/. The evidence is not incontrovertible and this is in that case the only native word with /kwe/.[1] EMJ *kwe-* reflects the OJ LB verb *kuwe-*. This change could be seen as an early beginning of the sweeping change of all bigrade verbs to monograde conjugation which in the main took place in NJ (15.1.2.1), but it is intriguing that this singular verb, which was not particularly frequent or otherwise prominent, changed in this way and maintained

[1] The evidence includes writings such as 化ろ which is interpreted as *kweru* rather than *keru* because 化 had the SJ (*go-on*) reading /kwe/.

Table 8.3 *Basic paradigm of EMJ* CVi- *and* CVe- *base verbs*

	UB	LB	LB	LB
Base	*kui-* //kuyi// 'regret'	OJ *taye* > *tae-* //taye// 'cease (intr.)'	OJ *kape-* > *kawe-* > *kae-* //kae// 'exchange'	*uwe-* > *ue* //ue// 'starve'
Infinitive	*kui*	*tae*	*kae*	*ue*
Conclusive	*kuyu*	*tayu*	*kau*	*uu*
Adnominal	*kuyuru*	*tayuru*	*kauru*	*uuru*
Exclamatory	*kuyure*	*tayure*	*kaure*	*uure*
Imperative	*kuiyo*	*taeyo*	*kaeyo*	*ueyo*

its separate conjugation for a relatively long period. It may be its unique stem shape which contributed to the stability of the conjugation. Another early beginning of the change of bigrade to monograde verbs is the migration of one-mora UB verbs to UM, which was under way in OJ and was completed in EMJ (cf. 3.4.3.1). Combined with the merger of *Ci* and *Cwi* (7.3.2.1) this resulted in a clear-cut distribution of *-i-* base verbs between UM and UB such that all one-mora *-i* base verbs were UM and all *-i* base of more than two moras verbs were UB. Despite the emergence of the LM class, a similar length-based distribution did not arise for the *-e* base verbs, as all other one-mora *-e* base verbs than *kwe-* (e.g. *e-* 'get', *pe-* 'pass') remained LB.

Several sound changes gave LB verbs of the shape *CVe-* (loss of /y/, /-p-/ > /-w-/ and loss of /-w-/; cf. 7.3.2.5). Verbs which reflected OJ *CVye-* remained //CVye// in underlying representation, just like UB *CVi* verbs (3.4.1.2), whereas those which reflected OJ *CVpe-* and *CVwe-* became //CVe// also in underlying representation. This is shown by the different conclusive, adnominal and exclamatory forms of these verbs, see Table 8.3. In LMJ all *CVe-* verbs became //CVye// underlyingly (see 12.3.1).

8.1.3 Verbal nouns

The class of verbal nouns, predicated by *se-* 'do', originates in EMJ, and in particular in the large intake of SJ loanwords, all of which were taken in morphologically as nouns. Some early examples are given later in 9.2.3.2. In OJ one use of *se-* was as a light verb in constructions with a verb infinitive (see 3.4.3.2.1), and the emergence of the verbal nouns may be seen as an extension of this use. *Se-* is still used as a light verb in NJ. After moraic nasals /Ĩ, Ũ, N/, *se-* became *ze-* (cf. 7.1.2.2). By far the greatest number of verbal nouns are SJ loanwords, but a few native forms are also included, especially verbs resulting from the derivational pattern ADJ-*nze-* reflecting the

OJ construction ADJ-*mi se-*, 'ADJ-ACOP.INF do; make ADJ of, take ADJ-ly', e.g. *karo-mi se-* > EMJ *karo-nze-* > NJ *karonji-* 'make light of, take lightly' (*karo-* 'light'), *omo-mi se-* > *omo-nze-* > *omonji-* 'value' (*omo-* 'heavy').

8.1.4 Consonant base verbs

The basic paradigms of the consonant base verbs are shown in Table 8.4. In the course of /-p-/ > /-w-/, OJ *-p* bases became /-w-/ bases around 1000. As /w/ was lost before /e/ around 1100, the exclamatory and imperative changed so that the base consonant, as in NJ, only appears before /a/. The *onbin* sound changes (7.1.4) affected the infinitive of *-p*, *-k*, *-b*, *-g*, *-m*, *-n* base verbs, ending in *-pi*, *-ki*, *-bi*, *-gi*, *-mi*, *-ni*, resulting in the emergence of variants of the infinitive when used as a stem for suffixation, generally referred to as the *onbinkei*. This variation was analogically extended also to consonant base verbs whose infinitives were not directly involved in the *onbin* sound changes, that is *-t*, *-s*, *-r* bases whose infinitives ended in *-ti*, *-si*, *-ri*, so that all consonant base verbs acquired a new conditioned, bound stem used for suffixation of some auxiliaries and inflectional endings. With *-m*, *-b*, *-p* bases, both vocalic and consonantal *onbin* stems resulted. These variants belonged to different varieties, but little concrete is known about their distribution. The consonantal variants are thought to be more typical of eastern dialects; the distribution in modern dialects corresponds largely with the major east–west dialect boundary, with vocalic variants prevalent on the western side and consonantal variants on the eastern side. The consonantal variants also appear later in the written sources than the vocalic ones. This is sometimes taken to reflect that the consonantal variants developed from the vocalic ones, but it is more likely due to a later and/or lesser representation in the written sources of the varieties in which they arose and were used.

The *onbin* stem was used with a variety of suffixes. As in NJ, a major use of the *onbin* stem was from early on with the gerund formant *-(i)te*, and it was also used regularly with the stative auxiliary (8.4.2) *-(i)tar-* (which originates in the OJ periphrastic construction *-(i)te ar-*), and with the respect auxiliary verb *-tamap-* (> *-tamaw-*). The initial /t/ of these suffixes changed to /d/ after stem-final nasals /Ĩ, Ũ, N/ (7.1.2.2). Rules for the distribution of the *onbin* stem must have differed between speakers and also changed over time. In EMJ *onbin* stems were also used before, for example, *-si* (adnominal of simple past *-(i)ki*) and somewhat more rarely before perfective *-(i)te-* and *-(i)n-*, modal past *-(i)ker-*, and the conclusive of simple past *-(i)ki*. In LMJ, perfective *-(i)te-* was regularly used with the *onbin* stem (until *-(i)te-* went out of use, see 12.1.3.1), and a general feature is that the suffixes regularly selecting the *onbin* stem have initial *t-*. However, the LMJ desiderative auxiliary *-(i)ta-* (12.1.5.3) was not used with the *onbin* stem and today reflexes of *-tamaw-* do

Table 8.4 *Basic paradigm of EMJ consonant base verbs*

	n-irr	*r*-irr	QD	QD
Base	*sin-*	*ar-*	*tor-* 'take'	*mot-* 'hold'
a- stem	*sina*	*ara*	*tora-*	*mota-*
onbin stem	*siN-*	*aQ-*	*toQ-*	*moQ-*
Infinitive	*sini*	*ari*	*tori*	*moti*
Conclusive	*sinu*	*ari*	*toru*	*motu*
Adnominal	*sinuru*	*aru*	*toru*	*motu*
Exclamatory	*sinure*	*are*	*tore*	*mote*
Imperative	*sine*	*are*	*tore*	*mote*

	QD	QD	QD
Base	*kak-* 'write'	*kog-* 'row'	*sas-* 'pierce'
a- stem	*kaka-*	*koga-*	*sasa-*
onbin stem	*kai-*	*koï-*	*sai-*
Infinitive	*kaki*	*kogi*	*sasi*
Conclusive	*kaku*	*kogu*	*sasu*
Adnominal	*kaku*	*kogu*	*sasu*
Exclamatory	*kake*	*koge*	*sase*
Imperative	*kake*	*koge*	*sase*

	QD	QD	QD pre-1000	QD 1000–1100	QD post-1100
Base	*yom-* 'read'	*yob-* 'call'	*kap-* > 'exchange'	*kaw-*	*kaw-*
a- stem	*yoma-*	*yoba-*	*kapa-*	*kawa-*	*kawa-*
onbin stem	*yoü- ~ yoN-*	*yoü- ~ yoN-*	*kau- ~ kaQ-*	*kau- ~ kaQ-*	*kau- ~ kaQ-*
Infinitive	*yomi*	*yobi*	*kapi*	*kawi*	*kai*
Conclusive	*yomu*	*yobu*	*kapu*	*kau*	*kau*
Adnominal	*yomu*	*yobu*	*kapu*	*kau*	*kau*
Exclamatory	*yome*	*yobe*	*kape*	*kawe*	*kae*
Imperative	*yome*	*yobe*	*kape*	*kawe*	*kae*

not select the *onbin* stem. In writing, *onbin* stems did not replace the infinitive in combination with these suffixes, but coexisted with it. To some extent this probably reflects variation in use, partly in rules for the distribution of the *onbin* stems, especially with those suffixes which were infrequent with the *onbin* stem, and partly in stylistic value. On the other hand, this is also thought to reflect norms which inhibited the use of *onbin* forms in writing; it is likely

that these forms were used much more widely in speech, as suggested for example by their frequent use in *kunten* materials. *Onbin* stems never became part of the classical literary norm, and they are usually not represented in the Classical Japanese verb paradigms in the school grammar. When noting inflectional endings or auxiliaries which regularly attach to the *onbin* stem, we will use the prefix '*(I)*', similarly to the '*(a)*', '*(i)*', and '*(e)*' which show that endings attach to the *a* stem, the infinitive, or the exclamatory (see 3.4.4). Thus, we will write *-(I)te* gerund formant, *-(I)tar-* stative auxiliary, but not **-(I)ker-*, because the modal past does not seem to have been used as regularly with the *onbin* stem. As with the other morphophonological prefixes, the '*(I)*' in *-(I)te* or *-(I)tar-* is not part of the phonemic shape of the suffixes, but merely shows that the formant attaches to the *onbin* stem.

In addition to the regularized new system of consonant base verb stems, the adnominal form of *r*-irr verbs or suffixes belonging to the *r*-irr conjugations generally assimilated to a following nasal or media in a grammatical morpheme, especially extensions such as *be-* necessitive, *nar-* hearsay, *mer-* evidential, e.g. *naru-besi* 'COP-NEC' => *naNbesi*, *paberu-meri* 'exist.HUM-EVID' => *pabeNmeri*, *aru-nari* 'exist-hearsay' => *aNnari*.

8.2 Adjectives and copula

In the course of EMJ both the regular and the adjectival copula acquired secondary, full verbal conjugations (8.2.1), while also retaining their primary paradigms. The primary paradigm for the regular copula was highly defective, and that for the adjectival copula lacked a number of the categories which verbs inflected for.

The primary forms of the adjectival copula underwent a great deal of change between OJ and EMJ. The forms in common use in EMJ are listed in Table 8.5, contrasting with those used in OJ shown above in 3.2.1. Apart from simplification, in the sense of elimination of variant forms for the same categories, the nominal form in *-kyeku*, which was frequent in OJ, was lost. Two formatives, infinitive-2 *-mi* and exclamatory-2 *-sa*, changed to become derivatives, as in NJ where they derive concrete and abstract nouns from adjectives. *-mi* is also reflected in the derivational pattern ADJ-*nze* mentioned above (8.1.3).

Also the adjectives acquired new variant forms as a result of the *onbin* sound changes (7.1.4): adnominal OJ *-ki* > EMJ *-i* (~ *-ki*) and infinitive OJ *-ku* > *-u* (~ *-ku*). These forms were not used in the poetry, but are found scattered throughout the prose texts. As with the verbs, their actual use in speech was probably somewhat wider than their textual occurrence suggests. As opposed to the verbal *onbin* stem, the adjectival *onbin* forms were not subject to grammatical conditioning.

Table 8.5 *EMJ adjectival copula forms*

Base		*taka-* 'tall'	*asi-* 'bad'
Finite			
Conclusive	*si*	*taka-si*	*asi*
Adnominal	*ki ~ i*	*taka-ki ~ taka-i*	*asi-ki ~ asi-i*
Exclamatory	*kere*	*taka-kere*	*asi-kere*
Non-finite			
Infinitive	*ku ~ u*	*taka-ku ~ taka-u*	*asi-ku ~ asi-u*
Gerund	*kute ~ ute*	*taka-kute ~ taka-ute*	*asi-kute ~ asi-ute*
Conditional	*kupa > kuwa*	*taka-kupa > taka-kuwa*	*asi-kupa > asi-kuwa*
Provisional	*kereba*	*taka-kereba*	*asi-kereba*
Concessive	*keredo(mo)*	*taka-keredo(mo)*	*asi-keredo(mo)*

Table 8.6 *EMJ copula forms*

Adnominal	*no*	
Infinitive	*ni*	*to*
Gerund	*nite > de; nisite*	*tosite*

The very defective EMJ primary copula paradigms are shown in Table 8.6; cf. 3.3.1 which has a few more forms. A new gerund, *de*, is found from the end of the period, originating in contraction of *nite*.

8.2.1 Secondary conjugations: ar- *extended forms*

In OJ the existential verb *ar-* was used with the infinitive of the adjectival copula (*-ku*), the regular copula (*ni*, *to*) and the negative auxiliary (*-zu*) to form analytic forms, and these combinations sometimes fused phonologically: ADJ-*ku ar-* => ADJ-*kar-*, NOUN-*ni ar-* => NOUN-*nar-*, NOUN-*to ar-* => NOUN-*tar-*, VERB-*zu ar-* => VERB-*zar-* (3.4.2.1–2). In EMJ the fused forms gave rise to secondary conjugations, all of which belong to the *r*-irr conjugation and have the basic paradigms in Table 8.7. Note, however, that the secondary conjugations did not replace the analytic forms which have continued in use through the history of the language (cf. 12.2.3).

8.2.2 Adjectival copula and negative

With the adjectival copula and the negative, the secondary conjugations supplemented the primary, basic paradigms and made morphophonologically

Table 8.7 *Basic paradigm of EMJ secondary conjugations of the copula, the adjectival copula and the negative auxiliary*

	Copula		Adjectival copula	Negative
Base	*nar-*	*tar-*	*-kar-*	*-(a)zar-*
a- stem	*nara-*	*tara-*	*kara*	*zara-*
onbin stem	*naQ-*	*taQ-*	*kaQ-*	*zaQ-*
Infinitive	*nari*	*tari*	*kari*	*zari*
Conclusive	*nari*	*tari*	*kari*	*zari*
Adnominal	*naru*	*taru*	*karu*	*zaru*
Exclamatory	*nare*	*tare*	*kare*	*zare*
Imperative	*nare*	*tare*	*kare*	*zare*

possible the formation of more inflected forms, for example an imperative, and free combination with tense and mood auxiliaries which did not combine with the primary forms. Frequent combinations (often with multiple occurrences of forms which incorporate *ar-*) include: *taka-kara-zu* 'isn't tall', *taka-kara-zari-ki* 'wasn't tall', *taka-kara-zari-keri* 'wasn't tall!', *taka-kara-mu* 'is probably tall', *taka-kara-zara-mu* 'probably isn't tall'. The main function of the secondary conjugations was combination with inflectional endings or auxiliaries, for which the *a-* stem, the *onbin* stem, and the infinitive were used, but although both adjectives and the negative auxiliary had a primary conclusive, adnominal and exclamatory (and the inflected forms built on the exclamatory), the secondary conclusive, adnominal and exclamatory were also used, although less frequently than the primary forms.

The secondary conjugation of the adjectival copula was also used with the extension *be-* necessitive, *be-kar-*, used in combination with auxiliaries such as negative, simple past, or modal past, e.g. *kurabu be-kara-zari-keri* 'shouldn't compare!'. A further development from *be-kar-* is the extension *bekasi-* necessitive, where *bek-*, or *beka-*, was treated as a verbal stem or root from which a *shiku* adjective was derived.

8.2.3 Copula

For the regular copulas, the new secondary forms became fully inflected copula forms which complemented the defective basic paradigm, not just to form extended combinations, but for simple predications: NP1 NP2-*nari* 'NP1 is NP2'. The basic forms in Table 8.6 continued in use, however, as adverbial and adnominal copula forms. Of the inflected copulas, *tar-* was restricted and mainly used with a subset of SJ adjectival nouns (8.3), as well as in *kanbun-*

kundoku, whereas *nar-* became the general copula. Like the secondary conjugations of the adjectival copula and negative auxiliary, the inflected copulas continued to coexist with the analytic forms: NP-*ni ari* 'is NP'. *Ni ar-* had the variant *nite ar-*, with *ar-* following the gerund instead of the infinitive. In the course of EMJ *nite ar-* became *de ar-*, which is the source of the cNJ copula *da* (cf. 12.2.2, 15.2, 16.2).

8.3 Adjectival nouns

The so-called *adjectival nouns* (*keiyōdōshi* 形容動詞, lit. 'adjectival verbs'), including words such as *siduka* 'calm, relaxed, quiet', *kiyora* 'clear', *yaparaka* (> *yawaraka*) 'soft', are usually regarded as having emerged as a separate word class in EMJ. The term 'adjectival noun' is widely accepted, and is adopted here, but 'nominal adjective' would be better, for these words are really a subclass of adjectives. Neither adjectives nor adjectival nouns occur with case particles. Whereas adjectives are adverbialized, adnominalized and predicated by the (restricted) adjectival copula, the adjectival nouns use forms of the regular copula in those functions, e.g. *siduka-nari* 'is calm', *siduka-ni* 'calmly', *siduka-narazu* 'isn't calm'.

Many of the EMJ adjectival nouns go back to OJ and many include derivational elements which are found already in OJ, especially *-ka*, *-raka*, *-yaka*. Usually, forms reflected as EMJ adjectival nouns are classified as adverbs in OJ. However, this seems to reflect only that they did not occur with case particles, but these forms do not conform to usual definitions of adverb in Japanese, which include adverbial use without a copula or other particle as a criterion, e.g. *ima* 'now'. These forms were in OJ adverbialized by *ni* (or *to*), adnominalized by *no*, and predicated by *ni* (or *to*) + existential verb, i.e. by combining with the OJ copula forms. Descriptively, therefore, there seems to be little reason not to recognize adjectival nouns as a distinct word class already in OJ: the emergence of an inflected copula (*nar-*) in EMJ seems to have been the impetus for recognizing this group of words as a separate word class from EMJ. The intake of a number of SJ loanwords as adjectival nouns may also have been a contributing factor, although the number of SJ adjectival nouns was not great in the first half of EMJ. Since then, however, the class of adjectival noun has been open, freely accepting loanwords (as opposed to the class of adjectives), so the vast majority of adjectival loanwords are borrowed as adjectival nouns (e.g. cNJ *yunīku-* 'unique' < English *unique*).

Thus most adjectival nouns are today SJ or western loanwords. A small number of SJ adjectival nouns used the rarer copula variant *tar-* (see 8.2.1), e.g. *daīdaī-tar-* 'grand'. Several new EMJ derivatives derive adjectival nouns, e.g. *-gati-* 'be prone, likely', *-ge-* 'seem, appear', *-yaī-* '-like'.

8.4 Auxiliaries

A number of the OJ auxiliaries went out of use or changed function in EMJ, whereas other new auxiliaries emerged. The OJ respect auxiliary *-(a)s-* was lost, but the OJ causative *-(a)sime-*, which was lost as a causative, was retained in its use to express respect. See 12.7 about honorific language. The OJ passive *-(a)ye-* was lost, but two new causative (*-sase-*) and passive (*-rare-*) auxiliaries appear which are still in use today (8.4.1). In the tense and aspect system a number of changes took place, including the appearance of a new stative auxiliary, *-(i)tar-* (8.4.2). Also a number of new modal auxiliaries are found in EMJ (8.5). Auxiliaries in productive use in EMJ are shown in (1):

(1)

Respect	*-(a)sime-*
Causative	*-sase-*
Passive	*-rare-*
Negative	*-(a)zu* ~ *-(a)zar-*
Stative	*-(i)tar-*
Stative	*-er-*
Perfective	*-(i)te-*
Perfective	*-(i)n-*
Simple past	*-(i)ki-*
Modal past	*-(i)ker-*
Conjectural	*-(a)m-*
Past conjectural	*-(i)kem-*
Intentional	*-(a)nze-/-(a)ũze-*
Optative	*-(a)maposi-*
Negative optative	*-(a)mau-*
Subjunctive	*-(a)masi*

The negative underwent some changes in the transition between OJ and EMJ which were outlined in 3.1.4.8.3. In the course of EMJ, the negative adnominal *-nu* (e.g. *kaka-nu* 'doesn't write') acquired a reduced variant *-n* (*kakan*) which is still used widely in Kansai today. As described above (8.2.1), the negative auxiliary acquired a secondary conjugation, *-(a)zar-*, which was used especially for combination with tense and mood auxiliaries.

8.4.1 Passives and causatives

During the EMJ period, the OJ passive *-(a)ye-* and causative *-(a)sime-* went out of use and were replaced by the alternative OJ passive *-(a)re-* and the causative *-(a)se-* which is only incipiently and partially attested in OJ (3.1.4.4). Passive *-(a)ye-* was lost from early on in EMJ, but is reflected in lexicalized forms such as cNJ *mie-* 'be visible' (< OJ *mi-ye-* 'see-PASS') and in fossilized

forms such as *arayuru* 'all' (< *ara-yuru* 'exist-PASS; can exist') (cf. 9.1.7); it was also used a little in *kanbun-kundoku*. Causative *-(a)sime-* gradually declined through EMJ, but was retained archaically in *kanbun-kundoku* and in *kanji-kana majiribun* writings which were heavily influenced by *kanbun-kundoku*. However, it was used as a respect auxiliary through the EMJ period (see 12.7). An important change in function of the passive is that from EMJ onwards it also came to be used to express respect; like OJ *-(a)sime-*, the new causative was also used to support respect auxiliary verbs (see 12.7.1.2).

In OJ, passives were only formed on consonant base and UM verbs, but not on LB, UB, *s*-irr and *k*-irr verbs, and that pattern holds also for the few examples of OJ *-(a)se-* (cf. 3.1.4.4). However, since EMJ, both causatives and passives are freely formed on verbs from all verb classes. The shape of the passive and causative auxiliaries which came to be used with LB, LM, UB, *s*-irr and *k*-irr verbs was new: *-rare-* and *-sase-*, and must have arisen by some sort of analogy. They were also extended to use with UM verbs, such that *-rare-* and *-sase-* were used with all vowel base verbs; the incipient OJ causatives attested on a few UM verbs: *kise-* 'make wear; dress' and *mise-* 'make see; show' were lexicalized with the meanings 'dress' and 'show', and productive EMJ causatives were formed on these verbs as *kisase-*, *misase-*. The traditional analysis of the formation is that each auxiliary has two basic variants, *-sase-~-(a)se-* and *-rare-~-(a)re-*, with the long forms attaching to the base of vowel base verbs (*ake* + *sase*; *ake* + *rare*) and the short forms attaching to the *a-* stem of consonant base verbs (e.g. *kaka* + *se*, *sina* + *se*; *kaka-re*, *sina-re*), but the analysis usually assumed for NJ is equally possible for EMJ, positing for each auxiliary a uniform shape, *-sase-* and *-rare-*, whose initial consonant is deleted after a base final consonant (e.g. *kak* + *sase* => *kakase-*, *sin* + *sase* => *sinase*; *kak* + *rare* => *kakare-*, *sin* + *rare* => *sinare*).

(2)

	Causative *-sase-*	Passive *-rare-*
'write'	*kakase-*	*kakare*
'die'	*sinase-*	*sinare-*
'be'	*arase-*	*arare-*
'open (tr.)'	*akesase-*	*akerare-*
'kick'	*kesase-*	*kerare-*
'arise'	*okisase-*	*okirare-*
'see'	*misase-*	*mirare-*
'come'	*kosase-*	*korare-*
'do'	*sesase-*	*serare-*

As a result, the passive and causative auxiliaries became entirely parallel, in distribution (now occurring with all verb classes), and in form and distribution

Table 8.8 *EMJ forms of causative and passive auxiliaries*

	Causative	Passive
Base	*-sase-*	*-rare-*
Finite		
Conclusive	*sasu*	*raru*
Adnominal	*sasuru*	*raruru*
Exclamatory	*sasure*	*rarure*
Imperative	*saseyo*	*rareyo*
Negative conjectural	*sasezi*	*rarezi*
Optative	*sasebaya*	*rarebaya*
Optative	–	*rarenamu*
Prohibitive	–	*na VERB-rare so*
Non-finite		
Infinitive	*sase*	*rare*
Gerund	*sasete*	*rarete*
Continuative	*sasetutu*	*raretutu*
Conditional	*saseba*	*rareba*
Provisional	*sasureba*	*rarureba*
Concessive	*sasuredo*	*raruredo*

of variants (differing only in their consonants: *s* for causative and *r* for passive); this is very different from OJ where there was no obvious morphological parallelism between causatives and passives. The EMJ causatives and passives seem to reflect a further morphologization of the derivational suffixes *-(a)s-* 'transitive' and *-(a)r-* 'intransitive'. The causative and passive auxiliaries belong to the LB conjugation; they exhibit the forms shown in Table 8.8, i.e. almost all the forms of lexical verbs.

As in OJ, combinations of causative and passive are not found when they are used in causative and passive functions, except for rare examples such as *awaserare-*, where *awase-* 'join' < *aw-* < *ap-* 'meet' could be taken as a lexicalized verb rather than a productively formed morphological causative. However, from early LMJ onwards they did combine when used in exalting function (see 12.7).

8.4.2 Aspect; stative and perfective

The main change which took place in the inventory of forms of the aspect and tense auxiliaries compared with OJ (3.1.4.5) was the change of the periphrastic stative *-(i)te ar-* (cf. 3.1.4.7.3) to a stative auxiliary, *-(i)tar-*. In OJ the periphrastic stative sometimes showed phonological fusion, VERB-*te ar-* => *-VERB-tar-*, and in the course of EMJ the fused shape was established as an independent

Table 8.9 *Basic paradigm of the EMJ stative auxiliary* -(i)tar-

a- stem	*tara-*
onbin stem	*taQ*
Infinitive	*tari*
Conclusive	*tari*
Adnominal	*taru*
Exclamatory	*tare*
Imperative	*tare*

stative auxiliary which belongs to the *r*-irr conjugation and has the basic paradigm in Table 8.9. Morphophonologically, this change is similar to the emergence of the secondary conjugations of the adjectival and regular copulas and the negative auxiliary. It should also be noted that periphrastic stative constructions remained in use through the MJ period, formed by a verb of existence (*ar-* or one of its synonyms) following the verbal gerund.

The establishment of stative *-(i)tar-* is an important part of a shift which began in EMJ in the system of tense and aspect, in which original aspectual markers, stative *-(i)tar-* and perfective *-(i)te-*, acquired tense uses, and eventually in subsequent periods supplanted the original tense markers, *-(i)ki* and *-(i)ker-* (< OJ *-(i)kyer-*), leading also to the loss of the category of perfective. This shift was completed in LMJ (12.1.3.2) where EMJ *-(i)tar-* changed to become the past tense inflectional ending *-(I)ta* still used in cNJ.

The earlier morphological stative *-er-* (< OJ *-yer-*) gradually went out of use and was eventually replaced by *-(i)tar-*, but not until sometime in LMJ. Until then they coexisted, but with different distribution: stative *-er-* is only found with consonant base verbs (3.1.4.7), whereas *-(i)tar-* is used with all verb classes. While the long-term picture clearly is one of replacement (accompanied by a change in function of *-(i)tar-*), there is no agreement among scholars about whether the two auxiliaries through the period of coexistence were (free or conditioned) variants or whether they expressed different categories, apart from the fact that whatever distinction they may have expressed was neutralized with vowel base verbs (where only *-(i)tar-* was used). It is clear that *-(i)tar-* interacted differently with other aspect auxiliaries than *-er-* did, and that *-(i)tar-* at some point acquired a function of *perfect* (understood as referring to the 'continuing present relevance of a past situation' (Comrie 1976: 52)), i.e. with some temporal reference, whereas *-er-* remained a simple, atemporal stative. In Takeuchi's detailed analysis, the state referred to by *-(i)tar-* is said to be temporally 'limited', but 'unlimited' in the case of *-er-* (1987: 167) (but with the distinction between the two neutralized with vowel base verbs).

In OJ, the perfective auxiliaries, *-te-* and *-n-*, did not combine with the stative *-yer-* and that pattern continued in EMJ. However, they did combine with the new stative/perfect auxiliary *-(i)tar-*. Significantly, they attached differently: *-ni-tar-*, but *-tari-te-*. This is related to uses of the perfectives not observed in OJ: their uses as perfectives as described above (3.1.4.6) continued in EMJ in what Takeuchi terms 'sequenced narration', but by contrast, in 'embeddings' *-(i)te-* marks 'recent past tense' whereas *-(i)n-* marks 'limited control', i.e. the accomplishment of an event 'with (considerable) difficulty *or* unintentionally' on the part of the agentive subject (Takeuchi 1987: 135). In these functions the distribution of *-(i)te-* and *-(i)n-* was markedly different from in OJ. They both (a) combined with *-(i)tar-*, and (b) each combined with all types of verb, irrespective of the transitivity of the host verb.

8.4.2.1 Progressives

OJ had an analytic progressive consisting of infinitive + *wor-* (the opaque lexicalized stative form of *wi-* 'sit down'), cf. 3.1.4.7.4. This construction was in EMJ paralleled by several similar constructions, in which a stative form of a verb of vertical motion, such as *pus-er-* 'be lying down' (cf. *pus-* lie down'), *tat-er-* 'be standing' (*tat-* 'stand up'), or, most frequently, *wi-tar-* 'be sitting' (*wi-* 'sit down'), was attached to the infinitive of a verb, e.g. *mi witari* 'is looking', *mi tateri* 'id.', *ipi witari* 'is saying', *katari puseri* 'is telling'. These constructions are both interesting in their parallelism with the OJ progressive and as links to later progressive/stative constructions as well as changes among existential verbs (cf. 12.4).

8.5 Modality

The conjectural auxiliary *-(a)m-* occupies a central position in the morphological system of modals in OJ and EMJ. In the course of the *onbin* sound changes it acquired variant forms for the conclusive and adnominal forms, see (3). The vocalic *onbin* form, *-ũ*, is not used much in the texts until towards the end of EMJ, but it is the form that eventually prevailed and which is used in LMJ and reflected in NJ (cf. 12.1.3).

(3)

Conclusive	*mu ~ n/ũ*
Adnominal	*mu ~ n/ũ*
Exclamatory	*me*
Concessive	*medo(mo)*

In EMJ, the already rich OJ inventory of grammatical modal expressions was further augmented. The new inflected optative verb form in *-(a)baya* was mentioned above (8.1.1). Some other new forms incorporate the OJ conjectural

-(a)m-: (a) The OJ combination of simple past and conjectural had at the EMJ stage contracted to a simple auxiliary, past conjectural *-(i)kem-*, inflecting like *-(a)m-*. (b) A new optative auxiliary *-(a)maposi-*, e.g. *mi-maposi* 'want to see' arose from truncation of the OJ construction *-(a)maku (no) posi*, e.g. *mimaku posi* or *mimaku no posi* 'want to see', that is, conjectural *-(a)m-* in the nominal form, functioning as a complement to the adjective *posi*. *-(a)maposi* (> *-(a)mawosi-* > *-(a)maosi-*) belongs to the adjectival *shiku* conjugation; it is very rare in the first half of EMJ, but used widely later. OJ *-(a)maku* was also used as the complement of other adjectives; with the adjective *u-* 'sad, undesirable' it was contracted to *-(a)mau-* described as a negative optative, but this auxiliary is almost entirely found in the *Genji monogatari*. (c) From truncation of the construction *-(a)mu (to) se-* 'CONJ (COMP) do; be about to, intend to' arose the auxiliary *-(a)nze-* ~ *-(a)ũze-* 'intentional' which belonged to the *s*-irr conjugation. As with the conjectural auxiliary, the vocalic *onbin* form *-(a)ũze-* appears later in the texts than the consonantal *-(a)nze-*, but was the form which won out and is reflected in LMJ. The intentional overlapped in function with the conjectural. Although it became frequent, it was also fairly short-lived and was used only in the second half of EMJ and in LMJ, but disappeared early in NJ (cf. 12.1.3).

Among the extensions, OJ *masizi-* negative potential was abbreviated to give EMJ *mazi-* which is usually described as the negative counterpart of *be-*. A new evidential extension *mer-* 'seems, appears' came into use in the first half of the period; it is usually thought to originate in a contraction of *mi* 'see' or *me* 'eye' and *ar-* 'be', and it belongs to the *r*-irr conjugation. The OJ presumptive extension *rasi-* lost its inflection and became a final particle, *rasi*, which, however, went out of use before the end of EMJ. In addition to *bekasi-* mentioned above (8.2.2), another short-lived formation on *be-* was *bera-* which belonged to the adjectival nouns and was used only in the first half of EMJ.

Of the OJ final particles (cf. 3.7.5), *kamo* was lost, but functionally replaced by the new exclamatory particle *kana*. Also evidential *miyu* was lost, functionally to some extent replaced by the new extension *mer-*. A new desiderative particle *gana* was used only after the combination of perfective and simple past auxiliaries, e.g. *mi-te-si gana* 'see-PERF-SPST.ADN DESID; they wanted to see her' (*Taketori*). Also a new emphatic particle *kasi* came into use.

Modal forms thus include inflected forms, auxiliaries and extensions. They were an important and conspicuous part of the EMJ language. The EMJ inventory is as in (4). No comprehensive analysis of the modal system(s) involved is available, so the forms are simply listed here with their traditional labels, but even in this form (4) gives an impression of the amount of morphologically encoded modality in the language.

(4) Inflected forms
-(a)baya optative, *-(a)namu* optative, *-(a)zi* negative conjectural, *-e/-yo* imperative.

Auxiliaries
-(a)m- conjectural, *-(a)nze-/-(a)ũze-* intentional, *-(a)maposi-* optative, *-(a)mau-* negative optative, *-(a)masi* subjunctive, *-(i)kem-* past conjectural, *-(i)ker-* modal past.

Extensions
be- necessitive, *bekasi-* 'id.', *mazi-* negative necessitive, *ram-* present conjectural, *mer-* evidential (sight), *nar-* evidential (hearsay, sound).

Final particles
gana desiderative, *kana* exclamatory, *kasi* emphatic, *mogana* optative, *na* prohibitive, *rasi* presumptive, *yo* emphatic.

8.6 Derivatives

A number of derivatives appeared during EMJ which to some extent overlap in meaning, expressing some kind of evidentiality or likelihood and likeness. They include those shown in (5). Most went out of productive use after EMJ, but *-gar-* and *-ge-* are still in frequent use today (cf. 12.1.5.3).

(5) *-gamasi-* 'seem like, resemble' (derives an adjective from nouns or the infinitive of verbs); *-gapasi-* (> *-gawasi-*) '-prone, be likely to' (derives an adjective from nouns or the infinitive of verbs); *-gar-* 'seem, appear, behave like' (derives a verb from adjectives, adjectival nouns, or nouns); *gati* '-prone, be likely to' (derives an adjectival noun from nouns or the infinitive of verbs); *-ge* 'seem, appear, behave like' (derives an adjectival noun from adjectives, adjectival nouns or nouns); *-mek-* 'seem, be like' (derives a verb from adjectives, adjectival nouns or nouns); *-yaũ* '-like' (derives an adjectival noun from nouns; SJ loan (様), cf. 9.2.3.2).

8.7 Particles

The case particles of EMJ are shown in (6), which is very similar to the inventory of case particles in NJ. Of the OJ case particles (see 3.7.1), genitive *tu* and *na (~da)* were lost, but reflected in lexicalized nouns (*ama-tu-sora* (heaven + sky) 'sky', OJ *ma-na-kwo* > EMJ *manako* (eye + ?child) 'eyeball'). The rare nominative particle *i* was lost in the transition between OJ and EMJ, although it continued to be used archaically – and quite frequently – in *kanbun-kundoku*.

Ablative *kara* and allative *pe* (< OJ *pye*) became firmly established in this period, and the OJ restrictive particle *made* came to be used as an allative particle. Of the different variants OJ ablative *ywo(ri)* ~ *yu(ri)*, only *yori* was used from EMJ onwards. OJ *ni* and *nite* became more specialized with *ni* being used more for arguments and *nite* more for adjuncts; from mid to late EMJ *nite* acquired the variant *de* which is still in use in contemporary NJ. Note that *ni*, *no*, *nite (de)* and *to* in addition to their uses as case and conjunctional particles remained forms of the copula, as they do in contemporary NJ.

(6)

Accusative	*wo* > *o*
Genitive	*ga, no*
Dative	*ni*
Comitative	*to*
Instrumental/locative	*nite* ~ *de*
Ablative	*yori*
Ablative	*kara*
Allative	*pe* > *we* > *e*
Allative	*made*

In the course of regular sound change (*-p-* > *-w-* and *-w-* > Ø, cf. 7.3), accusative *wo* > *o*, allative (OJ *pye* >) *pe* > *we* > *e* and topic *pa* > *wa*. This shows that these particles, and probably particles in general, were phonologically integrated with the host word, for the sound changes affected only /-p-, -w-/ in non-initial position.

8.7.1 Genitives

Genitive *ga* and *no* continued to mark subjects in subordinate clauses and nominalized main clauses as in OJ (cf. 3.7.1.1.1). *Ga*, which in OJ was used only with pronouns and nouns referring to humans, became even further restricted through EMJ, eventually occurring almost exclusively with pronouns, nouns used for 1st or 2nd person reference, and personal proper nouns. Conversely, in EMJ *no* seems to begin marking subjects in declarative, non-nominalized main clauses. See 12.6.2 about subject marking and the development of a nominative particle. However, as in OJ *ga* rather than *no* was used to case mark nominalized clauses whose predicate was in the adnominal form and this is reflected for example in the development of *ga* as a conjunctional particle (see 8.7.2).

Especially from EMJ, both *no* and *ga* have a function which is often referred to as 'pronominal', as in (7) and (8). There are a few such examples in OJ, but more from EMJ onwards. *Ga* was used in this function for as long as it was used as a genitive, i.e. into LMJ, and *no* is still used in this way in NJ,

see (9). This use of genitives is probably simply an instance of genitive noun modification where the modified noun is deleted, so that for example (7)–(9) derive from (10a–c).

(7) ima no aruzi mo saki **no** mo te
now GEN governor ETOP before **GEN** ETOP hand
tori-kapasite
take-join.GER
'The current governor and the previous one joined hands' (*Tosa*)

(8) ko no uta pa aru pito no ipaku
this GEN poem TOP some person GEN say.NOM
opotomo no kuronusi **ga** nari
Ôtomo GEN Kuronushi **GEN** COP.CONCL
'This poem is, according to some people, Ôtomo no Kuronushi's' (*Kokin*; remark after 17.899)

(9) kono hon wa Hanako **no** da
this book TOP Hanako **GEN** COP
'This book is Hanako's'

(10) a. ima no aruzi mo saki no aruzi mo te tori-kapasite
b. ko no uta pa . . . Opotomo no Kuronusi ga uta nari
c. kono hon wa Hanako no hon da.

8.7.2 Conjunctional particles

As in OJ, conjunctional particles followed the adnominal form of verbs and adjectives. The directional case particles *made*, *yori* and *kara* could be used after nominalized clauses in conjunctional functions 'until', 'from, since', 'after'. In the sense 'because' *kara-ni* came to be used in the second half of EMJ.

In addition to *wo*, *ni* and *to*, also *ga* came to be used as a conjunctional particle 'and, as, but', following the adnominal form of a verb. (11) from the *Konjaku monogatari-shū* is an early example:

(11) oti-iri-keru toki mi no toki bakari
fall-enter-MPST.ADN time hour.of.the.snake GEN time around
nari-keru **ga** pi mo yauyaku kure-nu
COP-MPST.ADN **GA** day ETOP gradually get.dark-PERF.CONCL
'It was around the hour of the snake (10 a.m.) when he fell in, and/but (now) it was gradually getting dark' (*Konjaku* 16.24)

This usage is thought to have developed through reinterpretation of headless nominalized clauses (cf. 12.6.1.3) marked (as subjects) by *ga*, as clauses linked by *ga* to the following clause (cf. Kinsui 2007). For example, (12) is open to the two interpretations shown. On interpretation (a) *kami ito kiyora-nite naga-karikeru* is a headless nominalized clause, of the kind often referred to as an internally headed relative clause (see 12.6.1.3), meaning 'hair, which had been very beautiful and long' and marked by *ga* as the subject for *oti posorite*; but on interpretation (b) it is a coordinate clause connected to the following clause by *ga*.

(12) kami ito kiyora nite naga-kari-keru **ga**
hair very beautiful COP.GER long-ACOP-MPST GA
wake-tori-taru yaũ ni oti posorite
divide-get-STAT.ADN appearance COP.INF fall become.thin.GER

a. 'Her hair, which had been very beautiful and long, had fallen out and become thin as if it had been taken away'

b. 'Her hair had been very beautiful and long, but (now) it had fallen out and become thin as if it had been taken away' (*Genji*: *Makibashira*)

8.7.3 Tote, nado

A new particle *tote* appears in this period, which functions as a variant of the particle *to* in its functions of complementizer and purposive conjunctional particle 'in order to'; like *to* it often occurs without a following governing verb of utterance or thinking, meaning 'saying, thinking', and it is for that reason often described as originating in a contraction of *to ipite* 'saying that' or *to omopite* 'thinking that', but it is more likely that it arose as a gerund-like variant of *to* in parallel with *ni* ~ *nite*, reflecting that most or all functions as a grammatical particle of *to* were grammaticalized from its primary copula function. This finds further support in the new EMJ forms *nado* ~ *nadote* 'why' which are etymologized as < *nani-to* ~ *nani-tote* 'being what?'.

8.8 Pronouns and demonstratives

The OJ simple system of personal pronouns was already in OJ supplemented by a number of terms of address (cf. 3.8.2), and it was lost in the course of EMJ. While the OJ 1st person pronoun *ware/wa-ga* continued in EMJ, the OJ 2nd person *nare/na-* went out of use, and around the middle of EMJ there is no longer a system of personal pronouns as such, but an inventory of terms of address and of self-reference. The development has aptly been described as one of 'de-pronominalization' (Vovin 2003: 95), accompanied by the extensive

Table 8.10 *EMJ demonstrative and interrogative forms*

	Proximal (speaker)	Mesial (hearer)	Distal	Interrogative
Short	*ko*	*so*	*ka/a*	–
Long	*kore*	*sore*	*kare/are*	*idure*
Location	*koko*	*soko*	*kasiko*	*iduku ~ iduko ~ idoko ~ doko*
Direction	*konata*	*sonata*	*kanata/anata*	–
Location/direction	*koti*	*soti*	–	*iduti/idura*
Manner	*kaku(te), kau*	*sa(te), sika*	*(ka-yaû)*	*ika, ikaga, ikani, ikade*
Time	*(ima)*	–	–	*itu*
Quantity/degree	–	–	–	*ikura*

use of terms referring to the speaker also as reflexives. The forms in (13) were used for 1st, 2nd and 3rd person personal and reflexive reference. Several of the forms involve OJ *onore/ono* 'reflexive' and the noun *mi* 'body' (< OJ *mwi*). The personal interrogative pronoun *tare* 'who' and the interrogative noun *nani* 'what' remained in use.

(13) 1st person: *maro*, *mi*, *onore-ga-mi/ono-ga-mi*
2nd person: *nandi* (< OJ *namuti*), *kimi*
3rd person: *watakusi* (originally 'private')
1st and 3rd person: *ware/wa-ga*, *wa-ga-mi*, *midukara*
1st, 2nd and 3rd person: *onore/ono-ga*, *onodukara*
2nd and 3rd person: *omape* (> *omawe* > *omae*), *gozen*[2]

8.8.1 Demonstratives and interrogatives

The OJ two-term demonstrative system of speaker versus non-speaker forms (3.8.3) changed into the three-term system familiar from NJ. Main forms, including the corresponding interrogative forms, are shown in Table 8.10.

It is thought that the distal terms originate as a subdivision of the OJ proximal, visual domain into a close (proximal) and distant (distal) distinction, and that the OJ non-speaker based terms subsequently were reinterpreted as mesial and hearer based. Whereas the non-speaker based forms in *so-* in the OJ system were mainly used anaphorically, the system from EMJ became clearly deictically focused on speaker, hearer and distal. However, the mesial forms in *so-* continued to be used extensively in anaphoric function, and the new distal

[2] *Omape* < OJ *opo-* 'great' + *mapye* 'front', but was reinterpreted as *o-* 'honorific' + *mape* and written as 御前; *gozen* originates in the *on*-reading of 御前 (cf. 9.2.4).

was also used (semi-)anaphorically in the sense 'that well-known', often in literary or scriptural references. Anaphoric usage is also found with the temporal deictic *ima* 'now' to mean 'then; at the time'. The morphological system is overall characterized by further regularization compared to OJ, but there was yet not the full parallelism within the demonstrative system, nor between the demonstrative and interrogative systems, which characterize NJ, although for example the new locational interrogatives *iduko*, *idoko* and *doko* clearly show some levelling in form.

Konata, *sonata*, *kanata* are thought to have arisen from reductions of *ko-no kata* 'this side', etc.; the distal forms in *a-* are somewhat less frequent than those in *ka-*. Both *kanata* and *anata* were used for 3rd person reference in EMJ and LMJ; *anata* only started being used for 2nd person reference towards the end of the eighteenth century, i.e. well into NJ.

8.8.2 Short and long forms

In OJ pronouns and demonstratives had a fairly productive system of short and long forms, e.g. *wa/ware* and *ko/kore* (3.8.1). In EMJ the short forms became restricted to use only with a genitive particle (*wa-ga* or *ko-no*).

8.9 *Kakari-musubi*

Kakari-musubi ('hanging-tying') is the name in Japanese grammar for a construction in which some constituent is marked by one of the '*kakari* particles' (a) *ka*, *ya*, *so/zo*, *namo/namu* or (b) *koso*, and the sentence predicate it relates to is in the (a) adnominal or (b) exclamatory form, rather than in the conclusive form generally used to conclude sentences. This construction is found in both OJ and EMJ, where it is a conspicuous and prominent feature of the language. It disappeared from the language in the course of LMJ and NJ as part of a complex of syntactic changes (12.6.1.2), but remained a normative feature of classical written Japanese. The construction remains in Ryukyuan and in a few mainland dialects to a more limited extent than in OJ and EMJ. *Kakari-musubi* has recently attracted a great deal of attention as part of the growing interest in the syntax of OJ and EMJ, and also in the context of reconstructing prehistoric Japanese syntax and morphology. Specifically, within mainly or purely formal syntactic analyses, *kakari-musubi* is sometimes said to involve *movement*, which has further been taken to support the analysis of OJ as having overt *wh* movement (cf. Watanabe 2002), but this is primarily of theoretical interest to formal linguistics.

Traditionally, *kakari-musubi* is thought of as an automatic agreement rule by which the presence of one element (a *kakari* particle) triggers or requires

the presence of another (a predicate (*musubi*) in the adnominal or exclamatory form), and the function of the *kakari* particles is said to be 'emphatic' (*zo*, *namu*, *koso*) or 'interrogative' (*ka*, *ya*). (14) gives some examples from OJ; see (8.9.2) about the differences between the *kakari* particles.

(14) a. wa ga kwopuru kimi ***so*** *kizo* *no* ywo
I GEN love.ADN my.lord SO last.night GEN night
ime ni mi-ye-***turu***
dream DAT see-PASS-PERF.ADN
'It was you, my beloved lord, that I saw last night in a dream' (*MYS* 2:150)

b. *sikama-gapa* *taye-mu* pi *ni* ***koso*** *a* *ga*
Shikama-river cease-CONJ.ADN day DAT KOSO I GEN
kwopwi *yama-**me***
love stop-CONJ.EXCL
'It is on the day when the Shikama River ceases its flow that my love for you will end' (*MYS* 15.3605)

c. *kagurwoki* *kami* *ni* *tuyu* ***so*** *oki-ni-**kyeru***
black hair DAT dew SO fall-PERF-MPST.ADN
'It is dew that has fallen on my black hair' (*MYS* 15.3649)

d. tukwi *wo* sirwotape *no* kumo ***ka***
moon ACC white.cloth COP.ADN cloud KA
kakus-***eru*** ama-*tu*-kwiri *kamo*
hide-STAT.ADN sky-GEN-mist KAMO
'Is it white clouds that are hiding the moon? Or is it the mist in the sky?' (*MYS* 7.1079)

More recent studies posit and emphasize fundamental functional and syntactic differences between sentences with and without *kakari-musubi*, viewing *kakari-musubi* as a *focus construction*, often with much the same semantic effect as *it*-clefts in English, as reflected in the translations in (14) above, but differing from cleft constructions by being mono-clausal. Thus, for example (14a) has the following focus-presupposition structure, with the focused constituent followed by the presupposition. This is structurally and semantically a quite different sentence from its non-focused counterpart(s) without *kakari-musubi*, which could be for example (16a) or (b):

(15)	**Focus**	**Presupposition**
	wa ga kwopuru kimi so	kizo no ywo ime ni miyeturu
	'It was you, my beloved lord	that I saw last night in a dream'

(16) a. kizo no ywo ime ni wa ga kwopuru kimi miyetu
b. kizo no ywo wa ga kwopuru kimi ime ni miyetu
'Last night I saw you, my beloved lord, in a dream'

The traditional term, *kakari-musubi*, which dates from the nineteenth century, makes reference to two sentence elements: *kakari* means 'hanging (down, out); relation, attachment, connection' and *musubi* refers to concluding ('tying up, together') a sentence; *kakari-musubi* may be thought of as that which is being 'hung out' (*kakari*) and that whereby it or the sentence is being 'tied together' (*musubi*). This conception of *kakari-musubi* is not dissimilar from that of the '*theme–rheme*' relation (cf. Greek *théma* 'that which is set up', *rhéma* 'that which is said'); note in particular that *kakari* and *théma* are very similar in meaning. In fact, Moto'ori Norinaga's original classification (in his *Kotoba no tama no o*, 1777/85) includes *wa* (contrastive) topic and *mo* inclusive topic, and zero (in zero marked topics) among the forms that take part in *kakari-musubi*, and today, the '*kakari* particles' of Classical Japanese in Japanese school grammar comprise *wa, mo, ka, ya, zo, namu, koso*. For zero, *wa* (< *pa*), and *mo* the *musubi* (predicate) takes the conclusive form, whereas it takes the adnominal or exclamatory form with the other *kakari* particles. '*Kakari-musubi*' was first used in this wider sense (in Togashi Hirokage's *Kotoba no tama-hashi*, 1826); current use of the term to refer only to the relationship between *ka, ya, zo, namu, koso* and the form of the predicate to which they relate is more recent. Today 'theme–rheme' has in most branches of linguistics been replaced by 'topic–comment', but there are differences in the original meaning of the terms, and topic–comment could be viewed as a subclass of theme–rheme. Pursuing the older conception of *kakari-musubi*, it is possible, without stretching the Japanese or European terms too much, to equate *kakari-musubi* and theme–rheme, as in (17), where topic and focus constructions are viewed as different subtypes of *kakari-musubi* (theme–rheme):

(17)	***Kakari*** (theme)	***Musubi*** (rheme)
	Topic	Comment
	Ø, *wa*, *mo*	conclusive
	Focus	Presupposition
	ka, *ya*, *zo*, *namu*	adnominal
	koso	exclamatory

This understanding of a relation between a *kakari* and a *musubi* constituent is, however, probably too broadly conceived, for the two types of *kakari*, topic and focus, are not mutually exclusive, and from a functional point of view

they belong to different levels of representation: topic–comment belongs to the *utterance* level, and focus-presupposition to the level of *proposition*. Sometimes the referent of a focused constituent will correspond to a kind of topic at utterance level (though not a backgrounded, 'old information'-type topic), as in (18) where the focused constituent is also marked as a contrastive topic.

(18) *ima* ***koso*** *pa* *wa-dori* *ni* *ara-me*
now KOSO TOP my-bird COP.INF exist-CONJ.EXCL
'(It is) now (that) I am/will be my own' (as opposed to later when I will be yours) (*KK*3)

Often, however, topic and focus are separately represented, with structures such as (19). For example, (14) above may in fact be interpreted in that way, see (20).

(19)

Topic	Comment	
	Focus	Presupposition

(20)

sikama-gapa [Ø] 'The Shikama river,	tayemu pi ni koso it is on the day when it ceases its flow	a ga kwopwi yamame that my love for you will end'
kagurwoki kami ni [Ø] 'On my black hair,	tuyu so it is dew	okinikyeru that has fallen on it'
tukwi wo [Ø] 'The moon,	sirwotape no kumo ka is it white clouds	kakuseru that are hiding it?'

A focus construction establishes a copular, predicative relation between focus and presupposition, like that between subject and nominal predicate, see (21), much more so than does for example the topic–comment relation, which is also a 'connecting' relation, but not a copular one. This shows one significant difference between topic and focus constructions. Like focus–presupposition, the subject–(nominal)predicate relation belongs to the level of proposition.

(21)

Focus	IS	presupposition
[wa ga kwopuru kimi so] you my beloved lord	IS	[kizo no ywo ime ni miyeturu] (what) I saw last night in a dream
[tuyu so] dew	IS	[okinikyeru] (what) has fallen

[tayemu pi ni koso] on the day it ceases	IS	[a-ga kwopwi yamame] (when) my love will end
[sirwotape no kumo ka] white clouds	IS	[kakuseru] (what) are hiding (it)

8.9.1 Uses of kakari-musubi

Kakari-musubi is structurally a focus construction, but in many cases, especially in poetry, the construction seems mainly to contribute exclamative or interrogative force to the utterance, or nuances of doubt or various kinds of emphasis; although we illustrated the basic structure by translating sentences with *kakari-musubi* by *it*-clefts above, it would in many cases be contrived or misleading to do so. See for example (22), a poem from the *Ise monogatari*, which has no fewer than three *kakari* particles. (The last two stanzas (*yuki . . . pa*) are a right-dislocated complement of *omopiki*.) Here the two uses of *kakari-musubi* in *yume ka to zo omopu* express 'emphatic doubt' ('what I wonder is, is it a dream!'), emphasizing the feeling of unreal-ness which the loyal servant experiences upon seeing his former master in humble, out-of-the way dwellings.

(22) wasurete pa yume **ka** to **zo** omopu
forget.GER TOP dream KA COMP ZO think.ADN
omopi-ki **ya**
think-SPST.CONCL YA
yuki pumi-wakete kimi wo mi-mu
snow stamp-divide.GER my.lord ACC see-CONJ.CONCL
to pa
COMP TOP

'Forgetting [that you have retreated from the world], this seems unreal! (I wonder, is this a dream?)! That I would have to stamp my way through the snow to see my lord – that I never thought [did I think that? (No!)]' (*Ise* 83)

8.9.2 The individual kakari *particles*

There are significant functional/semantic differences between the individual *kakari* particles and the sentences they constitute. Semantic tags (from Quinn 1987) and some salient distributional facts about the *kakari* particles are summarized in (23), which shows whether they occur in questions, are used with *wh-* words, and whether they can be used in sentence final (SF) position, in which case they focus the entire proposition.

(23)

	Q	*Wh-*	SF
ka 'doubted identity'	+	±	+
ya 'confirmation soliciting'	+	–	+
zo 'identifying'	±	±	+
namu 'confirmative'		–	–
koso 'unique identification'	–	–	–

Ka, *ya* are only used in questions, whereas *zo* may be used in questions; *ka*, *zo*, but not *ya*, may be used with *wh-* words. *Ka*, *ya*, *zo* are also used sentence finally (and are in this use usually classified as 'final particles' in Japanese school grammar); when sentence final *ka*, *ya*, *zo* follow a nominal predicate there is usually no overt copula between the noun and the particle. *Namu*, *koso* are not used in questions, with *wh-* words or sentence finally. In Quinn's terminology, *ka*, *zo*, *koso* are said to involve *identification*, that is, establishing (*zo*, *koso*) or establishing and questioning (*ka*) an identity relation between the focused constituent and the presupposition, whereas *ya*, *namu* either seek (*ya*) or provide (*namu*) *confirmation* of the relation. Semantically and in terms of distribution *ka/koso* seem to be paired as opposites, as are *ya/namu*. *Zo* is used most widely, overlapping with both *ka* and *koso*.

Ka ('doubted identity') is used only in questions, in order to focus a constituent, often a *wh-* word (24) ('who, what, when, how, . . . is it that . . .?'), but it can also be a lexical word singled out as the focus of a yes/no question, (25) ('is it X that . . .?'). When *ka* is used sentence finally, it is the predicate which is focused, (26).

(24) kakaru miti pa ikade **ka** **imasuru**
be.like.this.ADN path TOP why KA be.RESP.ADN
'Why is it that you are on a path such as this?' (*Ise* 9)

(25) *kurenawi no mo no suswo nurete ayu **ka***
crimson GEN robe GEN hem get.wet.GER sweetfish KA
*turu **ramu***
catch.CONCL PCONJ.ADN
'the hems of their crimson robes being wet, is it sweetfish that they are catching?' (*MYS* 5.861)

(26) siragiku pa pana **ka** ara-**nu** **ka**
white.chrysanthemum TOP flower KA exist-NEG.ADN KA
nami no yosuru **ka**
wave GEN approach.ADN KA
'The white chrysanthemums, are they flowers, or not? Are they breaking waves?' (*Kokin* 5.272)

Ya ('confirmation soliciting') is used in yes/no questions which seek the hearer's confirmation that an already identified state of affairs (the presupposition) is to be linked with the constituent inside the sentence marked by *ya* ('speaking of X, is it true that . . .?'), see (27); as here, *ya* is often used in disjunctions ('is it A, or is it B?). *Ya* can also be used in *wh-* questions to focus a non-*wh-* element, e.g. (28). *Ya* is often used in rhetorical questions, with exclamative force, (29). At the end of a sentence, *ya* solicits confirmation of the sentence in its entirety, that is, marks a neutral yes/no question. This is different from other uses of the *kakari* particles, for it involves no discernible focus, and this is reflected in the fact that the predicate preceding *ya* is in the conclusive form, not the adnominal, see (22) above. Sentence final *ya* is also often used in exclamative rhetorical questions, in OJ also after the exclamatory form, (30).

(27) kimi **ya** ko-**si** ware **ya** yuki-**kemu**
you YA come-SPST.ADN I YA go-PSTCONJ.ADN
'Was it you that came, or me who went, I wonder' (*Ise* 69)

(28) koko **ya** idoko
here YA where
'Where is this; where are we' (*Tosa*)

(29) tuki **ya** ara-**nu** paru **ya** mukasi no paru nara-**nu**
moon YA exist-NEG.ADN spring YA old.times COP.ADN spring COP-NEG.ADN
'Isn't the moon the same; isn't the spring that of old?' (*Ise* 4)

(30) sirakumo *mo* mipune *no* yama *ni* tayuru pi *ara-me* **ya**
white-cloud ETOP Mifune COP.ADN mountain DAT vanish.ADN day exist-CONJ.EXCL YA
'The white clouds will never disappear from Mt. Mifune! (Will there be a day when the white clouds disappear from Mt. Mifune? (No!))' (*MYS* 3.243)

Zo (so) ('identifying') identifies a constituent for focus, as new information on the background of the presupposition which follows ('it is X that . . .'), e.g. (31). Like *ka*, *zo* can focus an entire sentence, e.g. (32).

In OJ, the shape *so* is more common, the variant *zo* less so, but it is thought that *zo* in EMJ had replaced *so*, and the particle is commonly referred to as *zo*. We follow that practice here, but as most EMJ texts did not distinguish

between tenues and mediae in writing it is difficult to agree that the replacement of *so* by *zo* was as unquestioningly complete in EMJ as is usually assumed; note also for example that the *Ruiju-myōgi-shō* (6.2.3) have unambiguous notations of the particle as *so*.

(31) kyaũ pe kaperu ni wonnago no
capital ALL return when girl GEN
na-ki nomi **zo** **kanasibi-kopuru**
not.exist-ACOP.ADN only ZO grieve-regret.ADN
'Returning to the capital, it is only that my daughter is no more that I grieve and regret' (*Tosa*)

(32) wa ga koromo sur-eru ni pa
I GEN clothes colour-STAT.ADN COP.INF TOP
ara-zu
exist-NEG.CONCL
takamatu no nwobye yuki-sikaba pagwi no
Takamatsu GEN field go-SPST.PROV lush clover GEN
sur-***eru*** ***so***
colour-STAT.ADN SO
'It is not that I coloured my robe. It is that the bushclover rubbed off on it, when I was walking in the fields in Takamatsu.' (*MYS* 10.2101)

Namu (namo; nan) ('confirmative') is used to solicit agreement from the hearer and is typical of very 'engaged registers of speech' (Quinn 1987: 741) and is often used in explanations ('it is X, you see, that . . .'), e.g. (33), in several respects resembling NJ *ne(e)* 'you see'. It is therefore frequent in the literary prose, especially in conversations and in personal narrative styles (and narrator comments), but was not used in poetry. Thus, in OJ, *namo* (which was the OJ shape) is not frequent and is restricted to *Senmyō* (and a single example in *MYS* (12.2877)). In EMJ, OJ *namo* had become *namu* which in some cases was further reduced to *nan*, e.g. (34).

(33) pasi wo yatu watas-eru ni yorite **namu**
bridge ACC eight lay-STAT.ADN DAT depend.GER NAMU
yatupasi to ipi-**keru**
Yatsuhashi COMP call-MPST.ADN
'It is because there are eight bridges, you see, that it is called "Yatsuhashi"' (*Ise* 9)

(34) so no pito katati yori pa kokoro **nan**
that GEN person appearance ABL TOP heart NAMU
masari-tari-keru
excel-STAT-MPST.ADN
'That person, more than her appearance, it was her heart, you see, that excelled' (*Ise* 2)

Koso ('unique identification') is the only *kakari* particle whose predicate is in the exclamatory and not the adnominal. This is the case for all types of predicates from EMJ onwards, but in OJ, adjectival predicates of *koso* focused constituents were in the adnominal, e.g. (36). As opposed to *zo*, which also 'identifies', *koso* does so 'uniquely' ('it is X (and only X) that . . .'), corresponding to 'exactly' or 'just', see (35), and further (18), (20) above.

(35) wotoko pa ko no wonna wo **koso** e-**me**
man TOP this GEN woman ACC KOSO get-CONJ.EXCL
to omopu
COMP think.CONCL
'The man thought that it was this woman (and her alone) that he wanted' (*Ise* 23)

(36) nwo *wo* *piro-mi* *kusa* ***koso*** *sige-ki*
moor ACC wide.INF grass KOSO abundant-ACOP.ADN
'with the moor being wide, it is the grass which is abundant' (*MYS* 17.4011)

8.9.3 The musubi *predicate*

First of all it should be noted that a distinct form of the *kakari-musubi* predicate is not expressed in a number of cases: (a) If the final inflected element in the predicate belongs to the QD conjugation (including the frequent conjectural auxiliary *-(a)m-*), which has no (segmental) expression of the distinction between conclusive and adnominal (it is sometimes speculated that the adnominal was prosodically distinct from the conclusive form also outside its adnominal use, but that is not known; see 7.4.4.3). (b) If the sentence final predicate is incomplete, either abbreviated or left out. (c) If the *kakari-musubi* construction is within a non-finite clause, as conclusive, adnominal and exclamatory are finite verb forms. Such cases make up a non-neglible proportion of *kakari-musubi* constructions, suggesting that the distinctive form of the *musubi* predicate is of less importance than it is often afforded.

In most accounts of *kakari-musubi* within a focus interpretation of the construction, whether they are formally or functionally oriented, the use of the

adnominal or exclamatory form in the corresponding predicates is said to mark the *scope* of the focus constituent, that is, to show which is the predicate in the presupposition. This is in principle no different from the automatic rule of the traditional agreement description. A notable exception is the work of Quinn (1987) which divorces the two parts of the *kakari-musubi* construction, the *kakari* particle and the form of the *musubi* predicate, and gives each of them a functional, synchronic explanation. Quinn relates the use of the adnominal for the presupposed predicate to its overall function as a nominalized, and therefore 'referential' and 'grounded', verb form, suitable for a presupposed constituent.[3] This explanation of the use of the adnominal form in *kakari-musubi* constructions is attractive, but does not extend to the use of the exclamatory. The main uses of the exclamatory as a word form, distinct from its uses as a combinatory stem in word formation (cf. 3.4.4.3), were in *kakari-musubi* with *koso*, and in exclamations (mostly couched as rhetorical questions and followed by *ya*, *yamo*, *yapa*, *ka*, *kamo*), so it is not possible in the same way as with the adnominal to generalize from other uses of the exclamatory. Most scholarship agrees, though, that there are close diachronic and functional links between the adnominal and the exclamatory and that both exhibit nominal(ized) characteristics. Finally, in syntactic terms, the use of a nominalized verb form in the predicate of the presupposition ties in very well with the copular nature of the *kakari-musubi* relation, see (31) above.

8.9.4 Questions and exclamatives

Two other uses of the adnominal form should be mentioned in this context, namely in questions and in exclamatives. As outlined above, yes/no questions are constituted by the presence of *ya* or *ka* (with sentence final *ya* following the conclusive form making up a neutral, unfocused yes/no question). *Wh*-questions, on the other hand, do not require the presence of a question particle, but can be constituted simply by the presence of a *wh-* word, see (37). Also in such cases the predicate takes the adnominal form, despite the absence of a *kakari* particle. *Wh-* words can be considered focus constituents in their own right and that may account for the use of the adnominal form in such sentences, but it has also been suggested that the adnominal form more generally is used, as a marked form, in marked, non-declarative utterance types, including questions, but also in what are interpreted as exclamatives: (38) is one such example where the use of the adnominal form in the predicate is said to make the sentence exclamative. A potential problem with this kind of interpretation

[3] Similarly, but in a diachronic perspective, Wrona (2008) proposes that the *kakari-musubi* construction originated in the function of the adnominal form of marking non-declarative utterance types (interrogative or exclamative) and that the particles originally were secondary scope markers.

is the absence of any other evidence than the adnominal form that the sentence is in fact intended to be exclamative. There are not many examples of sentence final adnominal form not correlating with a *kakari* particle in OJ, but numbers increase through EMJ, e.g. (39).

(37) ide ika ni kokodaku kwopu*ru*
oh how COP.INF this.much love.ADN
'Why do I love so much?' (*MYS* 12.2889)

(38) wa ga koromode *no* puru toki *mo*
I GEN sleeve GEN dry.ADN time FOC
na-ki
not.exist-ACOP.ADN
'there is no(t even) time for my sleeves to dry!' (*MYS* 10.1994)

(39) suzume no ko o inuki ga nigasi-**turu**
sparrow GEN child ACC Inuki GEN let.go-PERF.ADN
'Inuki let my baby sparrows go!' (*Genji*: *Waka-murasaki*)

REFERENCES

General: Iwai 1970, Tsukishima 1969, 1987, Vovin 2003, Yamada 1913/1952. Aspect: Kinsui 2006, Sandness 1999, Takeuchi 1987. *Kakari-musubi*: Ohno 1993, Quinn 1987: 673–825, 934–49; Watanabe 2002, Whitman 1997, Wrona 2008: 195ff.

9 The sinification of Japanese

Just as early Japanese society received massive intellectual, cultural, political and social influence from Tang China, so the Japanese language was heavily influenced by Chinese, particularly through the MJ period. There is no doubt that prior to this, both in Nara and pre-Nara Japan, some intellectuals and clergy, as well as traders and fishermen, had some facility in varieties of spoken Chinese. As was set out in 4.2.2 above, a number of early loanwords into Japanese from Chinese, possibly mediated through the Korean peninsula, may be identified, and it is likely that there are more which we are not able to identify. However, the pervasive influence on Japanese from Chinese in the OJ and MJ periods which took place through the medium of text was of an altogether different order, affecting both usage and especially vocabulary to an extent which merits the designation *sinification*. This took place through two related, complementary modes of interacting with Classical Chinese text (*kanbun* 漢文), generally thought of as two ways of 'reading' the texts: *kanbun-kundoku*, the rendition of Chinese text in Japanese, which affected grammar and usage (see 9.1) and (*kanbun-*)*ondoku*, the vocalization of Chinese text as such, which paved the way for the intake of a large number of loanwords from Chinese (9.2). Both of these 'reading' practices have a long history in Japan, predating the Nara period and continuing into the present. It is convenient to treat them as one here, for it is from the Heian period we find the earliest direct evidence for the language of *kanbun-kundoku* and the beginnings of a large-scale adoption of SJ loanwords.

9.1 *Kanbun-kundoku*

Kanbun-kundoku (漢文訓読) is the interpretation, explication or translation in or into Japanese of Classical Chinese text. An important characteristic of *kanbun-kundoku* is the notion that it involves verbalizing the original Chinese text in Japanese, and it is popularly thought of as 'reading' Chinese text 'in Japanese' or 'with Japanese grammar'. The practice of *kanbun-kundoku*, understood as 'the "reading" of Chinese in a local vernacular language', is not restricted to Japan, but is a common feature of civilizations within the Sinitic cultural sphere, attested and described in the sixth and seventh centuries from places as far-flung as Japan, the Korean peninsula, Vietnam and Gao Chang (高昌, the site of an important oasis city on the Silk Road, in what is now the

Chinese province of Xinjiang). It is safe to assume that *kanbun-kundoku* in this wider sense predates these early attestations, and it may be thought that the advent of Chinese text in Japan from the Korean peninsula early in the fifth century was accompanied by this practice. In Korean, rendition of Chinese text in Korean is said to date back at least to the fifth century. It is likely that the strong logographic element in Chinese writing favoured the development and spread through East Asia of *kanbun-kundoku* (-like practices), together with the notion that it consists in 'reading' Chinese in another language. Although *kanbun-kundoku* certainly is a kind of 'translation', we use the broader term 'rendition' in the following in order to capture all of what the practice involves.

A basic feature of *kanbun-kundoku* is the *translation* of words and phrases in the Chinese text into Japanese. Chinese and Japanese are grammatically quite different: Chinese has no inflectional morphology, expresses grammatical relations by word order and has a large inventory of grammaticalized preposed verbs and adverbs (expressing amongst others categories such as negation and mood), as opposed to the fairly rich verbal and adjectival inflection, specification of syntactic roles by grammatical particles and free word order (except for verb-finality) of Japanese. Thus, *kanbun-kundoku* involves finding suitable translation equivalents in Japanese for content and function words in the Chinese texts. However, in addition *kanbun-kundoku* involves a number of processes in order to render ('read') the Chinese text in Japanese: *transposition* (change of word order) and *interpolation* or *specification* (of inflectional morphemes or grammatical particles).

9.1.1 Kunten

Kunten (訓点 'reading marks, glosses') is a cover-term for a variety of annotations added to Chinese text in order to aid these processes of its rendition in Japanese. The earliest extant *kunten*, from the late eighth century, are punctuation marks, showing phrasing and division of a text, and marks showing how to change the word order when rendering text in Japanese. The latter are collectively known as *kaeriten* 'reversal marks'; through time these have included numbers and other means of showing sequence. Especially from the Heian period onwards, more types of *kunten* are found which may roughly be divided into two classes: *kana glosses* and *okoto-ten*.

Kana glosses are *man'yōgana* or *kana* written next to a character, indicating its 'reading'. This could be a SJ word, in which case the gloss only had information about the pronunciation; or it could be the sound shape of a Japanese word used to render the Chinese word in Japanese, in effect constituting a translation or glossing in Japanese of the Chinese word. The development of *katakana* is closely linked to the practice of glossing, and *katakana* is traditionally viewed as originating as a subtype of *kunten*.

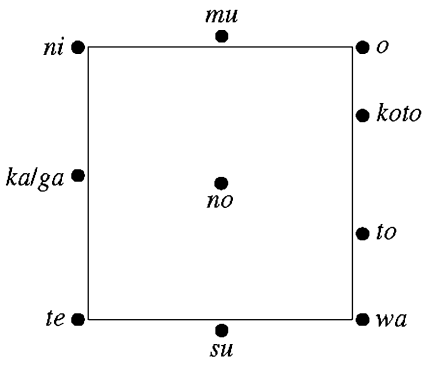


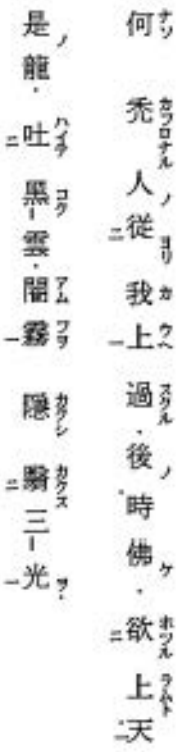
Figure 9.1 Example of *kunten* system

Okoto-ten (thus named after two frequently noted grammatical forms, the particle *wo* > *o* and the nominalizer/complementizer *koto*) or *tenioha* (named after the gerund formant *-te* and the particles *ni*, *o* and *pa* > *wa* (today written using the *kana* for *ha*)) are two common terms for diacritic marks which indicate grammatical morphemes. Graphically, *okoto-ten/tenioha* are lines, dots, circles, hooks or marks of other shapes which were placed next to or on *kanji*. *Okoto-ten/tenioha* are generally thought of as shorthand for grammatical particles or words, auxiliaries or inflectional endings; both the shapes and positions of marks are significant. Figure 9.1 is an example of part of a system which uses single dot marks for *-te*, *ni*, *o* and *wa* in the four corners of the space around a *kanji*, as well as marks for *o* and *koto*, exemplifying the frequent markings which gave rise to the two names for this kind of diacritic. It also includes marks for the particles *ka*, *ga*, *no* and *to*, as well as for *su* and *mu*.

Through the first half of the Heian period, a wealth of different *kunten* traditions developed, with individual scholars, sects or temples developing their own systems which grew increasingly complex and often secret, or at least exclusive. Tsukishima (1986) provides a large number of charts illustrating different systems, many of which are very elaborate. Within each school of reading, normative annotations and readings of individual texts became established, and from around the mid Heian period text annotations became the object of faithful, dogmatic tradition. From the Kamakura period, the introduction of neo-Confucianism was accompanied by new text interpretations, leading to some innovation and change in annotations of some Confucian texts. Today a simple system of *kunten*, confined largely to *kaeriten*, is taught in Japanese schools as part of the *kanbun* curriculum.

The following is an example of *kunten* and *kanbun-kundoku*. (1) is a Chinese text, a short passage from the *Fǎhuā wénjù* (法華文句, Japanese *Hokke mongu* 'The sentences and phrases in the Lotus sutra'), an explication of the Lotus Sutra by Zhi-yi (智顗, Japanese Chigi) from the second half of the sixth century, written down by one of his students. Zhi-yi was the founder of Tiantai Buddhism, the ancestor of Japanese Tendai Buddhism, and the *Fǎhuā wénjù* is an important text in Tendai Buddhism. (2) is the text annotated with *kunten*, adapted from Nishizaki (1992) which photographically reproduces the original annotated text and provides a precise transliteration in printed type (as well as a rendition in Japanese). The annotation dates from around the year 1000.

(1) 何禿人從我上過
後時佛欲上天
是龍吐黑雲閻霧隱騎三光

(2) 

In (2) we first of all see a great deal of *kana* glosses, both giving information about 'readings' of individual *kanji* and supplying grammatical information not represented in the Chinese text, such as particles and verbal auxiliaries and inflectional endings. The *kana* glosses do not distinguish between tenues and mediae, for example using 力 both for /ka/ in *kaburo* and for /ga/ in genitive *ga*, and inspection of the photographic reproduction of the original *kunten* text further shows that a number of variant *kana* shapes were used in the glosses. There are also some punctuation marks (a 'comma' after 過, 佛, 龍 and 雲; a 'full stop' after 天 and 光) and *kaeriten* in the form of the numbers 1 or 2 to the left of 従, 上, 欲, 天, 吐, 霧, 翳 and 光, indicating that the *kanji* (complex) marked '2' should be rendered after that marked '1'. Finally, there are two instances of *okoto-ten*, a single dot by the top left corner of 時 and 天, standing for *ni*.

(3) is the *yomi-kudasi* (lit. 'reading-down'), the Japanese text represented by the *kunten*. We have noted in CAPITALS readings not indicated in the *kunten* text, i.e., readings which we hypothesize in our interpretation. Note that 龍 appears earlier in the text where it is glossed as *riu*; 佛 is glossed simply by *ke*, which suffices to show that it is meant to be glossed *potoke* and not for example *putu*. Everything in lower case in (3) is directly represented in the *kunten* text. We use **boldface** for glosses giving grammatical information not represented in the Chinese text, and ***bold italics*** for such information given by *okoto-ten*, and we <u>underline</u> words which have been transposed, generally by a move to the right as instructed by numbers, but note also that the order of 欲 and 上 has been reversed without any overt instruction. Knowing and following the conventions used in this tradition of *kanbun-kundoku* allows a reader to render the Chinese text into Japanese.

(3)

何	禿		人		我		上	従	過
nazo	kaburo	**naru**	PITO	**no**	WA	**ga**	uwe	<u>yori</u>	sug**uru**,
why	bald	COP.ADN	person	GEN	I	GEN	above	ABL	pass.ADN

[The Dragon said:] 'Why is a bald man passing over me?

後		時		佛
NOTI	**no**	TOKI	***ni***	POTOke,
after	COP.ADN	time	DAT	Buddha

天		上		欲
AME	***ni***	<u>NOBOra-**mu**</u>	**to**	<u>possu</u>
heaven	DAT	ascend-CONJ.CONCL	COMP	want.CONCL

'Later, the Buddha wanted to ascend to heaven.'

是	龍	黒雲	闇霧	吐
KO **no**	RIU,	kokuUN-anbu **wo**		pai**te**
this GEN	dragon	black.cloud-dark.mist	ACC	breath.GER

三光	隠　翳
SANGWAŨ **wo**	kakusi-kakusu
three.lights (the sun, moon, and stars) ACC	hide-hide.CONCL

'The dragon hid the sun, moon, and stars by breathing black clouds and dark mist.'

Several questions arise from *kunten* texts such as (2). Should they primarily be considered 'annotated' Chinese text with more or less abstract instructions about their interpretation and rendition in Japanese, or are they rather orthographically severely underspecified Japanese text, superimposed on the still present Chinese source text? Is the use of *kunten* 'annotation' or is it 'writing'? *Kunten* texts present a multi-layered textuality of great complexity with non-trivial difficulties of interpretation.

In China itself, annotating text for interpretation or pronunciation is well established and has a long history. For example, the *shōten* ('tone marks') mentioned earlier (6.1.2.2) are similar to *kunten* and are part of the same overall phenomenon. Until recently it was thought that *kunten*, including *katakana*, were independent developments in Japan, if perhaps generally inspired by for example *shōten*; similar types of materials exist in Korea, but are somewhat later. In Korea, marks to annotate text are called *kugyŏl*; they are very similar to Japanese *kunten*, but the earliest have been thought to date from the ninth century. However, especially with the continuing discovery in both Japan and Korea of increasing amounts of *kunten* materials which are annotated not in ink, but by *stylus* (角筆 *kakuhitsu*) which leaves indentations or scratchings on the paper, but no colour, it is gradually becoming clear that techniques for annotation were used both on the Korean peninsula and in Japan at an earlier time than was previously thought, with the oldest such stylus materials in Korea dating from the late seventh century. It now in fact seems overwhelmingly likely that *kunten* techniques, too, like Chinese writing and text and *kanbun-kundoku*, were transmitted from the Korean peninsula to Japan. For example, the earliest Japanese materials are far more similar to the Korean materials than are later Japanese materials. Both *kanbun-kundoku* and *kunten* and their histories must be viewed in a pan-East Asian perspective, where, in particular, the spread of Buddhism and Buddhist canonical texts in Chinese translation and commentaries written in Chinese played an important role.

9.1.2 Kanbun-kundoku *and writing in Japanese*

A close relation holds between *kanbun-kundoku* and the development of writing in Japanese. In the course of *kanbun-kundoku*, fixed, habitual

renditions of individual *kanji* arose, resulting in conventional associations of many *kanji* with specific OJ words; or in other words, the establishment of conventional '*kun*-readings' of *kanji*. Once this association of decoding (reading) was established, the next step of reversing the relation to one of encoding (writing) was not a big one. For example, habitually using Japanese *ma- ~ me* 'eye' to translate into, that is *read* in, Japanese the Chinese word written by 目 established a representational relation between 目 and *ma- ~ me* 'eye':

(4) 目 => *ma- ~ me* 'eye'

This could now be reversed to have the word *ma- ~ me* 'eye' represented, that is, *written* by, 目, see (5), making possible logographic representation of Japanese. This is the origin of logographic writing of Japanese.

(5) *ma- ~ me* 'eye' => 目

Furthermore, by extension, once the encoding relation between *ma- ~ me* 'eye' and 目 was established, 目 could be used as a phonogram (*kungana*, cf. 1.1.2.5) to write the syllables /ma, me/:

(6) /ma, me/ => 目

Both the logographic and phonographic use of 目 shown in (5) and (6) are amply attested in the OJ sources. They provide indirect evidence that Chinese 目 in *kanbun-kundoku* in fact was rendered by *ma- ~ me* 'eye', for if 目 could be used to write OJ *ma- ~ me* 'eye' logographically and /ma, me/ phonographically, it is because *ma- ~ me* 'eye' habitually was used to translate Chinese 目 into Japanese in *kanbun-kundoku*.

The bidirectional reading–writing relationship between *kanji* and Japanese words and morphemes – and what appears to be an identification of the processes of reading and writing – is evident from uses of 訓 in the earliest sources from the Nara period. In Chinese the basic meaning of 訓 (EMC *xun^{h}) is 'instruct, teach; follow, obey', as is also reflected in many current SJ words, e.g. *kunren* 訓練 'training', *kyōkun* 教訓 'lesson'. It later came to be used in the sense of 'gloss, read, interpret (authoritatively)', cf. 訓詁 (SJ *kunko*) 'exegesis, interpretation, annotation, commentary'. This is the sense and use reflected in SJ *kundoku*, *kunten* etc. In Japan 訓 is used in this way in our earliest sources, for example in the 'reading' notes inside the main text of the *Kojiki*, to mean 'read (out) (a logographically written word)'. The reading notes are instructions, written in Chinese, about how to read the main text; they are not later additions, but part of the text. (7) is the first such note and

exemplifies this usage, instructing the reader to read the *kanji* 天 on this occurrence as *ama*, and not for example *ame*.

(7)

訓	高	下	天	云	阿	麻
read		below		say	/a/	/ma/

'***reading*** the 天 after the 高, say *ama*'

However, a quite different use of 訓 is found in the *Kojiki* preface, which is generally regarded as being written in Chinese, in the paragraph outlining the writing principles employed in the main text of the *Kojiki*. Here 訓 is used to mean 'logographic *writing*', a usage not found in Chinese. The passage is generally instructive, for it explains well the tension between logographic and phonographic writing of Japanese, and we already here see the juxtaposition of 音 and 訓 (SJ *on*, *kun*) which today are used about different 'readings' of individual *kanji*. (8) gives the text line by line together with Philippi's translation (1968: 43; emphasis added).[1]

(8)

然、上古之時、言意並朴、敷文構句、於字即難。
However, during the times of antiquity, both words and meanings were unsophisticated, and it was difficult to reduce the sentences and phrases to writing.

已因訓述者、詞不逮心。
If expressed completely in ***logographic writing***, the words will not correspond exactly with the meaning,

全以音連者、事趣更長。
and if written entirely ***phonographically***, the account will be much longer.

是以今、或一句之中、交用音訓、
For this reason, at times ***logographic and phonographic writing*** have been used in combination in the same phrase,

或一事之内、全以訓録。
and at times the whole matter has been recorded ***logographically***.

即、辞理叵見、以注明、意況易解、更非注。
Thus, when the purport is difficult to gather, a note has been added to make it clear; but when the meaning is easy to understand, no note is given.

[1] We change Philippi's 'ideographic' to the more current 'logographic', and 'phonetic' to 'phonographic'.

> 亦、於姓日下、謂玖沙詞、於名帯字、謂多羅斯、如此之類、随本不改。
> Again, in the case of surnames such as Kusaka, which is written 日下, and given names such as Tarasi, which is written 帯, the traditional way of writing has been followed without change.

In this way, all logographic writing of Japanese derives from *kanbun-kundoku* reversed from reading to writing. The basic mechanism is that any character or string of characters which could be rendered into Japanese could also be used to write the Japanese rendition, as illustrated above with 目 writing *ma- ~ me* 'eye'. Another simple example is the writing of single grammatical morphemes by single *kanji*, for example the use of 而 to write the flective *-(i)te* (gerund), e.g. 始而 *padime-te* 'beginning'. More complicated examples are the writing of morphologically complex forms, especially inflected verb forms, with logographic representation of grammatical elements, but with the order of the *kanji* reflecting the Chinese constituent order:

(9) 不有 ara-**zu** 'is not'
可有 aru **besi** 'should be'
所知在 sira-**re**-tari 'was known'

Such writing is very frequent in both *Norito* and *Senmyō* and is also found widely in the *Man'yōshū*. It is also a prominent feature of the *kanji-kana majiribun* way of writing which gained currency from the second half of EMJ. As mentioned above (6.1.1), Japanese writing today is a direct descendant of the *kanji-kana majiribun* of EMJ and LMJ, which in addition to the common principles of logographic writing of Japanese derived from *kanbun-kundoku* was particularly influenced by the use of *kana* glosses in *kunten* annotations. Thus, the way Japanese is written today may trace its origins directly back to *kanbun-kundoku* practices and *kunten* techniques.

9.1.2.1 Hentai kanbun; 'kanbun'

An extreme and complicated logographic way of writing Japanese has some, but not all, constituents placed in an order resembling Chinese constituent order and little specification of verbal inflection. Reading this type of text involves some of the same processes as *kanbun-kundoku*. (10) is a very simple example from the *Kojiki* which illustrates the main principles: transposing elements (here underlined) and supplying grammatical elements not represented in the text (in **bold**). Although it is a way of writing Japanese, this type of writing is confusingly known as *hentai kanbun* (変体漢文 'deviant Chinese text (or writing)'); it is called 'deviant' because it exhibits non-Chinese features (including word order and use), which is not surprising as it

is a way of representing Japanese, not Chinese. One well-known example is the use of 御 to write general honorific elements in Japanese (as in (10) where it represents the honorific prefix *mi-*), whereas 御 in Chinese is used to refer to the emperor.

(10) 1 2 3 4 5 6 7
於其嶋天降坐而

2		3	1	4 5 6 7
so	**no**	sima	ni	ama-kudari-masi-te
that	GEN	island	DAT	heaven-descend-RESP-GER

1 2 3 4 5 6
見立天之御柱

3	4	5 6	1 2
ame	no	mi-pasira **wo**	mi-tate
heaven	GEN	HON-pillar ACC	see-erect

1 2 3 4 5
見立八尋殿

3 4 5	1 2
ya-piro-dono **wo**	mi-tate-**tamapi-ki**
eight-hiro-palace ACC	see-erect-RESP-SPST.CONCL

'Descending from the heavens to this island, they erected a heavenly pillar and a spacious palace' (*Kojiki*, *Nihon koten bungaku taikei* 1, pp. 52–3; translation by Philippi 1968: 50)

Hentai kanbun is subsumed as a subtype under the more general term '*kanbun*', which has given rise to a great deal of confusion, because it is commonly used in a variety of meanings, to refer to quite different types of text, ranging from text written straightforwardly in Classical Chinese – this is the way we used the word above – over *hentai kanbun* to refer also sometimes to those portions of a Japanese text which are written in *kanji*.

In addition to *hentai kanbun*, a practice arose of writing a Japanese text by reversing fully, and not just partially as with *hentai kanbun*, the process of *kanbun-kundoku* and thus so to speak translating the Japanese text into Chinese with the purpose of it being re-translated into Japanese when read. Some texts written in this way were even supplied with *kunten* (both *kana* glosses and diacritics) in order to aid the interpretation and thus have the appearance of *kunten* texts. It may be thought that much, if not most, '*kanbun*' written in Japan since the late Heian period is not actually written in Chinese, although it looks that way, but is a cumbersome representation of Japanese. The writing

of Japanese 'in *kanbun*' – including but not limited to *hentai kanbun* – continued long into the modern period.

9.1.3 Orthographic overdifferentiation

'*Kun*-readings', that is habitual association of individual *kanji* with Japanese words, have occasionally imposed orthographic distinctions on Japanese which reflect distinctions in Chinese (cf. also 1.1.3.1 about polyvalence and equivalence in the use of *kanji* to write Japanese logographically). For example, the verb *yom-* must be written by (a) 読 or (b) 詠 depending on whether it means (a) 'read (silently), read (out), chant (a sutra)' or (b) 'compose, write a poem', leading to the popular belief that *yom-* is two different words. An extreme example is the word *oba* (< OJ *woba*) which simply means 'aunt', but which can be written 伯母 for 'older sister of father (or mother)' or 叔母 'younger sister of father (or mother)', imposing Chinese derived kinship differentiations on the writing of Japanese. It may well be that the split of OJ *mono* 'thing, being, person' into two separate words (a) 'thing' and (b) 'person' has been reinforced by the writing by two different *kanji* from early on: (a) 物 and (b) 者.

9.1.4 Kokuji

A final offspring of logographic writing of Japanese is the invention in Japan of *kanji* for Japanese words, the so-called *kokuji* (国字). Well-known examples which are in use today include those in (11). Note that the final three are used with '*on*-readings' to make up SJ (looking and sounding) vocabulary.

(11) 榊 *sakaki* 'sakaki-tree; sacred tree'; 辻 *tsuji* 'crossroad'; 凩 *kogarashi* 'fierce wind in late autumn and early winter'; 峠 *tōge* 'mountain pass; peak'; 畑, 畠 both *hatake* 'field'; 噺 *hanashi* 'talk'; 鱈 *tara* 'cod'; 躾 *shitsuke* 'discipline, manners'; 遖 *appare* 'splendid, brilliant'; 働 *hatarak-* 'to work', *dō* (労働 *rōdō* 'work'[2]); 腺 *sen* 'gland' (汗腺 *kansen* 'sweat gland'); 鋲 *byō* 'rivet' (画鋲 *gabyō* 'drawing pin').

9.1.5 Kuntengo

Although *kanbun-kundoku* was practised in Japan long before the spread of *kunten*, it is not until the appearance of *kunten* materials that we get direct

[2] 労働 was earlier written 労動, but in the Meiji period the current writing came to be used, following a period where 労動 was glossed *hatarak-* 'to work', which is the word 働 was made up to write. The *kanji* 働 itself was used already in the LMJ period.

evidence for the language used in *kanbun-kundoku*, although we do have some *indirect* evidence from written Japanese in the form of early identifiable influence from Chinese on Japanese (through *kanbun-kundoku*, see 9.1.6), and in the form of early well-established associations between individual *kanji* and Japanese words in the writing of Japanese, as mentioned in 9.1.2. The value of such indirect evidence is difficult to judge – the risk of circularity is obvious – and should only be taken to supplement direct evidence. The language used in *kunten* texts is referred to as *kuntengo* (訓点語). Since the beginning of the twentieth century, the study of *kunten* texts (see 6.2.2) and of *kuntengo* has become an important focus of study for Japanese linguists and philologists, more recently attracting overseas scholars as well. *Kunten* texts are important primary sources and it is thought that especially *kunten* texts from the first half of the Heian period reflect some form of vernacular language, and they thus contribute to filling the gap in our knowledge of early EMJ left by the lack of other types of sources between late OJ and *c.* 900. However, *kunten* texts are in some respects orthographically underspecified and on some points very difficult to interpret, so their evidential value is not entirely straightforward. Furthermore, strict, dogmatic norms of rendition and annotation arose, so that *kunten* texts from after the middle of the EMJ period generally cannot be taken to reflect contemporary language.

As may be expected, *kuntengo*, which originates in a form of translation, is specialized and special, differing from general language in a number of respects, in terms of *style* and *usage*. However, although *kuntengo*, like much translation language, most likely was stilted and in some respects formal, it first and foremost constitutes one use, or genre, of Japanese and does not differ in basic morphology or syntax from other genres of Japanese. For example, in *kuntengo* subject and object nouns generally have case particles to show their grammatical function, whereas subjects and objects often were left unmarked in general writing – and in spoken language. This is a difference in formality, not in grammar.

Through the early Heian period, a specialization can be observed, so that some expressions were used exclusively or predominantly in *kuntengo*, whereas other near-synonymous forms were used in general prose writing. Apart from differences in formality and genre, this to some extent reflects that *kuntengo* preserves archaic features of the language. One example is the particle *i* (3.7.1.3) which is only attested in a small number of examples in OJ and not used at all in general writing in EMJ, but which was used extensively in some schools of *kanbun-kundoku*. In other cases, however, it was a matter of tendencies that some frequent features of *kuntengo* would not be used much in general prose writing. (12) gives some examples from the EMJ period, of which the forms in (a) reflect differences in usage or genre, whereas those in (b) reflect retention of forms which had gone out of use in the contemporary

language. Towards the end of EMJ and in LMJ, *kuntengo* exerted a great deal of influence on formal writing, and some forms were revived in writing outside of *kuntengo*. This coincided largely with the spread and wider usage of *kanji-kana majiribun* (6.1.1).

(12)	General prose	*Kuntengo*
a.	*-(e)do* concessive	*-(e)domo*
	-nu, -ne, -de negative adnominal, exclamatory, gerund	*-zaru, -zare, -zusite*
	nar- copula	*tar-*
b.	*-sase-* causative	*-(a)sime-*
	yaũ-nar- 'be like'	*goto-*
	opase- (>owase-) 'exist.RESP'	*imas-, masimas-*

9.1.6 The influences of kuntengo *on the Japanese language*

More importantly, a number of usages in *kuntengo* originated in the course of attempting a faithful, literal rendition from Chinese. Some such features, which arose in the translation process and which thus reflect influence from Chinese, were carried over into and gained currency in general language used outside that context, for example through dissemination of sutra commentaries or in the related didactic *setsuwa* literature, or through the speech and writing of scholars and clergy familiar with *kanbun-kundoku*. This is not unlike the kind of influence the language used in Bible translations in Europe had on European languages, where many words and idioms arose in and spread through their use in Bible translations. Examples in English from William Tyndale's translation of the New Testament (1525) include *the powers that be*, and *eat, drink and be merry*, and words such as *busybody*, *castaway* and *zealous*.

Although it was during the EMJ period that the influence from *kanbun-kundoku* on Japanese gained momentum and increased, it seems clear that already OJ was influenced by *kanbun-kundoku*. This is evident in particular in the *Senmyō* and *Norito*, but also in some poetry from the *Man'yōshū*, but hints of this influence are found throughout the OJ text corpus. This means, conversely, that it is not possible to gauge the full extent or character of this influence, for we have no attestation of Japanese before it set in, but in the following we outline the type of influence *kanbun-kundoku* had on Japanese and give some examples of usage which permeated to general language.

Lexical loan translations originating in *kanbun-kundoku* abound in Japanese through the Old and Middle Japanese periods. Early examples include *ame-tuti*

'heaven and earth; the world' originating in the rendition of 天地; *ko no yo* 'this world (opposed to before- and after-life)' < 此世; *kaku-no-goto(ku, -si)* '(be(ing)) in this way' < a number of different expressions in Chinese: 如是, 如此, 如斯, 如, 若斯; *nana-kusa no takara* 'seven treasures; many treasures' < 七寶, *iki-sini* 'living and dying' < 生死; *toki-doki* 'sometimes' < 時時, *tokoro-dokoro* 'here and there' < 所所, 処処, and more generally the reduplicating pattern of the last two examples. In addition to loan translations it is very likely that *kanbun-kundoku* more indirectly has influenced the use of individual lexical and grammatical items in Japanese widely, but it is at present not possible to say much concrete about that. See below (9.3) for one example.

On a more general level, the language of *kanbun-kundoku* was characterized by being *expository*, with relatively sparse use of modals, and *explicit*, with for example far less drop of core case particles than in other genres. In that way *kanbun-kundoku* language undoubtedly encouraged the development and use of formal, expository styles and modes of discourse in Japanese. Possibly related to that is the influence on information structure in Japanese sentences, and this is perhaps the greatest structural influence which Chinese, through *kanbun-kundoku*, has exerted on general Japanese. In Japanese, clauses are typically connected by non-finite verb forms or post-verbal conjunctional particles which express a variety of conjunctional and modal categories, and other types of mood, including negation, are expressed by verbal auxiliaries or (post-verbal) clitics, i.e. particles or extensions. The function words which in Chinese express such categories generally occur earlier in the sentence, and in the process of *kanbun-kundoku* they were often rendered by adverbs, or nouns or verb forms drafted in to function adverbially, placed at the beginning of the sentence or clause. When carried over into general language, this may be thought to have contributed to the now widespread use of *conjunctions* (which in Japanese are sentence or clause initial adverbs) and sentence initial *modal adverbs* which introduce the overall modality of the sentence, both of them sometimes in correlation with specific modal verb forms. This type of usage has thus influenced the information structure of Japanese sentences; it may also have contributed to the decline in verbal modal auxiliaries in the language. Examples of such conjunctions and modal adverbs include *sikasite* 'and, then' (< *sika site* 'doing thus'), *sikaru ni* 'however' (< 'although it is thus'), both originally used to render 而 'then, and' (EMC *ɲɨ/ɲi). The adverb *imada* '(even) now' (etymologically *ima* 'now' + *-da* 'adverbial formant') came to be used in correlation with a negative verb form to mean 'not yet, never', e.g., *imada tokazute* 'not yet untying' (*KK* 2). This usage of *imada* originates in the rendition of the Chinese negation 未 'not yet' (EMC *muj^{h}) and continues in the modern language, where primarily the reduced shape *mada*, which is attested already from EMJ, is used. Yamada (1935) lists around

fifty examples of *kanbun-kundoku* usages which are preserved in the modern language, and the great majority are such conjunctions and adverbs.

A slightly different case, involving a more straightforward grammatical loan translation, is the rendition of Chinese 而後, which is a sentence connective 'and then'. In addition to *sikasite* and *sikaru ni* just mentioned, 而 on its own was also rendered by the gerund formant *-(i)te*, and was in turn used to write *-(i)te* as mentioned above; and 後 'after, afterwards' (EMC *ɣəw') could be rendered by *noti* 'end'. The combination 而後 was jointly rendered as (VERB)-*te noti* which found its way into general usage in the meaning 'after VERB-ing', e.g., *okosete noti* 'after sending', eventually to form the model of the NJ synonymous construction VERB-*te kara* (*tabete kara* 'after eating'). A noun such as *noti* 'end' lends itself to grammaticalizing to acquire conjunctional uses and that may well have contributed to the acceptance into general language of *-(i)te noti* as a conjunctional expression, but it remains syntactically unusual in Japanese to combine a gerund and a noun in this way.

A famous example of syntactic influence from Chinese, through *kanbun-kundoku* language, on Japanese is the use of a nominalized verb form to *introduce* reported speech or thought, leading to a framing construction, widespread already in OJ, where reported speech is both introduced and concluded by a verb of utterance, e.g. (13). (14) is an example from the *Man'yōshū*, repeated from 3.1.4.9.2 above.

(13) (X) *ipaku* 'Y' *to* *ipu*
say.NMNL COMP say
'(X) says "Y"'

(14) kamwiyo *ywori* **ipi-tute-*kuraku***
god.age ABL say-transmit-come.NMNL

swora*mitu*	yamato	no	kuni	pa
soaring	Yamato	GEN	land	TOP
sumyekamwi	*no*	*itukusiki*	kuni	
ruling.deity	GEN	august-ACOP.ADN	land	
kotodama	*no*	*sakipapu*	kuni	*to*
word.spirit	GEN	bless.ADN	land	COMP

katari-tugi *ipi-tugapi-kyeri*
tell-continue.INF say-continue-come.STAT.CONCL

'It has been recounted down through time since the age of the gods: that this land of Yamato is a land of imperial deities' stern majesty, a land blessed by the spirit of words' (Levy 1981) (*MYS* 5.894)

The *introduction* of reported speech is atypical of Japanese, where conclusion by a complementizer and verb of utterance is usual. It is thought to have originated in *kanbun-kundoku* renditions of phrases like Chinese 子曰 'the master says: ...', introducing sayings by Confucius in the Analects, which in *kanbun-kundoku* typically is glossed as *si ipaku* (> mid EMJ *iwaku*), using the nominal form of *ip-* 'say' (> *iw-*). The rendition of 子曰 'X' as *si ipaku 'X' (to ipu)*, rather than for example *si 'X' to ipu*, maintains the original word order, and again information structure, and keeps the rendition of 子 and 曰 together, at the expense of creating an unusual sentence construction. From EMJ, the nominal verb form disappeared from general use, but this construction continued to be used through the MJ period with other nominalizers, e.g. *koto* or SJ *yaũ* 'way, manner', as in (15). It is no longer used productively in the modern language.

(15) kaditori no **ipu** **yaũ**
oarsman GEN say manner

kurotori no moto ni siroki nami wo yosu
(black)scoter GEN base DAT white wave ACC break

to zo ipu
COMP FOC say

'The oarsman said: "below the black birds, the white waves are breaking" ' (*Tosa*)

Although the nominal form ceased to be productive and dropped from general use, a number of such forms continued to be used in *kanbun-kundoku* in these constructions, found their way into general language as sentence initial adverbs or nouns, and are retained into the modern language: *iwaku* mentioned above, which is now used as a noun 'reason, pretext; past' and also to introduce quotes or proverbs ('as the saying goes'); *negawaku wa* 'I pray, hope' (< OJ *negapaku pa* 'pray.NMNL TOP; what I pray'); *omoeraku* 'methinks' (< *omop-yeraku* 'think-STAT.NMNL; what I am thinking'); *omowaku* 'thought, opinion' (today written 思惑) (< *omopaku* 'think.NMNL; what I think'); *osimuraku wa* 'regrettably, unfortunately' (< *wosimuraku pa* 'regret.NMNL TOP; what I regret'); *osoraku* 'likely, probably' (< *oso(ru)raku* 'what I fear').

Other examples of usage retained in *kanbun-kundoku* language and subsequently carried over into general language include the OJ passive *-(a)ye-*, which dropped out of the language in the transition from OJ to EMJ, but is reflected in lexicalized modifiers such as *iwayuru* 'so-called' (< OJ *ipa-yuru* 'say-PASS.ADN'), *arayuru* 'all, every' (< *ara-yuru* 'exist-PASS.ADN'), both of

which are in use today. In *kanbun-kundoku*, *ipayuru* and *arayuru* were used in rendition of phrases such as 所謂 NOUN and 所在 NOUN, respectively. In Chinese 所 (EMC *ʂiǎʔ) is both a noun 'place' and also, as in these examples, a subordinator used in some relative clauses, and in these two cases 所 was rendered by adnominal verb forms, which subsequently passed into general language as lexicalized forms. Interestingly, a more literal way of rendering Chinese 所VERB之NOUN ('NOUN which VERBs'), instead of simply forming a normal Japanese relative clause, arose in *kanbun-kundoku* and was carried over into written language in the form VERB *tokoro-no* NOUN, where *tokoro-no* functions as a complementizer between the relative clause and the head noun, for example (16); see also 12.6.1.1.2. This usage is clearly motivated by the *kanbun-kundoku* rendition of 所 as *tokoro* 'place' and 之 as *no*, combined with the use of Chinese 所 in relative constructions. As with the quotative framing construction mentioned above, this construction is not ungrammatical in Japanese, but nor is it motivated internally within Japanese. In the Meiji period this usage was revived in *kanbun-kundoku*-like translations of relative pronouns in Dutch.

(16) tatekome-taru tokoro-no to
close-STAT.ADN door
'The door which had been closed' (*Taketori*)

9.2 *Ondoku*

Ondoku (音読 'pronunciation reading') is the reading and vocalization of Chinese text in Chinese, learned as a foreign language, without rendition or translation into Japanese. Over time a great many loanwords have entered Japanese based on this way of reading Chinese texts. Today somewhere between thirty-five and sixty per cent of words in running text, depending on genre, are SJ loanwords, and it is customary to speak of a distinct SJ vocabulary layer in the Japanese lexicon. The term 'Sino-Japanese' is ambiguous and that has given rise to several misunderstandings. There are three distinct, but interrelated issues, which are not usually distinguished explicitly: (a) Japano-Chinese: Chinese as a foreign reading language in Japan (9.2.1); (b) Sino-Japanese: nativized norms for pronouncing *kanji* (9.2.2); (c) Sino-Japanese loanwords: loanwords in Japanese deriving from J-Ch or SJ (9.2.3).

9.2.1 Chinese as a foreign (reading) language; Japano-Chinese

In the initial period of contact with Chinese text and language, the fifth and sixth centuries, this contact is thought to have been indirect and primarily to

have been mediated by immigrant or visiting scholars and later monks and nuns from the Korean peninsula, probably especially from the kingdom of Paekche, who taught and expounded Chinese texts: first Chinese classics, and later, after the introduction sometime in the sixth century of Buddhism into Japan, also Buddhist sutras and commentaries. This was done in part through *kanbun-kundoku*, but also through and in Chinese. An important element of studying texts in Chinese was the enunciation and vocalization of text, particularly in the recitation of Buddhist sutras. For that reason attention is often focused on the *pronunciation* of Chinese, hence the term *ondoku*. For lack of a more elegant term, we will refer to Chinese used in the study and reading of Chinese texts in Japan, that is to say, Chinese employed in *ondoku*, as 'Japano-Chinese', emphasizing that it was a variety of Chinese language.

It is often thought that the Chinese taught and learned in Japan in this early period was based on southern Chinese varieties, but this is not really known. It seems clear (a) that it does not reflect any single variety of Chinese, but is a cumulative and multi-layered conglomerate of varieties of Chinese, most likely transmitted by scholars and monks from the Korean peninsula in the main in the fifth through early seventh centuries; and (b) that it is not a single uniform norm, but exhibited variation between different schools and sects within which conventions for reading and reciting texts, especially sutras, became fixed, with orally transmitted pronunciation norms, which gradually became increasingly removed from Chinese spoken in China.

However, through the seventh and eighth centuries there was extensive direct contact with Tang China, with envoys and students dispatched to visit and study in China. In the course of their studies they acquired contemporary Chinese as spoken in the Tang capital Chang'an and brought this back with them to Japan, introducing new, competing pronunciation norms and readings of texts. This led to tension between the old and the new ways of pronouncing Chinese. Between 792 and 806 (during the reign of emperor Kanmu (737–806, *r.* 781–806)), several imperial decrees were issued that the proper Chinese pronunciation (漢音 Japanese *kan-on* 'Han (= Chinese) pronunciation' or 正音 *sei-on* 'correct pronunciation') be used, both in the study and reading of Chinese classics, which by then had become all important for civil service exams, and in official and public recitations of Buddhist sutras. The decrees used phrases such as 皆令讀漢音; 勿用呉音 'make everybody read in Chinese pronunciation; do not use Wu pronunciation' (cited from Yuzawa 1996: 47). The older established norms of J-Ch have variously been referred to as 和音 (*wa-on* 'Japanese pronunciation'), 対馬音 (*tsushima-on* 'Tsushima pronunciation' reflecting an early putative route of transmission of Chinese), or 呉音 (*go-on* 'Wu-pronunciation'), which is often interpreted as showing that early J-Ch was based on south-eastern Chinese from the region around present-day

Shanghai which was the seat of the ancient, barbarian Kingdom of Wu. We here adopt the most widely used terms, *go-on* for the older (composite) norm(s) and *kan-on* for the new norm.

The necessity of issuing decrees shows that the replacement of the older well-established pronunciation norms did not proceed without resistance, which was particularly strong from Buddhist sects to whom the familiar, well-established vocalization of sacred texts was understandably important. This may be compared to the widespread resistance against reformations of Bible translations (or even earlier against translating the Bible into vernacular languages) in Europe. Consequently, the older pronunciation norms were retained in many Buddhist sects and traditions and also used for some sutras introduced after the issue of the decrees. It is worth emphasizing that these decrees concerned J-Ch, that is to say, the norms for reading Chinese text(s); they did not concern the shape and pronunciation of SJ loanwords which at the time had already been taken into the language.

Kan-on was at first used both for study and reading of text and for practical communication. It was maintained both by instruction by native speakers, through regular intercourse with China (also after the official envoys to China ceased towards the end of the ninth century), and by study of Chinese rhyme tables and pronunciation guides. Efforts were made to maintain correct standards of pronunciation, and, as in China, the correct, normative 'reading' of individual characters became an independent object of study. As a result, *kan-on* in the course of the ninth and tenth centuries became a fossilized norm and over time as removed from Chinese spoken in China as the older *go-on* had been. Later, competing ways of pronouncing Chinese, representing different varieties of Chinese with different phonetics and phonological categories, came to Japan, especially in the context of Japanese Zen Buddhism from the late twelfth century onwards, said to be based on southern Chinese varieties, leading to the establishment of J-Ch pronunciation norms within some Zen sects which were different from the *kan-on*, referred to as *sō-on* ('Song pronunciation', 宋音, after the Song dynasty, 960–1279) or more commonly as *tō-on* ('Tang pronunciation', 唐音), which we use here. Note that *tō* 唐 was used in the sense of '(real) Chinese', like *kan* 漢 had been centuries earlier when the *kan-on* pronunciation was being promulgated as the real Chinese pronunciation, and that it does not refer to the Tang dynasty (618–907).

Thus, three different main varieties of J-Ch can be distinguished, defined in relation to *kan-on*: (a) pre-*kan-on* comprises the varieties collectively referred to as *go-on* and sometimes *wa-on*; (b) *kan-on* is the mainstream norm which arose out of J-Ch based on the Chinese of Chang'an in the late Tang period; finally, (c) post-*kan-on* are the subsequent *tō-on* pronunciation norms. Although we speak of three main varieties, there must have been considerable variation within them, depending on the traditions of individual schools and Buddhist sects.

9.2.1.1 The phonetics and phonology of Japano-Chinese

The several varieties of J-Ch have played important parts in the history of Japanese: each has contributed loanwords to Japanese, and *go-on* and *kan-on* formed part of the basis for determining the choice and sound values of *ongana* (1.1.2.5). In essence, however, J-Ch varieties are varieties of Chinese – even if they were functionally restricted – and the study of phonetics and phonology of J-Ch is best thought of as a branch of Chinese historical phonology, of less concrete importance for the study of Japanese.

Apart from reflexes in Japanese loanwords or in the sound values of *man'yōgana*, we know little about the phonetics of J-Ch. It is inevitable that some amount of phonological approximation to Japanese took place in the course of acquisition and over time as contact with the source language(s) was lost, and perhaps also as a result of wider popular participation in sutra recitation. It is, however, also clear that J-Ch sound systems were foreign, deriving and maintaining their prestige in part from being identifiably foreign in sound texture. An illustrative analogy might be the pronunciation of French taught in English schools (until recently, at least) which is elaborate, incorporates stereotypical and often exaggerated features of French phonetics, but also is characterized by a great amount of phonetic approximation to English; which in short sounds neither English nor French.

The most important material we have to study the phonetics and phonology of J-Ch are pronunciation glosses and glossaries for individual texts and dictionaries (see 6.2.3), which give instructions for the pronunciation of characters or words. However, much of this material is unpublished or not easily accessible, and in any case it is difficult to interpret, for (in addition to *fǎnqiè* glosses, see 6.1.4) it employs *man'yōgana* or *kana*. Generally the material is interpreted and presented in terms of the orthographic categories of the *kana* syllabary, which was used to record Japanese and whose orthographic categories reflect Japanese phonology. This makes it very difficult to acquire knowledge about differences between Japanese and J-Ch. A basic question is: Did the phonograms used in pronunciation glosses for J-Ch have the same reference (sound values) as when used in writing Japanese? This issue is further confounded by the fact that glosses in texts or dictionaries are often cited today as if they give information about the pronunciation of SJ loanwords or of individual characters used in SJ loanwords, rather than as pronunciation guides to J-Ch. In addition to glosses, living traditions of sutra recitation potentially hold important clues to the phonetics and phonology of earlier J-Ch, although they are said not to constitute unbroken traditions, but over time to have been subject to normative correction in terms of SJ *kanji* readings.

9.2.1.2 Early Middle Chinese, Japano-Chinese and ongana

By 'Early Middle Chinese' (EMC) we refer to the Chinese language reflected in the *Qièyùn* (切韻), a dictionary which by means of *fǎnqiè* glosses (cf. 6.1.4)

records the correct pronunciation of Chinese characters. It was compiled in the late sixth century by a group of scholars from different parts of China and published in 601 by Lu Fayan; although there are no extant copies, it is thought to be faithfully reflected in the eleventh-century dictionary *Guǎngyùn* (広韻). The pronunciations recorded in the *Qièyùn* seem to be an overdifferentiating compromise between the scholars involved, who were also not quite sober when they tried to agree on the pronunciation of individual characters. The phonology of the *Qièyùn* thus does not represent any one variety of early Chinese, although it is often taken to reflect the language of Chang'an. However, EMC is well reconstructed, and it therefore constitutes a convenient reference point for most of Chinese historical phonology, and also for studies of the sound values of *man'yōgana*. Even so, it is important to keep in mind that none of the varieties of J-Ch directly reflects EMC, as *kan-on* postdates EMC and *go-on* both predates EMC and may be based on a somewhat different variety of Chinese.

Ongana were written representations of OJ sounds. The *ongana* were to a large extent chosen to represent OJ sounds on the basis of a perceived *similarity* between the OJ sound and the J-Ch pronunciation associated with a *kanji*, although we saw above that use of some *ongana* must have been based on other criteria than their sound values in J-Ch, for example the use of 売 to represent OJ /mye/ (cf. 1.1.2.5). All of this shows that the relation between the reconstructed EMC sound values and and the sound values of the *ongana* was very indirect.

9.2.2 Sino-Japanese

As mentioned, the distinction between J-Ch and SJ is not a common one and SJ is usually equated with J-Ch, as if they were one and the same thing. However, they are fundamentally different: SJ is a nativization of J-Ch, removing it from the realm of a foreign language and providing a nativized pronunciation norm of *kanji*, which derives from J-Ch, but which in contrast to J-Ch is in full conformity with Japanese phonology and can be used within Japanese.

SJ is not a language, less so than J-Ch, but essentially a norm for pronouncing *kanji*. It may in many respects be likened to latinate English. SJ made available for easy vocalization in Japanese the full range of words from Chinese. SJ was used primarily in *ondoku*, that is reading out Chinese text without rendering it into Japanese, but it is not the case that SJ has replaced J-Ch. For example, whereas *ondoku* of most Chinese texts today is no longer done in J-Ch but in SJ, some schools of sutra recitation even today maintain the use of J-Ch, and that was much more so through the pre-modern period. As SJ derives from J-Ch, different layers of SJ resulted, reflecting the three main varieties of J-Ch, and we here use the same names for them, speaking of SJ *go-on*, SJ *kan-on* and SJ *tō-on*.

As the distinction between J-Ch and SJ is not usually made explicitly, it is not possible to say much in detail about the history and development of SJ. The common term for what we today call *ondoku* seems earlier to have been *kowe* (> late EMJ *koe*) 'voice', whereas *kundoku* was mostly referred to by *yomi* 'reading', although *kuni* or *kun* (訓) were also used. There are several references in both EMJ and LMJ texts to reading out of Chinese text or words, using the term *kowe* 'voice' > *koe*, but it is not clear whether they refer to J-Ch or to SJ. From early LMJ we find references to different types of *on* (音), variously named but reflecting the three layers of J-Ch and SJ.

It seems likely that nativizations of J-Ch to give SJ became established towards the very end of the EMJ period when we see an increase in the use of SJ vocabulary written in *kanji*, in the *kanji-kana majiribun* texts. The greater variability in sound shapes and even morphology of SJ loanwords until the second half of EMJ (see 9.2.3.2) indicates that loans from the first half of EMJ were taken in from J-Ch, but mostly from SJ thereafter, and this lends further support to this dating of the establishment of SJ. The existence of SJ was also a prerequisite for the occasional emergence of SJ words by using the SJ readings of *kanji* which were used logographically to write native words (see 9.2.4) and this is something we find from LMJ. For example, in early LMJ the written form 御前, which was used to write EMJ *omae* (< OJ *opo-mapye*), gave rise to *gozen* by using the SJ reading; and that would not have been possible without a SJ reading norm. At the end of LMJ, Rodrigues in *Arte da lingoa de Iapam* (see 10.2.2.2) makes reference to what are clearly three layers of SJ, 'coye', i.e. *koe*. He exemplifies different readings of *kanji*, citing 'govon' (*go-on*), 'canvon' (*kan-on*), and 'tŏin' (*tō-on*), of for example 行 (*guiŏ* /gyɔɔ/, *cŏ* /kɔɔ/, *an*) which since then have been used as the stock example in most exemplifications of the three different types of readings. His examples and explanations make clear that the notion of 'coye' first of all pertains to *ondoku*. He thus gives by way of illustration five different versions of the name of the Lotus Sutra (Sanskrit *Sad-dharma-puṇḍarīka-sūtram* 'true-law-lotus-sutra, i.e. sutra of the lotus of the true law/teaching', which in Chinese is rendered 妙法蓮華経) (*Arte*, p. 666, following Doi's emendations): in (17a) 'bongo' (Sanskrit; an adapted pronunciation which is not very far from the original Sanskrit)), (b) 'govon', (c) 'canvon', (d) 'tŏin', and (e) 'vago' (Japanese, i.e. a *kanbun-kundoku* rendition of the Chinese, here glossed in (18)). Thus, at the very end of LMJ there were three well-established layers of SJ, known by the same names we use today and used in different traditions of *ondoku*. Note, however, also that they were thought of as different languages, on a par with Sanskrit and Japanese; and finally that the SJ version of the name of the Lotus Sutra in use today (*Myōhōrenge-kyō*) is the one identified as *go-on*. It is interesting that *ikada* 'raft (/small boat)' is used to render Chinese 経, which translates Sanskrit *sūtra* 'thread, string, rule'.

(17) a. Satarama fundariquia sotaran.
b. Meô fô rengue qiuŏ (/myoo foo renge kyɔɔ/)
c. Beô fŏ renga quei. (/byoo fɔɔ renga kei/)
d. Beô fa renga quin. (/byoo fa renga kin/)
e. Taye naruya norino fachisuno fana icada.

(18)	tae	naru ya	nori no	fatisu no	fana ikada
	wonderful is	EMPH	law GEN	lotus GEN	flower raft

9.2.2.1 Present-day Sino-Japanese

Present-day dictionaries of *kanji* as used in Japan and in Japanese, so-called *kan-wa* (漢和) dictionaries, give for almost all characters listed in them one or in most cases several SJ readings, '*on*-readings', often described as 'Chinese readings', or readings which derive from Chinese. SJ readings are usually classified into four overall classes, which are defined in reference to the three main varieties of J-Ch: *go-on*, *kan-on*, *tō-on*, and *kan'yō-on* ('idiomatic readings' 慣用音, readings in common use, which do not conform to regular assignment to the other classes).

The SJ readings listed in present-day dictionaries are an intriguing conglomerate of prescriptive readings and of actual recorded usage, observed in SJ loanwords and in SJ reading traditions. The prescriptive SJ readings in present-day dictionaries were to a large extent established deductively by philologists (such as Moto'ori Norinaga) in the Edo period through study of the Chinese rhyme tables, in particular by application of the phonological categories of the *Yùnjìng* (韻鏡, Japanese *inkyō*). The *Yùnjìng* is thought to reflect the pronunciation of the second half of the Tang period (618–907), i.e. largely the Chinese on which *kan-on* J-Ch was based, and itself to date from the late ninth century or the first half of the tenth century. The *Yùnjìng* came to Japan in the thirteenth century and since then came to form an important basis for the study of *kanji* readings, first in J-Ch, but later also for the study and codification of SJ *kanji* readings. The effort in the Edo period aimed at establishing correct SJ reading and pronunciation norms, as well as norms for correct etymological spellings of the readings. Almost all characters are assigned a *kan-on* and most also a *go-on* reading which is constructed as a 'correct' pronunciation in correspondence with the categories in the *Yùnjìng*. These readings thus need never have been used in SJ or in a SJ loanword, but nor were they ever intended to provide a record of current readings.

On the other hand, the SJ readings listed in dictionaries also do include observed usage. While it would be easy enough to construct 'correct' *tō-on* readings for all *kanji*, *tō-on* readings included in dictionaries are generally those attested in SJ loanwords. And the *kan'yō-on* listed in dictionaries are by

Table 9.1. Kanji *readings*

	EMC	*Go-on*	*Kan-on*	*Tō-on*	*Kan'yō-on*
尺	tɕʰiajk	*syaku* (*shaku*)	*seki*		
行	*ɣaiŋ$^{(h)}$ /ɣɛ:jŋ$^{(h)}$	*gyaũ* (*gyō*)	*kaũ* (*kō*)	*an*	
杏	*ɣəiŋ' /ɣɛ:jŋ'	*gyaũ* (*gyō*)	*kaũ* (*kō*)	*an*	*kyaũ* (*kyō*)
灯	*təŋ	*toũ* (*tō*)	*toũ* (*tō*)	*ton*	
		tyaũ (*chō*)	*teĩ* (*tei*)	*tin* (*chin*)	
密	*mit	*miti* (*michi*)	*bitu* (*bitsu*)		
		mitu (*mitsu*)			
子	*tsi'/tsi'	*si* (*shi*)	*si* (*shi*)	*su*	
双	*ʂaiwŋ/ ʂœiwŋ	*soũ* (*sō*)	*saũ* (*sō*)		
請	*dziajŋ	*syaũ* (*shō*)	*seĩ* (*sei*)	*sin* (*shin*)	
		zyaũ (*jō*)			
熱	*ɲiat	*netu* (*netsu*)	*zetu* (*zetsu*)		
		neti (*nechi*)			
輸	*ɕuə̆	*su*	*syu* (*shu*)		*yu*
石	*dʑiajk	*zyaku* (*jaku*)	*seki*		*koku*
					syaku (*shaku*)
肉	*ɲuwk	*niku*	*ziku* (*jiku*)		
辱	*ɲuawk	*niku*	*zyoku* (*joku*)		
		noku			
描	*miaw	*meu* (*myō*)	*beu* (*byō*)		
明	*miajŋ	*myaũ* (*myō*)	*meĩ* (*mei*)	*min*	
立	*lip	*ripu* (*ryû*)	*ripu* (*ryû*)		*ritu* (*ritsu*)
篦	*pɛj	*pai* (*hai*)	*pei* (*hei*)		*pi* (*hi*)
争	*tʂəijŋ/ tʂɛ:jŋ	*syaũ* (*shō*)	*saũ* (*sō*)		
捜	*ʂuw	*syu* (*shu*)	*sou* (*sō*)		
挿	*tʂʰəip/ tʂʰɛ:p	*sepu* (*shō*)	*sapu* (*sō*)		
経	*kɛjŋ	*kyaũ* (*kyō*)	*keĩ* (*kei*)	*kin*	
京	*kiajŋ	*kyaũ* (*kyō*)	*keĩ* (*kei*)	*kin*	
匹	*pʰjit	*piti* (*hichi*)	*pitu* (*hitsu*)		*piki* (*hiki*)
打	*tajŋ'	*tyaũ* (*chō*)	*teĩ* (*tei*)	*da*	
徳	*tək	*toku*	*toku*		
常	*dʑiaŋ	*zyaũ* (*jō*)	*syaũ* (*shō*)		
実	*ʐit	*ziti* (*jichi*)	*situ* (*shitsu*)		*zitu* (*jitsu*)
英	*ʔiajŋ	*yaũ* (*yō*)	*eĩ* (*ei*)		
竹	*truwk	*tiku* (*chiku*)	*tiku* (*chiku*)	*situ* = /siQ-/ (*shitsu*)	

definition (part of) the sound shapes of SJ words which do not conform to the normative *go-on* or *kan-on*, and are not *tō-on*. It should finally be noted that not all SJ loanword shapes are captured by the SJ readings in dictionaries. That applies especially to SJ loanwords from the Heian period (see 9.2.3.2).

Table 9.1 gives some examples of SJ *kanji* readings classified in this way, giving for reference also the reconstructed EMC readings, and in 9.2.2.2–3 we

Table 9.2 *EMC syllable initial consonants*

EMC	*Go-on*	*Kan-on*
*m	m	b
*m(Vŋ)	m	m
*n	n	d
*n(Vŋ)	n	n
*b	b	p
*p	p	p
*d	d	t
*t	t	t
*g	g	k
*k	k	k
*z	z	s
*s	s	s

give a brief overview of some of the commonly observed correspondences between *go-on*, *kan-on* and EMC. As mentioned above, it must be emphasized that none of the layers of J-Ch or of the SJ character readings directly derives from EMC, but with this caveat in mind, we will say that SJ readings 'reflect' EMC. Although the phonological adaptation of SJ loanwords in the EMJ and LMJ periods was not as regular as these correspondences suggest, they do to a large extent also hold for SJ loanwords, and for that reason we give the SJ readings in the sound shape we reconstruct for EMJ on the basis of their historical spelling and the EMC sound value. This is strictly speaking anachronistic, as many readings date from NJ, but this will give an impression of the sound changes undergone by SJ loanwords.

9.2.2.2 Syllable initial consonants

EMC seems to have had thirty-eight different syllable initials (reconstructed by Pulleyblank as /p, p^h, b, m, w; ʈ, $ʈ^h$, dʐ, ʂ, ʐ; ts, ts^h, dz, s, z; tr, tr^h, dr, nr; t, t^h, d, n, l; tɕ, $tɕ^h$, dʑ, ɲ, ɕ, ʑ, j; k, k^h, g, ŋ, x, ɣ; ʔ/) which is a somewhat more complex system than Japanese. Only some are given here as representative. *Go-on* nasals correspond to *kan-on* mediae, except usually when followed by *-Vũ* or *-Vĩ* which are the reflexes of EMC *-Vŋ, and *go-on* mediae correspond to *kan-on* tenues. Regular sound changes have changed /p-/ > NJ /h-/ (7.3.1, 11.3, 14.3) and /d-/ to /z-/ before /i, u/ (14.1).

9.2.2.3 Syllable final consonants

EMC had six main syllable final consonants (/p, t, k, m, n, ŋ/) with fairly straightforward correspondences. *Go-on* and *kan-on* largely agree in their

Table 9.3 *EMC syllable final consonants*

EMC	*Go-on*	*Kan-on*
*-p	pu	pu
*-t	ti	tu (t)
*-k	ku / {e, a, o, u} __ ki / {i} __	ku / {i, a, o, u} __ ki / {e} __
*-m	N	N
*-n	N	N
*-ŋ	ũ / {i, a, o, u} __ ĩ / e __	ũ / {i, a, o, u} __ ĩ / e __

reflexes of syllable final consonants, except with EMC *-t, *-k. Regular sound change has changed /-pu/ > /-u/ (7.3.1.1) and /Ũ, Ĩ/ > /u, i/ (11.1.2). As shown in (Table 9.3), EMC *-t is usually said to be reflected in *go-on -ti* and *kan-on -tu*, but in fact, at the end of the LMJ period most *kan-on* and some *go-on* reflexes of EMC *-t were simply /-t/, which mostly became /tu/ (sometimes /ti/) in the NJ period. Syllable final /-t/ and the reflexes in SJ vocabulary of EMC *-t will be discussed further in 11.4; see also 7.2.

EMC *-ŋ was reflected differently, as /ũ/ or /ĩ/ depending on the preceding vowel. Combined with a difference between *go-on* and *kan-on* in the reflex of the vowels and glides in EMC *(C)iaj(ŋ) as *go-on* /ya/ and *kan-on* /e/ (realized with a phonetic onglide or strong palatalization of the preceding consonant: [ʲe] or [C$_j$e], cf. 2.3, 7.3.2.4), this has resulted in a frequent correspondence between *go-on* and *kan-on*: *go-on* EMJ *(C)yaũ* > LMJ *(C)yɔɔ* > NJ *(C)yoo* > cNJ *(C)yō* :: *kan-on* EMJ *(C)eĩ* > NJ *(C)ei*. For example: 英 *yō* :: *ei*, 生 *shō* :: *sei*, 京 *kyō* :: *kei*.

Many readings reflecting EMC *-p, *-t, *-k have variants in final /Q/, which are not usually listed as separate readings. They result from assimilatory processes in combinations with a following consonant: *-pu(-C-)* > *-Q(-C-)*; *-ti(-C-)/-t(-C-)* > *-Q(-C-)*; *-ki(-k-)/-ku(-k-)* > *-Q(-k-)*. Finally, there is a small group of *kanji* with readings where EMC *-p is reflected as *-tu* > cNJ *-tsu*, e.g. 立 *ritsu* (as in cNJ 独立 *dokuritsu* 'independence', cf. EMC *lip) or 雑 *zatsu* (複雑 *fukuzatsu* 'complicated', cf. EMC *dzəp/dzap).[3] These readings are thought to have arisen as back formations from assimilated forms such as those above, with the final *-Q* analogically interpreted as deriving from *-tu*. Like the *kan-on* reflexes of EMC *-t, these forms had final *-t* at the end of

[3] The relevant *kanji* are (giving in brackets the *kan'yō-on* and the EMC sound values): 立 (*kan'yō-on*: *ritsu*; EMC *lip), 雑 (*zatsu*; *dzəp/dzap), 接 (*setsu*; *tsiap), 摂 (*setsu*; *ɕiap), 執 (*shitsu*; *tɕip), 湿 (*shitsu*; *ɕip), 蟄 (*chitsu*; *drip), 颯 (*satu*; *səp/sap); see Komatsu (1956).

LMJ (11.4), suggesting that the back formation was from *-Q* to *-t* which later became *-tu*, that is: *-pu-(C-) > -Q-(C-) > -t > -tu*.

9.2.3 Sino-Japanese loanwords

A number of identifiable Chinese loanwords were used already in OJ (4.2.2). Some of them are undoubtedly very old, predating the study of Chinese text, and it is likely that some entered Japanese through the Korean peninsula. However, that early trickling of loanwords from Chinese is insignificant in comparison with the great number of loanwords which entered the language through the medium of J-Ch or SJ, by being used outside the context of reading Chinese text and eventually carried over into everyday language. It is such words we refer to as 'SJ loanwords'.

As mentioned above, the distinction between J-Ch and SJ is not usually made explicitly and it is therefore difficult to determine for many loanwords whether their proximate source is J-Ch or SJ. However, it is noteworthy that among SJ loanwords until the second half of EMJ, we see greater variability in and adaptation of sound shapes than later, and we also have small but noticeable numbers of examples of morphological adaptation where SJ loanwords become verbs or adjectives (9.2.3.2), whereas that is much rarer later. All of this indicates that loans from the first half of EMJ were taken in from J-Ch, whereas most loans from the end of EMJ onwards were taken in from the nativized SJ.

Through time many SJ loanwords must have entered the language through writing, but it is clear that the majority of SJ loanwords, especially in the first half of the EMJ period, had gained currency in everyday spoken language. However, it is equally clear that there was some consciousness about their provenance, for example because they were successfully avoided in poetry; that would not have been possible if they had been fully assimilated. As with loanwords from European languages centuries later, SJ loanwords had, and today to a large extent still have, a socio-linguistic status well captured by German *fremdwort* ('alien word'; cf. also 17.1.1). The intake of SJ loanwords was significant enough to affect phonology, especially phonotactics (7.2), and to some extent grammar, especially word classes (8.1.3).

9.2.3.1 Sino-Japanese loanwords in Old Japanese

OJ is thought not to have had many SJ loanwords. Very few are found in the OJ poetic texts: in the *Man'yōshū* a small number of SJ loanwords is thought to be found in a series of poems in volume 16 (3827–58), see (19). Many of them are logographically written, so it is not always clear that they in fact are SJ loanwords and not rather logographically written native words. It is on the other hand also possible that some of them are not SJ loanwords, but (Japano-)

Chinese words used for stylistic effect. In any case, the phonological shape of the logographically written words is guesswork. In the first poem in that series (*MYS* 16.3827) the *kanji* 一二三四五六 appear, but there is no agreement about whether they were intended to represent SJ or native Japanese numerals; 五 and 六 appear elsewhere in these poems, clearly representing SJ numerals and traditionally read as *go* and *roku*.

(19) 波羅門 ***baramoni*** 'Brahman' (EMC **pa la mən*, Skt. *Brāhmaṇa*). This is the traditional reading of the 門 in the single occurrence of this word in the *Man'yōshū*; in EMJ a more regular SJ version came to be used: *baramon*, which is the form the word has today.

檀越 ***daniwoti/danawoti*** 'benefactor' (EMC **dan wuat*, Skt. *dānapati* 'benefactor; giver; giving-lord'). The two versions are the traditional readings of the single occurrence in the *Man'yōshū*; in EMJ a more regular SJ, *dan.wotu*, came to be used, but the nativized version survived in a reduced form, *danna* 'benefactor', eventually giving NJ *danna* 'husband, master'.

五位 ***gowi*** 'fifth rank'.

香 ***kaũ*** 'fragrance' (in some reading traditions read *kori* 'fragrance' which is a native word or a naturalized loan).

功 ***kuũ*** 'accomplishment, merit'.

無何有(-乃-郷) ***mugau(-no-satwo)*** 'Mugō village; not-even-anything village, village of nothingness' (legendary place of natural emptiness, void of human artifacts; (both this and the following are from *Chuang-Tzu*, a Chinese Taoist classic (fourth or third century BC) which exerted great influence on later Chinese Buddhism);

藐姑射(-能-山) ***pakwoya(-no-yama)*** 'Mt. Hakoya' (legendary dwelling of sage hermits).

法師 and 僧 ***popusi*** '(Buddhist) monk, priest' (in the *Nihon shoki* 法師 is glossed as *popusi*).

力士(舞) ***rikizi(-mapi)*** 'strong-man(-dance)' Buddhist ritual dance.

旨菜 ***saukepu*** 'Saikachi' (name of a tree; some reading traditions have *pudi-no-kwi*, *kapara-pudi*).

生死 ***syaũzi*** 'living and dying' (in some traditons read *ikizini/ikisini* which is a loan translation of 生死, cf. 9.1.6).

塔 ***tapu*** 'stupa' (the Chinese is a loan from Skt. *stūpa*).

Outside that series of poems, only 過所 *kwaso* (15.3754 'travel pass') and 朝参 *teusan* (18.4121 'coming to court' (other readings include *mawiri*, *mikadwo-mawiri*, *miyade*)) are found in the *Man'yōshū*.

By contrast, the *Senmyō* had a significant number of loanwords from Chinese: Vovin 2005: 60–2 lists sixty-six SJ loanwords found in the *Senmyō*; they are all written logographically and we do not know their sound shape. Many of these words are fairly specialized terms (護法 'protect the Buddhist law', 菩薩 'bodhisattva', 職事 'office manager', 順孫 'obedient grandchild') and it is quite possible that some of these were J-Ch words used within Japanese, but others are more common words which probably were established loanwords (楽 'music', 謀反 'rebellion, treason', 経 'sutra', 斬 'beheading'). Taken together with the fact that SJ loanwords were normatively excluded from poetry in EMJ, this leads to a questioning of the usual assumption of absence to any significant extent of SJ loanwords in OJ: It is quite likely that their scarcity in the written sources from OJ reflects that this norm for poetry was already active in OJ, but that at least learned or official language had far more SJ loanwords than is usually thought.

9.2.3.2 Sino-Japanese loanwords in Early Middle Japanese

Also in the Japanese poetry from the EMJ (and LMJ) periods we find almost no SJ loanwords. As mentioned earlier, this absence from poetry confirms their sociolinguistic status as identifiable *fremdwörter*. In prose writing, however, the use of SJ loanwords can be seen to increase steadily through the period. The following word counts show the proportion of SJ loanwords in some well known sources (extracted from Tsukishima 1969: 588–9, 1987: 277–8).

(20) Text frequency of SJ loanwords (% of words in running text)

Ise monogatari	6.2%
Tosa nikki	4.0%
Kokin wakashū (*kana* preface)	11.3%
Genji monogatari	12.6%
Hamamatsu chūnagon monogatari	14.3%
Daijionji sanzōhōshi-den	43.0%

(21) Lexical frequency of SJ loanwords (% of different words)

Genji monogatari	4.8%
Konjaku monogatari	14.0%
Sangōshiiki-chū	>50.0%
Daijionji sanzōhōshi-den	85.8%

In the prose written mainly in *hiragana*, the proportion of SJ loans out of distinct words remains well under ten per cent through the period, but it is somewhat higher in *kanji-kana majiribun*, as for example in the *Konjaku monogatari*. This is related both to orthography and subject matter: First, as such texts include a high proportion of logographic writing it was easy to use SJ words or phrases written in *kanji*. Second, *kanji-kana majiribun* was often *setsuwa* literature about or inspired by Buddhism, and it included much Buddhist terminology and many Indian or Chinese proper names. Thus in the *Konjaku monogatari* the stories taking place in India and China which have a strong Buddhist element have a much higher proportion of SJ loanwords, whereas in the stories which take place in Japan, the proportion of SJ loanwords is more or less as in the *Genji monogatari*. It is not surprising that some *kanbun-kundoku* texts, such as the *Daijionji sanzōhōshi-den* annotations, contain a large proportion of words which were left untranslated as either SJ loanwords or J-Ch words.

There is, to be sure, much vocabulary relating to Buddhism and politics and philosophy among the SJ loanwords, in addition to words relating to the life, positions and ranks at the sinified court, conforming to the picture of the use of SJ or Chinese words in OJ. However, the variety and number of everyday vocabulary, including emotional and expressive vocabulary, is striking, showing that a significant number of SJ loanwords had become well integrated into everyday language, at least of the court nobility and educated classes. Another remarkable feature is that in terms of SJ loanwords the vocabulary composition of EMJ does not differ much from that of NJ, apart from frequencies of usage. Morphologically, SJ loanwords were invariably taken in as nouns, but in addition to general nouns (22), we already at the beginning of EMJ have a well populated class of verbal nouns used with the light verb *se-* 'do' (23), and a sizeable class of adjectival nouns/adverbs used with forms of the copula (*ni* infinitive, *no* adnominal, *to* infinitive, or extended *ni ar-* or, more rarely, *to ar-*), see (24).

(22) Nouns
sauyaku (草薬) 'a herbal medicine'; *daitoko* (大徳) 'monk of great virtue'; *inmyaūzi* (陰陽師) 'fortuneteller', *nikki* (日記) 'diary', *sipu* (集) 'anthology', *neti* (熱) 'fever'

(23) Verbal nouns
gu-se- (具) 'furnish, be furnished (with)'; *si-se-* (死) 'die'; *wen-ze-* (怨) 'resent'; *ron-ze-* (論) 'discuss'; *rongi-se-* (論議) 'debate'; *penge-se-* (変化) 'transform'; *kuyaũ-ze-* (供養) 'make offerings and prayers'; *buku-se-* (服) 'drink'; *kei-se-* (啓) 'speak HUM'

(24) Adjectival nouns/adverbs
gokuneti-no 'very hot' (極熱); *pizaũ-nari* 'extreme' (非常)
maũ-ni/no 'strong, fierce' (猛)
toni-ni 'swiftly, suddenly' (頓); *ziti-ni* 'really' (実); *beti-ni/no* 'particular(ly)' (別); *gen-ni* 'really' (現); *seti-ni* 'definitely' (切); *zinen-ni* (自然) 'spontaneously'; *sizen-ni* (自然) 'of, by itself'
rinrin-to 'very cold' (凛凛)

We also find compounds of SJ and native words, see (25), in some cases with some phonological integration, as well as native grammatical material used with SJ words, see (26). Borrowed SJ prefixes, such as *ko-* 'late, deceased' (故), *sai-* 'most' (最), combined with native lexical material. Also, derivational morphology was borrowed from SJ, e.g. *-yaũ* '-like' (様) which derives an adjectival noun from a noun (cf. 8.6). Finally, we see a few examples of SJ loanwords used as inflected adjectives, (27), or even verbs, (28); these are few in number and may to some extent be the result of literary playfulness and inventiveness, but they do conform to well-established native morphological patterns, such as reduplication to form *shiku* adjectives. All of this shows that SJ loanwords formed a well established and integrated part of the EMJ lexicon, even if still recognizable as *fremdwörter*. The examples of SJ loanwords listed in (25)–(27) are all from the first half of the EMJ period. Many are written in *hiragana* in the texts, but here we also give the source *kanji* for reference.

(25) Lexical compounds:
nama-zuryaũ 'governor with no real power' (*nama-* 'unripe, immature' + 受領 'governor'); *aigyaũ-duk-* 'be charming' (愛敬 'charm' + 'attach (v.intr.)'); *nana-moji* 'seven letters, characters' (*nana-* 'seven' + 文字 'letter, character'); *saũzi(n)-mono* 'vegetarian food' (精進 'devotion to Buddhism; abstention' + *mono* 'thing, stuff'); *bakuti* 'gambling; gambler' (< *baku-uti*; 博 'gamble' + *uti* 'striking'); *setimi* 'day/period for abstention (from eating meat) and devotion' (< *seti+imi*; 節 'time' + *imi* 'abstention')

(26) Early examples of native prefix or suffix with SJ lexical word:
mi-kesiki 'appearance' (*mi-* 'beautification' + 景色); *daiji-domo* 'great things' (大事 + *-domo* 'plural'); *taisyaũ-tati* 'generals' (大将 + *-tati* 'plural'); *sau-gati-* 'have much (grass style) cursive writing' (草 + *-gati* 'be frequent, likely to be/have' which derives an adjectival noun); *raũ-gawasi-* 'noisy, disorderly, ill-mannered' (乱 + *-gawasi* (< *-gapasi*) '-like' which derives an adjective)

Table 9.4 *Native Japanese and SJ numerals*

	Native	SJ
1	*pito*	*iti*
2	*puta*	*ni*
3	*mi*	*san*
4	*yo*	*si*
5	*i*	*go*
6	*mu*	*roku*
7	*nana*	*siti*
8	*ya*	*pati* (> NJ *hati*)
9	*kokono*	*kiu* (> NJ *kyuu*)
10	*towo, to, -so*	*zipu* (> NJ *zyuu*)
20	*pata*	
30	*miso*	
40	*yoso*	
50	*i*	
100	*momo, -po*	*pyaku* (> NJ *hyaku*)
1000	*ti*	*sen*
10,000	*yorodu*	*man*

(27) Adjectives

sipune- 'persistent, stubborn' (執念); *kotigoti-si-* 'unrefined' (骨骨); *raũraũ-zi-* 'refined, talented' (労労); *taidai-si-* 'inconvenient' (怠怠; sometimes a different etymology is suggested for this word, namely the OJ adjective *tagitagi-si-*); *zaezae-si-* 'of learned appearance' (才才)

(28) Verbs

saũzok- 'dress up' < *sauzoku* 'dress' (装束); *saisik-* 'colour, paint' < *saisiki* 'colouring' (彩色); *saudok-* 'make a fuss' < *saudo(u)* 'disturbance' (騒動), adding *-k-* to the SJ noun to create a verb stem; *sarugaw-* 'joke, jest' < *sarugau* < *saru-gaku* 'Sarugaku farce (the precursor of *kyōgen*)' (猿楽)

9.2.3.2.1 Numerals

The native system of numerals is simple and partly based on vowel alternations to show doubling: *pito* '1' ~ *puta* '2'; *mi* '3' ~ *mu* '6'; *yo* '4' ~ *ya* '8'. However, the system does not provide easily for formation of higher numbers. We saw above that some SJ numbers were used in OJ (9.2.3.1), but the intake of SJ numerals is usually thought not to have taken place until EMJ.

9.2.3.3 Sino-Japanese loanwords in Late Middle Japanese

During the LMJ period the use of SJ loanwords in the texts increased. This is probably in part related to the genres represented in the sources, including more *kanji-kana majiribun*, but the establishment of SJ was a major factor, making use of originally Chinese words more freely available in Japanese and thereby facilitating both intake and use of SJ loanwords. In *Esopo* from the end of the period (see 10.2.2), we find a sizeable proportion of SJ vocabulary, and also *Vocabulario* lists a large amount of SJ words. This shows that SJ words had become a well-integrated part of general vocabulary and language use by then.

In addition to the increased intake and use of loanwords taken in from *kan-on* and *go-on* SJ, a new layer of J-Ch came to Japan during the first half of the period, used especially in some Zen Buddhist sects. This is the *tō-on* variety of J-Ch (9.2.1), which also gave rise to loanwords, and also eventually to a *tō-on* SJ. Examples of such loanwords are *anzu* 'apricot' 杏 or 杏子; *andon* 'lantern' 行灯; *isu* 'chair' 椅子; *fusin* 'construction' 普請 (note that EMJ *p* > LMJ *f*, see 11.3). We did not above comment on correspondences between *kan-on/go-on* and the third, minor layer of SJ, *tō-on*, but a look at Table 9.1 shows a stereotypical feature of *tō-on*, the reflection of EMC *- ŋ as /N/, giving correspondences such as *kan-on/go-on* /ĩ, ũ/ (> /i, u/) :: *tō-on* /N/, as in 京 *keĩ/kyaũ* > *kei/kyau* > NJ *kei/kyoo* :: *kin*. As mentioned above, J-Ch *tō-on* is said to be based on southern Chinese varieties, but the loanwords taken in during the LMJ period also include words deriving from contact between Japanese fishers and traders with their continental colleagues. In that sense, some of the words characterized as *tō-on* are direct loans from Chinese, rather than SJ loanwords (which are based on J-Ch or SJ).

9.2.3.4 Sino-Japanese loanwords, Japano-Chinese and Sino-Japanese

It is clear from the examples of EMJ SJ loanwords given above that the shapes of some SJ loanwords are not captured in the SJ *kanji* readings, e.g. *daitoko* (大徳), *pizaũ* (非常) whose regular SJ readings are *daitoku* and *hizyoo* (< *pizyaũ*). This reflects first of all on the non-descriptive, rationalizing nature of the established SJ readings (9.2.2.1). However it also illustrates another point, viz. that OJ and EMJ SJ loans must be thought to have been taken into Japanese from J-Ch, not from SJ. In that connection, it should also be noted that most SJ loanwords from the EMJ period which were in everyday use derive from the pre-*kan-on* (that is, *go-on*) norm of J-Ch (and therefore often correspond to the SJ *go-on* readings), showing the persistence of the J-Ch *go-on* also after the decrees promoting the use of

kan-on. Also a number of everyday SJ loanwords still in common use today, such as *niku* (肉) 'meat', *netsu* (熱) 'fever', or *konnichi* (今日) 'today' are based on *go-on*, as are the SJ numerals, except *kiu* (> NJ *kyuu*) '9' which is *kan-on*, but *go-on ku* is also used for example in counting out the numbers.

Most SJ loanwords taken in during the LMJ period, on the other hand, are reading loans mainly based on SJ *go-on* and *kan-on*. The establishment of SJ made vocalization of originally Chinese words far more freely available in Japanese. This did not mean that any Chinese word could be used in Japanese, but it did mean that unfamiliar or unknown words did not sound alien and therefore they would more easily gain acceptance and currency. This also forms the background to the modernization of the Japanese lexicon during the Meiji period (17.3.2) which would not have been as smooth without the ready availability of SJ.

As we saw above (9.2.3.3) a number of SJ loanwords were taken in from the new *tō-on* variety of J-Ch which arose in the early LMJ period, and others were direct loans from Chinese. Nonetheless, since LMJ and into NJ, the major donor of SJ loanwords has not been Chinese or J-Ch, but SJ. Furthermore, many of the words originally taken in through J-Ch have since been reformed to conform to SJ pronunciation norms. Thus, *daitoko* and *hizoo* (< *pizaũ*) are no longer in use but have been replaced by *daitoku* and *hizyoo*. However, the so-called *kan'yō-on* to some extent reflect J-Ch, in the sense that they are current SJ readings which derive from J-Ch, but which do not conform to the normative SJ readings in dictionaries. For example, 動 has the following SJ readings listed in modern *kanji* dictionaries: *tō* (*kan-on*); *zū* (*go-on*); *dō* (*kan'yō-on*). Of these the only one used in the many SJ words whose writing includes this character is *dō*, that is, the *kan'yō-on*. In this case the *kan'yō-on* was also the reading used when new vocabulary was coined in the Meiji period (e.g. cNJ *jidōsha* 'automobile').

9.2.4 Sino-Japanese words arising through on*-reading of a* kun*-writing*

Some SJ words originate not in borrowing from J-Ch or SJ, or from coinage of elements borrowed from J-Ch or SJ, but rather from applying SJ readings ('*on*-readings') to character combinations which were originally used as a logographic representation ('*kun*-writing') of native words. Thus for example, NJ *kazi* (< *kwazi*) 'fire', originates in the *on*-reading of 火事, which was originally a *kun*-writing (i.e. a logographic representation), of *pi-no-koto* 'fire; fire-GEN-thing'. A few examples are given in (29).

(29)	Earlier word	*kun*-writing	*on*-reading; SJ word
	early EMJ *pi-no-koto* 'fire; fire-GEN-thing'	火事	*kazi* (late LMJ; < *kwazi*)
	early EMJ *mono-no-na(-no uta)* 'acrostic poem; thing-GEN-name(GEN-poem)'	物名(歌)	*butumei(ka)* (late LMJ)
	(OJ *opo-mapye* 'great-front' >) early EMJ *omape* (> *omawe* > late EMJ *omae*) used as a term of address and reinterpreted as 'RESPECT-front'	御前	*gozen* (early LMJ) term of address
	LMJ *toki-fakari* (*toqifacari* in *Vocabulario*) 'clock; time-measuring'	時計	*tokei* (NJ) (but cf. also *tokei* 土圭 'sun-dial')
	early EMJ *owas-* and *owasimas-* 'exist.RESP'	御座	*goza* (early LMJ), used in *goza-ar-* > NJ *gozar-* 'be.POL'

9.3 The case of 者

A single example illustrates some of the complexities involved in *kanbun-kundoku* and *ondoku* and their influence on Japanese. Chinese 者 (EMC *tɕia') is traditionally said to be used as a pronominal head of a relative clause, 'he who ..., the fact that' (often specifically 'the sort of person who ...'), but 者 has also been shown to be used to mark topics and conditionals (Harbsmeier 1981: 210–28).

Reflecting these uses in Chinese, 者 is in *kanbun-kundoku* rendered as:

(30) a. a noun particle *pa* 'topic';
b. the inflectional verb endings *-(a)ba* 'conditional' and *-(e)ba* 'provisional';
c. a noun *mono* 'person'.

Conversely, in writing Japanese, 者 was used logographically to write:

(31) a. *pa* 'topic';
b. *-(a)ba* conditional 'if' and *-(e)ba* provisional 'as, when';
c. *mono* 'person' (but not *mono* in the meaning 'thing' which is usually written 物).

These renditions in Japanese of Chinese 者 and the writings by 者 of the same Japanese forms in fact provide strong support for the analysis of Chinese 者 and more generally show how old Japanese *kun*-readings and -writings can reflect on and contribute to the understanding and analysis of Chinese.

In writing Old Japanese phonographically, 者 was used in the *MYS* as a *kungana* (although very rarely) for the syllable /pa/ (deriving from the *kun*-reading and -writing noted above) and as an *ongana* for the syllable /sa/ (on the basis of its sound value in J-Ch, ultimately deriving from the EMC sound value).

者 also at least from early LMJ gave the SJ suffix *-sya* which designates agent nouns '-person; -er, -or', as reflected in for example cNJ *isha* 'doctor' 医者, *ekisha* 'fortuneteller' 易者, *kagaisha* 'assailant' 加害者, *gakusha* 'scholar' 学者, *kisha* 'reporter' 記者.

It is also likely that the use of *mono* as a pronominal relative clause head functioning as a kind of topic marker in a cleft-like construction, as in (32) from *Makura no sōshi*, reflects influence on the use of *mono* from the use in Chinese of 者, brought about by the rendition of 者 as *mono*. This construction is used in many of the beginnings of the chapters in the *Makura no sōshi*.

(32)	susamazi-ki	**mono**	piru	poyuru	inu
	depressing-ACOP.ADN	MONO	daytime	howl.ADN	dog

'what is depressing: a dog howling in the daytime'

Perhaps finally also some uses of the provisional verb form were influenced by the use in Chinese of 者, mediated by the rendition of 者 by the provisional form, as in (33) (from the *Hōjōki*), where *oreba* (*or-* 'be') seems to mean 'those who were' rather than 'when, as they were'.

(33)	ie	no	naka	ni	ore**ba**	tatimati-ni
	house	GEN	inside	DAT	be.PROV	instantly
	pisige-na-mu		to	su		
	be.crushed-PERF-CONJ		COMP	do		

'those who were indoors would be crushed instantly'

REFERENCES

Kanbun-kundoku: King and Whitman forthcoming, Kōno 1969, Tsukishima 1977, 1987, Yoshida *et al.* 2001. Influence on Japanese from Chinese: Miller 1967: 245–6; Yamada 1935. *Ondoku*: Tsukishima 1993. *Kan-on* decrees: Yuzawa 1996: 46ff. SJ readings: Kamei 1954, Nakada and Hayashi 2000: 331ff., Numoto 1986, Tsukishima 1977: 23ff. On EMC *-p reflected as *-tu*: Komatsu 1956, Unger 1988, Vance 1986. SJ loanwords: in OJ: Vovin 2005: 59ff.; in EMJ: Vovin 2003: 25ff., Tsukishima 1969: 588ff, Yamada 1913/1952: 588ff.

Part III

Late Middle Japanese

The vast majority of sources in Japanese script from this period are written in *kanji-kana majiribun* (cf. 6.1.1). At the end of the period a number of texts in or about Japanese written in alphabet writing appear, produced by Christian missionaries, mainly Jesuits. Table 10.1 on p. 298 is a chronological list of important sources.

10.1 Early Late Middle Japanese

The sources from the first half of the LMJ period are in the main written in the increasingly fossilized classical norm which arose out of the language of the end of the EMJ period. This does not mean that innovations are not at all reflected in those sources, but they are not systematically represented. The personal *hiragana* literature (see 6.2.1) did not continue into the LMJ period. There is, however, a body of *gunki monogatari* (軍記物語 epic 'war tales'): *Heiji monogatari* (平治物語 'The tale of the Heiji war'), *Hōgen monogatari* (保元物語 'The tale of the Hōgen war'), *Heike monogatari* (平家物語 'Tale of the Heike'), *Genpei jōsuiki* (源平盛衰記 'An account of the Genpei wars'), *Soga monogatari* (曾我物語 'The tale of the Soga brothers'), *Masukagami* (増鏡 'The larger mirror'), *Taiheiki* (太平記 'The record of the great peace'). Two important texts in the essayistic *zuihitsu* style are the *Hōjōki* (方丈記 'An account of my hut') by Kamo no chōmei and *Tsurezuregusa* (徒然草 'Essays in idleness') by Yoshida Kenkō. The *setsuwa* literature which appeared in the EMJ period flourished in particular in early LMJ, with works such as the *Ujishūi monogatari* (宇治拾遺物語 'Stories gleaned at Uji'); *Hōbutsushū* (宝物集 'A collection of treasures'); *Jikkinshō* (十訓抄 'A treatise of ten rules'); *Shasekishū* (沙石集 'Sand and pebbles'). Often in the *setsuwa* literature, dialogue parts reflect (features of) the contemporary vernacular not reflected in the frame narrative. This is a pattern which is also found in the popular literature from the Edo period. Other important materials from early LMJ include the *Sangōshiiki-chū* (三教指帰注), a commentary on Kūkai's *Sangōshiiki* (三教指帰), written in *kanji-kana*

Table 10.1 *Important LMJ sources*

Thirteenth century
Hōjōki (1212)
Ujishūi monogatari (1190–1242)
Sangōshiiki-chū (*c.* 1220)
Gukanshō (1220)
Heiji monogatari (early thirteenth century)
Hōgen monogatari (early thirteenth century)
Heike monogatari (early thirteenth century)
Jikkinshō (1252)
Shasekishū (1283)
Izayoi nikki (second half of thirteenth century)
Fourteenth century
Genpei jōsuiki (early fourteenth century)
Tsurezuregusa (*c.* 1330)
Jinnō shōtōki (*c.* 1339–43)
Soga monogatari (mid fourteenth century)
Masukagami (*c.* 1370)
Taiheiki (*c.* 1372)
Fifteenth century
Setsuyōshū (mid fifteenth century)
Rongo-shō (before 1475)
Shūyaku-shō (1477)
Shiki-shō (1480)
Irop'a (1492)
Sixteenth century
Mōshi-shō (1516)
Mōgyū-shō (1534)
Shiganikkai (1534)
Feiqe monogatari (1593)
Esopono fabulas (1593)
Dictionarium Latino–Lusitanicum ac Japonicum (1595)
Rakuyōshū (1598)
Vocabulario da lingoa de Iapam (1603–4)
Arte da lingoa de Iapam (1604–8)
Arte breve da lingoa Iapoa (1620)
Ars grammaticae iaponicae linguae (1632)
Dictionarium sive thesauri linguæ Iaponicæ compendium (1632)

majiribun, the *Gukanshō* (愚管抄 'The future and the past') by Jien, the *Izayoi nikki* (十六夜日記 'The diary of the waning moon') a travel diary from the second half of the thirteenth century by Abutsu, and the *Jinnō shōtōki* (神皇正統記 'An account of our divine sovereigns and true royal line'), by Kitabatake Chikafusa.

10.2 Late Late Middle Japanese

Two sets of sources from the LMJ period are of particular importance: the *shōmono* from the middle of the fifteenth to the middle of the sixteenth century, and the Christian sources from the end of the sixteenth and the beginning of the seventeenth century. The Korean *Irop'a* (1492), a basic textbook of Japanese for Korean officials, is another important external source which contributes to dating of some phonological changes (see e.g. 11.6).

10.2.1 Shōmono

From the second half of the fifteenth century, the so-called *shōmono* (抄物) appear. These are commentaries, or lecture notes, on Chinese classics and Buddhist scriptures. Important texts include the *Rongo-shō* (論語抄), *Shiki-shō* (史記抄), *Shūyaku-shō* (周易抄), *Mōgyū-shō* (蒙求抄), *Shiganikkai* (四河入海), and *Mōshi-shō* (孟子抄). In these texts we find language which has changed greatly since the end of the EMJ period and which is very different from the language reflected in the sources from the first half of the LMJ period. It is in most respects very similar to the language reflected in the Christian sources from the very end of the period, showing that most of the changes reflected in the Christian sources had taken place by the middle of the sixteenth century. A stereotypical feature of the didactic style in the *shōmono* is the very frequent use of the emphatic, or perhaps rather admonishing or impressing, sentence final particle *zo*.

10.2.2 Christian sources

Around the middle of the sixteenth century the first European we know about reached Japan, namely a Portuguese cast-away who landed in 1543 on the island of Tanegashima, south of Kyushu. Soon thereafter, in 1549, Jesuit missionaries arrived in Japan, led by Francis Xavier. Between this time and 1639 when all Europeans with the exception of the Dutch were expelled from Japan – a policy of national seclusion known as *sakoku* (鎖国) 'closed country' that lasted until the 1850s – the Jesuits carried out missionary work with great zeal. For that purpose they produced and published a large amount of material, including texts, grammars and dictionaries, especially after they brought a printing press to Japan in 1591. They acquired an impressive competence in and knowledge about the Japanese language and collaborated closely with Japanese converts. The works produced by the missionaries and their Japanese collaborators are collectively referred to as *kirishitan shiryō* in Japanese ('Christian materials') while books produced on their movable-type press are referred to in the history of Japanese printing as *kirishitan ban*

('Christian imprints'); we adopt a similar usage here and refer to the 'Christian sources'.

The work of the Jesuit missionaries constitutes one of the greatest feats on record of linguistic and cultural penetration of an unknown and alien culture. Within a short span of time the missionaries learnt both the contemporary vernacular and the classical written language, and also familiarized themselves with classical literature and scholarship. They published a number of texts, both in alphabet writing and in Japanese script. Some texts were in a simple form of the classical written language, sometimes even employing an archaizing *kanbun-kundoku*-like style, such as the opening of the *Contemptus mundi* (a translation published in alphabet writing in 1596 of Thomas à Kempis' *Imitatio Christi* (1418)), which uses the OJ nominal form *notamauaqu* /notamawaku/ to introduce a saying by Jesus, in the same way that sayings by Confucius were introduced in *kanbun-kundoku* renditions of the Analects (*si iwaku* or *si notamawaku*, cf. 9.1.7). Both the original spelling and a phonemic transcription are given in (1):

(1) Von arujino notamauaqu: QUI SEQITUR ME, NON AMBULAT IN TENEBRIS, SED HABEBIT LUMEN VITÆ. Ioan. 8. Vareuo xitŏ monoua yamigiuo yucazu: tada jumiŏ no ficariuo motçu bexi to nari.

on-aruzi	no		notamawaku: ...			ware	o	sitɔɔ
HON-lord	GEN		say.NMNL:			I	ACC	follow
mono	wa	yamidi	o	yukazu:				
person	TOP	dark.road	ACC	walk-NEG				
tada	zyumyɔɔ	no				fikari	o	motu
but	life	GEN				light	ACC	hold
besi	to	nari						
NEC	COMP	COP						

'The Lord says: 'Whoever follows me will never walk in darkness but will have the light of life.'

The texts written in the contemporary vernacular are valuable sources for the language at the end of the LMJ period. The missionaries also wrote and compiled grammars and dictionaries which included information about sociolinguistic and regional differences. Their descriptions and analyses of Japanese are incisive and sophisticated, and still of practical use as well as of scholarly interest. Unfortunately their work remained largely unknown in Europe until the late nineteenth century. Had that not been the case, western Japanology would most likely have progressed significantly beyond its current stage.

The work on Japanese done by the missionaries was a collaborative effort which for the most part did not identify the individuals who took part in it.

However, two figures may be singled out for mention: The first is the Portuguese João Rodrigues (1561–?1633/34) who came to Japan in 1577 at the age of 16 and entered the Jesuit order in 1596. He was expelled from Japan by the shogunate in 1610 and spent the rest of his life in Macao. He is responsible for two of the most important of the Christian materials, the *Vocabulario da lingoa de Iapam* and the *Arte da lingoa de Iapam*, published in Nagasaki between 1603 and 1608. The second is the Spanish Dominican Diego Collado (d. 1638) who came to Japan in 1619, but only stayed until 1622. His *Ars grammaticae iaponicae linguae* and *Dictionarium sive thesauri linguæ Iaponicæ compendium* were published in 1632 in Rome. To some extent they incorporated Rodrigues's work.

The material is eminently accessible for study: It is available in excellent reproductions and annotated editions, much of it in recent years also in electronic form, and the grammars and dictionaries have been published in annotated translations into modern Japanese and other languages. Issues of text tradition do not arise as these are printed texts, and the material presents no particular philological difficulties.

The Christian sources have been extensively studied, both by historians and by linguists. The most extensive use of the materials has been in the areas of phonology and lexicology. The contemporary phonology, including main allophonic variation, has been well and exhaustively described on the basis of these materials, and the dictionaries form an extremely important source of the contemporary lexicon, with good exemplification of usage and with word definitions in other languages (Latin and Portuguese). The Christian sources have also been studied as part of the history of the study of Japanese, and also as part of the discipline of 'missionary linguistics'. Particular attention has been paid to the descriptive approach of applying traditional Latin grammar to the description of a typologically quite different language. Although these materials have been studied intensely, they hold a wealth of unmined information and important data.

10.2.2.1 Texts

The two most valuable texts as linguistic source material are the *Feiqe monogatari* (*The tale of the Heike*), an abbreviated translation or retelling of the popular classic originally from the thirteenth century, and the *Esopono fabulas* (*Aesop's fables*), a translation into Japanese from a Latin version of Maximus Planudes' introduction to and version of the fables – incidentally the first piece of 'western' literature to be translated into Japanese. Both are written in the contemporary vernacular. *Feiqe* seems to retain some older features of the language not found in *Esopo*, which may be the result of influence on *Feiqe* from the original, either unintentional or deliberate retention of classical features. *Feiqe* is, however, also a single text with longer narrative passages

and more varied language represented, as opposed to the short fables in *Esopo*, and linguistic differences between *Feiqe* and *Esopo* can also be taken to reflect such differences.

Texts of Christian doctrine and devotional literature published by the missionary press were written in a simple form of classical Japanese which presumably lent the texts more authority. They include *Contemptus mundi* (1596, cf. 10.2.2), *Doctrina Christan* (1600) and *Giya do pekadoru* (1599), an abridged translation into Japanese written in Japanese script of Fray Luis de Granada's *Guía de pecadores* ('Sinners' guide' from 1555) which became popular and was widely read also after the expulsion of the Christians from Japan. It is the first published text in Japanese to employ the *handakuten* (6.1.2.3) – after its appearance in the *Rakuyōshū* dictionary (10.2.2.3) the previous year – in order to spell /p/ in loanwords.

10.2.2.2 Grammars

Two grammars, of which unfortunately no known copies survive, are reported to have been produced shortly after the arrival of the Jesuits: *Arte da lingoa Japonesa* (1551, by D. da Silva) and *Grammatica da lingoa Japonesa* (1564, by J. Fernandez). However, the *Arte da lingoa de Iapam* is a monumental grammar and manual of use, edited or directed by Rodrigues and published in Portuguese in 1604–8 in three volumes over 239 folios, covering most aspects of the language to a very high standard; it focuses on the vernacular, but also addresses the classical written language. In terms of scope and quality it was probably not superseded until Samuel E. Martin's *A reference grammar of Japanese* (1975), but regrettably it was largely forgotten soon after the closure of Japan to foreigners and was only rediscovered and made an object of study in the late nineteenth century. Thus, it never had much impact on Japanology in the west. Its translation into modern Japanese by Doi Tadao published in 1955 makes the *Arte* easily available; note that references to *Arte* in this book are to Doi's translation which is clearly indexed with reference to the original. In Japanese it is usually referred to as *Nihon (dai)bunten* (日本(大)文典). In 1620 a revised, abbreviated one-volume edition, *Arte breve da lingoa Iapoa* (*Nihon shōbunten* 日本小文典) was published. Finally, *Ars grammaticae iaponicae linguae* is a fine grammar of Japanese written in Latin by Collado and published in 1632 in Rome; it is available in a translation into English (Spear 1975).

10.2.2.3 Dictionaries

The Christian missionaries published four dictionaries: *Dictionarium Latino–Lusitanicum ac Japonicum* (1595) is a Latin–Portuguese–Japanese dictionary, which was based on Calepino's dictionary of Latin which had come to serve as a template for making dictionaries since its publication in 1502; it is not

known who directed the compilation of this dictionary. The *Dictionarium sive thesauri linguæ Iaponicæ compendium* (1632) is a Latin–Spanish–Japanese dictionary, compiled by, or under the direction of, Collado. The *Rakuyōshū* (1598; 落葉集) is a dictionary of more than 2,200 *kanji*, organized in three parts, by *on*-readings and *kun*-readings (in *Iroha* order, cf. 6.1.2.4), and character shape. It is in this dictionary that the *handakuten* (6.1.2.3) was first used, in order to reflect /p/ in SJ vocabulary.

By far the most important of the dictionaries is the *Vocabulario da lingoa de Iapam* (1603–4, henceforth '*Vocabulario*'), a Japanese–Portuguese dictionary compiled under the direction of Rodrigues. It contains over 32,000 words in a main dictionary and a supplement. It is the most comprehensive dictionary of Japanese compiled before the end of the nineteenth century and the major source of information about premodern Japanese, in particular LMJ, vocabulary. It is a sophisticated dictionary, with both word definitions and extensive examples of usage; it includes and identifies dialectal, archaic, literary, poetic, Buddhist and vulgar words, as well as words particular to women's or children's language. Verbs are listed under their infinitive, followed by the nonpast and past in abbreviated form, as exemplified in (2), thus giving the essential lexical information about verbs needed to identify their conjugational class and to form other inflected forms. *Vocabulario* is available in a translation into modern Japanese (Doi *et al.* 1980) and is in Japanese known as the *Nippo jisho* (日葡辞書).

(2)	'open'	*aqe, uru, eta*	*(= ake, akuru, aketa)*
	'emerge'	*ide, zzuru, eta*	*(= ide, iduru, ideta)*
	'write'	*caqi, u, aita*	*(= kaki, kaku, kaita)*
	'call'	*yobi, u, ôda*	*(= yobi, yobu, yooda)*

REFERENCES

Shōmono: Yuzawa 1929/1970. Korean materials: Hamada 1970. Christian materials: Doi *et al.* 1980, Sugahara 2000, Sugimoto 1989.

11 Phonology

The Christian sources in alphabet writing provide extremely valuable information about the phonology at the end of the LMJ period, including a number of features of pronunciation which were not reflected in writing in Japanese script. In addition to the system as it may be inferred from the transcription in the texts and dictionaries, the grammars (Rodrigues's *Arte* and *Arte breve* and Collado's *Ars grammaticae*) have notes on pronunciation. Table 11.1 shows the sound inventory at the end of the period and includes also the transcription used in *Vocabulario* (the transcriptions used in the various Christian sources differ on points of detail, but they reflect the same phonological system). Unless otherwise noted, LMJ forms cited in this section are from the Christian sources, mostly *Vocabulario*, in the transcription used therein followed by a phonemic transcription if that is significantly different, e.g. *chôzzu* /tyoodu/ 'washing'.

Table 11.1 shows both free moras (/CV, CyV, CwV/) and long syllables with a long vowel, which in the Christian sources were noted by a single vowel with a diacritic (/CVV, CyVV, CwVV/); see (11.5) about the phonemic analysis of long vowels. Table 11.1 does not include long syllables whose second mora was a consonant or a high front vowel (which formed a long diphthong with the preceding nuclear vowel), as these were noted by individual segments in the Christian sources. The inventory of bound moraic phonemes is shown in (1); see in particular 11.1.2 about changes from the EMJ to the LMJ system of bound moraic segments.

(1) Bound moraic phonemes

Vowels	/I, V (or U)/
Consonants	/Q, N, t/

The sound changes which took place during the LMJ period are not many or complicated. In addition to outlining those changes, we will note features of pronunciation reflected in the Christian sources. In several cases they are the only indication of a dating of features we still find in NJ, or a concrete indication of features we posit for earlier stages of the language.

Table 11.1 *LMJ sound inventory*

Short CV, CyV and Cwa syllables

/a/ *a*	/ka/ *ca*	/sa/ *sa*	/ta/ *ta*	/na/ *na*	/pa/ *pa*	/fa/ *fa*	/ma/ *ma*	/ya/ *ya*	/ra/ *ra*	/wa/ *va, ua*
/i/ *i, j, y*	/ki/ *qi, qui*	/si/ *xi*	/ti/ *chi*	/ni/ *ni*	/pi/ *pi*	/fi/ *fi*	/mi/ *mi*		/ri/ *ri*	
/u/ *v, u*	/ku/ *cu, qu*	/su/ *su*	/tu/ *tçu*	/nu/ *nu*	/pu/ *pu*	/fu/ *fu*	/mu/ *mu*	/yu/ *yu*	/ru/ *ru*	
/e/ *ye*	/ke/ *qe, que*	/se/ *xe*	/te/ *te*	/ne/ *ne*	/pe/ *pe*	/fe/ *fe*	/me/ *me*		/re/ *re*	
/o/ *vo, uo*	/ko/ *co*	/so/ *so*	/to/ *to*	/no/ *no*	/po/ *po*	/fo/ *fo*	/mo/ *mo*	/yo/ *yo*	/ro/ *ro*	
	/ga/ *ga*	/za/ *za*	/da/ *da*		/ba/ *ba*					
	/gi/ *gui*	/zi/ *ji*	/di/ *gi*		/bi/ *bi*					
	/gu/ *gu, gv*	/zu/ *zu*	/du/ *zzu*		/bu/ *bu*					
	/ge/ *gue*	/ze/ *je*	/de/ *de*		/be/ *be*					
	/go/ *go*	/zo/ *zo*	/do/ *do*		/bo/ *bo*					
	/kya/ *qia*	/sya/ *xa*	/tya/ *cha*	/nya/ *nha*	/pya/ *pia*	/fya/ *fia*	/mya/ *mia, mea*		/rya/ *ria, rea*	
		/syu/ *xu*								
	/kyo/ *qio, qeo*	/syo/ *xo*	/tyo/ *cho*	/nyo/ *nho*		/fyo/ *fio*			/ryo/ *rio, reo*	
/gya/ *guia*	/zya/ *ja*	/dya/ *gia*		/bya/ *bia*						
	/zyu/ *ju*									
/gyo/ *guio*	/zyo/ *jo*	/dyo/ *gio*								
	/kwa/ *qua*									
	/gwa/ *gua*									

Table 11.1 (*cont.*)

Long CVV, CyVV and kwɔɔ syllables									
	/kuu/ *cŭ, cû*	/suu/ *sŭ, sû*	/tuu/ *tçŭ, tçû*	/nuu/ *nŭ, nû*	/puu/ *pŭ, pû*	/fuu/ *fŭ, fû*	/muu/ *mŭ*	/yuu/ *yŭ, yû*	/ruu/ *rŭ*
/oo/ *vô, uô*	/koo/ *cô*	/soo/ *sô*	/too/ *tô*	/noo/ *nô*	/poo/ *pô*	/foo/ *fô*	/moo/ *mô*	/yoo/ *yô*	/roo/ *rô*
/ɔɔ/ *vŏ, uŏ*	/kɔɔ/ *cŏ*	/sɔɔ/ *sŏ*	/tɔɔ/ *tŏ*	/nɔɔ/ *nŏ*	/pɔɔ/ *pŏ*	/fɔɔ/ *fŏ*	/mɔɔ/ *mŏ*	/yɔɔ/ *yŏ*	/rɔɔ/ *rŏ*
	/guu/ *gŭ, gû*	/zuu/ *zŭ*	/duu/ *zzŭ*						
	/goo/ *gô*	/zoo/ *zô*	/doo/ *dô*		/boo/ *bô*				
	/gɔɔ/ *gŏ*	/zɔɔ/ *zŏ*	/dɔɔ/ *dŏ*		/bɔɔ/ *bŏ*				
	/kyuu/ *qiŭ, qiû*	/syuu/ *xŭ, xû, xiŭ*	/tyuu/ *chŭ, chû*	/nyuu/ *nhŭ, nhŭ niŭ, niû*		/fyuu/ *fiŭ*			/ryuu/ *riŭ, riû*
	/kyoo/ *qeô, qiô*	/syoo/ *xô*	/tyoo/ *chô, teô*	/nyoo/ *nhô, neô*	/pyoo/ *peô, piô*	/fyoo/ *feô, fiô*	/myoo/ *meô, miô*		/ryoo/ *reô, riô*
	/kyɔɔ/ *qiŏ, qeŏ*	/syɔɔ/ *xŏ*	/tyɔɔ/ *chŏ*		/pyɔɔ/ *piŏ*	/fyɔɔ/ *fiŏ, feŏ*	/myɔɔ/ *miŏ, meŏ*		/ryɔɔ/ *riŏ, reŏ*
	/gyuu/ *guiŭ, guiû*	/zyuu/ *jŭ, jû*	/dyuu/ *giŭ, giû*		/byuu/ *biŭ*				
	/gyoo/ *gueô, guiô*	/zyoo/ *jô*	/dyoo/ *giô*		/byoo/ *beô, biô*				
	/gyɔɔ/ *guiŏ, gueŏ*	/zyɔɔ/ *jŏ*	/dyɔɔ/ *giŏ*		/byɔɔ/ *biŏ, beŏ*				
	/kwɔɔ/ *quŏ*								

11.1 Nasality and medial voicing

In the course of LMJ changes took place to two interrelated phonetic features: *medial voicing* of non-initial tenues and *prenasalization* of the mediae were both lost. These two allophonic phonetic features were prominent elements in the sound texture of OJ and EMJ (2.2.2, 7.1.2.4). They are no longer found in

most varieties of NJ, but are retained in some dialects, and a reflex of prenasalization is found in the nasal allophone [ŋ] of intervocalic /g/ in NJ which until recently was a regular phonetic feature, but today a normative feature of elegant diction, though not spontaneously productive for most speakers. There is no positive evidence for medial voicing in the Christian sources and judging from the fact that it is not reflected in transcription or mentioned in the notes to pronunciation in the grammars, it seems that medial voicing was lost from the language by the end of the LMJ period. It also appears that the loss of prenasalization of mediae was well under way by the end of the LMJ period. *Arte* and *Ars gr.* note that vowels are nasalized before /d, g/ and sometimes /b/ (*Arte*, p. 637) and /b, g/ (*Ars gr.*, p. 5); this reflects prenasalization of these phonemes, perceived mainly as nasalization of the preceding vowel, but also indicates that prenasalization was in the process of being lost.

It was through the loss of medial voicing and eventually early in NJ of prenasalization and nasality harmony that the sound texture of Japanese changed from its premodern to its modern form. This is well illustrated by (2), repeated from 2.2.2 above, which shows how different are the OJ, EMJ, LMJ versus NJ phonetics of *tanabata* which did not change phonemically between OJ and NJ.

(2) /tanabata/
OJ [tãnãᵐbada]
NJ [tʰanabatʰa]

Loss of medial voicing and of prenasalization could well be said to be the two major phonetic changes in the history of Japanese. Apart from the direct effect on the sound texture of Japanese, loss of medial voicing was closely related to two phonological changes concerning nasality: the loss of postnasal neutralization (11.1.1) and the ensuing changes regarding phonemic nasality among bound moraic phonemes (11.1.2). Recall also that before it was lost, medial voicing was also directly involved in both *onbin* (7.1.4) and in the change of intervocalic /p/ to /w/ (7.3.1.1). Finally, it is possible that the loss of medial voicing was accompanied by the introduction of aspiration of tenues, which must have been an important factor in the change of /p/ to /f/ (11.3) and possibly also in the affrication of /t/ before /i, u/ (11.6).

11.1.1 Postnasal neutralization

The rule of postnasal neutralization (7.1.2.2) ceased to apply as an automatic phonological rule during LMJ. This is shown clearly by the occurrence of /N/ before all consonants, giving minimal pairs such as *anxin* /ansin/ 安身 'being safe' versus *anjin* /anzin/ 安心 'spiritual peace'. *Anjin* is a Buddhist term,

which must have been lexicalized and univerbated (i.e. /anzin/ rather than /an/+/sin/), before the loss of postnasal neutralization, whereas /ansin/ 安身 'being safe' must have been lexicalized and univerbated after the loss of postnasal neutralization. This is also the case for 安心 used in the general sense of 'relief' in the shape /ansin/, also reflecting lexicalization after the loss of postnasal neutralization. It is likely that it was the loss of medial voicing of tenues which resulted in the loss of postnasal neutralization.

Although lost as an automatic phonological rule, postnasal neutralization was reflected both in morphophonological rules applying in the formation of some inflectional forms (see 12.3.2 below) and in many words which were lexicalized while the rule was still active. In addition to the few forms mentioned in 7.1.2.2, some examples are given in (3). In (3b) are shown a few examples which have variant shapes, reflecting early and late lexicalization, or later (reading) reformations, like 安心 /anzin/ versus /ansin/.

(3) a. *andon* 行灯 'candle' < *an* + *ton*
caronzuru 'takes lightly' < *karon* + *suru*
conjiqi /konziki/ 金色 'golden colour' < *kon* + *siki*
xŏjo /syɔozyo/ 生所 'place of birth' < *syaũ* + *syo*
sanbô /sanboo/ 三宝 'three treasures' < *san* + *pou*
nanbocu /nanboku/ 南北 'south and north' < *nan* + *poku*.

b. *vonju* /onzyu/ 飲酒 'drinking as a (Buddhist) sin' < *on* + *syu*; the shape *vonxu* /onsyu/ was also used, reflecting later lexicalization; contrast NJ *inshu* 'drinking (in general)'
sendacu 洗濯 'washing' < *sen* + *taku*; today reformed to *sentaku*
dôjucu /doozyuku/ 同宿 'young men serving priests in temples' < *doũ* + *syuku*; contrast NJ *doosyuku* 'cohabitation'
nindẽ /ninden/ 人天 'human and heavenly beings' < *nin* + *ten*; contrast *zinten* (which is *kan-on*)

11.1.1.1 'Vmu no xita nigoru'

There is clear evidence, in the form of an explicit pronunciation rule, that postnasal neutralization and its lexicalized effects were part of metalinguistic consciousness. The rule is mentioned in at least two *shōmono*, *Nindengenmokushō* (人天眼目抄, *c.* 1471–3) and *Gyokujin* (玉塵, 1563), as a rule for reading out, but it is also clearly explained by Rodrigues in *Arte*: in a section on the 'use of "*sumi*" [tenues] and "*nigori*" [mediae]', Rodrigues gives a rule whereby tenues are pronounced as mediae after 'ŏ, ô, ŭ, n', that is to say, after /ɔo, oo, uu, N/, giving examples such as *xŏji* /syɔozi/ 'living' (< *syaũ* (生)+ *si* 'do.INF'), saying that '*ji*' /zi/ originally is '*xi*' /si/. After intro-

ducing the rule (and noting that it has exceptions) Rodrigues goes on to say 'ao que elles dizem, *Vmu no xita nigoru*' (*Arte*, p. 633), 'which they [the Japanese] refer to as "Vmu no xita nigoru" [pronounce as a media below, i.e. after, *Vmu*]'. As Rodrigues explains, '*Vmu*' refers to the *kana* letter for *u* (う), which is used in the spelling in Japanese script of the long vowels *ŏ*, *ô*, *ŭ* (/ɔɔ, oo, uu/), and the *kana* letter for *mu*, which is often used for *n* (/N/) (cf. 6.1.2.5.1). Thus, Rodrigues does not describe his own observation, but reports a named rule, *vmu no xita nigoru*. Note that *u* was noticeably nasalized [ũ] before *m* (as for example explicitly described in *Arte*, p. 639) and that accordingly *Vmu* may be a bit more subtle than immediately apparent, referring, or having originally referred, specifically to [ũ] (at some earlier point to /ũ/) and not simply to the *kana* letter for *u*.

11.1.2 Redistribution of phonemic nasality

Among the moraic phonemes, the distribution of nasality changed from the EMJ system, where nasality was distinctive in morpheme final position for both consonants and vowels, but not distinctive in morpheme nonfinal position (7.1.2), to a system where nasality was distinctive in the moraic consonants, but not in the vowels, regardless of position in the word, see (4), that is much like NJ. In NJ word final /N/ is pronounced as an unreleased uvular or velar nasal, [ɴ] or [ŋ], but in LMJ it was an alveo-dental [n], as shown by *renjō* (11.4.2).

(4) /Q, N, I, V (or U)/

There were two parts to this change: First, nasality was lost in morpheme final moraic vowels, merging /Ũ, Ĩ/ with /U, I/. In the native vocabulary, nasal moraic vowels occurred almost exclusively in the *onbin*-stem of verbs (see 7.1.2.2), (5a) below, and overall most nasal moraic vowels were found in SJ vocabulary, as reflexes of Chinese *-ŋ (cf. 9.2.2.3), see (5b).

(5) /Ũ, Ĩ/ > /U, I/

a. *koĩ-* 'row' ≠ *koi-* 'thresh, strip off' > *koi-*
toũ- 'jump' ≠ *tou-* 'ask' > *tou-*

b. *keĩ* (京) ≠ *kei* (系) > *kei*
koũ (公) ≠ *kou* (口) > *kou*

Second, in morpheme nonfinal position nasality became phonemic in the moraic consonants. This second part of the change only affected the phonemic underlying representation and had little immediate effect as the phonetic

nasality of individual forms remained as before. Until that change nasality in morpheme nonfinal moraic consonants had been assigned by the phonetic nasality of the following consonant, for example a prenasalized media: /aki**Cd**o/ => [aki**C^{n}d**o] => [aki**nd**o] (cf. 7.1.2.1), but as part of this change the phonetic nasality was phonemically redistributed and assigned to the moraic consonant: /akiNdo/. Conversely, in forms such as /taCto-/ 'precious', the moraic consonant was reinterpreted as distinctively oral: /taQto-/.

(6)

/C/ >	/N/ /__ {b, d, g, z, m, n}
	/Q/ /__ {p, t, k, s}
/akiCdo/ >	/akiNdo/
/taCto-/ >	/taQto-/

In the Christian sources, /Q/ is transcribed by doubling the following consonant, e.g. *maccura* /maQkura/ 'very black', *maxxiro* /maQsiro/ 'very white', etc. (or before the letter 'q' by 'c', e.g. *facqua* /faQkwa/ 白花 'white flower'); /N/ was mostly transcribed by 'n', sometimes before labials by 'm', and sometimes by a tilde over the preceding vowel, e.g. *nindẽ* /ninden/ 'human and heavenly beings'. /I/ (< EMJ /Ĩ, I/) was transcribed by 'i, j'. /U/ (< EMJ /U, Ũ/) was involved in a set of changes which produced the long vowels in Table 11.1 above, see 11.5, and was transcribed as part of those long vowels.

11.2 Loss of /w/

Loss of /w/ in the main took place in EMJ, where syllable initial /w/ was lost before /o/ and before /i, e/ in word internal position (7.3.2.3). The loss of remaining /w/ before /i, e/, that is, mainly in word initial position (but including also word medial /w/ in newly lexicalized compounds), seems to have been completed during the first part of the LMJ period, probably around 1300. The change was manifested as mergers of /.wi/ ≠ /.i/ as /.i/, and of /.we/ ≠ /.e/ as /.e/ which as in EMJ was phonetically realized as [^{j}e] (cf. 11.8). This change also includes loss of /w/ after /k/, and the merger of /kwe/ ≠ /ke/ as /ke/. However, the mora /kwe/ was in any case marginal and probably not well integrated in EMJ; apart from the single EMJ word *kwe-* 'kick' it was found only in some SJ vocabulary.

(7) /w/ > Ø / __ {i, e}

EMJ	LMJ
wi- 'sit down; exist' ≠ *i-* 'shoot' >	*i* /i/
we 'picture' ≠ *e* 'inlet' >	*ye* /e/ [^{j}e]
kwe- 'kick' ≠ *ke* 'hair' >	*qe* /ke/

Loss of /w/ in EMJ and early LMJ left /w/ only before /a/, both in syllable initial position, /.wa/, and after /k-, g-/, /kwa, gwa/, which although restricted to SJ vocabulary were more stable and widespread than /kwe/. Both /kwa, gwa/ are clearly reflected in the Christian sources, e.g. *facqua* /faQkwa/ 白花 'white flower', *quaxi* /kwasi/ 菓子 'fruit (for dessert)', and were not lost until NJ; they are still preserved in some dialects (see 14.4). Furthermore, as part of the change of /au/ > /ɔɔ/ (see 11.5), syllable initial /.w/ was lost before /ɔɔ/ < /aU/. That is to say, /.wau/ and /.au/ merged as /.ɔɔ/, which was pronounced with a labial onglide [ʷɔ:] (11.8). (However, /kau/ > /kɔɔ/ ≠ /kwau/ > /kwɔɔ/ remained distinct.)

(8) /.w/ > Ø / __ {ɔɔ}

EMJ	LMJ
wau 'king' ≠ *au* 'meets' >	*vŏ* /ɔɔ/ [ʷɔ:]

11.3 Fricativization of /p/; /p/ > /f/

Through the history of Japanese, two sets of changes have affected OJ /p/. The first one, intervocalic /-p-/ > /-w-/, took place in the second half of the tenth century (see 7.3.1.1). The second one, /p/ > /f/ > /h/, started in LMJ. At some point during the LMJ period, /p/ changed to /f/, except in the position after /N, Q/ (cf. 9). The dating of this sound change is difficult for it resulted in no mergers and was not reflected in Japanese script: the *kana* which had been used for /pV/ (はひふへほ) simply came to be used for /fV/, as well as for /pV/ and for /bV/ (cf. 6.1.3).

(9) /p/ > /p/ / {N, Q} __
/f/ elsewhere

Some time during NJ, /f/ became /h/ which is the main NJ reflex of OJ /p/ in initial position. Also that change is difficult to date, because it, too, found no orthographic expression in Japanese script (the *kana* はひふへほ then, as today, being used for /hV/). However, the transcriptions in the Christian sources make clear that /p/ > /f/ was complete by the end of the period, but that the stage /h/ had not yet been reached. (10) shows some forms from *Vocabulario*:

(10) *fa* 'leaf' (OJ *pa*; NJ *ha*)
fi 'day' (OJ *pi*; NJ *hi*)
fune 'boat' (OJ *pune*; NJ *hune*)
fe 'passing' (OJ *pe*; NJ *he*)
foka 'outside, beside' (OJ *poka*; NJ *hoka*).

The precise sound value of /f/ at the end of the LMJ period is not easy to establish. Remarks in Collado's *Ars gr.* (p. 4) indicate that it was not in all varieties of Japanese a straightforward [f], but perhaps rather [ɸ], possibly on the way towards [h].

(11) The letter *f* in some provinces of Japan is pronounced just as in Latin, but in others as if it were approximately *h* – somewhere between *f* and *h*, by folding and closing the mouth and lips, but not completely.[1]

This change introduced the phoneme /f/ into the phonological system of Japanese and resulted in severe restrictions on the occurrence of /p/ in the lexicon. The change of /p/ > /f/ is often said to have applied to word initial /p-/, complementing the change of intervocalic /-p-/ > /-w-/, but that does not take account of the time lag between the two changes. The change /-p-/ > /-w-/ took place in the second half of the tenth century, leaving /p/ only in initial position and after /Q/, but between that time and the change of /p/ > /f/ during the LMJ period, words with non-initial /p/ came into being, particularly through lexicalization of compounds, and non-initial /p/ in those words was also affected by the change. Thus, *Vocabulario* (and NJ) has many words with internal /f/ < /p/. A few are given in (12). For these words we cannot know whether they were lexicalized before or after /p/ > /f/, but they may all be thought to have been firmly lexicalized by the end of LMJ.

(12) *yedafa* /edafa/ 'branches and leaves; digression' (*yeda* /eda/ 'branch' + *fa* 'leaf'), *safodo* /safodo/ 'so, that' (*sa* 'that' + *fodo* 'extent'), *mŏfaya* /mɔɔfaya/ (NJ *mohaya*) 'already' (*mɔɔ* 'already' (?< *mau* < *ima wa* 'now TOP') + *faya* 'early')

Three well-known words with internal LMJ /f/ > NJ /h/ are *afiru* (NJ *ahiru*) 'duck' which is first attested late LMJ, *afure-* (NJ *ahure-* > cNJ *afure-*) 'overflow' which was OJ *abure-* but appears to have changed to *apure- after the change of /-p-/ > /-w-/, but before /p/ > /f/ by which it became *afure-*, and (early LMJ *apau* >) late LMJ *afɔɔ* (NJ *ahoo* > cNJ *aho*) 'fool(ish)', first attested early in the thirteenth century and of unknown etymology although it looks like a SJ word. An interesting example is *asafi* 'morning sun' (NJ *asahi*) which is attested as OJ *asapi*, but must have been a transparent compound to escape the change of /-p-/ > /-w-/. Also SJ vocabulary which retained non-initial /p/ because of transparent constituency or morphemic independence (cf. 7.3.1.2) took part in /p/ > /f/. Thus – whatever the history of individual

[1] I am grateful to Stephanie West for the translation from the Latin.

words – /p/ > /f/ was not constrained to word initial position, but applied regularly to /p/ which did not occur after /N, Q/ (with a few non-phonologically defined exceptions discussed in 11.3.4).

11.3.1 Retention of /p/ after /Q/

Just like the change of internal /-p-/ > /-w-/ in EMJ did not apply to /-p-/ after /Q/ (7.3.1.2), the change of /p/ > /f/ did not apply to /p/ after /Q/, thus leaving *-pp-* (= /Qp/) sequences in addition to *appare* 'splendid' and *moppara* 'entirely' mentioned in (7.3.1.2), for example *yappari* (not in *Vocabulario*) /yaQpari/ 'as expected, sure enough' (thought to be related to the root *yapa-* 'soft, pliant' which also has reflexes in *yawa-*; *yappari* today has the variant *yahari* which is a hypercorrect back formation, probably based on the spelling やはり), *yoppodo* /yoQpodo/ 'very, much' ((?< *yopodo*) < *yoki podo* 'good extent'; is today used mainly in the shape *yohodo* which like *yahari* is a hypercorrect back formation), or *cappato* /kappa-to/ 'suddenly'.

By far the most occurrences of /Qp/ are found in SJ vocabulary where retention of /p/ after /Q/ has been analogically extended so that all SJ morphs with initial NJ /h-/ (< LMJ /f/ < EMJ /p/) have automatic allomorphs in /-p-/, if they are used after /Q/. Similar lexical alternations are also found, although to a lesser extent, in the native vocabulary when prefixes of the shape CVQ- combine with forms with initial LMJ /f-/, e.g. (13) with intensifying *maQ-*:

(13) *maQ-* +

fadaca 'naked'	*mappadaca* 'stark naked'
fajime /fazime/ 'beginning'	*mappajime* 'origin, start'
fira 'flat'	*mappira* 'without fail'
firu 'daytime'	*mappiru* 'middle of the day'
fucura 'core'	*mappucura* 'core'

There are other verbal prefixes of the shape *CVQ-*, e.g. EMJ *piki-* 'pull.INF' > *piC-* > LMJ *fiC-* > NJ *hiQ-* after which verbs with initial /h/ appear with /p/ (e.g. LMJ *fippar-* 'pull, drag' > NJ *hippar-*). All such forms may now be regarded as lexicalized and some may have been formed analogically, but the point remains that they reflect preservation of /p/ after /Q/. This is also the case with a small number of compounds which have second components with initial /pp/, alternating with initial NJ /h/, e.g. LMJ *asafara* 'morning-belly; stomach of someone who has not eaten breakfast' > NJ *asahara* ~ *asa-ppara* (not attested until the eighteenth century), the reduplicated form NJ *happa* 'leaf' (childish) <= *ha* 'leaf' (< OJ *pa*), or a few NJ forms which are compounds with *de-* 'protrude'(e.g. *deppa* 'buck tooth' (*ha*) 'tooth'), *deppana* 'long nose' (*hana* 'nose').

11.3.2 Retention of /p/ after /N/

In EMJ, /p/ was neutralized as /b/ after /N/ by the postnasal neutralization rule, i.e. //Np// => /Nb/ (7.1.2.2), but as discussed above (11.1.1) this rule was lost in the course of LMJ. As after /Q/, /p/ was preserved after /N/ and did not change to /f/, and from LMJ onwards /Np/ became a frequent sequence, as in the few examples in (14) from *Vocabulario*. Most words in this group are SJ, as in (14a), but a few examples are also found which include native vocabulary, see (14b):

(14) a. *anpu* 'safety' (安否), *bonpon* 'book from India' (梵本), *monpa* 'sect' (門派).

b. *binpigue* /binpige/ 'hair and beard at the temples' (*bin* 'temple' + *pige* 'beard'), *bimpima* /binpima/ 'right moment' ~ *bin fima* (*bin* 便 'convenient' + *pima* 'time, moment')

Compounds such as SJ *xinfôracu* (神法楽) 'entertainment for the gods/spirits', which is one of the few words with the /Nf/ in *Vocabulario*, is a transparent productive compound from <= *sin* 'god/spirit' + *fooraku* 'entertainment'. On the whole, words with /Np/ must have been lexicalized after the loss of postnasal neutralization, but before /p/ > /f/ (or have been analogically formed). Words such as those in (14) which retain /p/ after /N/ contrast with words such as *sanbô* and *nanbocu* in (3) above which have /p/ reflected as /b/ after /N/ and must have been lexicalized before the loss of postnasal neutralization.

11.3.3 Alternations arising from changes of /p/

The changes affecting /p/ have left traces in lexical alternations, see (15) which shows the main alternations within NJ:

(15) a.	p- > f- > h- *hyaku* 'hundred'	~	-p- *(rop)-pyaku* 'six hundred'
b.	p- > f- > h- *hyaku* 'hundred' *hara* 'stomach'	~	-b- <= -p- *(san)-byaku* 'three hundred' *(biiru)-bara* 'beer-belly'
c.	p- > f- > h- *hara* 'plain'	~	(-ɸ- <) -w- < -p- *(fuji-)wara* 'wisteria plain; proper name'

d.	-p- > -pp- *appare* 'splendid!'	~	-w- < -p- *aware* 'evocative of deep feeling')
e.	-p- > -w- *uwa-* 'top'	~	-Ø- < -w- < -p- *ue* 'top'
f.	-p- => -b- *yuube* 'evening' (< *yupu* 'evening + *pye* 'side')	~	(-Ø- <) -w- < -p- *mae* 'front' (< *ma-* 'eye' + *pye* 'side')

Of these (15a) holds for all SJ morphemes in initial /h/ which have (potential) variants for use after /Q/, for the few native words where /p/ is preserved after /Q/, and for forms with /p/ preserved after /N/. (b) holds for native vocabulary in *rendaku* and in forms with /Nb/, as well as for SJ vocabulary which lexicalized the effects of postnasal neutralization before it was lost (11.1.1). Some morphemes take part in both (a) and (b), for example *hyaku* ~ *-pyaku* ~ *-byaku* 'hundred'. (c)–(f) are all quite rare.

11.3.4 /p/ in expressive vocabulary

Mimetics and other expressive adverbs seem to retain /p-/, as in many NJ mimetics, e.g. *pika-pika* 'sparklingly' (cf. OJ *pikar-*, NJ *hikar-* 'shine') or *puka-puka* 'puffingly' (cf. OJ *puk-*, NJ *huk-* 'blow (intr.)'). However, the history of mimetics and their phonemic shapes is not well described. That is especially so with regard to mimetics with /p/ because until the beginning of the seventeenth century there were no orthographic means to note /p/ to the extent this was different from /f/ after the change /p/ > /f/ had taken place (cf. 6.1.3). *Vocabulario* lists the following words with initial *p-*:

(16) *pappato* 'the rising of dust, waves, flames, etc.'; *pararito*, *fararito* (a) 'the sound of grain or the like falling', (b) 'all, every'; *patto* 'the rising and scattering of things like smoke (or spurting blood)'; *paxxito* (/passito/) 'for example hitting or striking the target with arrow or spear'; *pinpin(to)* 'the kicking and jumping of domestic animals (e.g. horses)'; *pixxito* (/pissito/) 'being gathered, jammed tightly'; *ponpon*, *po(n)ponto* 'the sound of beating a Japanese hand drum or letting off a rifle shot'; *poppoto* 'blazing up of flames etc.'.

These are all expressive adverbs, whose meanings were evidently not easy to define. The question remains, however, whether use of [p] for expressive purposes actually reflects preservation of OJ /p/, or whether these are expressive

innovations. The fact that we for some mimetics, as with *pika-pika* and *puku-puku* above, are able to identify cognate OJ verbal roots with initial /p-/ supports the former view, but the systematic productivity and elaborate sound symbolic conventions in the formation of Japanese mimetics must also be taken into account. Note for example that the equivalent in *Vocabulario* of NJ *pika-pika* is '*ficaficato*' (*fika-fika-to*), i.e. with *f* rather than *p*, suggesting that NJ *pika-pika* may be an innovation. Another interesting example is '*foroforoto*' in *Vocabulario* which, amongst other uses, is used about continuous, steady crying, as in *namidaga foroforoto vochita* (/namida ga foro-foro-to otita/) 'his tears fell *foro-foro*'; this seems to be a descendant of EMJ *poro-poro*, in turn probably related to the verb ***kobore***- 'spill, flow, drop' together with which it is sometimes used, as in *namida poro-poro-to kobore-tamawi-nu* 'his tears dropped *poro-poro*' (from *Genji monogatari*, *Yūgiri*). Now, NJ has both *horo-horo* and *poro-poro*, in addition to *boro-boro*, which seem to note ascending magnitude or flow of tears, and it is not at all clear what the relation is between EMJ *poro-poro*, LMJ *foro-foro* and NJ *horo-horo* ~ *poro-poro*. The LMJ form suggests that regular sound change applied to EMJ *poro-poro* > LMJ *foro-foro* > NJ *horo-horo*; that, however, runs counter to the usual presumption of preservation of /p/ in mimetics and leaves NJ *poro-poro* stranded. Is NJ *poro-poro* then a direct reflex with retention of /p/ of EMJ *poro-poro* which was not noted in the Christian sources, or is it in fact a NJ innovation?

11.3.5 Phonemic split: /p/ > /f, p/

It would be possible to view [p] after /N, Q/ as an allophone of /f/, as [f] did not occur in those two contexts in a (lexicalized) word, although we might find the phonetic difference between [p] and [f] too great to consider them plausible allophones of one phoneme. However, as it is, [p] was also found outside that phonological context, and we even have a rare minimal pair such as *pixxi(to)* (/pissito/) 'being gathered, jammed tightly' versus *fixxi* (/fissi/) 'pen and paper' (筆紙), clearly showing that a phonemic split had occurred: /p/ > /f, p/. In addition, /p/ was independently needed in loanwords (including proper names) introduced by the Christian missionaries, such as *paraizo* 'paradise', *purutaruko* 'Plutarch', or *pekadoru* 'sinner'. These three words are written in *hiragana* (ぱらいぞ, ぷるたるこ, and ぺかどる) in *Giya do pekadoru* (1599; cf. 10.2.2.1).

11.4 Sino-Japanese syllable final /-t/

A famous feature of the Christian sources is the use of syllable final *-t* in the writing of SJ words both in word final and nonfinal syllables, e.g. (17). There is little doubt that this reflects contemporary pronunciation, as also shown for example by *renjō* (11.4.2).

(17) *fotnet suru* 'get fever' (発熱)
connit 'today' (今日)

In NJ the words which in late LMJ had final /-t/ have an epenthetic vowel, usually /u/ (*netsu* 'fever'), but sometimes /i/ (*konnichi* 'today'). In earlier sources in Japanese script, these words are written with *kana* for *tu* (つ) or *ti* (ち), but it has recently been shown that a distinction also was made in some LMJ *kana* sources between *-tu* and *-t*, by using variant *kana* (*hentaigana*, see EMJ 6.1.2) which were originally used as equivalents for /tu/.

11.4.1 Sources of final /-t/

Final /-t/ is found among SJ words reflecting EMC *-t. Whereas EMC *-p and *-k were reflected in SJ vocabulary by /-pV/ and /-kV/, especially in SJ readings as *-pu* (> *-u*), *-ku*, and *-ki* (see 9.2.2.3), EMC *-t was in a number of SJ words reflected as final /-t/, with no epenthetic vowel. The orthographic categories of the *kana* do not provide for representation of /-t/ and seem to have used *kana* for *tu* or *ti* for /-t/. The use of *hentaigana* to represent /-t/ is a fairly new discovery and it has yet to be seen how widespread this practice was. In NJ SJ *kanji* readings EMC *-t is conventionally rendered as *-ti* in *go-on* and *-tu* in *kan-on*, and it is usually thought that SJ loanwords taken in from (Japano-Chinese or Sino-Japanese) *go-on* had final /-ti/, whereas words originating in *kan-on* had /-t/, written by *kana* for *tu* and later acquiring an epenthetic /u/ to give /tu/. This is almost certainly an overly simplistic view. For example, the SJ loanword for 'fever' 熱 was late LMJ *net* as shown by the Christian sources. In EMJ this was written as *neti* (cf. 9.2.3.2) whereas the NJ shape is *netu*. This indicates that the EMJ writing as *neti* was meant to represent /net/, and significantly, the initial /n-/ of these forms clearly show that they derive from *go-on*, corresponding to EMC *ȵiat, cf. 9.2.2. (For 熱 modern *kanji* dictionaries give *go-on* as *netsu/nechi* and *kan-on* as *zetu*.) All this strongly suggests that this *go-on* based SJ loanword was *net* until the beginning of NJ. Another similar example is late LMJ *jitni* /zit-ni/ 'really' which was written *ziti-ni* in EMJ (cf. 9.2.3.2), but which is cNJ *jitu-ni*. Again, this suggests that the form was *zit* from EMJ until early NJ, and also this is clearly *go-on* based, corresponding to EMC *ʑit. (For 実, modern *kanji* dictionaries give *kan'yō-on* as *jitsu*, *go-on* as *jichi* and *kan-on* as *shitsu*, showing, compared to the readings for 熱, some arbitrariness in the assignment of current forms to *kan'yō-on* or to one of the other layers of readings.)

Final /-t/ is almost entirely limited to SJ vocabulary, but it is also found in a very small number of variant shapes of native words in the Christian sources, whose regular shapes, however, have final vowels: *ximot* 'cane, rod (used in criminal punishment)' (~ /simoto/), *carafit* 'big box; lit. Chinese chest'

(~ /karafitu/); *tetdai* 'help' (~ /tetudai/). Spelling variants by *hentaigana* in sources in Japanese script confirm that these words had final /-t/.

11.4.2 Renjō

Final /-t/ is also reflected in *renjō*, the process in which a final consonant is copied as the onset of a following syllable with initial vowel or glide (onset creation; see 7.2.1). *Renjō* after final /-t/ is in *Arte* (p. 637) said to be regular before *va* /wa/ which becomes *ta* after /-t/, so that for example *taixetta* /taiset-ta/ is from *taixetua* /taiset-wa/ 'importance-TOP'. The only examples given in *Arte* involve the topic particle *wa*, but in both *shōmono* and the early NJ *kyōgen* texts (13.3) there are also frequent examples of *renjō* after /-t/ before the accusative particle *o*, which becomes *to*. Through the period several lexical items are also found, for example /settyoosyuu/ (*Setchōshū*) which is an alternative name for the mid fifteenth-century dictionary today known as *Setsuyōshū* /setuyoosyuu/, showing that it must earlier have been /set.yousyuu/, giving /set**u**yoosyuu/ by vowel epenthesis and /set**t**yoosyuu/ by *renjō*). However, they are not many; the only two lexical forms included in *Vocabulario* with *renjō* after final /-t/ are *xecchin* /settin/ 'toilet; "tight spot"' (雪隠) <= *set.in* (mentioned also in 7.2.1 and still used in NJ) and *xŏjenbattacu* /syɔɔzenbattaku/ 'reward good, punish evil' <= *syaũzen-bat.aku* (賞善罰悪) which is no longer used.

By contrast, *renjō* after /N/ appears to have been far more widespread and regular. Thus, *Arte* (pp. 636–7) sets it out as a general rule of pronunciation, saying that for example *sanya* /san.ya/ 'mountains and fields' should be pronounced *sannha* /san.nya/; other examples are *xinnhô* /sin.nyoo/ 'trust' <= *xinyô* /sin.yoo/, *ninguenna* /ningen-na/ 'person-TOP' <= *ninguenua* /ningen-wa/. Also *Vocabulario* includes a number of forms with *renjō*, e.g. *annon* 'peace and quiet' (安穏, *an* + *on*) or *innen* 'destiny' (因縁, *in* + *en*), but sometimes notes that this is the actual pronunciation, although they are written differently, or 'correctly' should be pronounced without *renjō*. Both *Arte* and *Vocabulario* demonstrate awareness of *renjō* as a general, productive rule after /N/ and of the relationship between underlying and surface form. In addition, *Vocabulario*, while noting actual pronunciation, recommends avoiding such forms in writing. In *Esopo* topic and accusative particles are always written *ua*, *va* and *vo*, *uo*, also after /-t/ and /N/, i.e. without *renjō*, e.g. *xujinuo* /syuzin-o/ 'master-ACC' or *funbetuo* /funbet-o/ 'judgement-ACC', although these words would have been pronounced as from /syuzinno/ and /funbetto/.

Renjō thus reflects final /-t/ and also shows that word final /N/ phonetically was [n] (as opposed to NJ where word final /N/ is an unreleased uvular [ɴ] or velar or [ŋ]). The difference in application of *renjō* after /N/ and /-t/ also indicates a difference between those two phonemes in the language, suggesting that final /-t/, which was (almost) only found in SJ vocabulary, was not as

stable as other syllable final phonemes, and that it was in the process of acquiring an epenthetic vowel.

11.5 Long vowels

Through the LMJ period, long diphthongs of the shape /Vu/ changed as shown in (18a), which also gives the further developments in NJ. These changes are in Japanese known as *chōon-ka* (長音化 'change to long sound'), exemplified in (18b), which gives the spelling in *Vocabulario* followed by a phonemic notation; see further below about the phonemic interpretation of the long vowels.

(18) a.

(late) EMJ	>	LMJ	>	NJ
/iu/		/yuu/		/yuu/
/eu/		/yoo/		/yoo/
/au/		/ɔɔ/		/oo/
/ou/		/oo/		/oo/

b. Examples

(late) EMJ	LMJ	NJ
iu 'says'	*yŭ* /yuu/	*yuu*
uresiu 'long ago'	*urexŭ* /uresyuu/	*uresyuu*
akiũdo 'trader'	*aqiŭdo* /akyuudo/	*akyuudo*
eu 'gets drunk'	*yô* /yoo/	*yoo*
keu 'today'	*qeô, qiô* /kyoo/	*kyoo*
teũdu 'wash(ing)'	*chôzzu* /tyoodu/	*tyoozu*
auti 'Japanese bead tree'	*vŏchi* /ɔɔti/ [ʷɔːtʃi]	*ooti*
waũ 'king, emperor' (SJ 王)	*vŏ* /ɔɔ/ [ʷɔː]	*oo*
yaũ 'way of doing' (SJ 様)	*yŏ* /yɔɔ/	*yoo*
taUto-'precious'	*tŏto-* /tɔɔto-/	*tooto-*
ougo 'protection'	*vôgo* /oogo/ [ʷoːgo]	*oo*
osou 'slowly, late'	*vosô* /osoo/	*osoo*
koũdi 'lane, alley'	*côgi* /koodi/	*koozi*

These changes can be said to have introduced long vowels in the phonological system, although it must be thought that /iI/ and /uU/ (which arose in EMJ both through *onbin* (7.1.4) and through loss of /-p-/ before /u/ and of /-p-, -w-/

before /i/ (7.3.1.1, 7.3.2.3)) were pronounced as long vowels all along. We assume here that the changes to long vowels took place after the loss of distinctive nasality in the moraic vowels (see 11.1.2). There is, however, no evidence either way and both orders are possible.

The difference between the LMJ and NJ long vowels is that EMJ /aU/ and /oU/ gave two distinct long vowels in LMJ, which later merged as NJ /oo/. They are consistently transcribed differently in the Christian sources, as (EMJ /aU/ >) *ǒ* and *ô* (< EMJ /oU/). The difference in vowel quality is usually thought to have been (/aU/ >) [ɔ:] versus [o:] (< /oU/). Both were rounded vowels and had a phonetic onglide in syllable initial position, as shown by the transcriptions *vǒ*, *uǒ* [ʷɔ:] and *vô*, *uô* and [ʷo:] (like short syllable initial /.o/, transcribed by *vo*, *uo*), see below (11.8)).

11.5.1 Phonemic analysis

In (18) and in Table 11.1 above the long vowels are transcribed by double vowels. More abstractly long vowels in Japanese are often analysed phonemically as consisting of a nuclear vowel followed by an 'empty' moraic vowel or length phoneme, noted variously as /V/, /R/, or /H/, e.g. *kyoo* 'today' /kyoV/, kyoR/, or /kyoH/. A problem in applying that analysis directly to the long vowels in LMJ is that *ǒ* [ɔ:] then would be /ɔV/ (or /ɔR, ɔH/), but that would be structurally unusual as the short vowel /ɔ/ was not a part of the system. It is alternatively possible to view *ǒ* [ɔ:] as the realization of a long *a*, i.e. as being a realization of /aV/ (or /aR, aH/). Support for that may be found in forms such as the exclamatory final particle *nǒ* /nɔɔ/ which is usually thought to be from a lengthened version of *na*, or the mesial demonstrative adverb *sǒ* /sɔɔ/ which comes from *sa* (see 12.5), but on the other hand these are singular developments, possibly analogically motivated. Against the analysis of *ǒ* [ɔ:] as /aV/ speaks the fact that *ǒ* [ɔ:] clearly was a rounded vowel, with a phonetic onglide when in syllable initial position (see 11.8). It is also possible to view *ǒ* [ɔ:] as a phonetic realization of underlying /aU/. This has the advantage of providing a source in the phonemic representation of the phonetic rounding. It may find support in the fact that the changes to long vowels also gave rise to morphophonological rules for verb and adjective forms ending in *-u*, whereby for example the conclusive of /-w/ base verbs was realized with a long vowel: *aw-u* 'meets' => *vǒ* /ɔɔ/, *ow-u* 'carries' => *vô* /oo/ (see further 12.3.3). On that analysis, the changes to long vowels reflected in the Christian sources at the end of the LMJ period were mainly changes in the phonetic realization of diphthongs whose outcomes were later phonemicized as reflected in NJ. In any case, *ǒ* [ɔ:] seems to have been fairly short-lived in the language and merged with [o:] in the first part of the NJ period. In (19) we show two different phonemic analyses of the long vowels:

Table 11.2 *Sources of SJ* koo *and* yoo

		EMJ		LMJ		NJ
	EMC	-p- > Ø /__ u		denasalization	monophthongization	ɔɔ > oo
交	*kaɨw/kɛ:w	*kau*			*kɔɔ*	*koo*
甲	*kaɨp/kɛ:p	*kapu*	*kau*		*kɔɔ*	*koo*
香	*xɨaŋ	*kaũ*		*kau*	*kɔɔ*	*koo*
光	*kwaŋ	*kwaũ*		*kwau*	*kwɔɔ*	*koo*
口	*k^{h}əw’	*kou*			*koo*	*koo*
劫	*kɨap	*kopu*	*kou*		*koo*	*koo*
公	*kəwŋ	*koũ*		*kou*	*koo*	*koo*
様	*jɨaŋ^{h}	*yaũ*		*yau*	*yɔɔ*	*yoo*
用	*juawŋ^{h}	*youũ*		*you*	*yɔɔ*	*yoo*
要	*ʔjiaw	*eu*			*yoo*	*yoo*
葉	*jiap	*epu*	*eu*		*yoo*	*yoo*

(19)

EMJ	>	LMJ	>	NJ
/iu/		/yuu/		/yuu/
[iu]		[yu:]		[yu:]
[C$_{j}$iu]		[Cyu:]		[cyu:]
/eu/		/you/ or /yoo/		/yoo/
[^{j}eu]		[yo:]		[yo:]
[C$_{j}$eu]		[Cyo:]		[Cyo:]
/au/		/au/ or /ɔɔ/		/oo/
[au]		[^{w}ɔ:]		[o:]
[Cau]		[Cɔ:]		[Co:]
/ou/		/ou/ or /oo/		/oo/
[^{w}ou]		[^{w}o:]		[o:]
[Cou]		[Co:]		[Co:]

Regardless of the phonemic analysis of the changes in (18) and their outcomes, they consist of a decrease in contrast in long /VU/ diphthongs between the peak vowel and the moraic vowel following it and may for that reason be characterized as *monophthongizations*.

11.5.2 Sources of long vowels in Modern Japanese Sino-Japanese

Before the monophthongizations, a sizeable proportion of SJ vocabulary had /Vu/ diphthongs, themselves from a number of sources, and the monophthongizations therefore resulted in widespread merger within SJ vocabulary of

previously distinct syllables. Table 11.2 shows the sources of SJ *koo* and *yoo*. Changes preceding the monophthongizations include loss of /-p-/ before /u/ (EMJ) and loss of nasality in /-ũ/ (LMJ).

11.6 Assibilation and palatalization

Through the LMJ period, /t, d/ became phonetically assibilated before the high vowels, /i, u/, giving allophonic rules as in (20), which also shows the transcription in the Christian sources.

(20)	/t/ =>	[t^s] / _ u	/tu/ *tçu*
		[$t^ʃ$] / _ i	/ti/ *chi*
	/d/ =>	[d^z] / _ u	/du/ *zzu*
		[$d^ʒ$] / _ i	/di/ *gi*

The onset of the phonetic assibilation is difficult to date precisely as it was not reflected in writing in Japanese before the Christian sources, but the Korean *Irop'a* from 1492 transcribes Japanese /ti, tu/ by 디 /ti/ and 두 /tu/, showing that assibilation had not yet set in by then. The phonetic assibilation of /d/ eventually resulted in the merger of /d/ and /z/ before /i, u/, but that phonemic change is not thought to have been completed until NJ (see 14.1), as the Christian sources keep /d/ and /z/ distinct before /i, u/, noting /zi/ as *ji*, /di/ as *gi*, /zu/ as *zu*, and /du/ as *zzu*.

11.6.1 Palatalization

Consonants remained palatalized before the front vowels /i, e/, as in EMJ (21):

(21) C => [C_j] / __ {i, e}

It is directly reflected in the transcriptions used in the Christian sources (see Table 11.1) that /s, z, t, d/ were palatalized before /i/ and /s, z/ before /e/: /si/ *xi*, /zi/ *ji*, /ti/ *chi*, /di/ *gi*, /se/ *xe*, /ze/ *je*. Palatalization before /i, e/ is not as clearly and directly reflected in the transcription of other short syllables, e.g. short /mi/ *mi*, /ni/ *ni*, /me/ *me*, /ne/ *ne*. However, on closer inspection the facts point to palatalization of all consonants before /i, e/: Short syllables with palatal glide, /CyV/, are transcribed by *CiV* or in some cases alternatively by *CeV*, e.g. /kyo/ *qio, qeo*, /mya/ *mia, mea*, /rya/ *ria, rea*, showing that both *Ci-* and *Ce-* were used to represent /Cy-/ and suggesting that the consonant was palatalized in *Ci* and *Ce* used to represent single moras, e.g. /mi/ *mi*

[2] The alternative transcriptions of /ki/ by *qui* (rather than *qi*) and /ke/ by *que* (rather than *qe*), which could be taken to suggest non-palatalized /k/ before /i, e/, are very rare. Note that as *gi* was used to transcribe /di/ [$d^ʒi$], the use of *gui* for /gi/ and *gue* for /ge/ should not be taken to show that /g/ was not palatalized before /i, e/; in particular, *gui-* and *gue-* are used to transcribe /gy-/.

[m$_{j}$i], /me/ *me* [m$_{j}$e].[2] In the transcription of long syllables /Cyoo/ or /Cyɔɔ/, equivalence between *Ci-* (*qi-*, *gui-*, *fi-*, *bi-*, *pi-*, *mi-*, *ri-*) and *Ce-* (*qe-*, *gue-*, *fe-*, *be*, *pe-*, *me-*, *re-*) is even more widespread. Also the transcription of /nyuu/ by *nhŭ*, *nhû*, *niŭ*, *niû* and of /nyoo/ by *nhô*, *neô* shows equivalence between *nh-*, *ni-*, *ne-* to note /ny-/; likewise, /ty-/ is transcribed by both *ch-* and *te-*. Finally, also the fact that /Ciu/ and /Ceu/ changed to /Cyuu/ and /Cyoo/ (see 11.5) shows that consonants were palatalized before /i, e/.

11.7 Final vowel deletion

A phonetic feature described by Collado in *Ars gr.* (p. 5) which is familiar from NJ is the weakening and deletion of word final /i, u/. Collado describes this as a general phenomenon and gives examples such as *fitotçu* 'one' pronounced as *fitotç* [fitots], *axi no fara* 'reed plain' => *ax* [aʃ] *no fara*, and *gozaru* 'exist, polite' => *gozar* (which interestingly shows word final vowel deletion after /r/, which does not happen in cNJ). Affrication and palatalization of the consonant before /i, u/ show that final vowel deletion was a very low level phonetic rule which took place after other allophonic variation had been assigned. It should be noted that final vowel deletion thus is not related to, or involved in, the occurrence of syllable final /-t/, as the many forms with syllable final /-t/ did not have palatalization or affrication.

11.8 Onglides

At the end of LMJ, syllable initial /.e/ and /.o/ were still, as in EMJ, pronounced with an onglide.

(22) /.e/ => [^{j}e]
/.o/ => [^{w}o]

This is clearly shown by the transcriptions in the Christian sources, *ye* and *vo*, *uo*, and also by the fact that /.eu/ > /.yoo/. Also the new long vowels [o:] and [ɔ:] (11.5) had labial onglides in syllable initial position, as shown by the transcriptions *vŏ*, *uŏ* and *vô*, *uô*. Functionally, the onglides served to signal syllable boundaries, creating a phonetic onset in a vowel initial syllable. This is for example shown by the distinction, clearly reflected in the transcriptions, between /o.o/ belonging to two different syllables and /oo/ in a single long syllable (whether it is ultimately analysed phonemically as /oV/ or /oU/, see 11.5.1), as in (23):

(23)	Hetero-syllabic /o.o/	Homo-syllabic /oo/
	touoi [to^{w}oi] /to.oi/ 'distant'	*tôi* [to:i] /too.i/ 'Eastern barbarian' (東夷)
	qiouo [k$_{j}$o^{w}o] /kyo.o/ (虛) 'falsehood + ACC'	*qeô, qiô* [k$_{j}$o:] /kyoo/ 'today'

While in general the onglides served to maintain the syllabic independence of /.o/ and /.e/, there are a few sporadic changes where /.o/ or /.e/ have been incorporated as a second mora into a preceding syllable, e.g. *vô* [^{w}o:] /oo/ 'big' (< EMJ *o.o* < *owo-* < *opo-*), *tôca* [to:ka] /too.ka/ 'ten days' (< EMJ *to.o* < *towo*; the word 'ten' itself is listed separately as *touo* [to^{w}o] /to.o/, i.e. as dissyllabic); or *cairu* /kairu/ 'frog' (< EMJ *kaweru* < *kaperu*; the NJ form is *kaeru* and *Vocabulario* gives *cayeru* as an alternative form, but says that *cairu* is the colloquial form).

A similar phenomenon is observed when syllable initial /a/ followed short syllables ending in /i, e/: The transition between the two syllables seems to have produced a palatal glide which became the onset in the second syllable so that /i.a, e.a/ became /iya, eya/; this was sometimes noted in writing, both in the Christian sources and in materials in Japanese script, for example *miya-cæxi* /miyakasi/ 'votive light' (< *mi-akasi* 'HON-light'). Most examples are verbal compounds with the reciprocal auxiliary verb *-aw-* 'do together' further changed to *-yɔɔ* (< *(-i/-e)-au*): *caqeyŏ* /kakeyɔɔ/ 'matches, is equal to' (< *kake-au*), *cami-yŏta* /kami-yɔɔta/ 'bit each other' (< *kami-auta*).

11.8.1 Diphthongal asymmetry

The monophthongizations (11.5) are the last in a number of sound changes which occurred since OJ which have resulted in an asymmetrical distribution of diphthongs, favouring /yV, Vi/ diphthongs over /wV, Vu/.

From EMJ the language acquired /VU/ and /VI/ syllables, with both of /I, U/ occurring after all vowels (7.1.1). However, in the course of the monophthongizations in (18), monomorphemic /Vu/ disappeared from lexical representation; it remained in morphologically complex forms, which, however, always have a morpheme boundary: /V-u/, e.g., *kau* 'buys' <= *kaw-u* (see 12.3.3). In NJ, long syllables ending in /i/ are often monophthongized in casual speech (and regularly in all registers in some dialects), for example in standard NJ as follows (with other outcomes in some dialects): *keito* 'wool' => [keeto], *nai* 'non-existent' => [ne:], *hidoi* 'terrible' => [hide:], *samui* 'cold' => [sami:]. It is likely that monomorphemic /ei/ has changed phonemically to /ee/, so that for example the phonemic shape of *sensei* 'teacher' is /sensee/, but monophthongization of the other diphthongs is in most NJ dialects probably to be

regarded as phonetic realizations from underlying /ai, oi, ui/. Thus, /Vu/ diphthongs disappeared within a fairly short span of time after coming into being, whereas /Vi/ diphthongs are affected by such changes at a much slower pace, if at all.

Conversely, whereas /Vu/ diphthongs were eliminated in the course of the changes in (18), syllables with a post-consonantal palatal glide (/Cy-/) which through EMJ had been restricted to SJ vocabulary, gained in lexical frequency through the changes *Ciu* > *Cyuu* and *Ceu* > *Cyoo*, which introduced these syllable types also in native vocabulary.

OJ had most combinations of both of the onset glides, /w, y/, and nuclear vowels, /i, e, a, o, u/, the two – symmetrical – exceptions being the absence of /wu, yi/ (cf. Table 2.4); on the phonemic interpretation of the *kō-otsu* distinctions given in 2.1.4, OJ also had /Cwi, Cwo, Cye/. However, since the end of OJ, /w/ has gradually been lost both in /Cw-/ and /.w-/ contexts (cf. 7.3.2, 11.2, 14.4). This has resulted in limitations on the occurrence of the labial glide /w/, such that it at the beginning of NJ had disappeared before all vowels except /a/, whereas the palatal glide /y/ remained before all non-front vowels, /a, o, u/ (see Table 11.1).

This pattern of change and distribution has to some extent been upset by very recent loanwords in contemporary NJ (e.g., *wirusu* 'virus', *wesutan* 'western (movie)', *wokka* 'vodka'), but also there it has been observed that /wV/ and /Vu/ are less stable than /yV/ and /Vi/.

REFERENCES

Doi *et al.* 1980, Imaizumi 1971, Mabuchi 1971, Martin 1987, Wenck 1959. Final /-t/: Morita 1993: 196ff.; final /-t/ and *hentaigana*: Sugahara 2000: 21–82. *Renjō*: Morita 1993: 209ff. Diphthongal asymmetry: Kubozono 2008.

Whereas the main phonological changes between OJ and NJ took place during the EMJ period, it was during the LMJ period that most of the significant grammatical changes took place which transformed Japanese from its premodern to its contemporary shape in both morphology and syntax. The course and precise dating of some changes is difficult to trace through the written sources; they are mainly observable in the sources dating from the end of the period. It may be no coincidence that sweeping changes took place during a period of civil war and great social upheaval and change which also would have resulted in a relaxation of social and linguistic norms.

12.1 Verb morphology

12.1.1 Inflected verb forms

Table 12.1 shows the main inflected verb forms at the end of the LMJ period.

Comparing this inflectional paradigm with that of EMJ in 8.1.1 above shows that among the finite forms from the EMJ paradigm the only surviving category is the imperative, which acquired a new variant shape in *-i* with the vowel base verbs, alternating with the older ending *-yo*; see 12.3 below for the shapes with all conjugation classes. Other EMJ forms were either lost or reinterpreted, and new forms had appeared. The loss of the exclamatory, conclusive and adnominal verb forms will be discussed in 12.1.2 and the emergence of the nonpast, past, intentional and volitional forms in 12.1.3.

Other lost forms were used through most of LMJ, but were either entirely or almost lost by the end of the period, as reflected in *Feiqe* and *Esopo*: The prohibitive *na… so* was used through LMJ, but was on its way out towards the end of the period. It had by then to a large extent functionally been replaced by the prohibitive final particle *na* which is still in use in NJ. The negative conjectural *-(a)zi*, too, is rare and had largely been replaced by *mai~mazii* 12.1.4.1). Optative *-(a)namu* was lost during LMJ, whereas the newer EMJ optative *-(a)baya* is found sporadically even in *Esopo*, but on the whole it has been replaced by the new desiderative auxiliary *-(i)ta-* (12.1.5.3). Among the

Table 12.1 *LMJ inflected verb forms*

	QD	LB
Basic stem	*kak-*	*ake-*
Finite		
Nonpast	*kaku*	*akuru*
Past	*kaita*	*aketa*
Intentional	*kakɔozuru*	*akyoozuru*
Volitional	*kakɔo*	*akyoo*
Past conjectural	*kaitarɔo*	*aketarɔo*
Imperative	*kake*	*akei ~ akeyo*
Non-finite		
Infinitive	*kaki*	*ake*
Gerund	*kaite*	*akete*
Conditional	*kakaba (~ kaitewa)*	*akeba (~ aketewa)*
Provisional	*kakeba*	*akureba*
Concessive	*kakedomo ~ kaitemo*	*akuredomo ~ aketemo*
Past conditional	*kaitara(ba)*	*aketara(ba)*
Past provisional	*kaitareba*	*aketareba*
Past concessive	*kaitaredomo*	*aketaredomo*
Intentional provisional	*kakɔozureba*	*akyoozureba*
Intentional concessive	*kakɔozuredomo*	*akyoozuredomo*

non-finite forms, continuative *-(i)tutu* was no longer used productively; its uses in cNJ, where it amongst others is used to form a (written) progressive on telic verbs (*shinitsutsu ar-* 'be dying'), seem to be revived from the classical written language. Conditional *-(a)ba* was still in use at this time, but was in decline. Conversely, incipient conditional uses of the provisional *-(e)ba* are found at the end of the period and in addition, the conditional form of the copula, *naraba*, was starting to be used as a conditional particle, reflected in NJ as *nara*. Finally, a number of new tensed forms had appeared, see (12.1.3.3), including a new past conditional *-(I)tara(ba)*, which eventually gave the NJ conditional *-(I)tara*.[1]

12.1.1.1 Gerund

In addition to its continued use as a subordinate verb form, the gerund came to be used in grammaticalized serial verb constructions during LMJ. In OJ and EMJ, the infinitive was used in a few such constructions, for example with *mi-* to form experiential expressions (*mi miru* 'tries looking'), or with positional verbs in a stative form to form progressives (*mi witari* 'is looking'; see

[1] It is sometimes alternatively suggested that the NJ conditional *-(I)tara* orginates in the LMJ past provisional *-(I)tareba (> -(I)tarea > -(I)tara)*, and likewise that the NJ conditional particle *nara* is not from *naraba*, but from provisional *nareba > narea > nara* (Iwai 1973: 153, 181).

8.4.2.1), but with the exception that existential verbs were used with a verbal gerund to form periphrastic or analytic statives (3.1.4.7.3) and with the copula gerund *nite* (> *de*) to form extended copula forms (8.2.3), the gerund was not used much in such constructions. However, from LMJ a number of other verbs habitually came to combine with the gerund in grammaticalized serial verb constructions, for example VERB-*te mi-* 'try VERB-ing' (*mi-* 'see') and VERB-*te ok-* 'do and leave; do in advance' (*ok-* 'put') which are still used in cNJ, and benefactive constructions such as VERB-*te kure-* 'VERB for me/my sake' (*kure-* 'give (me)'), VERB-*te kudasare-* 'VERB.RESP for me/my sake' (*kudasare-* 'give (me) (RESP)'), VERB-*te tamawar-* 'id.', VERB-*te tabasim-* 'id.'. This type of analytic construction is widely used in NJ, with a number of different verbs.

12.1.2 Loss of exclamatory, conclusive and adnominal

The exclamatory form was already in the second half of EMJ used only as the predicate in *kakari-musubi* constructions with the *kakari* particle *koso* (8.9.2) – as opposed to OJ and parts of early EMJ where it was used outside *kakari-musubi* with *koso* – and the form had already then lost independent function. Although the actual shape is found through most of the LMJ period, it thus already by late EMJ did not represent an independent morphological category, but was simply a conditioned variant of the conclusive, used in agreement with *koso*. In the following we refer to this as the 'exclamatory variant' (of the conclusive). In the course of the loss of *kakari-musubi* (12.6.1.2), the exclamatory variant came not to be used at all and disappeared from the language as a word form. However, it continued to be used as a stem on which some inflected forms were built (see 12.3).

In both *Feiqe* and *Esopo*, the exclamatory variant is found in agreement with *koso*. In *Esopo*, the predicates used with *koso* are mainly grammatical forms which belong to or diachronically reflect the *r*-irr conjugation (such as *are* 'exist', *gozare* 'exist.POL', *nare* copula, *-(I)tare* 'past') or the intentional *-(a)uzure*, but also with those morphemes there are exceptions, where the predicate is not in the exclamatory variant, but the nonpast (see 12.6.1.2.3). In *Feiqe* there are more cases than in *Esopo* of lexical verbs with an exclamatory variant in a predicate correlating with *koso*, but there are also a great deal of exceptions, that is, predicates correlating with *koso* which are not in an exclamatory variant. The use of *koso* in *shōmono* texts is like that in *Esopo*. That suggests that the use of the exclamatory variant in *Feiqe* may have had archaizing features, originating perhaps from the fact that it was a translation or retelling of an early LMJ text. *Kakari-musubi* with *koso* is no longer found in the early NJ *kyōgen* texts.

The loss of the ***conclusive*** and ***adnominal*** categories took the course of a merger between the two forms (henceforth 'concl/adn merger'), in which the

Table 12.2 *Merger of EMJ conclusive and adnominal*

	EMJ		LMJ
	Conclusive	Adnominal	Nonpast
r-irr	*ari*	*aru*	*aru*
n-irr	*sinu*	*sinuru*	*sinuru*
LB	*aku*	*akuru*	*akuru*
UB	*oku*	*okuru*	*okuru*
s-irr	*su*	*suru*	*suru* (~ *su*)
k-irr	*ku*	*kuru*	*kuru*
QD	*kaku*	*kaku*	*kaku*
UM	*miru*	*miru*	*miru*
LM	*keru*	*keru*	*keru*

adnominal increasingly came to be used to conclude main clauses, outside *kakari-musubi*. Such conclusive uses of the adnominal are found in small numbers in OJ and increasingly in the first half of EMJ, where, however, the use of the adnominal in this position is said functionally to have been exclamative (cf. 8.9.4). However, through late EMJ and LMJ, the use of adnominals in conclusive position increased greatly. In the course of this the exclamative effect faded and was lost and eventually the adnominal replaced the conclusive altogether, resulting in the loss of the morphological distinction between the two forms and their functions, see Table 12.2. This may be thought to have happened in the first half of LMJ. The resulting form took over most of the syntactic functions of both the conclusive and the adnominal, e.g. as predicate in main clauses and in adnominal clauses; this will be discussed in more detail below (12.6.1).

For the *r*-irr verbs, the conclusive was the only form which was distinct from the QD verbs, and as a result of the concl/adn merger, the *r*-irr verbs merged into the QD verbs.

As opposed to other verb classes, the *s*-irr verbs used both conclusive (*su*) and adnominal (*suru*) shapes in variation through LMJ, with both variants being used in both adnominal and main clauses, but towards the end of the period *suru* was settled upon as the main variant. However, a reflex of the conclusive is still found as a variant at the end of the period, especially when followed by negative conjectural *mazii* ~ *mai* (*su mazii*, *su mai* 'probably doesn't do') and prohibitive *na* (*su na* 'don't do', which however alternated with *suru na*); these usages of the variant *su* are still found in cNJ. Particularly the EMJ intentional auxiliary *-(a)ũze-* (8.5), which belongs to the *s*-irr conjugation, used both original conclusive (*-uzu*) and adnominal (*-uzuru*) shapes until the end of LMJ; from the second half of LMJ both shapes were used as

Table 12.3 *Loss and changes of EMJ auxiliaries*

EMJ	LMJ	NJ
-(a)zu negative	*-(a)n-* negative	–
-(i)n- perfective	–	–
-(i)te- perfective	–	–
-er- stative	–	–
-(I)tar- stative >	*-(I)ta* past/perfect	*-(I)ta* past
	-(I)tara(ba) past conditional	*-(I)tara* conditional
	(*-(I)tari* representative)	*-(I)tari* representative
-(i)ki simple past >	–	–
-(i)ker- modal past >	–	–
-(a)m- conjectural >	*-(a)u* volitional	*-(y)oo* volitional
-(a)ũze- intentional >	*-(a)uzuru* intentional	–
-(I)tara-m- stative-conjectural	*-(I)tarɔɔ* past conjectural	*-(I)taroo* past conjectural

variants in both main and adnominal clauses. As late as in *Esopo, -uzu* is used in sizeable numbers, although the main variant is *-uzuru*, and there are tendencies in the use of the variants which seem to reflect the original distribution.

12.1.3 Loss and change of auxiliaries

In the course of LMJ, most of the remaining auxiliaries described in 3.1.4.5 were either lost, or changed from auxiliary to flective (recall that subjunctive *-(a)masi* was lost already in EMJ). The only exception is the negative, which remained an auxiliary into NJ (12.1.5.1). Of the OJ auxiliaries in 3.1.4.5, most were lost entirely from the language in the course of LMJ, leaving productive reflexes of only two of these auxiliaries which were so important to the language in OJ and EMJ: (a) conjectural *-(a)m-* gave the volitional ending *-(a)u* which is the reflex of the *onbin* variant of the EMJ conjectural conclusive *-mu/ũ*. In NJ *-(a)u* > *-(y)oo* (see 15.1.1). And (b) EMJ stative *-(I)tar-* is reflected in the past/perfect flective *-(I)ta* (> NJ *-(I)ta*), in the past conditional *-(I)tara(ba)* (> NJ *-(I)tara* conditional), and in the NJ representative *-(I)tari* of which there are a few examples in LMJ, but which does not become established as an inflected form until NJ. The combination of stative and conjectural, *-(I)tara-m-* is reflected in the LMJ past conjectural *-(I)tarɔɔ* (> NJ *-(I)taroo* past conjectural). In addition, the (late) EMJ intentional *-(a)ũze-* (see 8.5) was reflected as an inflectional ending *-(a)uzuru*. The lost auxiliaries are individually reflected sporadically in some dialects, but the OJ and EMJ systems of auxiliaries are not reflected in any dialect.

12.1.3.1 Morphological categories

The perfective auxiliaries, and the category of perfective, were lost altogether. Already in EMJ the two perfectives had partially acquired different functions and distribution (see 8.4.2), with *-(I)te-* being used as a recent past in some contexts; through LMJ both auxiliaries declined in use and were eventually lost. At the end of the period, *-tu* (the conclusive of *-(I)te-*) is reflected in two uses: (a) in alternative or enumerating constructions V_1*-tu* V_2*-tu se-* 'do V_1 and V_2', functionally corresponding to the NJ representative in *-(I)tari* (see 15.1); and (b) in the flective *-(I)turɔɔ* '(recent) past conjectural' (see further 12.1.4).

The EMJ auxiliaries expressing past tense were lost in the course of LMJ and the past tense function shifted to the reflex of the EMJ stative *-(I)tar-*. The two EMJ past tense auxiliaries, *-(i)ki* simple past and *-(i)ker-* modal past, and the use of *-(I)te-* as a 'recent past' auxiliary, continued in the first half of LMJ, and after the concl/adn merger the EMJ adnominal shapes of the simple past, *-si*, the modal past, *-keru*, and the recent past, *-turu*, were used, but they were lost before the end of the period. As mentioned, the recent past was reflected in late LMJ *-(I)turɔɔ* '(recent) past conjectural'. The simple past is reflected in some dialects, but is lost from standard NJ. The modal past is reflected in the emphatic particle *-kke* which is used in cNJ after past *-(I)ta* (*-(I)ta-kke*) and copula *da* (*da-kke*), reflecting LMJ *-takkeru* (< *tari-keru*) and *dearі-keru*, with drop of final *-ru* (see 12.4.1).

Alongside the decline of the earlier past tense markers, the reflex of the EMJ stative *-(I)tar-* was reinterpreted from an aspectual marker to combine the functions of past tense and perfect. The timing and course of this are difficult to establish, as is the causal relationship, if any, between the decline and loss of the EMJ past tense marker and the acquisition of past tense function of the EMJ stative. Already in EMJ, *-(I)tar-* was spreading at the expense of the older OJ stative *-yer-*, which started losing its productivity in the EMJ period, although it remained a part of the written classical language. With the loss of *-yer-* and the change of *-(I)tar-* past/perfect, the functions associated with the statives (see 3.1.4.7) came to be expressed solely by analytic constructions, mainly gerund *-(I)te* + existential verb (see 12.4). Identical constructions (*-(i)te ar-*) were the OJ origin of the EMJ auxiliary *-(I)tar-* (8.4.2), and clearly this is a pervasive and recurring constructional pattern in the language.

The OJ and EMJ conjectural auxiliary *-(a)m-* gave the LMJ volitional form *-(a)u*. The OJ/EMJ conjectural had both volitional and conjectural uses, but these were now divorced so that *-(a)u* generally came to be used in volitional and future functions. However, with existential verbs and the regular and adjectival copula, the reflex of the EMJ conjectural retained conjectural function. The combination of stative (or past) and conjectural, *-(I)tara-m-* became

a past conjectural, *-(I)tarɔɔ*; see further 12.1.4 about the expression of conjectural meaning. Note that even in cNJ the volitional is occasionally used in conjectural function in classical-flavoured writing.

The EMJ intentional auxiliary *-(a)ũze-* (see 8.5) > early LMJ *-(a)uze-* gave the late LMJ intentional form *-(a)uzuru* which overlapped functionally to some extent with volitional *-(a)u*. The intentional disappeared in the course of NJ and is no longer found in cNJ.

12.1.3.2 From auxiliary to flective; inflection for tense

A significant part of the changes affecting the EMJ morphological system was a shift of EMJ *-(I)tar-*, *-(a)m-*, *-(a)ũze-* and *-(I)tara-m-* from auxiliary to flective: LMJ *-(I)ta*, *-(a)u*, *-(a)uzuru*, *-(I)tarɔɔ*, and the concomitant change of the outcome of the concl/adn merger to a nonpast tense.

The pivot in this change was the concl/adn merger. Conclusive and adnominal were the only finite categories these auxiliaries inflected for – they had no imperative and the exclamatory was at the beginning of LMJ no longer a distinct morphological category, but a variant of the conclusive (12.1.2). Thus, once the conclusive and adnominal merged, these auxiliaries no longer had distinct forms expressing distinct finite morphological categories, and they became flectives.

This was, however, not simply a shift in morpheme type of these individual morphemes. It first and foremost constituted a change in the categories expressed by basic inflected forms in the system. Before the concl/adn merger both lexical verbs and auxiliaries inflected for conclusive and adnominal; when that distinction merged and the auxiliaries changed to flectives, the outcome of the concl/adn merger in the lexical verbs, the flective *-u* ~ *-uru*, no longer *combined* with the reflexes of *-(I)tar-*, *-(a)m-* and *-(a)ũze-* to form one of their inflected forms, but instead came to *form oppositions* with them – it was reinterpreted as a nonpast, in opposition with past, intentional, volitional and imperative.

This is shown in Table 12.4. (See further about the shapes of the LMJ past in 12.3.2 and 12.3.3, and volitional and intentional in 12.3.3.) In this way, verbs came to inflect for categories which earlier had been expressed by auxiliaries, first of all tense; thus the representation of these categories became obligatory, whereas it earlier had been optional (cf. 3.1.4).

12.1.3.3 Non-finite forms

The new inflected tense distinctions were also represented among the non-finite forms, see (1). This extensive system reflects the origin of the tense markers in auxiliaries and arose in the transition between auxiliarization

Table 12.4 *Changes of EMJ auxiliaries to LMJ flectives*

		EMJ			LMJ	
kak-	Imperative	*kake*	>		*kake*	Imperative
'write'	Conclusive	*kaku*	>			
	Adnominal	*kaku*	>		*kaku*	Nonpast
kai-tar-	Conclusive	*kai-tari*	>			
+ STAT	Adnominal	*kai-taru*	>	*kaitaru* >	*kaita*	Past
kaka-m-	Conclusive	*kaka-ũ*	>			
+ CONJ	Adnominal	*kaka-ũ*	>	*kakau* =>	*kakɔɔ*	Volitional
kaka-ũze-	Conclusive	*kaka-ũzu*	>			
+ INT	Adnominal	*kake-ũzuru*	>	*kakauzuru* =>	*kakɔɔzuru*	Intentional
kai-tara-m-	Conclusive	*kai-tara-ũ*	>			
+ STAT-CONJ	Adnominal	*kai-tara-ũ*	>	*kaitarau* >	*kaitarɔɔ*	Past conjectural
ake-	Imperative	*akeyo*	>		*akei*	Imperative
'open'	Conclusive	*aku*	>			
	Adnominal	*akuru*	>		*akuru*	Nonpast
ake-tar-	Conclusive	*ake-tari*	>			
+ STAT	Adnominal	*ake-taru*	>	*aketaru* >	*aketa*	Past
ake-m-	Conclusive	*ake-ũ*	>			
+ CONJ	Adnominal	*ake-ũ*	>	*akeu* =>	*akyoo*	Volitional
ake-ũze-	Conclusive	*ake-ũzu*	>			
+ INT	Adnominal	*ake-ũzuru*	>	*akeuzuru* =>	*akyoozuru*	Intentional
ake-tara-m-	Conclusive	*ake-tara-ũ*	>			
+ STAT-CONJ	Adnominal	*ake-tara-ũ*	>	*aketarau* >	*aketarɔɔ*	Past conjectural

and inflection. It was greatly simplified through early NJ, to comprise only a non-tensed provisional and a conditional, see 15.1. In LMJ the past conditional was found both in its full shape, *-(I)taraba*, and in the abbreviated form *-(I)tara*.

(1)

	Nonpast	Past	Intentional
Provisional	*kakeba*	*kaitareba*	*kakɔɔzureba*
Conditional	*kakaba*	*kaitara(ba)*	–
Concessive	*kakedomo*	*kaitaredomo*	*kakɔɔzuredomo*
Provisional	*akureba*	*aketareba*	*akyoozureba*
Conditional	*akeba*	*aketara(ba)*	–
Concessive	*akuredomo*	*aketaredomo*	*akyoozuredomo*

12.1.4 Conjectural

As mentioned above, the OJ/EMJ conjectural auxiliary *-(a)m-* changed to a flective *-(a)u* which only had volitional function with most verbs, hence the emergence of volitional forms of verbs (12.1.3.2). However, with existential verbs, and constructions involving existential verbs, OJ/EMJ *-(a)m-* only had conjectural function, and that also held for the new flective *-(a)u* when used with existential *ar-* and with forms built on *ar-*. Thus, EMJ *ara-ũ* 'exist-conjectural' > LMJ *arau* > *arɔɔ* which was used in conjectural function, both as an independent existential verb, and in its grammaticalized uses to form analytic adjectival and regular copula forms and after verb gerunds to form an analytic stative. Similar developments hold for the EMJ stative conjectural *-tara-ũ* > LMJ past conjectural *-(I)tarau* > *-(I)tarɔɔ*, and for the secondary adjectival and regular copula conjugations: *-kara-ũ* > *-karau* > *-karɔɔ* and *nara-ũ* > *narau* > *narɔɔ*. With *ar-* and the secondary adjectival and regular copula conjugations, the intentional, *-(a)ũze-*, was also used, and was reflected in forms such as EMJ *ara-ũzu(ru)* > LMJ *arauzu(ru)* > *arɔɔzu(ru)*, and adjectival *-karɔɔzu(ru)* and copula *narɔɔzu(ru)*, but with no discernible functional difference from the conjectural.

In addition to the conjectural auxiliary *-(a)m-*, the extension *ram-* already in OJ provided an alternative, unambiguous expression of the present conjectural (3.6). In the course of the *onbin* sound changes the conclusive and adnominal *ramu* changed to *ran* ~ *raũ* > early LMJ *rau* > late LMJ *rɔɔ*, which eventually became a particle due to the loss of the exclamatory as a distinct morphological category (12.1.2). It was used with the nonpast of verbs in general and with intentional *-(a)uze-* (*-uzu rɔɔ*), and with recent past *-(I)te-*; here the combination was univerbated as a past conjectural *-(I)turɔɔ*, which continued in use after the loss of recent past *-(I)te-*. In late LMJ *rɔɔ* was frequent in the *shōmono* texts and was also used in *Esopo* and *Feiqe*, and it became the main expression of the category of conjectural.

In late LMJ there were thus two sets of expressions of conjectural which seem to have been more or less synonymous: (a) use of *rɔɔ* as a final particle (and in an inflectional past conjectural ending), and (b) a number of specific inflected forms, reflecting OJ *ara-mu*, see (2).

In NJ, *rɔɔ* is no longer used, but has been functionally replaced by *darō* or its polite variant *deshō* (both originating as conjectural forms of the copula) following a finite verb (*kaku darō*, *kaita darō*) or adjective form (*akai darō*), or directly following a nominal predicate (*hon darō*). In addition to *darō* (and other copula forms, such as *de arō*), adjectival copula *-karō* and past conjectural *-(I)tarō* are still used, in old fashioned writing and productively in some dialects.

(2)

	a. *rɔɔ* (< *rau* < *raũ* < *ramu)*	b. *arɔɔ (< ara-u < ara-ũ < ara-mu)*
Verb, nonpast	*(kaku) rɔɔ*	–
ar-	*aru rɔɔ*	*arɔɔ*
Adjectival copula	*-ku aru rɔɔ (~ -karu rɔɔ)*	*-karɔɔ*
Copula	*de aru rɔɔ*	*narɔɔ*
Verb, past	*-(I)turɔɔ*	*-(I)tarɔɔ*

cNJ *darō* does not seem to be a direct reflex of a form like *dara-u, but rather an analogical formation. Part of the background for that is probably the extensive use of *rɔɔ* and also that the synchronic segmentation of the past conjectural, *-(I)tarɔɔ*, in late LMJ was not obvious: was it past *-(I)ta* + *rɔɔ*, conforming to the pattern with nonpast verbs (LMJ *kaku rɔɔ*); or was it an unanalysable single form (conforming to diachronic formations of past conjectural forms through the history of the language: OJ *-(i)kye-m-* > EMJ *-(i)kem-*, replaced by LMJ *-(I)tu(-)rau* > *-(I)turɔɔ*, further replaced by *-(I)tara-u* > *-(I)tarau* > *-(I)tarɔɔ* > NJ *-(I)taroo* > cNJ *-(I)tarō*). Note also in this connection early LMJ *yarau* (> LMJ *yarɔɔ*) which originates in a contraction of EMJ *ya ara-ũ*, the *kakari* particle *ya* followed by *ar-* + conjectural. *Yarɔɔ* is a particle with the meaning 'is it, I wonder?' used after nominal predicates and after verbs, both nonpast and past, and thus functionally corresponding to a conjectural particle and copula form. *Yarɔɔ* could also have served as a model for analogical formation of a form such as *daroo*; further, *yarɔɔ* is possibly the source of the modern Kansai copula *ya* which could have arisen in a reanalysis of *yarɔɔ* as the conjectural form of *ya*, with further forms such as past *yatta* and negative *ya na-* being analogical formations.

12.1.4.1 Negative conjectural

Negative conjectural was expressed by reflexes of the EMJ extension *mazi-* (8.5): *mazii* which directly reflects the EMJ adnominal *mazi-i* and a further reduced form *mai*. Both forms were used in attributive and predicative function, though usually the latter. Like *rɔɔ*, *mazii* ~ *mai* changed from extension to particle in the course of the concl/adn merger. *Mazii* ~ *mai* followed the nonpast of consonant base verbs (e.g. *fairu mai*, *fair-* 'enter'); with vowel base verbs (including the passive and causative auxiliaries which belonged to the LB conjugation) it usually followed the base (*uke mai*, *uke-* 'receive'; *fatasare mai*, *fatas-are* 'fulfil-PASS'), or with *se-* 'do' and *ko-* the old conclusive (*su mai*, *ku mai*). *Mai* survives into cNJ, but mainly in conservative writing. Negative conjectural is instead generally expressed by *darō/deshō* following a negative form.

Table 12.5 *Main forms of the negative at the end of the LMJ period*

		kak-	*ake-*
Nonpast	*nu*	*kakanu*	*akenu*
Past	*nanda*	*kakananda*	*akenanda*
Past conjectural	*nandarɔɔ*	*kakanandarɔɔ*	*akenandarɔɔ*
Infinitive	*zu*	*kakazu*	*akezu*
Past infinitive	*nande*	*kakanande*	*akenande*
Gerund	*ide*	*kakaide*	*akeide*
Concessive	*nedomo*	*kakanedomo*	*akenedomo*
Provisional	*neba*	*kakaneba*	*akeneba*
Past provisional	*nandareba*	*kakanandareba*	*akenandareba*
Conditional	*zuwa, zunba*	*kakazuwa, kakazunba*	*akezuwa, akezunba*

12.1.5 Other auxiliaries

12.1.5.1 Negative

Of the OJ auxiliaries discussed in 3.1.4.5, only the negative has remained an auxiliary through the history of the language, attaching to the *a-* stem of consonant base verbs and the base of vowel base verbs. Main forms at the end of the period are shown in Table 12.5. The negative inflects for many of the categories that verbs inflect for. Note that the paradigm of the negative was completely replaced in the standard language in NJ, see 16.3.

The late LMJ negative paradigm is more suppletive than the EMJ negative paradigm (8.2.2). The nonpast directly reflects the EMJ adnominal and several other forms are direct continuations of the EMJ forms. However, the paradigm also includes a number of new forms. Interestingly, their etymologies are all unknown and they all went out of use during NJ in the common language (but are retained dialectally): The gerund *-ide* is similar to the EMJ negative gerund *-de*, but any derivational relationship is unclear. A number of negative past tense forms (past *-nanda*, past conjectural *-nandarɔɔ*, past infinitive *-nande*, past provisional *-nandareba*) reflect a past negative auxiliary, *-(a)nandar-*, which came into use in the first half of LMJ. It has the basic paradigm shown in (3a) below. Alongside other auxiliaries it changed from an auxiliary to give a number of inflected forms in the second half of LMJ, but its basic paradigm already shows signs of that process being under way. Past *-nanda* continued in use into the nineteenth century. Etymologically, *-(a)nandar-* seems to involve a combination of negative *-(a)n-* and the stative (past) auxiliary *-(I)tar-*, but the details of the composition of the form are entirely unclear.

The secondary negative auxiliary *-(a)zar-*, whose basic paradigm is repeated in (3b), was frequent in EMJ and in the first half of LMJ; for example, the depiction of the three monkeys which 'see no evil, hear no evil, and speak no evil' (*mizaru, kikazaru, iwazaru*) involves a pun on the word *saru* 'monkey'

and the negative *-zaru*. The adnominal *-zaru* came to be used in both adnominal and conclusive function as the two forms merged (12.1.2). Combinations with the reflex of the EMJ simple past adnominal *-zassi* (< *zari-si*) and with the new past tense *zatta* (< *-zari-ta(ru)*) were used through LMJ, but from the middle of the period started being replaced by *-(a)nandar-* which came into the language then. Use of *-(a)zar-* decreased during late LMJ and it is not reflected in the paradigm in Table 12.5. It was, however, retained in use in classical-flavoured writing.

(3)

	a. LMJ negative past *-(a)nandar-*	b. Secondary negative *-(a)zar-*
a- stem	*nandara-*	*zara-*
onbin stem	–	*zaQ-*
Infinitive	*nande (~ nandari)*	*zari*
Conclusive	*nanda*	*zari*
Adnominal	*nanda (~ nandaru)*	*zaru*
Exclamatory	*nandare*	*zare*
Imperative	–	*zare*

12.1.5.2 Passive and causative

Passive *-rare-* and causative *-sase-* remained in use, attaching to host verbs in the same way as in EMJ and belonging to the LB conjugation, that is, with the same forms as lexical LB verbs, e.g. nonpast *-sasuru*, *-raruru*; past *-saseta*, *-rareta*; volitional *-sasyoo*, *-raryoo*, etc.; especially in late LMJ, a number of irregular 'analogical' shapes of these two auxiliaries are found, e.g. *-sasase-* or *-rarare-*. Causative and passive basically had the same functions as in EMJ (8.4.1), with the notable exception that their exalting function became independent so that both passive and causative were now used to express respect on their own (cf. 12.7.1.2), as opposed to EMJ where they supported other respect morphemes. In late LMJ causative + passive, *-sase-rare-*, was often used to express a higher level of respect. Probably related to that are other multiple occurrences of these two auxiliaries seen in the texts, though an example such as *ut-are-rare-sase-rareta* 'was stricken.RESP' (with a double passive, *-are-rare*, expressing passive voice, and causative-passive, *-sase-rare-*, expressing respect) is very rare.

12.1.5.3 New auxiliaries

A number of new auxiliaries came into use during LMJ. New auxiliaries in use at the end of the period are summarized in (4):

(4)

Potential	*-e-*
Desiderative	*-(i)ta-*
Evidential	*-(i)sɔɔ-*
Evidential	*-(i)ge-*
Polite	*-(i)marase-*

The ancestor of the cNJ potential auxiliary *-re-* makes its appearance in the second half of LMJ. It is found with consonant base verbs in the shape *-e-*, e.g. *yom-e-nu* 'cannot read' in the *Shiki-shō* from 1477, and initially had the same functions as the passive: passive, potential, and respect. *-E-* is usually said to have developed as a reduced variant of the passive *-rare-* and for example in the *Shiki-shō* we find both *yom-e-nu* and *yoma-re-nu* 'read-PASS-NEG; cannot read'. Alternatively, *-e-* may have developed from the auxiliary verb *-e-* 'be able to' (< *e-* 'get'). During NJ this auxiliary became specialized as an independent potential auxiliary, but only recently in cNJ was it analogically extended to be used with vowel base verbs, e.g. cNJ *tabe-re-* 'be able to eat'.

Several other new auxiliaries involve in their history a shift of a lexical item to grammatical function. Desiderative *-(i)ta-* (thought to derive from the adjective *ita-* 'intense, much, wonderful') became firmly established as an auxiliary during early LMJ, eventually replacing EMJ *-(a)maosi-*. As in cNJ, *-(i)ta-* had adjectival inflection and attached to the infinitive of verbs, e.g. early LMJ conclusive *kaki-ta-si, ake-ta-si*, adnominal *kaki-ta-i, ake-ta-i*. During LMJ, VERB-*ta-* expressed both 'want to VERB' and 'want somebody else to VERB', but in NJ specialized to mean '(I) want to VERB'. Already in LMJ *-(i)ta-* was used in combination with the evidential derivative *-gar-*: *kaki-ta-gar-* 'he seems to want to write, he wants to write'.

From late LMJ, the early LMJ noun *sau* (which is thought to be a reduced form of either *sama* 'appearance, shape, state' or SJ *sau* < *saũ* 'appearance, shape' (相, EMC **siaŋ/siaŋ^{h} 'look at, inspect /observe, judge the appearance') became used as an evidential auxiliary about appearance of state or imminence: 'looks/seems as if/as if about to', attaching to the infinitive of verbs and the stem of adjectives, forming an adjectival noun, i.e., combining with the copula: *-(i)sau-nar-* > *-(i)sɔɔ-na* (>NJ *-(i)soo-da*), e.g. *taka-sɔɔ-na* 'it looks tall', *kaki-sɔɔ-na* 'he looks, seems about to go'. This is still used widely in NJ. In NJ *soo-na* further developed uses as a hearsay extension (15.3).

Also two polite style auxiliaries have lexical sources (see 12.7.2): *-(i)marase-*, which is the ancestor of cNJ *-(i)mas-*, developed during late LMJ; it is from *mawirase-* 'give.HUM', a lexicalized causative of *mawir-* 'come.HUM, go.HUM'. Before then, *-(i)sɔɔrɔɔ* was used as a polite auxiliary but had gone out of use by the end of LMJ; it is from *sauraw-* (a suppletive polite

equivalent of the existential verb *ar-*) < early EMJ *saburap-*, said to be from OJ *samorap-* 'serve, be in attendance')

The EMJ derivational morpheme *-ge-(nar-)* which derived an adjectival noun from adjectives (8.6), came to be used more widely as an evidential auxiliary, *-(i)ge-(na)* 'appears, seems', attaching to the infinitive of verbs and to the stem of adjectives. It also developed uses as an extension, with the same meaning, but following finite verb forms, see for example (7b) in 12.2.2.

12.1.6 Extensions

Of the EMJ extensions (see 8.5), evidential *nar-* (sound, hearsay) went out of use after EMJ, and evidential *mer-* (sight) disappeared in the first half of LMJ. Present conjectural *ram-* and negative necessitive *mazi-* changed to particles as mentioned above (12.1.4). Necessitive *be-* was used through the period: there are tendencies already during LMJ to use nonpast *bei/bee* (< adnominal *beki*) and infinitive *byoo* (< *beu* < *beku*) as particles, but on the whole *be-* still at the end of the period had productive adjectival inflection. In addition to EMJ *bekasi-* (8.2.2, 8.5), a number of other analogical forms are found through the period, e.g. *bera-nar-*, *besii-*.

The SJ noun EMJ *yaũ* 'appearance, shape' (様 EMC **jɨaŋʰ*), which already in EMJ had conventionalized uses, early in LMJ acquired fully grammaticalized functions as an extension: early LMJ *yau-nar-* > late LMJ *yɔɔ-na* (> NJ *yoo-da*), belonging to the class of adjectival nouns. Initially its uses were similar to *goto-* '(is) like', but eventually it became an evidential 'seems (as if), appears'. It is still used as an evidential in cNJ, *yō-da*, and its infinitive, *yō-ni*, has further been grammaticalized as a complementizer and a conjunctional particle.

12.2 Adjectives and copula

In EMJ the adjectival copula and the regular copula both had a highly suppletive paradigm of primary forms, as well as secondary, fully inflected verbal forms: adjectival copula *-kar-* and copula *nar-/tar-* which arose from lexicalization of phonological fusion of *-ku ar-* and *ni/to ar-* (8.2.1). By the end of LMJ the two paradigms of primary and secondary forms had, for both the adjectival and the regular copula, merged, or were in the final stages of merging, into one suppletive paradigm which reflects both primary and secondary forms and which incorporates some of the new categories for which verbs came to inflect, see 12.2.1 and 12.2.2 below. Like the verbs, the adjectival and regular copula at this stage both still retained direct combination with the negative auxiliary. As in EMJ, the frequent use of analytic forms, *-u/-ku ar-* and *de/nite/ni ar-*, continued, supplementing the new paradigms, using also

Table 12.6 *Late LMJ adjectival copula forms*

Finite	
Nonpast	*i (~ karu) ~ kere*
(Past)	*(katta)*
Conjectural	*karɔɔ ~ karɔɔzuru*
Imperative	*kare*
Non-finite	
Infinitive	*u ~ ku*
Gerund	*ute ~ kute*
Conditional	*kuwa ~ kunba ~ utewa ~ kutewa*
Provisional	*kereba*
Concessive	*keredomo ~ utemo ~ kutemo*
Negative	*kara-n-*

other existential verbs than the neutral *ar-*, particularly polite *gozar-*. Note that towards the end of the period the negative adjective *na-* had become regularly used as a suppletive negated form of *ar-* and that analytic negation generally used *na-*, rather than a negative form of *ar-* (such as *aran*), thus negative adjectival copula *-u/-ku na-*, negative copula *de(wa) na-*; cf. also 16.3.

12.2.1 Adjectives

For the adjectival copula, the concl/adn merger left nonpast *-i* as the only reflex of the OJ conclusive and adnominal formants, (5a). With no reflex of the conclusive form, which is what distinguished the so-called *ku* and *shiku* adjectives (3.2.4), the difference between these two classes of adjective disappeared and was reflected only as a difference in stem shape, i.e., whether the stem ends in *-si* or not, (5b):

(5)		OJ	EMJ		LMJ	
a.	Conclusive	*si ~ Ø*	*si ~ Ø*	>		
	Adnominal	*ki*	*i*	>	*i*	Nonpast
b.	Conclusive	*taka-si*	*taka-si*	>		
	Adnominal	*taka-ki*	*taka-i*	>	*taka-i*	Nonpast
	Conclusive	*utukusi-*	*utukusi-*	>		
	Adnominal	*utukusi-ki*	*utukusi-i*	>	*utukusi-i*	Nonpast

The main late LMJ forms and variants of the adjectival copula are given in Table 12.6.

At this stage, the adjectival copula had most of the categories for which verbs inflect. However, past tense was usually formed analytically, e.g. *-u/-ku*

atta or *-u/-ku gozatta*. Past *-katta* was not yet well established; it was rare and is found mostly with *na-* 'not exist' which in LMJ seems to have been close to having been lexicalized as a verb *nakar-*, from the secondary conjugation (*na-kar-*). It never made the transition fully, but did exhibit forms which were not generally found for other adjectives such as *nakatta* or *nakaru*. The nonpast variant *-karu* reflects the EMJ secondary adnominal and was rare in late LMJ; it was used in both attributive and predicative function, and also with the conjectural particles *rɔɔ* and *mazii* (12.1.4). Also imperative *-kare* reflects the secondary conjugation. *-kere* reflects the EMJ exclamatory and was used as a variant of the LMJ nonpast in agreement with *koso*. Morphologically *-karɔɔ* ~ *-karɔɔzuru* reflect combinations of the secondary conjugation *-kar-* with the EMJ auxiliaries conjectural *-(a)m-* and intentional *-(a)ũze-* and thus correspond to the volitional and intentional in the verbal conjugations, but functionally *-karɔɔ* ~ *-karɔɔzuru* are both conjectural. Of the two infinitive variants, *-u* and *-ku*, the *-u* variant was used more, but both are found in the same environments, for example in analytic constructions with an existential verb or the negative adjective *na-*, e.g. past *-u atta* ~ *-ku atta* or negative *-u na-* ~ *-ku na-*, or as adverbial or non-finite forms. The main exception seems to have been when used as a stem in the formation of the two conditionals where only *-ku* was used: *-kuwa* ~ *-kunba*; of these, *-kunba* is more typical of *kanbun-kundoku*. Like the verbs, the adjectival copula had new conditional and concessive forms built on the gerund: *-utewa* ~ *-kutewa* and *-utemo* ~ *-kutemo*.

The adjectival copula still combined with the negative as *-karan-*, but increasingly negation was expressed by analytic forms, e.g. *-u/-ku na-*, as is the case in NJ.

Note finally that the concessive adjectival form *-keredomo* already in late LMJ came to be used as an independent conjunctional particle *keredomo* 'but, however', with later frequent NJ reduced forms *kedo*, *keredo*.

12.2.2 Copula

Main inflected forms and variants of the regular copula at the end of LMJ are given in Table 12.7.

The inflected copula *tar-*, which was restricted in use all along (8.2.3), was lost altogether outside fossilized expressions (e.g. in some adnouns such as *kaku-taru* 'certain'), but infinitive *to* remained. The forms in Table 12.7 reflect EMJ primary forms, EMJ secondary forms (from *nar-*) and new secondary forms which arose in late LMJ: *dyar-* (< *de ar-*). The two secondary conjugations thus differ in whether they are built on the infinitive *ni* or on the gerund *de* (< *nite*).

The nonpast forms in the LMJ paradigm are abbreviated reflexes of adnominal secondary forms, with loss of final *-ru* (see 12.4.1): *na* < *naru* and *dya* <

Table 12.7 *Late LMJ copula forms*

Nonpast	*na ~ dya (~ dea ~ da) ~ nare*
Attributive	*no*
Conjectural	*narɔɔ*
Infinitive	*ni ~ to*
Gerund	*de ~ nite*
Conditonal	*naraba (~ dewa)*
Provisional	*nareba*
Concessive	*naredomo ~ demo*
Negative	*nara-n-*

dyaru < *de aru* (and *dea* < *de aru*). Nonpast *da* is also thought to derive from *de aru*; it is a marginal form attested but a few times in *shōmono* texts and only included here because it is the form found in cNJ (15.2, 16.2). Although there was a tendency for *dya* to be used in main clauses and *na* adnominally, both could be used in either function, see (6) and (7). As in OJ/EMJ (and today in cNJ), LMJ had a distinct attributive copula form *no*, which was restricted in use and differed from the verbal OJ/EMJ adnominal and LMJ nonpast forms in being used only in adnominal function (cf. 3.3.1). A reflex of the EMJ exclamatory *nare* was used as a variant of the nonpast in agreement with *koso* until the end of the period (12.6.1.2.3).

(6) a. wasi wa syotyoo no ɔɔ **dya**
I TOP birds GEN king COP.NONPST
'I am king of the birds' (*Esopo*)

b. kono koto wa asa-kara-nu
this thing TOP shallow-ACOP-NEG.NONPST
fusin **dya** fodo ni
suspicion COP.NONPST extent COP.INF
'because this is highly suspicious' (*Esopo*)

(7) a. kata-me **na** sika
one-eye COP.NONPST deer
'the one-eyed deer'

b. kono mono wa nao isyu ga
this person TOP furthermore concern NOM
gozaru ge-**na**
exist.POL EVID-COP.NONPST
'It appears that this person furthermore has some concern' (*Feiqe*)

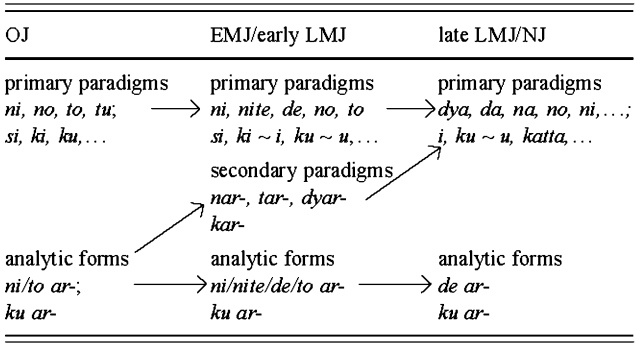
Naraba, *nareba*, *naredomo* reflect the EMJ inflected copula *nar-*, as do combinations with the negative auxiliary (12.1.5.1), such as nonpast *naranu* 'it isn't' or past provisional *naranandareba* 'as it wasn't'. Of the two gerunds *de* was used far more widely, but its source *nite* is still used in some texts, e.g. *Feiqe*. Etymologically *de* and *nite* are built on infinitive *ni*, but already from late EMJ, *ni* was getting more restricted and *de* or *nite* were used more generally as non-finite copula, 'is and . . .'; this tendency continued into LMJ where *de* also took over the function of being the form on which analytic forms were built: *de ar-* (which as mentioned gave rise to a short-lived secondary form: *dyar-*). The LMJ copula paradigm underwent significant changes in NJ, but the analytic forms were much as they are in cNJ, which for example uses polite negative *dewa arimasen* (or contracted *ja arimasen*), or formal written style copula *de ar-*. In addition to the neutral existential verb *ar-*, other existential verbs were also used to form analytic copula forms, e.g. *de gozar-* which is also used in cNJ, especially as a superpolite copula *de gozaimas-*. While negation could be expressed by the negative auxiliary, *naranu* 'isn't', it was more usually analytically formed, using either a negated form of an existential verb, or the adjective *na-* which in LMJ became the regular suppletive negative form of *ar-* also in these constructions, e.g. *de na-* or *dewa na-*, as in cNJ (cf. 16.3).

Several lexicalized forms reflect the secondary conjugation: *nari-tomo* 'be it, even' (< conclusive *nari* + conjunctional concessive particle *tomo)*, *nara-dewa* 'only, limited to' (< negative gerund *narade* + *wa*), and the conditional particle *naraba*.

12.2.3 Summary of the development of regular and adjectival copula paradigms

The development of the morphology of the regular and adjectival copula through time is strikingly similar; to some extent this is shared by the negative auxiliary. First of all, both regular and adjectival copula have through the attested history of Japanese had analytic constructions consisting of infinitive or gerund and an existential verb. This also holds for the negative in OJ, EMJ and early LMJ. In OJ, regular (3.3) and adjectival (3.2) copula and the negative auxiliary (3.1.4.8) each had a paradigm of primary inflected forms, which was irregular, highly suppletive and, for the copula, very defective. These primary paradigms were supplemented by analytic forms, built on the infinitives *-ku*, *ni/to* and *-zu*. In EMJ both primary and analytic forms continued, with a few changes in the primary paradigms, but in addition lexicalized, fully verbal secondary paradigms emerged, which originated in the analytic forms (8.2). Finally, in late LMJ, the primary and secondary adjectival and regular copula paradigms merged into single suppletive primary paradigms whose forms reflect both the primary OJ/EMJ paradigms and the secondary EMJ paradigms

Table 12.8 *Development of the copula and adjectival copula paradigms*



(12.1.5.1, 12.2.1-2). This final change coincided with and to a large extent paralleled the change of the verbal auxiliaries to inflected forms. The overall developments are summarized in Table 12.8.

12.3 Basic paradigms

The basic paradigms of the different verb classes are shown in Table 12.9. As opposed to the morphological system, the basic paradigms did not change much in the course of LMJ. The exclamatory was lost as an independent, distinct word form, but remained a stem used in the formation of the provisional and concessive verb forms and we now list it as a stem ('*e*- stem'), alongside the *a*-stem and the *onbin* stem of the consonant base verbs, but also for the vowel base verbs. This mode of presentation is an adaptation of the *katsuyōkei*-system used in Japanese school grammar, with the forms slightly rearranged and the *onbin* stem added (cf. 3.4.6.3). For OJ and EMJ this presentation at once captures core inflected verb forms and stems, providing a basis for forming other forms (cf. 3.4.6), but for LMJ (and NJ) it is much less useful. It mainly has the latter function of providing a basis for the formation of morphophonologically more complex forms, but is in that respect more cumbersome than necessary, and for example the presentation of verbs in *Vocabulario* (10.2.2.3) is simpler and does the same job. Finally, if used diachronically, in particular comparing EMJ and LMJ, the presentation in terms of these basic paradigms gives the wrong impression of morphological continuity across these periods and does not capture the sweeping changes which the morphological system underwent between EMJ and LMJ. However, as mentioned before, this mode of presentation is ingrained in traditional Japanese scholarship so we include it for LMJ as well.

Table 12.9 *LMJ basic verb paradigms*

	N-irr	QD	QD	QD
Basic stem	*sin-* 'die'	*mot-* 'hold'	*tor-* 'take'	*ar-* 'exist'
a- stem	*sina-*	*mota-*	*tora-*	*ara-*
onbin stem	*sin-*	*moQ-*	*toQ-*	*aQ-*
e- stem	*sinure-*	*mote-*	*tore-*	*are-*
Infinitive	*sini*	*moti*	*tori*	*ari*
Nonpast	*sinuru*	*motu*	*toru*	*aru*
Imperative	*sine*	*mote*	*tore*	*are*

	QD	QD	QD
Basic stem	*kak-* 'write'	*kog-* 'row'	*sas-* 'pierce'
a- stem	*kaka-*	*koga-*	*sasa-*
onbin stem	*kai-*	*koi-*	*sai-*
e- stem	*kake-*	*koge-*	*sase-*
Infinitive	*kaki*	*kogi*	*sasi*
Nonpast	*kaku*	*kogu*	*sasu*
Imperative	*kake*	*koge*	*sase*

	QD	QD	QD
Basic stem	*yom-* 'read'	*yob-* 'call'	*tow-* 'request'
a- stem	*yoma-*	*yoba-*	*towa-*
onbin stem	*you-* ~ *yon-*	*you-* ~ *yon-*	*tou-* ~ *toQ-*
e- stem	*yome-*	*yobe-*	*toe-*
Infinitive	*yomi*	*yobi*	*toi*
Nonpast	*yomu*	*yobu*	*tou*
Imperative	*yome*	*yobe*	*toe*

	LM	LB	LB	LB
Basic stem	*ke-* 'kick'	*ake-* 'open, (trans.)'	*tae-* (//taye//) 'cease, (v.intr.)'	*kae-* (//kaye//) 'exchange'
e- stem	*kere-*	*akure-*	*tayure-*	*kayure-*
Infinitive	*ke*	*ake*	*tae*	*kae*
Nonpast	*keru*	*akuru*	*tayuru*	*kayuru*
Imperative	*keyo/kei*	*akeyo/akei*	*taeyo/taei*	*kaeyo/kaei*

Table 12.9 (*cont.*)

	UM	UB
Basic stem	*mi-* 'see'	*oki-* 'arise'
e- stem	*mire-*	*okure-*
Infinitive	*mi*	*oki*
Nonpast	*miru*	*okuru*
Imperative	*miyo*	*okiyo*
	s-irr	*k*-irr
Basic stem	*s(e)-* 'do'	*ko-* 'come'
e- stem	*sure-*	*kure-*
Infinitive	*si*	*ki*
Nonpast	*suru/su*	*kuru*
Imperative	*seyo/sei*	*koyo/koi*

12.3.1 Verb classes

As mentioned above, *r*-irr was no longer a separate class, but had become a part of the QD class. Other than that, the verb classes of LMJ are the same as those for EMJ.

Innovating combinations of passive and causative with *s*-irr verbs arose, coexisting with the older combinations, for example of *se-* 'do': do.PASS *sare-* (~ non-innovating *se-rare-*) or do.CAUS *sase-* (~ *se-sase-*). This, along with the persistence of the conclusive shape *su* (12.1.2), is usually interpreted to mean that the *s*-irr verbs were tentatively, at least by some speakers, reinterpreted as consonant base verbs, with the basic stem *s-*.

In EMJ, there were two types of *CVe* LB verbs, underlying //CVye// and underlying //CVe//. The palatal glide in the first type only surfaced when followed by /u/, never when followed by /e/ (because of the loss of /y/ before /e/, see 7.3.2.2). This difference reflects three OJ sources of EMJ LB verbs in *CVe*, as shown in (8), with developments including loss of /y/ (8a), /-p-/ > /-w-/ (8b) (7.3.1.1), and loss of /w/ (8b, c) (7.3.2.3).

At the end of EMJ, type (a) was underlyingly //(C)Vye// (cf. conclusive *tayu*, infinitive *tae*), whereas types (b and c) were //CVe// (conclusive *kau*, *uu*, infinitive *kae*, *ue*). However, by the end of LMJ all three types had become //(C)Vye//, see (9), as shown by uniform nonpast *tayuru*, *kayuru*, *uyuru*. This levelling of all EMJ *CVe* verbs as //CVye// was probably prompted by the automatic phonetic palatal onglide of /.e/, manifested in the infinitive of all forms, regardless of diachronic source, as shown in (9).

(8)	Basic stem	Conclusive	Infinitive
a.	OJ *CVye-* OJ *taye-* > EMJ *tae-* //taye-//: 'cease (intr.)'	*tayu*	*tae* ([ta^{j}e])
b.	OJ *CVpe-* OJ *kape-* > EMJ *kawe-* > EMJ *kae* //kae//: 'exchange (v.tr.)'	*kau*	*kae* ([ka^{j}e])
c.	OJ *CVwe-* OJ *uwe-* 'starve' > EMJ *ue* //ue//:	*uu*	*ue* ([u^{j}e])

(9)	LMJ *CVe* //CVye//	Nonpast	Infinitive
	tae //taye//	*tayuru*	*tae* ([ta^{j}e])
	kae //kaye//	*kayuru*	*kae* ([ka^{j}e])
	ue //uye//	*uyuru*	*ue* ([u^{j}e])

12.3.2 Onbin *stems*

Onbin stems were through LMJ regularly used with simple past adnominal *-si*, prohibitive *na -(I)so*, perfective *-(I)te-* and its reflexes (12.1.3), stative *-(I)tar-* and its reflexes, including past tense *-(I)ta*, and with gerund *-(I)te*. Both vocalic and consonantal *onbin* forms were found with *-w*, *-b*, *-m* base verbs, sometimes of the same verbs within one text, for example in the *shōmono*; however, the Christian sources from the end of the period use almost exclusively the vocalic forms (see 16.1). Following the loss of phonemic nasality in moraic vowels and of the phonological rule of postnasal neutralization (11.1), morphophonological rules arose, like those of cNJ, which changed initial /t/ in the suffixes above to /d/ when used after *onbin* stems from *-b-*, *-g-*, *-m-*, *-n-*stems, i.e. those stems that gave nasal *onbin* in EMJ. Thus, the phonemic shapes of the gerunds did not change, but they were now produced by morphophonological rules which made reference to the final consonant of the basic stem of the verb, rather than to the final moraic consonant or vowel in the *onbin* stem.

(10)	Basic stem	*Onbin* stem	Gerund
a.	*yob-* 'call'	*you-* ~ *yon-*	*yoode* ~ *yonde*
	kog- 'row'	*koi-*	*koide*
	yom- 'read'	*you-* ~ *yon-*	*yoode* ~ *yonde*
	sin- 'die'	*sin-*	*sinde*

b.	*tow-* 'ask'	*tou-* ~ *toQ-*	*toote* ~ *totte*
	mot- 'hold'	*moQ-*	*motte*
	kak- 'write'	*kai-*	*kaite*
	kos- 'cross'	*koi-*	*koite*
	tor- 'take'	*toQ-*	*totte*

12.3.3 Morphophonological rules

The monophthongizing sound changes described above in 11.5 gave rise to the morphophonological rules in (11), applying to verb and adjectival copula forms in *-u*, see (12)–(15).

(11) a. /-iu/ => /-yuu/
b. /-eu/ => /-yoo/
c. /-au/ => /-ɔɔ/
d. /-ou/ => /-oo/
e. /-uu/ => /-uu/

(12) Verbal *u- onbin* stem

	Basic stem	*Onbin* stem	Gerund	
a.	*iw-* 'say'	*iu*	*iu-te* =>	*yuute*
	kanasim- 'be sad'	*kanasiu*	*kanasiu-de* =>	*kanasyuude*
b.	*ew-* 'get drunk'	*eu*	*eu-te* =>	*yoote*
c.	*kaw-* 'buy'	*kau*	*kau-te* =>	*kɔɔte*
	kam- 'bite'	*kau*	*kau-de* =>	*kɔɔde*
d.	*tow-* 'ask'	*tou*	*tou-te* =>	*toote*
	yom- 'read'	*you*	*you-de* =>	*yoode*
e.	*kuw-* 'eat'	*kuu*	*kuu-te* =>	*kuute*
	kum- 'draw (water)'	*kuu*	*kuu-de* =>	*kuude*

(13) Nonpast of *-w-* bases

	Basic stem	Nonpast	
a.	*iw-* 'say'	*iu* =>	*yuu*
b.	*ew-* 'get drunk'	*eu* =>	*yoo*
c.	*aw-* 'meet'	*au* =>	*ɔɔ* [^{w}ɔ:]
d.	*tow-* 'ask'	*tou* =>	*too*
e.	*kuw-* 'eat'	*kuu* =>	*kuu*

These rules still, with a few changes brought about by the NJ change of /ɔɔ/ > /oo/ (see 14.2), apply to *u- onbin* stems in dialects which use those stems, that is, particularly in Kansai; for example Kyoto *kaw-* 'buy' has the past *koota* (< *kɔɔta*). However, they no longer apply to the nonpast of *-w-* bases, so for example *aw-* 'meet' today has the nonpast form *au* (not *oo* < *ɔɔ*). Note that cNJ *yow-* 'get drunk' is the result of a reanalysis of LMJ *ew-* on the basis of the monophthongized *onbin* stem and nonpast: *eu-te* 'drunk' => *yoote* and *eu* 'gets drunk' => *yoo*, reinterpreted as being from *yow-*. Likewise, *iw-* 'say' was reinterpreted as *yuw-* and listed in *Vocabulario* by the infinitive *yui* (although in standard NJ this verb today has the infinitive *ii-*, but other forms reflecting *yuw-*, e.g. nonpast *yuu* (spelled *iu*) or past *yutta*).

The volitional flective had the shape /u/ and, like its ancestor the conjectural auxiliary *-(a)m-*, attached to the *a-* stem of consonant base verbs and the basic stem of vowel base verbs. In NJ the shape of the volitional changed with vowel base verbs (see 15.1.1).

(14) Volitional *-(a)u*

	Basic stem	Volitional	
a.	*oki-* 'arise'	*oki-u* =>	*okyuu*
	mi- 'see'	*mi-u* =>	*myuu*
b.	*ake-* 'open'	*ake-u* =>	*akyoo*
	ne- 'sleep'	*ne-u* =>	*nyoo*
	se- 'do'	*se-u* =>	*syoo*
c.	*kaw-* 'buy'	*kawa-u* =>	*kaɔɔ* ([kaʷɔ:])
	mot- 'hold'	*mota-u* =>	*motɔɔ*
	kak- 'write'	*kaka-u* =>	*kakɔɔ*
	kos- 'cross'	*kosa-u* =>	*kosɔɔ*
	yob- 'call'	*yoba-u* =>	*yobɔɔ*
	kog- 'row'	*koga-u* =>	*kogɔɔ*
	yom- 'read'	*yoma-u* =>	*yomɔɔ*
	sin- 'die'	*sina-u* =>	*sinɔɔ*
	tor- 'take'	*tora-u* =>	*torɔɔ*
	ar- 'exist'	*ara-u* =>	*arɔɔ*
d.	*ko-* 'come'	*ko-u* =>	*koo*

The adjectival copula infinitive *-u* is today widely used in Kansai dialects (sometimes with further changes in form), e.g. Kyoto *taka-* 'tall' + NEG => *tako nai* 'isn't tall' (< *takoo nai* < *takɔɔ nai*). Standard cNJ generally does not use the adjectival *-u* infinitive, except in the so-called super-polite formation (a late borrowing from Kyoto, not used much today) which also incorporates

the monophthongization rules, e.g. *utsukushi-* 'beautiful' => *utsukushū gozaimasu*, *taka-* 'tall' => *o-takō gozaimasu*, and in a few forms lexicalized from that formation: *ohayō* (*gozaimasu*) 'good morning' (< *haya-* 'early') or *omedetō* (*gozaimasu*) 'congratulations' (< *medeta-* 'joyous').

(15) Adjectival copula infinitive *-u*

	Stem	Infinitive	
a.	*utukusi* 'beautiful'	*utukusi-u* =>	*utukusyuu*
b.	*be* 'necessitive'	*be-u* =>	*byoo*
c.	*taka* 'tall'	*taka-u* =>	*takɔɔ*
d.	*kuro* 'black'	*kuro-u* =>	*kuroo*
e.	*usu* 'thin'	*usu-u* =>	*usuu*

12.4 Existential verbs; *ar-*, *i-*, *or-*

The main, neutral verb of existence in Japanese is *ar-* which has been a prominent feature of the language since OJ, in more or less unchanged form and function. In addition to the uses as an existential verb, *ar-* has through the history of the language had important grammatical functions (3.4.2.1.1), namely (a) to form stative constructions and (b) to form extended analytic forms with the adjectival and regular copula and the negative auxiliary. Since OJ *ar-* has had a number of exalted or polite variants, e.g. OJ *mas-*, *imas-*, *pab(y)er-* (not phonographically attested in OJ, cf. 3.4.2.1), EMJ *imase*, *opase-* (> *owase-*), LMJ *odyar-*, *oryar-*, *goza(a)r-* (cf. 12.7.2), which have also been used in the grammatical functions of *ar-*.

NJ has two more existential verbs, *i-* and *or-*. In standard NJ, *i-* is used mostly with animate subjects, whereas *ar-* is used with inanimate subjects, and *or-* is a humble synonym of *i-* (although *or-* is used in some inflected forms instead of *i-*, e.g. infinitive *ori*). In other varieties of cNJ these verbs are used and distributed differently. For example, Kansai Japanese tends to use *or-* as the neutral verb of existence with animate subjects.

Statives have through the history of Japanese been formed by a verbal infinitive or gerund + *ar-* (or another verb of existence), and in OJ, EMJ and early LMJ also by stative auxiliaries (*-yer-* and *-(I)tar-*), both of which incorporate *ar-*. As mentioned above, the EMJ stative auxiliary *-(I)tar-* (< OJ *-(i)te ar-*) is reflected in the late LMJ and NJ past tense flective *-(I)ta* (12.1.3). In standard NJ statives are generally formed by gerund + *i-* (*kaite iru* 'is writing, has written', usually abbreviated *kaiteru*), whereas gerund + *ar-* is only used as a restricted resultative stative construction (*kaite aru* 'has been written'). Again, other varieties form statives differently, in Kansai Japanese mostly by gerund + *or-* (*kaite oru*, *kaitoru*). In some varieties of Japanese, statives

include progressive function, but other dialects have separate expression of progressive, e.g. *kakyoru* (<= *kaki oru*) 'is writing' in some varieties of Kansai, reflecting the OJ analytic progressive (3.1.4.7.4).

The two existential verbs NJ *i-* and *or-* derive from OJ *wi-* 'sit down, settle down' and *wor-* 'be sitting', respectively. Kinsui's detailed study (2006) has shown (a) that these two verbs were not used as simple existentials until LMJ, and (b) that *wor-* is the old lexicalized stative form of (an ancestor of) *wi-*, paralleling the regular morphological OJ stative in *-yer-*, e.g. *sakyer-* 'be blooming' (< *saki ar-), *kyer-* 'be wearing' (< *ki ar-), but reflecting a slightly different formation (e.g. *(w)u ar- or *wo ar-, attaching *ar-* to the diachronic root of *wi-*, not its infinitive). Thus, in OJ and EMJ *wi-* and *wor-* were used as lexical verbs, *wi-* 'sit down' being the antonym of *tat-* 'rise, arise, stand up', and *wor-* 'be sitting' the antonym of *tater-* 'be standing', the morphological stative of *tat-*. In addition, from EMJ a new regular stative was formed on *wi-* by the EMJ stative auxiliary to give *wi-tar-* 'be sitting' whose use gradually increased over *wor-*.

In OJ, *wor-* was used to form progressives, e.g. *tomosi wor-* 'be lighting' (3.1.4.7.4). This use continued in EMJ, where *wor-* was also, though more rarely, used with the gerund rather than the infinitive. From EMJ progressives were also formed with *wi-tar-*, usually on the infinitive, but sometimes on the gerund (8.4.2.1). The fact that this use of *wor-* and *wi-tar-* to form progressives precedes their use as free existential verbs, with animate subjects, suggests that the use in progressive constructions formed the basis for the reinterpretation of *wor-* and *wi-* as existential verbs, that is to say, developments as shown in (16a).

(16)

	Lexical verb	Stative lexical verb form	Use in progressive constructions	Existential verb
a.	*wi-* 'sit'	*wor-* 'be sitting down'	*mi wor-* 'be seeing'	*or-* 'be'
		wi-tar- 'be sitting down'	*mi witar-* 'be seeing'	*itar- > ita > i-* 'be'
b.	*tat-* 'stand up'	*tat-er-* 'be standing'	*mi tater-* 'be seeing'	
	pus- 'lie down'	*pus-er-* 'be lying down'	*mi puser-* 'be seeing'	

Note that the morphological statives of also *tat-* 'stand up' and *pus-* 'lie down', *tater-* 'be standing' and *puser-* 'be lying', in EMJ were used in progressive constructions, but that they were not reinterpreted as existential verbs, (16b). This offers strong support for the development proposed here, from a stative lexical verb to use in progressive constructions and then further to an existential verb (as opposed to a course of development where the use as existential verb precedes the use in progressive constructions).

Such a use of stative positional verbs in progressive constructions is found in other languages, for example Danish, where *ligge* 'be lying', *sidde* 'sit, be sitting', *stå* 'be standing' are used as exemplified in (17), with the positional meaning bleached or lost. Outside such constructions these verbs are lexical stative verbs, which cannot be used as existentials.

(17)	han	ligger	og	kører	rundt	i	sin	store	bil
	he	is lying	and	drives	around	in	his	big	car
	'he is driving around in his big car'								

Already in OJ, *wor-* was univerbated, incorporating the original morphology as part of the lexical stem. As stative *-(I)tar-* changed from auxiliary to a flective (12.1.3.2), EMJ *wi-tar-* > LMJ *i-tar-* gave late LMJ *ita*, and it appears to have been the form *ita* which was initially reinterpreted as a nonpast existential verb. In late LMJ there are uses of *ita* in nonpast existential function and this is still found in cNJ dialects in northern Japan, but *ita* was generally reformed to *i-*, presumably because the morphology remained transparent and *ita* does not conform to the basic stem shape of verbs.

By the end of LMJ, *i-* was firmly established as an existential verb with animate subjects and was also widely used in stative constructions, mostly following the gerund, but sometimes following the infinitive. At that point, however, it had not entirely replaced *ar-* and its exalted and polite synomyms in these functions. *Ar-* was still used with animate subjects in late LMJ, alongside *i-*, but was eventually replaced by *i-* to give the distribution we find today. The specialized use of *i-* and *or-* as existential verbs with animate subjects is probably a reflection of their original semantics, which were mainly agentive and volitional, and of their use in progressives. The spread of the originally progressive construction with *(w)ita(r-)* to replace stative constructions in *ar-* may be thought to be related to the change to a past tense marker of the stative auxiliary *-(I)tar-* which incorporated *ar-* and also had analytic variants with *ar-* and its synonyms.

Since *wor-* started being replaced by the regular stative form *wi-tar-* in early EMJ, *wor* > *or-* has been subject to various reinterpretations in different varieties, to a large extent in socio-linguistic terms, but *or-* has survived both in standard cNJ and in Kansai cNJ where, as mentioned above, it is in many varieties used as the neutral existential verb with animate subjects.

12.4.1 Loss of -ru

During LMJ a number of grammatical forms built on or diachronically incorporating *ar-* dropped the final *-ru* in reflexes of the adnominal form, see (18). Also the ending *-ka* stereotypical of Kyushu dialects, e.g. *yo-ka* 'is good', in the same way derives from the secondary conjugation of the adjectival copula: *yo-karu* > *yo-ka*.

(18) *na* COP.NONPST < *naru* COP.ADN
dya COP.NONPST < *dyaru* COP.ADN < *de aru*
da (very rare) COP.NONPST < *daru < *dearu*
-(I)ta Past/Perfect tense < *-taru* STAT.ADN
-(k)ke EMPH (cf. 12.1.3.1) < *-keru* MPST.ADN

12.5 Pronouns and demonstratives

Some terms of self-reference, address and 3rd person reference are shown in (19); these forms include demonstratives used for personal deictic reference. Among the 1st person terms, *watakusi* was earlier used for 3rd person (EMJ) and *ore* for 2nd person (OJ).

(19) 1st person: *watakusi, ware, ore, koti, kore, soregasi*
2nd person: *watono, nusi* (originally 'owner, master'), *onusi, onosi* (< *o-nusi*), *soti, konata, sonata, nandi, onore, kisyo* (SJ 貴所 'noble, august place'), *sonofɔɔ* (*fɔɔ* > *foo* > cNJ *hō* 'direction, side')
3rd person: *kare, are, yatu, aitu, soitu, kore, sore*
Interrogatives: *tare, dore*

By the end of the period the system of demonstratives and interrogatives was close to the one we find in NJ.

(20)

	Proximal	Mesial	Distal	Interrogative
Nominal	*kore*	*sore*	*are*	*dore*
Modifier	*kono*	*sono*	*ano*	*dono*
Location	*koko*	*soko*	*asoko (/kasiko)*	*doko*
Location/ direction	*koti/kotira*	*soti*	*ati/atti*	*doti*
Location/ direction	*konata*	*sonata*	*anata*	*donata*
Manner	*kau* > *kɔɔ*	*sau* > *sɔɔ*	–	*dɔɔ*

The demonstrative forms are parallel, but the most remarkable change is the levelling of the interrogative forms to conform to the demonstratives; the interrogative adverb *dɔɔ* is from the very end of the period. The interrogative pronoun remained *tare*, however, and only changed to *dare* in late NJ; the use of *donata* as an interrogative pronoun is from early NJ. *Anata* was used for 3rd person reference in NJ, but changed towards the end of the eighteenth century to 2nd person reference.

12.6 Syntactic changes

Two sets of major syntactic changes have taken place in Japanese. One comprises the loss of the concl/adn distinction and includes also the loss of *kakari-musubi* (12.6.1) and the emergence of *no* as a nominalizer (12.6.1.3). The other is the change of *ga* from a genitive to a nominative case particle (12.6.2). These changes are mainly manifested through the LMJ period, but are initiated or anticipated in late EMJ.

12.6.1 Loss of the distinction between conclusive and adnominal

As set out above, the loss of the distinction between conclusive and adnominal was an important part of the changes which took place in the morphological system. However, the distinction was functionally syntactic and the loss of it also forms part of the complex of syntactic changes which took place during this period. The main functions of the conclusive and adnominal forms in OJ and EMJ were as follows (cf. 3.1.3.1):

(21) Conclusive
- a. predicate in declarative main clauses
- b. with extensions, sentence final particles, and some conjunctional particles

(22) Adnominal
- a. predicate in adnominal clauses
- b. a nominalized form
- c. the predicate in exclamative or interrogative main clauses
- d. *musubi* predicate in *kakari-musubi* with *zo*, *namu*, *ka*, *ya*
- e. with some conjunctional and other particles

The concl/adn merger took the course of the adnominal being used increasingly to conclude declarative main clauses from late EMJ. This eventually resulted in the (descendants of the) adnominal taking over the functions earlier associated with the conclusive. This was not limited to the nonpast, but held also for the past, intentional, volitional and past conjectural, that is to say, all

the inflected forms which arose as a result of the concl/adn merger (cf. 12.1.3.2). These forms thus at the end of LMJ had the following main functions:

(23) a. predicate in all types of main clauses (except imperative)
b. predicate in adnominal clauses
c. a nominalized form
d. use with extensions and particles

However, it is interesting to see that new distinct morphological marking of some of the functions in which the adnominal was used distinctively from the conclusive was acquired, or at least attempted, at the same time as, or after the concl/adn merger, namely when used as the predicate in *kakari-musubi* and exclamatives and interrogatives (12.6.1.2) and in nominalizations (12.6.1.3).

12.6.1.1 Adnominal clauses

As in NJ, by far the most common type of adnominal clause in premodern Japanese was as shown in (24), where a head noun is directly modified by a clause. This has been the main type of adnominal clause through the history of Japanese, but as opposed to NJ, premodern Japanese used the distinct adnominal verb form in adnominal clauses. The concl/adn merger thus meant the loss of the expression of the difference between a predicate in an adnominal clause and in declarative main clauses. However, there was from OJ onwards no segmental distinction among the QD verbs (or from EMJ onwards UM and LM verbs) between conclusive and adnominal forms, and as NJ shows, the absence of a conclusive versus adnominal distinction poses no parsing problems of adnominal clauses for speakers. In (24) there are two embedded adnominal clauses; the first has its predicate (*su naru* 'said to do') in an identifiable adnominal form, whereas the second does not, because the verb is a QD verb (*ip-* 'call') whose adnominal and conclusive were (segmentally) identical.

(24) [wotoko mo su naru] [nikki to ipu] mono
man ETOP do.CONCL EVID.ADN diary COMP call.ADN thing
'The thing called diary which men are said to keep' (*Tosa*)

12.6.1.1.1 Complementizers Japanese adnominal constructions generally do not have complementizers, i.e. linking elements, between an adnominal clause and a head noun. However, constructions such as (25) with a genitive particle between a modifying clause and a modified noun are found through premodern Japanese, although they are considered ungrammatical in NJ.

(25) OJ

a. [*wagimokwo* *ni* *mise-mu*] ***ga*** *tame* *ni*
my.beloved DAT show-CONJ.ADN GA sake COP.INF
'In order to show (them) to my beloved' (*MYS* 19.4222)

b. [taye-*mu*] ***no*** kokoro
end-CONJ.ADN NO heart
'The intention to end it (our relationship)' (*MYS* 12.3071)

EMJ

c. [kabakari no yononaka o
this COP.ADN world ACC
omowi-sute-mu] **no** kokoro
think-discard-CONJ.ADN NO heart
'The intention of abandoning this world' (*Genji: Agemaki*)

Early LMJ

d. [tokoro o omoi-sadame-zaru] **ga** yue ni
place ACC think-decide-NEG.ADN GA reason COP.INF
'Because I hadn't settled on a (permanent) place' (*Hōjōki*)

In such constructions *ga* or *no* can be said to function as complementizers.[2] An important feature of this construction is that it was used only in 'gapless' adnominal clauses, as in the examples in (25), that is to say, noun modifying clauses where the modified noun has no syntactic function within the adnominal clause.

In OJ there are not many phonographically attested examples, although the reading tradition of logographically written text posits quite a few more. In EMJ and early LMJ, the construction was used in small but steady numbers. It is not now possible to present a detailed account of the distribution and development of this construction, but the following tendencies may be noted: (a) in OJ, EMJ and early LMJ the construction – although never frequent – is clearly found in more examples than can be attributed to error; (b) in these periods, it is used more with *ga* and less with *no* as the particle between the clause and the noun; and (c) it is restricted to gapless constructions.

[2] There are other examples of adnominal clauses in OJ with a particle between the adnominal clause and its head noun such as (i), but it is doubtful whether *ya* (or similarly used *i si*, *ya si*, *yo*), which is usually glossed 'emphatic' in such examples, can be said to function as a complementizer here.

i. [*wotomye* *no* *nasu*] *ya* *ita-two*
maiden GEN sleep.RESP EMPH plank-door
'The door behind which the maiden is sleeping' (*KK* 2)

From the second half of LMJ two changes take place: the frequency of use of this construction increases, and *no* comes to be used regularly instead of *ga*, e.g. (26a). Although saying it was characteristic of written language, Rodrigues (*Arte*, p. 506) describes this use with nouns such as *yoxi* /yosi/ 'manner, way', *aida* 'while', *yuye* /yue/ 'reason', *tocoro* 'place', giving examples such as (26b).

(26) Late LMJ

a. [kuru o tanomu] **no** kari-no-tamadusa
come ACC ask NO goose-GEN-letter (= letter)
'The letter asking me to come' (*Shingoshûi wakashû*, 1384)

b. "*xitagŏno tocoroni*"
[sitagɔɔ] **no** tokoro ni
follow.NONPST NO place COP.INF
'When following'

Both of these changes in use are traditionally ascribed to a new *kanbun-kundoku* practice in the Zhū zǐ (朱子) school of Confucianism, in which Chinese 之 was rendered as *no*, also in these constructions where the Japanese syntax does not require it; an early precursor of this view was held by Moto'ori Norinaga, who also denounced such constructions as incorrect and uncharacteristic of the Japanese language. The idea is that use of the construction would have spread from the use in this school of *kanbun-kundoku* into common (written) language. However, as we saw above, the construction has been in use through all of premodern Japanese, and it is perhaps more likely that use of the construction increased through late LMJ in response to the merger of the distinction between the conclusive and adnominal verb forms. The construction was, however, never fully generalized and it is today considered ungrammatical. Even so, an informal internet search easily turns up cNJ examples such as (27a). It is also noteworthy that adnominal clauses with a genitive particle used as complementizer are regularly produced by children in the course of their acquisition of Japanese as a first language (as well as by second language learners), but here not limited to gapless adnominal clauses, e.g. (27b) by a $2\frac{1}{2}$-year-old girl.

(27) a. [omoshiroi burogu ni suru] **no** tame
interesting blog COP do NO purpose
ni wa
COP.INF TOP
'in order to make an interesting blog'

b. [chigau] **no** kami
be.different NO paper
'a different piece of paper'

12.6.1.1.2 ***Tokoro-no*** A different type of complementizer is found in (28), with *tokoro-no* between adnominal clause and head noun. As opposed to adnominal clauses used with *ga/no*, *tokoro-no* can be used with relative clauses, that is, adnominal clauses with a syntactic gap, as shown in (28). The use of *tokoro-no* as a complementizer originates in *kanbun-kundoku* (cf. 9.1.7) and seems only to have been used in written language, including its vocalized versions. It was revived in NJ in *kanbun-kundoku*-like renditions of Dutch texts (cf. 17.3.3) and may still be used in written language.

(28) [$\emptyset_i$ tatekome-taru] tokoro-no **to**$_i$
close-STAT.ADN TOKORO-NO door
'The door which had been closed' (*Taketori*)

12.6.1.2 Loss of kakari-musubi

The *kakari-musubi* construction gradually declined in the course of LMJ and had all but disappeared from the language at the end of the period, when it was still found to a limited extent with *koso*, but no longer with *namu*, *ya*, *zo*, *ka*. *Kakari-musubi* is today only retained in a few mainland and some Ryukyuan dialects. Of the *kakari* particles, *ka* (interrogative) and *zo* (exclamative) are widely used as sentence final particles in cNJ, and *koso* is used as an emphatic particle; *koso* and *zo* are also lexicalized in common expressions like *yookoso* 'welcome' (< *yo-u* < *yo-ku* 'good-INF' + *koso*) and *doozo* 'please; here you are' (< *doo* 'how' + *zo*).

The loss of *kakari-musubi* is sometimes presented as a direct result of the concl/adn merger, but that is probably somewhat simplistic. First, that would leave unexplained the fact that *kakari-musubi* with *koso* was also lost (even if that happened a little later than the loss of *kakari-musubi* with an adnominal *musubi* predicate). In that case it seems clear that the loss of the distinct exclamatory verb form was a direct result of the loss of *kakari-musubi* with *koso*, for by the beginning of LMJ use in *kakari-musubi* with *koso* was the only function of the exclamatory. Second, there are mainland dialects (e.g. in Wakayama) which maintain a concl/adn distinction, but in which *kakari-musubi* is also lost. Third, any view of causal relations between the concl/adn merger and the loss of *kakari-musubi* depends on the view taken of what *kakari-musubi* actually was. If *kakari-musubi* were viewed as a simple agreement relation between one form (a *kakari* particle) and another (an adnominal or exclamatory *musubi* predicate), then naturally the loss of the distinctive predicate form would mean the end of that relation. However, we saw above

(8.9.3) that the form of the *musubi* predicate was not the defining feature of *kakari-musubi*: in many cases the *musubi* predicate did not have a distinct shape, namely, if it was in a subordinate clause, had a nominal predicate with zero copula, or the predicate was a QD or UM verb (which had no (segmental) distinction between conclusive and adnominal). If, on the other hand, as on Watanabe's analysis (8.9), *kakari-musubi* was basically a movement operation, then the loss of movement equalled the loss of *kakari-musubi*. This is what Watanabe suggests, saying further that this took place in OJ, leaving behind a simple agreement relation between *kakari* particle and predicate form which gradually disappeared because it was unmotivated and vacuous after the loss of movement. Quinn's functional account (8.9.2, 8.9.3) invests the *kakari* particle and the shape of the *musubi* predicate with separate, but combining functions, and on that view, the concl/adn merger should not affect the function of the *kakari* particle. It does seem intuitively clear that the loss of *kakari-musubi* and of the concl/adn distinction form part of the same complex of changes, but the precise nature of the relation is difficult to establish. Below we discuss the possibility that the exclamative and interrogative functions of the adnominal form retained some morphological marking also after the concl/adn merger (see 12.6.1.2.2). In any case, it is clear that the *kakari-musubi* construction – whatever the formal analysis of it – was an important focus construction in OJ and EMJ, and that the loss of it is among the most important syntactic changes to have affected Japanese.

12.6.1.2.1 ***Namu* and *ya*** *Namu* was lost entirely early in LMJ and is no longer found in the materials from the end of the period. *Ya* largely replaced *ka* inside yes/no questions through EMJ and was used through LMJ, but seems gradually to have been reinterpreted as expressing uncertainty, 'I wonder', rather than simply indicating yes/no questions; through LMJ the function of indicating yes/no questions was taken over by sentence final *ka*. *Ya* was at the end of the period used mostly in combination with various modal forms of *ar-*. Some of these combinations were lexicalized: Early LMJ *yaraũ* (> LMJ *yarɔɔ*) is a copula like particle, 'is it, I wonder?', originating in a contraction of *ya ara-ũ* (< OJ *ara-mu*) 'exist-CONJ'. *Yarɔɔ* is possibly the source of the modern Kansai copula *ya* (see 12.1.4). Early LMJ *yaraũ* has also given the particle *yara*, still used in NJ with the meanings 'and so on; I wonder' and also with *wh-* words to form existential quantifiers (*doko-yara* 'somewhere'). In addition, *ya* had by the end of LMJ developed uses as a particle 'and, or, or the like', often used together with *nando* or *nado* 'such as, etc.'.

12.6.1.2.2 ***Ka* and *zo*** The use of both *zo* and *ka* changed in the course of LMJ. *Zo* is only and *ka* mainly used as sentence final particles at the end of LMJ, e.g. (29)–(31). The main pattern shown in these examples, which are all

taken from *Esopo*, is that *zo* is used in exclamative main clauses (29), and in *wh-* questions (30), whereas *ka* is used in yes/no questions (31). This use of *zo* is for example also found in the *shōmono* materials.

(29) a. kurusyuude kuyamu wa tikusyɔɔ no waza zo
suffer.GER regret.NONPST TOP beast GEN job ZO
'suffering and regret is the burden of the beast!' (*Esopo*)

b. ookame no kuru zo
wolf GEN come.NONPST ZO
'The wolf is coming!' (*Esopo*)

(30) a. ware wa doko e yuku zo
I TOP where ALL go.NONPST ZO
'where do I go?' (*Esopo*)

b. nandi ware o-ba tare to
you I ACC-TOP who COMP
omoo zo
think.NONPST ZO
'Who do you think I am? (*Esopo*)

c. nandi wa naze ni sita bakari
you TOP why COP.INF tongue only
o-ba kɔɔte kuru zo
ACC-TOP buy.GER come.NONPST ZO
'Why did you only buy tongue?' (*Esopo*)

(31) kisyo wa yakusoku wa wasureta ka
you TOP promise TOP forget.PST KA
'Did you forget your promise?' (*Esopo*)

This distribution suggests that one important function of the adnominal verb form, namely to mark sentences as exclamative or interrogative, was taken over after the concl/adn merger by two former *kakari* particles, first of all *zo*, but in yes/no questions by *ka*. The use in (32) is quite rare, but noteworthy in this connection: it is a *wh-* question, marked by *zo*, but in addition, the *wh-* word is marked by *ka*, as opposed to the usual pattern shown in (30). (32) can thus be thought of as a late reflex of the OJ and EMJ *kakari-musubi* pattern in which *wh-* words were often marked by *ka* (more so in OJ than in EMJ) correlating with distinct sentence final morphology, in OJ and EMJ the adnominal form, but in late LMJ the particle *zo*. This suggests that *zo* was at once nominalizing and predicating; *zo* could also predicate nominals as in (29a). On this interpretation, expression of the functional differentiation between

declaratives on the one hand and exclamatives and questions on the other was maintained also after the concl/adn merger.

(32) tarebito ka si ni syoo zo
who KA teacher COP.INF do.VOL ZO
'Who shall we make the teacher?' (*Esopo*)

The use of *ka* as a general sentence final interrogative particle in NJ is quite different from the LMJ pattern of distribution of *ka* and *zo* exemplified above and is thought to be an eastern dialect feature which spread after the ascendancy of the language of Edo/Tokyo.

Note finally that the use of *ka* after interrogatives to form existential quantifiers (*dare-ka* 'someone', *nani-ka* 'something', etc.) is a NJ development which took place after *ka* entirely lost uses such as that exemplified in (32).

12.6.1.2.3 ***Koso*** As opposed to the other *kakari* particles, *koso* is still used today as an emphatic particle, although to a limited extent. In mainland dialects that retain *kakari-musubi*, this involves reflexes of *koso*. At the end of the LMJ period *koso* is still used in *kakari-musubi* focus constructions. It is not frequent, but it is found in both *Esopo* and *Feiqe*, e.g. (33a) which is an unambiguous focus construction and which has an exclamatory variant in the predicate (see 12.1.2 about the exclamatory). As mentioned above (12.1.2), the predicates used with *koso* in *Esopo* are mostly grammatical forms belonging to or reflecting *r*-irr verbs, but there are exceptions, even with those morphemes, where the predicate is not in the exclamatory variant. Thus, the use of *koso* in *kakari-musubi* at the end of the LMJ period represents the final stage in the loss of *kakari-musubi*. (33b) is an example from the first half of the sixteenth century (1534) where the predicate correlating with *koso* is not in the exclamatory.

(33) a. sore o-ba Esopo **koso** nusunde
that ACC-TOP Aesop KOSO steal.GER
tabete **gozare**
eat.GER exist(POL).EXCL
'It is Aesop who has stolen and eaten that' (*Esopo*)

b. kono kotowari wa kami **koso**
this reason TOP god KOSO
siru to omoeba
know.NONPST COMP think
'If you think that it is god who knows the reason for this' (*Shiganikkai*)

12.6.1.3 Nominalizations; emergence of nominalizing no

In addition to the uses in adnominal clauses and in *kakari-musubi*, the OJ/EMJ adnominal functioned to form headless nominalizations (cf. 3.1.3.1), see (34), (35) and (36). Of these, examples like (34) are sometimes analysed as 'internally headed (or circum-nominal) relative clauses', with the head noun represented inside, but not after the modifying clause, so that in (34a) *kiku no pana* would be interpreted as a head noun modified by *uturoperu* 'which has faded', or in (34b) *tori* would be interpreted as a head noun modified by two clauses, *pasi to asi to akaki* 'whose beak and feet are red' and *sigi no opokisa naru* 'which is the size of a snipe'. Examples like (35) have been analysed as 'headless relative clauses', meaning 'that which..., the one who...', which have representation of a head noun *neither* after *nor* inside the adnominal clause. As mentioned in 3.1.3.1, such constructions were rare in OJ and almost exclusively used in pseudo-cleft constructions like (34a), but from EMJ the construction came to be used more widely, e.g. (35b). It has been proposed that such analyses of the constructions in (34) and (35) violate basic notions of relative clauses as being noun modifiers (Shibatani, p.c.). We here simply treat them neutrally as a subgroup of headless nominalizations. The pattern exemplified in (36) is often referred to as a complement clause. Use of the adnominal in that function was rare in OJ, which generally used the nominal form (cf. 3.1.3.3), but it became frequent in EMJ.

(34) a. [kiku no pana no
chrysanthemum GEN flower GEN
uturop-eru] wo worite
fade-STAT.ADN ACC break.GER
'picking some faded chrysanthemum' (*Ise* 18)

b. [[siro-ki tori no pasi to
white-ACOP.ADN bird GEN beak and
asi to aka-ki] sigi no
feet and red-ACOP.ADN snipe GEN
opokisa naru]
size COP.ADN
midu no upe ni asobitutu iwo
water GEN top DAT frolic.CONT fish
wo kupu
ACC eat.CONCL
'A white bird which has a red beak and feet and is the size of a snipe was eating fish while frolicking on the water.' (*Ise* 9)

(35) a. [*kadi* *no oto* *suru*] *pa*
oar GEN sound do.ADN TOP
ama-wotomye *kamo*
fisher-girl Q
'the ones making the oar-sounds, is that the fisher-girls?'
(*MYS* 15.3641)

b. [Usiromi to iu] wa kami
Ushiromi COMP call.ADN TOP hair
naga-ku wokasige nareba
long-ACOP.INF beautiful COP.PROV
'As the one called Ushiromi had long hair and was beautiful' (*Ochikubo*)

(36) a. [imizi-u naku pito aru]
terrible-ACOP.INF cry.ADN person exist.ADN
wo kikitukete
ACC hear.GER
'hearing that there was a person who was crying terribly'
(*Ise* 6)

b. [yo pito-yo sira-nu koto
night one-night know-NEG.ADN thing
ni yori utipiki-tamawi-turu] koso
DAT depend torment-RESP-PERF.ADN FOC
ito warina-kari-ture
very unbearable-ACOP-PERF.EXCL
'It was unbearable that you tormented me through the night because of something I didn't know about' (*Ochikubo*)

In late LMJ, after the concl/adn merger, the inflected forms which resulted from the merger (the nonpast, past, volitional and intentional) came to be used in these nominalizing functions, as in (37):

(37) a. [tiisai **ko** no naku] o sukasu
small child GEN cry ACC calm.down
tote sono fawa kamaete nakaba
PURP that mother really! cry.COND
ookame ni yarɔɔzu to yuu
wolf DAT give.INT COMP say
'In order to calm down a small child who was crying, his mother said "if you keep crying, I'll give you to the wolf".' (*Esopo*)

b. [sisi no yuuta] wa ware wa arufodo
lion GEN say.PST TOP I TOP all
no kedamono no ɔɔ nareba …
COP.ADN animal GEN king COP.PROV
'(What) the lion said: "As I am the king of all animals…"'
(*Esopo*)

c. [kurusyuude kuyamu] wa tikusyɔɔ
suffer.GER regret.NONPST TOP beast
no waza zo
GEN job ZO
'It is the job of the beast to suffer and feel regret!' (*Esopo*)

However, in early NJ *no* emerged as a nominalizer which came to be used in these functions. (38) contains two early examples:

(38) a. [soregasi ga suite yomu] **no**
I NOM like.GER read.NONPST NMLZ
wa seisuiki o suite yomu
TOP Seisuiki ACC like.GER read.NONPST
'What I like reading is (I like reading) the (*Genpei*) *Seisuiki*'
(*Kyōgen-ki*, *c.* 1660)

b. [fara no tatu] **no** wa warui
stomach GEN rise.NONPST NMLZ TOP bad.NONPST
'It is bad that he is angry' (*Kabuki jūhachibanshū*, *c.* 1700)

Thus, nominalizing *no* took over the nominalizing functions of the OJ/EMJ adnominal and came to be used in all contexts where the OJ/EMJ adnominal formed headless nominalizations, see (39)–(41), but only after a hiatus with no distinct morphological expression of these functions, between the concl/adn merger and the emergence of *no* as a nominalizer. In formal syntax, NJ nominalizing *no* is often given entirely different analyses in different contexts: as a pronoun in (40b), and as a (verb selected) complementizer in (41b). However, formal syntactic analysis does not agree about whether *no* in examples such as (38b), which are often analysed as internally headed relative clauses, should be interpreted as a pronoun or as a complementizer.

(39) a. [kiku no pana no uturop-eru] wo worite (=34a)

b. Taroo wa [ringo ga sara no
Taro TOP apple NOM plate GEN
ue ni atta] **no** o totta
top DAT was] NMLZ ACC took
'Taro picked up an apple which was on a plate'

(40) a. [Usiromi to iu] wa kami naga-ku wokasige nareba (=35b)

b. [kinō katta] **no** o tabeta
yesterday bought NMLZ ACC ate
'I ate the one I bought yeserday'

(41) a. [imizi-u naku pito aru] wo kikitukete (=36a)

b. [kinō kita] **no** o wasureta
yesterday came NMLZ ACC forget.PST
'I forgot that I came yesterday'

There is a sizeable literature which presents different scenarios for the development of these nominalizing functions of NJ *no*. The pronoun-like use of *no*, as in (40), predates the complementizer-like use (Kinsui 1995), but it has been pointed out by Wrona (forthcoming) that there is little need to posit hypothetical developmental steps between these functions if they are considered subfunctions of a general nominalizing function. As for the emergence of *no* as a nominalizer, Wrona points out that *no* emerges as a nominalizer *after* the increase in the use of *no* as a complementizer in adnominal clauses (see 12.6.1.1.1), proposing that this use as complementizer in adnominal clauses formed the basis for the generalization of *no* as a nominalizer.

12.6.1.4 Summary of the developments of the functions of the Old Japanese/ Early Middle Japanese adnominal

The expression in late LMJ and NJ of the functions which the OJ/EMJ/ early LMJ adnominal had can be summarized as in (42). Overall expression of the functions of the adnominal has been replaced by the final particles *zo*, *ka* (12.6.1.2) and by nominalizing *no* (12.6.1.3). It is conspicuous that the only function of the OJ/EMJ adnominal which does not have distinct morphological expression in late LMJ and NJ is the adnominal function, apart from the marginal use of *no* as a complementizer in ungapped adnominal clauses (12.6.1.1.1).

(42) LMJ and NJ expressions of the functions of the EMJ adnominal

	Late LMJ	NJ
kakari-musubi	(*zo*)	–
exclamatives	*zo*	*zo*
wh- questions	*zo*	*ka*
yes/no questions	*ka*	*ka*
adnominal clauses	(*no*)	(*no*)
(other) nominalizations	–	*no*

12.6.2 The genitive particles, subject marking and the emergence of a nominative case particle

In OJ *ga* and *no* were genitive particles, used with some morphological, syntactic and semantic specialization (see 3.7.1.1.1). Both *ga* and *no* were in OJ and EMJ used to mark subjects of adnominal clauses, other subordinate clauses, and exclamative or interrogative main clauses (including in *kakari-musubi*, see 8.9).

In OJ, EMJ and early LMJ *ga* was much more restricted than *no* as a genitive, both in adnominal and in subject marking function. In OJ *ga* was mainly used with pronouns and nouns referring to human beings, and through EMJ and early LMJ *ga* became increasingly restricted to use only with pronouns and nouns referring to 1st and 2nd person and with personal proper nouns. On the other hand, only *ga*, not *no*, was used to case mark headless nominalizations (as reflected in the development of *ga* into a conjunctional particle, cf. 8.7.2), and *ga* was used as a complementizer more than *no* (see 3.7.1.1.2, 12.6.1.1.1). A final minor context where *ga* was used rather than *no* was in fractions, e.g. *sanbun ga iti* 'one third'; this usage is found at least from late EMJ.

No, by contrast, had no restrictions with regard to the type of noun it could mark, both in adnominal and in subject marking function, and through EMJ and early LMJ the use of *no* to mark subjects expanded. However, *no* did not make the full transition to marking subjects of declarative main clauses.

However, at the end of LMJ the situation is markedly and remarkably different: *ga* had entirely lost its productive adnominal uses (being almost exclusively used to modify nouns in idioms and in lexicalized possessive pronouns such as *waga* 'mine') and also its function as a complementizer, but had expanded its use as a subject marker dramatically. Other than its independent use as a conjunctional particle, *ga* was by the end of LMJ only used productively to mark subjects, both in declarative main clauses, e.g. (43), as well as in other clauses. Thus, *ga* may be thought by the end of LMJ to have completed a shift from a severely restricted genitive to nominative case particle, although *ga* even today in cNJ remains restricted in declarative main clauses, marking only focussed subjects and subjects in existential or presentational sentences. The emergence of a nominative case particle may well be considered one of the few major syntactic changes to have taken place in the attested history of Japanese.

(43)	Amonia to	yuu	sato	**ga**	odyaru
	Amonia COMP	call	village	NOM	exist.POL
	'There is a village called Amonia'				

No, on the other hand, remained a genitive particle. In addition to its adnominal functions, *no* was still used to mark subjects in subordinate and nominalized clauses, e.g. (44). The circumscription of the use of *no* to mark subjects has continued into contemporary NJ where *no* still marks subjects in short adnominal clauses, but not in other subordinate clauses.

(44) a. sisi **no** yuuta wa: ware wa arufodo
lion GEN say.PST TOP I TOP all
no kedamono no ɔɔ nareba ...
COP.ADN animal GEN king COP.PROV
'The lion said: "As I am the king of all animals ..."' (*Esopo*)

b. ookame **no** kuru zo
wolf GEN come.NONPST zo
'The wolf is coming' (*Esopo*)

Conversely, *no* in late LMJ became increasingly used as a complementizer in certain adnominal clauses (see 12.6.1.1.1) and also developed its general nominalizing function (12.6.1.3). The development of the functions of OJ genitive *no* and *ga* are summarized in Table 12.10. The shifts in use of *ga* and *no* between early and late LMJ are remarkable, but their course has not been charted in detail. It appears as if *no* and *ga* around the middle of LMJ swapped or flipped important syntactic functions between them: subject marking and use as complementizer. It is a major challenge facing future studies in Japanese diachronic syntax to provide a plausible account of this, as well as of the development of the functions of *ga* and *no* in different dialects (where they have the same or similar functions, but distributed differently over the two particles, cf. also 3.7.1.1.1).

12.6.2.1 Socio-linguistic differentiation of no *and* ga

Through EMJ and LMJ *no* and *ga* are famously said to have been differentiated in socio-linguistic terms, so that *ga* was deprecating or humble, whereas *no* was deferential, respectful or neutral. Rodrigues is quite explicit about this in the description in *Arte* of *ga* and *no*, which are both included among the particles expressing the nominative ('que servem ao Nominativo')[3] and are given as the only two genitive particles.

ga is used in careful speech and it indicates that what is referred to by the noun preceding it is deprecated. It is used with a first-person pronoun or a third-person pronoun referring to a low-ranked person. Also it is used to slight or deprecate other persons.

[3] The other articles listed as serving that function (that is, marking a phrase that can be thought of as the subject) are *wa, yori, kara, ni, niwa, ni oitewa*.

Table 12.10 *Summary of the development of the uses of* ga *and* no

		OJ	early EMJ	late EMJ (Insei)	early LMJ (Kamakura)	late LMJ (Muromachi)	NJ (Edo)
Adnominal marking	***no***	+	+	+	+	+	+
	ga	+	+	+	(+)	–	–
Subject marking in subordinate clauses and in non-declarative main clauses	***no***	+	+	+	+	+	(+)
	ga	+	+	+	+	+	+
Subject marking in declarative main clauses	***no***	–	–	–	–	–	–
	ga	–	–	–	–	+	+
Complementizer (connecting a clause to a nominal(ized) head)	***no***	(+)	(+)	(+)	(+)	+	+
	ga	+	+	+	+	–	–
Nominalizer (as pronoun and verb selected complementizer)	***no***	–	–	–	–	–	+
	ga	–	–	–	–	–	–
Conjunctional particle	***no***	–	–	–	–	–	–
	ga	–	–	+	+	+	+

No is normally used in the relative clause marking the second- and third-person pronouns, and it connotes deference or at least non-deprecation of the referent.

In the attributive function there are two [particles]. *No* is used with the second- and third-person pronouns to refer to respected persons, and *ga* is used with the first-person pronoun and the third-person pronoun to refer to a low-ranked person, and sometimes with the second-person pronoun when that person is to be deprecated.

(Shibatani 1990: 356, extracting and translating from *Arte*, pp. 501–3)

This passage is interesting, but it is first of all puzzling, for it is at odds with the use of *ga* and *no* in for example *Esopo* and other late LMJ texts, where *ga* is used to mark most kinds of subjects. It is difficult therefore to know what to make of this description in *Arte*. It suggests that this was a conspicuous feature of metalinguistic consciousness at the time, representing contemporary linguistic folklore or a rationalization of usage in EMJ or early LMJ. However, as long as there were actual personal pronouns in the language (cf. 8.8), for example, *ga* was used with 1st, 2nd and interrogative personal pronouns, so it is also difficult to recognize this often claimed difference in EMJ or in OJ. In any case, it is difficult to link the proposed socio-linguistic differentiation of *ga* and *no* to the syntactic specialization which we find at the end of LMJ.

12.7 Honorific language

Through its history Japanese has had grammatical expression of what is very generally characterized as *honorific language* (*keigo* 敬語). It is common to distinguish two independent parameters: *exaltation* (12.7.1) and *politeness* (12.7.2).

12.7.1 Exaltation

Exaltation characterizes the social relation between those spoken about, or between the speaker and those spoken about, and exaltation is sometimes referred to as *social deixis*. Two overall parameters are distinguished: (a) 'respect' (*sonkei* 尊敬), also referred to as 'esteemed subject', 'respect for subject', 'subject exaltation'; and (b) 'humility' (*kenjō* 謙譲), which in the Japanese descriptive tradition, which focuses on the subject, may be termed 'humble subject', but which in more recent scholarship in English is referred to as 'respect for object' (or indirect object), or 'object exaltation'; here we use the English equivalents of the Japanese terms, *respect* and *humility*, for these two categories.

Of the two categories, respect (marking of subject as exalted) has always been more developed than humility. Often a number of levels of respect are set up. This may be exemplified by Rodrigues's list in *Arte* (p. 61) of socially differentiated ways of expressing a command, from the least to the most respectful, see (45). The 'give!' forms employ a number of morphological formants, while the forms for 'come!' are highly suppletive, using verbs that are also used as existential verbs and as polite forms (12.7.2).

(45) a. 'give!'

aguei, *agueyo*	(*agei*, *ageyo*; IMP)
aguesaxime	(*age-sasime*; *-sasim-*, IMP)
aguesai	(*agesai*; RESP.IMP)[4]
aguesaxemaxe	(*age-sase-mase*; CAUS-*mas-*, IMP)
aguerarei	(*age-rarei*; PASS.IMP)
voagueare	(*o-age-are*; *o-VERB-ar-*, IMP)
voaguearō	(*o-age-arɔɔ*; *o-VERB-ar-*, VOL)
aguesaxerarei	(*age-sase-rarei*; CAUS-PASS.IMP)
voague nasarei	(*o-age-nasarei*; *o-VERB-nasar-*, IMP)
voague nasareô	(*o-age-nasaryoo*; *o-VERB-nasar-*, VOL)

[4] A 'slightly respectful' imperative was in late LMJ and early NJ expressed by *-sai* (the irregular imperative of a respect auxiliary *-(i)sar-*) attached to regular vowel stem verbs, or *-i* attached to the *a-* stem of consonant base verbs, e.g. *yomai* 'read!', or *irai* under 'come!' in (45b).

b. 'come!'

coi	(*koi*; IMP)
irai	(*irai*; *ir-* 'enter', RESP.IMP)
voriare	(*oryare*; *oryar-* 'be', IMP)
vogiare	(*odyare*; *odyar-* 'be', IMP)
gozare	(*gozare*; *gozar-* 'be', IMP)
gozarō	(*gozarɔɔ*; *gozar-* 'be', VOL)
voidenasarei	(*oide-nasarei*; *oide* 'go out' *nasar-* 'do.RESP', IMP)
voidenasareô	(*oide-nasaryoo*; *oide* 'go out' *nasar-* 'do.RESP', VOL)

In this way, it is in most descriptions and in metalinguistic consciousness, as reflected in the Japanese and English terms, the social aspects of this grammatical marking and the social norms it refers to which attract the most attention. However, especially in the frequent absence of lexical or pronominal expression of subject, object or indirect object, the basic system of grammatical marking of respect and humility to some extent serves to identify those roles vis-à-vis discourse participants, as in cNJ *o-yomi-ni nar-* '(someone esteemed reads;) you read', or *o-yomi-se-* '(someone humble reads;) I read'. The system of exaltation thus grammatically has a similar function to inflection for grammatical 2nd (respect) or 1st (humility) person.

12.7.1.1 Noun exaltation

Exaltation mostly concerns clause predicates, as will be discussed immediately below, but in addition nouns and nominal words, including verbal or adjectival infinitives, may be exalted by prefixes. The primary exalting prefix is OJ *mi-* which is usually glossed 'honorific'. It is reflected in lexicalized forms such as *miya* 'palace' (*ya* 'house'), *mikadwo* 'palace; emperor' (*kadwo* 'gate (to palace)'), *mikwo* 'prince, princess, shrine maiden' (*kwo* 'child'). Already in OJ, the combination of *opo-* 'great' with *mi-* was conventionalized as expressing a higher degree of honorification than simple *mi-*, often referring to the emperor, e.g. *opomi-ki* 'saké for the emperor' (*ki* 'saké') or *opomi-koto* 'imperial order' (*koto* 'word, speech'). In the course of EMJ, OJ *opomi-* was reduced to *oon-* (mid-EMJ) and *on-* (late EMJ), which replaced *mi-* as the general honorific noun prefix, and eventually *o-* which since LMJ has been the main honorific prefix. It is used widely today in a variety of functions, including formation of the 'honorific infinitive' (*oyomi* 'read' <= *yom-*; *ohayoo* 'early' <= *haya-*), with nouns as a kind of 2nd person prefix (*o-namae* 'your name'), or for simple politeness (*o-kane* 'money' is more polite than *kane*), often referred to as *beautification*, or fully lexicalized (e.g., *onaka* 'stomach'). Also in the course of EMJ, the SJ prefix *go-* (御 EMC 'imperial' *ŋiăʰ) gained currency, first of all with SJ words; note that 御 in Chinese is not used as a general honorific, but means 'imperial'.

12.7.1.2 Predicate exaltation

Overall, exaltation on predicates has been expressed in three main ways, namely by (a) auxiliaries, e.g. OJ *-(a)s-* RESP, *-(a)sime-* CAUS (enhancing a following respect expression); EMJ/LMJ *-rare-* PASS (used for respect), *-sase-* CAUS (enhancing a following respect expression), (b) auxiliary verbs, e.g. *-tamap-* RESP, *-mawir-* HUM, *-tatematur-* HUM, and (c) lexical suppletion, e.g. *imas-* 'be, come, go, RESP', *kakur-* 'die' ('hide oneself'), *tukapematur-* 'serve. HUM', *mawos-* 'say.HUM, speak.HUM', *mawir-* 'enter.HUM', and many, many others. As mentioned, the expression of humility has always been less developed than respect, and this is also reflected in the fact that humility has never been expressed by auxiliaries, only by auxiliary verbs and suppletive lexical verbs.

The inventories of forms used in these functions have changed over time, as have rules for their combinations. However, with a few exceptions the basic system of respect and humility has not changed much between OJ and cNJ. In Old Japanese, as opposed to later stages, 'respect for subject' could be used by the speaker about himself, e.g. *wa ga tata-**s**-ere-ba* I GEN stand-**RESP**-STAT.PROV 'as I stand' (*KK* 2), spoken by the god *Yachihoko no kami no mikoto*. This is referred to as 'absolute [as opposed to relative] respect' and disappeared in the transition to EMJ. Exaltation is ubiquitous in the literary prose texts from EMJ and early LMJ. Not only were multiple, mutually reenforcing combinations of respect expressions used, such as *-sase-rare-tamap-* CAUS-PASS-RESP, expressing a high degree of respect, but humility and respect was often expressed on one verb, usually expressing the respect of the speaker (writer) towards the subject of the verb, but expressing the humility of the subject towards the (indirect) object, for example in the combination *-tatematuri-tamap-* HUM-RESP. The combination of respect and humility went out of use in late LMJ and is no longer part of the system. The varied and frequent use of exalting expressions is so prominent in the literary texts from EMJ and early LMJ, especially the *monogatari*, that it is a distinctive feature of the appearance and image of these texts. It is, however, not likely that exalting expressions were as widely used in common language as they were in the refined and elegant language of the court ladies and nobles. Through LMJ, the use of exalting expressions in the texts declined somewhat, probably reflecting more everyday and less elegant language use, rather than structural changes. In cNJ, exalting language has been regularized to mainly morphological expression, with a small number of auxiliaries and auxiliary verbs and far fewer suppletive verbs than in EMJ or LMJ. It should finally be mentioned that despite the prominence of the use of exalting expressions in metalinguistic consciousness and the importance assigned to it in normative discourse about Japanese, a number of Japanese dialects do not use exalting language.

12.7.2 Politeness

As opposed to exaltation, *politeness* (*teineigo* 丁寧語 'polite language') is a matter of style or register and expresses relations between speaker and hearer without reference to the content of the discourse. Politeness is a prominent feature of cNJ where polite style is expressed by the auxiliary *-(i)mas-* or the polite copula *des-*, but, as with exaltation, there are NJ dialects which do not have polite style. The grammatical expression of polite style is a relatively late addition to the language: OJ did not have it at all, and while there are incipient uses in EMJ, it was not thoroughly established until early LMJ. Polite style originates in exalting expressions through a process of shifting the target of respect or humility from the subject of a sentence to the hearer (respect) or speaker (humility), eventually interpreted as a characterization of the speech situation, or the relationship between speaker and hearer. Thus, all material used to express polite style originates in respectful or humble forms, and, as with exalting auxiliaries and auxiliary verbs, the development of the individual polite style markers is often complex.

Around the middle of EMJ, *paber-*, which until then was used as a humble verb 'be in attendance' or a humble equivalent of the existential verb *ar-*, started being used as a *polite* equivalent of *ar-* in its various uses. Thus, the expression of politeness was at first limited to contexts where *ar-* was used (existential sentences, extended regular and adjectival copula forms, stative constructions; cf. 12.4). However, in early LMJ, *paber-* was in this function replaced by *sauraw-*, which in addition to being used as a suppletive polite equivalent of *ar-* came to be used as an auxiliary verb, that is, attaching to the infinitive of verbs to express polite style, e.g. *mi-saurawabaya* 'see-POL.OPT; if only I could see', which is an early example from the *Heike monogatari* (early thirteenth century). This development made possible the expression of polite style with all predicate types and established politeness as a grammatical category in the language. *Sauraw-* comes from early EMJ *saburap-* which is said to be a changed form of OJ *samorap-* 'serve, be in attendance', also reflected in the word *samurai*.

During LMJ *-sauraw-* was used as the general polite marker; it changed to *-sɔɔraw-* by regular sound change (11.5), but also gave a number of other more reduced shapes with irregular inflected forms, such as *sɔɔ(-)* which was both used on its own as the nonpast form and as a stem for attaching other morphemes: *sɔɔta* past, *sɔɔnu* negative. This divorced its polite-style marking function from its use a lexical verb. Thus, *sɔɔrɔɔ* and its different variants might best be understood as an auxiliary, rather than as an auxiliary verb, but it should be noted that they continued to function as polite variants of *ar-* in all its functions, including combining with the copula gerund *de* to form polite copula forms, e.g. *de sɔɔrɔɔ* or *de sɔɔ*. The cNJ polite copula *des-* is by some

scholars thought to descend from *de sɔɔ*. *Sōrō-bun* is the name of a formal style of writing, which developed during NJ, which is characterized by extensive use of NJ *sooroo* to conclude sentences.

LMJ had several more polite style synonyms of *ar-*: *odyar-* (< *o-ide-ar-*; *ide-* 'go out') and *oryar-* (< *o-iri-ar-*; *ir-* 'enter') both originate in respect verbs for 'come, go' and further 'exist'. They came to be used widely in late LMJ as polite style verbs (especially as *ar-* equivalents), but disappeared during NJ. Also *gozar-*, which is still used widely in cNJ as a suppletive (super-)polite existential verb, originated in LMJ as a respectful equivalent of *ar-*. It was originally a compound of *goza* and *ar-*, and through LMJ had the alternative shape *goza ar-* and the negative form *goza na-*. *Goza* arose in early LMJ as a SJ reading of 御座 which had been used as a *kun*-writing of the respectful existential verbs *owas-* and *owasimas-* (see 9.2.4). At the end of LMJ, *gozar-* was the more frequent shape and the verb was in frequent use.

Finally, in the second half of LMJ, *-(i)sɔɔrɔɔ* was replaced as the general marker of polite style by the auxiliary *-(i)marase-*, which is the source of the present-day polite style auxiliary *-(i)mas-*, e.g. from *Esopo*: *i-marasuru* exist-POL.NONPST 'is', *osie-marasyoozu* teach-POL.INT '(I) will teach'. As opposed to the other polite style verbs and auxiliaries mentioned above, *-(i)marase-* does not originate in a suppletive respectful or humble existential verb. Its OJ source is the humble verb *mawir-* 'come.HUM, go.HUM' (which itself is reconstructible as *maw + ir- 'humble prefix + enter'), through lexicalization of the causative *mawira-se-* to a suppletive humble verb *mawirase-* 'give.HUM' > (by regular sound change, 7.3.2.3) *mairase-*, which came to be used as a humble auxiliary verb *-mairase-*, which in turn in late LMJ was further reduced phonologically and became an auxiliary *-(i)marase-* which shifted from humble to polite, eventually being reduced even further to its present-day shape *-(i)mas-*; this is thought to have taken place early in NJ, but there are examples from LMJ which suggest that it may have happened already in late LMJ and that the two shapes *-(i)marase-* and *-(i)mase-* coexisted for some time during LMJ, but that the latter was only sporadically reflected in writing.

REFERENCES

General: Imaizumi 1971, Iwai 1971, 1973, Rodrigues 1604–8. *Shōmono*: Yuzawa 1929/1970. *Feiqe*: Yuzawa 1928. *Esopo*: Ōtsuka and Kita 1999 (includes a classified index to *Esopo*). Concl/adn merger: Yamauchi 2003. Existential verbs: Kinsui 2006. *No* as subject marker: Yanagida and Whitman 2009. Adnominal clauses and nominalizations: Frellesvig and Whitman forthcoming, Kinsui 1995, Shibatani 2009, Wrona 2008: 237ff., forthcoming.

Part IV

Modern Japanese

13.1 The Edo period: Linguistic diversity and common language

In the Edo period, geographical and social mobility was very low, and most people stayed all their life in the same place within provinces which were in effect isolated feudal states. This situation resulted in a large number of fairly small and self-contained speech communities. The linguistic diversity with lack of mutual intelligibility between many dialects within Edo-period Japan is famous and is well illustrated by the following extract from Furukawa Koshōken's *Tōyū zakki* (東遊雑記 'Notes from a journey to the east'), an account of a journey in 1788 to northern Honshu and Hokkaido, accompanying inspectors from the central government in Edo, where he describes experiences with the local dialects they encountered (cited from Shibata 1998/1965: 184). Of the dialect in Tajima (in present-day Fukushima prefecture), Furukawa writes:

> Both sides can understand only half of the other's language. There was nothing one could do but laugh about it even at the inns. When we asked for *cha-zuke* 'tea on boiled rice' they brought *yu-zuke* 'hot water with boiled rice' instead, and we would go into the inn kitchen each time and make it ourselves.

And about the dialect in Nanbu (in present-day Iwate prefecture):

> The language of both men and women was gibberish, with only two or three words out of ten being comprehensible. The local lords usually give each inspector, in addition to his guides, two or three people well-informed about the castle town for the duration of his stay. The lord of Morioka Castle gave us two interpreters in addition because the speech of this part of the Nanbu domain is notoriously incomprehensible. But even the interpreters often could not understand what people say here.

13.1.1 Common language

Although the centre of political power shifted away from Kyoto with the establishment of Edo as the *de facto* capital by the Tokugawa shogunate at the beginning of the seventeenth century, Kyoto retained its position of the

prestige centre of high culture through the Edo period, and its language the position of prestige language. While not having the standing of a 'standard' language, a form of Kyoto Japanese was in use throughout Japan and functioned at the beginning of the Edo period as a common language among the political and military elites, high ranking officials, clergy, etc., in addition to the many dialects of great diversity all over Japan used by local people.

13.1.2 The language of Edo

Edo was a small castle town at the beginning of the seventeenth century. Its growth since then has taken place mainly by immigration. Through the Edo period, the system of 'alternate attendance' (*sankin-kōtai*) brought a constant flow of local feudal lords and their entourages to Edo, which also attracted significant numbers of unemployed warriors and which had a growing merchant and artisan population. Life in Edo was characterized by frequent and diverse social intercourse between the inhabitants and visitors and by a lively popular culture and entertainment industry. Through the second half of the Edo period, Osaka developed similarly to Edo as a socially diverse centre for commerce and trade, also with a rich popular culture. These two settings thus made for dynamic linguistic environments which were quite different from the rural settings in which most people lived, or from the socially segregated life in Kyoto, and the language of both Edo and Osaka developed accordingly.

The language of Edo/Tokyo took form during the Edo period, but the process of its formation is difficult to trace in detail. It is clear, however, that Edo Japanese is no direct continuation of any particular local dialect of Japanese, but came into being and developed through close contact between speakers of different varieties in an urban setting. In its formative phase, the language of Edo involved a significant component of the Kyoto-based common language, which was influenced by various dialects and by its use in the urban setting of Edo. The colloquial local language of Edo which developed through the period and which is attested particularly from the middle of the eighteenth century is certainly characterized by a wealth of occasional, fashionable, jocular or specialized vocabulary, by formal styles employing different honorifics, as well as by informal speech styles with noticeable phonological reductions; a stereotypical example is *arinsu* which was a polite auxiliary verb reduced from *arimasu*, used mainly by prostitutes in the red-light district of Yoshiwara. However, these are in many cases ephemeral features of ever-changing and renewing language usage, typical of a growing urban setting, and many of the special features which may be found in the colloquial language of Edo have not survived into the contemporary language.

Alongside the developing urban popular idiom, Edo Japanese, a form of the Kyoto-based common language continued to be used in Edo by the upper

classes and towards the end of the period also by the emerging educated middle class. This is the variety which eventually came to form the basis for standard cNJ. Although it was influenced by the local Edo language, it changed slowly, and it maintained its contiguity with the language in Kyoto. Differences between the versions of the common language used by the upper classes in Kyoto and in Edo developed, but the users of the common language continued to make up a socio-economically constituted, but geographically discontinuous, large speech community, with some internal variation, which played an important role through the NJ period, at least until the beginning of the twentieth century.

13.2 The Meiji period: Unification and standardization

Meiji was a period of unification in political terms: the creation or building of a unified, modern nation state. This unification effort also affected language, in two main ways: one was the unification, or alignment, of the spoken (i.e., contemporary) and the written language (13.2.1); and the other was the unification of the different varieties of Japanese spoken through the country as, or under, one national language, that is to say, the creation of a national, standard language (13.2.2).

13.2.1 Genbun'itchi

As mentioned above, Japanese was throughout the Edo period written in a variety of ways which had in common that they reflected the contemporary language very indirectly or not much at all, mostly being very convoluted and involving some form of the classical written language. In addition to the fact that there was no established way of representing the contemporary language in writing, the effort and time required to become literate were considerable. In the second half of the nineteenth century, with the opening of Japan and establishment of commercial and diplomatic contact with European countries and the US, an appreciation grew among educators, reformers and modernizers that the way in which Japanese was then written hindered literacy, education and, more generally, modernization. From the 1860s through to the first decade of the twentieth century, there were fierce debates over whether, and how, to write the contemporary language, with lines sharply drawn between proponents and opponents of reform. From the mid 1880s authors who wanted to write realistic novels joined the movement, or campaign, for vernacularization of the written language, which came to be known as *genbun'itchi* ('unification of speech and writing', 言文一致). The serial publication in 1887–9 of Japan's first modern novel *Ukigumo* by Futabatei Shimei (1864–1909), which was written in the vernacular, was an important landmark which gave the campaign

additional momentum. From then onwards more literature written in the vernacular became published, and also newspaper editorials gradually adopted a vernacular-based written form. The first school textbooks with large proportions of vernacular language were issued by the Ministry of Education in 1903–4. Although a form of the classical written language was still used in many newspaper articles until the 1920s, and until the end of WWII in official documents and government decrees, and even in some academic publications (for example Yamada Yoshio's grammars of OJ and EMJ from 1913), the reformers had for all practical purposes won by the mid 1910s, with the vernacular adopted for education, literature and much public written discourse. As mentioned in 6.1.3, reform of the *kana* orthography with its etymological spelling principle (for example spelling /kyoo/ 'today' as けふ "ke.fu"), did not take place until after WWII.

13.2.2 A national language; standardization and dialect eradication

Part of the modernization effort and the establishment of a unified nation-state was also the notion of a unifying national language. Especially after the publication of an essay collection entitled *Kokugo no tame* 'For the national language' by Ueda Kazutoshi in 1895, in which he drafted in the SJ word *kokugo* (国語 'country-language') for this notion of Japanese as national language, the word *kokugo* has become widely used to refer to the Japanese language, and the word still has strong emotional value for some. For example, the renaming in 2001 of the 'Society for Japanese Linguistics' from *Kokugo gakkai* to *Nihongo gakkai* caused rifts in Japanese academia which have yet to be healed. *Kokugo* is an exclusive, somewhat nationalistic term, for it can only be used to refer to the Japanese language by and for the Japanese. However, both the notion of national language and the term for it were in the twentieth century exported to Korea, where 国語 *kug-ŏ* means 'the Korean language'.

Both the notion of *kokugo* and the practical purposes of the vernacularization of the written language (spreading literacy and education) pointed in the direction of establishing a standard language, that is to say, privileging one variety over others for use in public life, including education, and as an emblem of the nation. Thus, in 1901, the Ministry of Education decreed that the Japanese language taught in schools should be that of Tokyo, whereby was meant *not* the language of the common people of the downtown area, but the language of the middle and upper classes of the Yamate area, in other words the descendant of the Edo-influenced variety of the common language (13.1). One example from phonology of a well-attested Edo feature which did not pass into the standard language is the monophthongization of /ai/ as [e:]. This is attested from the early eighteenth century and is still today a colloquial feature of Tokyo speech, but is not sanctioned in the standard language.

In 1902, the *Kokugo-chōsa-iinkai* (国語調査委員会 'National Language Research Council') was set up, charged with surveying the state of the national language and making recommendations, amongst other things, for establishing a standard language. The *Kokugo-chōsa-iinkai* was dissolved as a body in 1913, but its recommendations regarding a standard written language were published in *Kōgohō* (口語法 'Grammar of the vernacular', 1916) and *Kōgohō bekki* (口語法別記 'Supplement to Grammar of the vernacular', 1917). It was in these volumes that the normative grammar of standard Japanese as we know it today was set out and where it was clearly said that standard Japanese was based on the speech of the educated middle and upper classes of Tokyo. The chief editor for these two volumes was Ōtsuki Fumihiko (1847–1928), so if any individual is to be credited with the creation, or at least codification, of modern standard Japanese, it is he.

While standardization and vernacularization of the written language has had many positive effects in terms of spread of literacy and education and increased popular participation in public life and eventually also in the political process, the flipside was a directed and quite successful effort to eradicate dialects, especially through the first half of the twentieth century, leading to a marked loss of linguistic diversity in Japan which is regrettable, at least to linguists and ethnographers. In the pre-war years, users of dialect in schools were subjected to public ridicule and even punishment. For example, in some schools those overheard using the local dialect rather than the standard language were forced to wear *hōgen fuda* 'dialect tags'. It was only after WWII that the active discouragement of use of dialects was relaxed, but even today many people are ashamed of speaking their local dialect in public. Recently, an Osaka-based variety of Kansai Japanese is getting wider public exposure, especially in popular culture, and is even thought fashionable and imitated by some young people. Even so, many speakers of Tokyo Japanese will profess – somewhat disingenuously – an inability to understand Kansai Japanese.

13.3 Sources

Present-day Japanese is observable and well described. We have a large amount of material from the Meiji period onwards (second half of the nineteenth century), both in Japanese script and in alphabet writing, which shows that the language has not changed much since then. Before Meiji, throughout the Edo period, Japanese was written in a variety of ways which had in common that they reflected the contemporary language very indirectly or not much at all, mostly being very convoluted and involving some form of the classical written language. These ways of writing ranged from *(hentai) kanbun* over various forms of Classical Japanese to highly formalized versions of post-Classical Japanese, which was, however, still heavily influenced by the

classical written language. There is, however, also a large body of popular fiction and drama in which the contemporary vernacular was used, especially in renditions of spoken language, namely in lines in plays and in dialogue parts of novels. Until the middle of the period, the language reflected in these sources was mostly that of Kyoto, but from the second half of NJ, literature which reflects the emerging urban idiom of Edo was increasingly published. The Edo sources are far too numerous to name.

In addition to *kabuki* and puppet play scripts, the drama material includes *kyōgen*, which are lively short comedies or sketches, originally performed in between *noh* plays as a kind of comic relief. *Kyōgen* plays gained independent popularity from the Muromachi period and at that time they were performed in the contemporary spoken language. However, over time a formulaic style of language evolved which is reflected in the great majority of the surviving texts which date from the mid seventeenth century onwards. As the plays were originally handed down in performing traditions, the NJ *kyōgen* texts in fact reflect many features of the language of the Muromachi period.

13.3.1 Material in alphabet writing

We have a wealth of material in alphabet writing from the second half of the nineteenth century which is very close to the contemporary language. Two earlier decriptions of Japanese are, however, worth mentioning. The first is a seventeen-page article, published in 1792 in Uppsala, 'Observationes in Linguam Japonicam' by the Swedish botanist Carl Peter Thunberg who stayed in the Dutch trading post in Dejima off Nagasaki for around a year and a half during 1775–6. In it Thunberg gives a brief description of the Japanese language. The language he describes seems to be common educated Japanese, including, however, identifiable Kyushu dialectal features. Thunberg is clearly no linguist, but his little piece contains valuable information on late eighteenth-century Japanese. Phillipp Franz Balthasar von Siebold was a German physician who taught medicine in Japan from 1823 to 1829 (when he was expelled from Japan accused of being a spy because he had collected maps). His seventy-page long *Epitome linguae japonicae* was written in Nagasaki in 1824 and, again, has valuable information about Japanese from the early part of the nineteenth century.

13.4 From Late Middle Japanese to standard contemporary Modern Japanese

As outlined above, the colloquial urban idiom of Edo and Tokyo, Edo Japanese, is not a direct descendant of the language of Kyoto OJ, EMJ or LMJ reflected in the written sources, because of its multiple sources and influences (although

its initial main component was the common language). However, the situation is different for the version of the common language used in Edo/Tokyo upon which the standard language came to be based. Although it was undoubtedly influenced by Edo-Japanese, it represents a fairly unbroken tradition of the Kyoto-based common language. Thus, comparing the language of for example *Esopo*, which is from the very end of LMJ and which reflects the common language of Kyoto, with cNJ, what is really striking is how relatively speaking few and small the structural differences are. In fact, there are very few features of cNJ which are not straightforwardly derivable from LMJ as reflected in the sources from the end of the period. In the following the main phonological and morphological differences between LMJ and cNJ will be outlined. In a few cases differences between LMJ and cNJ seem to reflect influence from eastern dialects (mediated through Edo Japanese). However, most phonological and morphological changes took place through the common language and are reflected in both Kyoto and Tokyo cNJ, demonstrating the coherence and influence of the common language speech community at least until the beginning of the twentieth century. These changes, outlined in Chapters 14 and 15, are describable as simple linguistic changes between LMJ and cNJ. Features which are now part of the standard language, but were not used in Kyoto LMJ or NJ, and which are thought to reflect eastern dialect influence, will be discussed below (in Chapter 16).

REFERENCES

General: Iwai 1974, Matsumura, 1998, Sugimoto 1988, Yuzawa 1957, 1970. Women's language: Endō 2006, Sugimoto 1985. Unification, standardization and dialect eradication: Ramsey 1991, Shibata 1998/1965, Twine 1991. Western sources: Sugimoto 1989.

Table 14.1 gives the by now very conservative sound inventory of free moras in cNJ. The phonemic representation in Table 14.1 is used in the *kunreishiki* (訓令式 'official directive system') system of transcription, which in a modified version is used by most linguists, whereas the more broadly used Hepburn system reflects allophonic variation (see 14.5) and uses letter combinations which appeal to English readers: /si/ *shi*, /ti/ *chi*, /tu/ *tsu*, /hu/ *fu*, /zi/ *ji*, /sya/ *sha*, /syu/ *shu*, /syo/ *sho*, /tya/ *cha*, /tyu/ *chu*, /tyo/ *cho*, /zya/ *ja*, /zyu/ *ju*, /zyo/ *jo*. Some of this originally allophonic variation has since become phonemic in recent cNJ (see 14.6) and that is the main reason we have chosen to use the modified Hepburn system in transcription of cNJ in this book.

During NJ a few sound changes took place. The precise dating of them is not easy for there are no extensive materials in alphabet writing until the end of the Edo period. The sound changes which took place between the end of LMJ and conservative cNJ are outlined in the following. All the distinctions which were lost in the course of these changes are preserved dialectally.

14.1 Merger of /d/ and /z/ before /i, u/

Through the latter part of the LMJ period /t, d/ became phonetically assibilated before the two high vowels, /i, u/, as reflected in the Christian sources (11.6). Early in the NJ period, this further led to the merger of /d/ and /z/ before /i, u/; the outcome of the merger is considered to be /z/, which however continues to have affricated realizations:

(1) /d/ > /z/ / __ /i, u/

/du/ [dᶻu] ≠ /zu/ [zu] > /zu/ [dᶻu ~ zu]

/di/ [dʒi] ≠ /zi/ [ʒi] > /zi/ [dʒi]

midu 'water' ≠ *mizu* 'not seeing; doesn't see' > *mizu*

midika- 'close' ≠ *mizika-* 'short' > *mizika-*

This change is often thought of as a merger between the moras *di* ≠ *zi* and *du* ≠ *zu*, and in Japanese it is referred to in terms of the *kana* letters representing

Table 14.1 *Free moras in conservative cNJ*

a	ka	sa	ta	na	ha	pa	ma	ya	ra	wa
i	ki	si	ti	ni	hi	pi	mi		ri	
u	ku	su	tu	nu	hu	pu	mu	yu	ru	
e	ke	se	te	ne	he	pe	me		re	
o	ko	so	to	no	ho	po	mo	yo	ro	
	ga	za	da		ba					
	gi	zi			bi					
	gu	zu			bu					
	ge	ze	de		be					
	go	zo	do		bo					
	kya	sya	tya	nya	hya	pya	mya		rya	
	kyu	syu	tyu	nyu	hyu	pyu	myu		ryu	
	kyo	syo	tyo	nyo	hyo	pyo	myo		ryo	
	gya	zya			bya					
	gyu	zyu			byu					
	gyo	zyo			byo					

these moras, the *yotsu-gana* 'four *kana*', viz. じぢずづ. The Christian materials keep the four moras distinct, noting them as *ji* /zi/, *gi* /di/, *zu* /zu/ and *zzu* /du/, respectively, but the merger seems to have taken place in the first half of the seventeenth century. As mentioned in 6.1.3 above, spelling conventions still today keep じ versus ぢ, and ず versus づ distinct in compound contexts, for example spelling *mi<u>z</u>ika-* 'close, familiar' as みぢか, reflecting its origin in a compound *mi* 'body' + *<u>t</u>ika-* 'close', or *kana-<u>z</u>ukai* '*kana*-usage, spelling' as かなづかい, cf. *<u>tsu</u>kai* 'usage'.

14.2 Merger of /ɔɔ/ and /oo/; /ɔɔ/ > /oo/

Also early in the NJ period, the distinction between /ɔɔ/ (< /au/) and /oo/ (< /ou, eu/) (cf. 11.5) was lost:

(2) /ɔɔ/ > /oo/

tɔɔto- 'precious' ≠ *imooto* 'little sister' > *tooto-*, *imooto*

This change affected in particular the shapes of a number of SJ vocabulary items (see 11.5.2), and the morphophonological rules applying to inflected adjective and verb forms ending in *-u* (12.3.3), which changed so that both /au/ and /ou/ gave /oo/, see (3). See below (15.1.1) about the volitional whose

formation changed so these rules became irrelevant, and see 14.2.1 about the nonpast of *-w-* base verbs.

(3)

LMJ		NJ
/au/ => /ɔɔ/	>	/au/ => /oo/
/ou/ => /oo/	>	/ou/ => /oo/

(4) a. Verbal *u- onbin* stem

kaw- 'buy'	*kau-te* =>	*koote*
kam- 'bite'	*kau-de* =>	*koode*
yow- 'get drunk'	*you-te* =>	*yoote*
yom- 'read'	*you-de* =>	*yoode*

b. Adjectival copula infinitive *-u*

taka- 'tall'	*taka-u* =>	*takoo*
kuro- 'black'	*kuro-u* =>	*kuroo*

14.2.1 /Vu/ diphthongs

In the course of the LMJ contractions of /Vu/ sequences to long vowels, /Vu/ diphthongs were eliminated at the phonemic surface level by sound changes (cf. 11.5) and by morphophonological rules such as those discussed immediately above which contracted sequences of /V + u/ which arose in the formation of some inflected forms (12.3.3). However, in the course of NJ, this rule stopped applying to the formation of the nonpast of *-w-* base verbs, so that in cNJ the nonpast of *aw-* 'meet' is *au* and *yow-* 'get drunk' is *you*, not *oo* or *yoo*. This, combined with the intake of loanwords such as *mausu* 'mouse (especially for use in medical experiments)' (from German *Maus*), reintroduced /Vu/ sequences in the language. Apart from the very recent relaxation of restrictions on the occurrence of /Q/ (see 14.6 below), this was the only change which affected long syllables in the NJ period.

14.3 Delabialization of /f/; /f/ > /h/

The final change to affect the reflex of OJ /p/ was the change of /f/ (see 11.3) to /h/ which occurred at some point during NJ. This change was primarily a phonetic change with no mergers or other immediate phonemic impact and it was therefore not reflected in general writing. Thunberg (1792) mostly writes words reflecting LMJ /f/ with 'f' (e.g. *fanna* 'flower'), but says that words with 'f' sometimes are pronounced ('read') with 'h' (e.g. *hanna*) and Siebold (1826) uses 'h' in his transliteration of the *kana* letters はひふへほ, but notes that they are often pronounced with 'f'. However, by the middle of the nineteenth century, sources in alphabet writing generally write 'h', except before

'u' where it is 'f', reflecting that /h/ is pronounced [ɸ] before /u/. Hepburn's transcription system from the end of the nineteenth century adopted this allophonic notation which since then has dominated transcriptions in the west. Very recently, due to a large intake of loanwords from English, [ɸ] has become phonemic /f/ before vowels other than /u/ (14.6).

14.4 Loss of phonetic onglides

Two changes affected glides: one was the loss of the automatic onglides in the syllables /.o/ and /.e/, see (5). This happened earlier to /.o/ which is transcribed 'o' in Thunberg (1792) and Siebold (1824), whereas /.e/ is transcribed *ye* into the Meiji period, as reflected in old spellings such as *Yedo* for the capital, as well as spellings in English dating from that period, e.g. *yen* for the Japanese currency, *Yebisu* for the beer brand, or *ue* spelled *uye* in for example *Inouye* and other personal names, especially by Japanese-Americans.

(5)

LMJ		NJ
/.o/ => [^{w}o]	>	/.o/ => [o]
/.e/ => [^{j}e]	>	/.e/ => [e]

Second, the labial glide which had survived in /kwa, gwa/ in SJ words such as *kwaiwa* (会話) 'conversation', *kwasi* (菓子) 'cookie', or *gwannen* (元年) 'first year in a dynastic period' was lost, so that these syllables merged with /ka, ga/ (*kaiwa*, *kasi*, *gannen*). Again, this change was completed late in the nineteenth century, and the glide is reflected in occasional spellings such as *Kwannon* (Kannon, the Goddess of Mercy) or in *Kwansei Gakuin Daigaku*, the name of a private university in Kobe.

14.4.1 Loss of palatalization before /e/

Allophonic palatalization of consonants before /e/ was lost, probably related to the loss of the automatic glide before /.e/. However, palatalization of all consonants is retained before /i/. Thus *siri* [ʃiri] 'buttocks', but *seri* [seri] 'parsley' < LMJ [ʃeri].

14.5 Summary of main allophonic variation

The most conspicuous allophonic variation in conservative cNJ is summarized in (6). We do not exemplify palatalization of all consonants or aspiration of /p, k/. Intervocalic nasalization of /g/ is on the way out and is for many speakers today a normative feature requiring conscious effort. As in LMJ, vowels in final position are often weakened, as are /i, u/ between consonants, especially, but not exclusively, if these are phonetically unvoiced.

(6)

/t/ =>	[tʰ] /__ /e, a, o/
	[ts] /__ /u/ [affrication]
	[tʃ] /__ /i, y/ [affrication, palatalization]
/s/ =>	[s] /__ /e, a, o, u/
	[ʃ] /__ /i, y/ [palatalization]
/h/ =>	[h] /__ /e, a, o/
	[ɸ] /__ /u/ [labialization]
	[ç] /__ /i, y/ [palatalization]
/z/ =>	[z] /__ /e, a, o/
	[z, ᵈz, dz] /__ {u} [affrication]
	[ʤ] /__ /i, y/ [affrication, palatalization]
/g/ =>	[g] / #__
	[ŋ] /V__V

14.6 Recent phonemic changes due to loanwords

In the course of the intake of loanwords from the beginning of the Meiji period, but particularly after WWII, some phonemic changes have taken place, which are either phonemicization of the phonetic variants outlined immediately above, and/or new combinations of existing phonemes. Thus, in recent cNJ, the following sounds which in conservative cNJ are conditioned variants are now distinctive before /e/, in addition to before /a, o, u/ (7a). This was made possible by the loss of automatic palatalization before /e/. Furthermore, [t] has become phonemic before /i, u/ (7b), and, marginally, [ts] is now found before other vowels than /u/, (7c). Finally, [ɸ] is now used widely before other vowels than /u/, (7d).

(7) a. [ʃ] *sherī* 'sherry' [ʃeri:] versus *seri* 'parsley' [seri]

[tʃ] *chēn* 'chain' [tʃe:N] versus *teinai* (邸内) 'premises' [te:nai]

[ʤ] *jerī* 'jelly' [ʤeri:] versus *zero* 'zero' [zero]

b. [t] *pātī* 'party' [pa:ti:] versus *pachinko* 'pachinko' [patʃiŋko]

Tūru 'Tours' [tu:ru] (city in France) versus *tsūru* 'tool' [tsu:ru]

c. [ts] *tsaitogaisuto* '*zeitgeist*, spirit of the times' [tsaitogaisuto]

kantsōne 'canzone' [kantso:ne]

d. [ɸ] *fan* 'fan' [ɸaN] versus *han-* 'half' [haN]

Due to sound changes which took place in EMJ and LMJ, /w/ was lost before other vowels than /a/ (11.8.1), but recent loanwords have re-introduced /w/ to a limited extent before /i, e, o/, e.g. *wirusu* 'virus', *wesutan* 'western (movie)', *wokka* 'vodka'. Finally, [v] is used by some educated speakers in free variation with [b] or [w] in *some* loanwords reflecting foreign [v]: *vandaru-zoku* (~ *bandaru-zoku*) 'the Vandals (Germanic tribe)' (SJ *-zoku* 族 'tribe'), *venetsia* ~ *benetia* 'Venice', *vaimāru* ~ *waimāru* 'Weimar'.

In long syllables /Q/ was until cNJ restricted to occurring before tenues (cf. 11.1.2), but it now also occurs before mediae e.g. *beddo* 'bed' /beQdo/, *baggu* 'bag' /baQgu/, *karejji* 'college' /kareQzi/. Long /aa/ (/aV/) seems not to have been used outside expressive forms until the analogical formation of the distal adverb *ā* 'that way' (corresponding to *kō* 'this way' and *sō* 'that way'), which is attested in the late eighteenth century. From the Meiji period /aa/ is found in loanwords such as *gāden* 'garden'.

REFERENCES

Mabuchi 1971, Martin 1987, Vance 1987, Wenck 1959.

15.1 Verbs

The main inflected standard cNJ verb forms are shown in Table 15.1. In terms of morphological categories, the main differences between the LMJ (12.1.1) and NJ paradigms are the loss of the tensed non-finite forms (12.1.3.3) and the loss of the intentional (12.1.3). The new representative reflects the EMJ stative auxiliary *-(i)tar-*; there are examples of this use of *-(i)tari* in LMJ, but it does not become established as an inflected form until NJ. The past conjectural is no longer a productive form in cNJ, but is used in old-fashioned writing and in some dialects. See below regarding the shape of the nonpast (15.1.2.1), the past, gerund, conditional and representative (16.1), the volitional (15.1.1), and the imperative (16.4).

15.1.1 Formation of the volitional

In LMJ the volitional was formed by *-u*, attaching to the *a-* stem of consonant base verbs and the basic stem of vowel base verbs, e.g. *kaka-u* => *kakɔɔ*, *ake-u* => *akyoo* (12.3.3). In NJ the shape and formation of the volitional changed to give the NJ forms shown in (1):

(1)

	LMJ	NJ
kak- 'write'	*kaka-u* => *kakɔɔ*	*kakoo*
kaw- 'buy'	*kawa-u* => *kaɔɔ*	*kaoo*
ake- 'open'	*ake-u* => *akyoo*	*akeyoo*
ne- 'sleep'	*ne-u* => *nyoo*	*neyoo*
oki- 'arise'	*oki-u* => *okyuu*	*okiyoo*
mi- 'see'	*mi-u* => *myuu*	*miyoo*
ko- 'come'	*ko-u* => *koo*	*koyoo*
se- 'do'	*se-u* => *syoo*	*shiyoo*

Table 15.1 *Standard cNJ inflected verb forms*

	QD	QD	LB
Basic stem	*kak-*	*kaw-*	*ake-*
Finite			
Nonpast	*kaku*	*kau*	*akeru*
Past	*kaita*	*katta*	*aketa*
Volitional	*kakoo*	*kaoo*	*akeyoo*
Imperative	*kake*	*kae*	*akero*
(Past conjectural)	*(kaitaroo)*	*(kattaroo)*	*(aketaroo)*
Non-finite			
Infinitive	*kaki*	*kai*	*ake*
Gerund	*kaite*	*katta*	*akete*
Conditional-1	*kaitara*	*kattara*	*aketara*
Representative	*kaitari*	*kattari*	*aketari*
Conditional-2	*kaitewa*	*kattewa*	*aketewa*
Provisional	*kakeba*	*kaeba*	*akereba*
Concessive	*kaitemo*	*kattemo*	*aketemo*

Apart from the merger of /ɔɔ/ and /oo/ (14.2), which also affected the shape of the volitional of the consonant base verbs, e.g. *kakɔɔ* > *kakoo*, the only change was in the shape of the vowel base verb volitional. This change is probably related to the levelling and greater stem transparency in the vowel base verb conjugations. Whether the volitional forms are analysed as (a) *kak* + *yoo* => *kakoo* and *ake* + *yoo* => *akeyoo*, with deletion of the initial glide in the suffix with the consonant verbs, or (b) *kak* + *oo* => *kakoo* and *ake* + *oo* => *akeyoo*, with automatic glide insertion with the vowel base verbs, is mainly a theoretical question. In any case, the new formation resulted in greater uniformity and transparency in the morphological composition of this form across the conjugation classes, with suffixation of a flective (*-yoo* or *-oo*) directly to the basic stem of verbs from all conjugation classes (except the irregular *s*-irr class).

15.1.2 Verb classes

The number of conjugational verb classes was reduced during the NJ period, to give the five classes of cNJ.

15.1.2.1 Levelling of vowel base verb conjugations; merger of monograde and bigrade verbs

From OJ through LMJ most vowel stem verbs belonged to the bigrade conjugations (LB and UB), which had stem alternations between a basic stem, e.g. *ake-*, and a derived form, *aku* (<= *ake* + *u*), which in OJ and EMJ was the

conclusive form also functioning as a stem for formation of further forms (3.4.1.3). With the concl/adn merger and the loss of the exclamatory form in LMJ (12.1.2), the derived form was no longer an independent word form, but only a derived stem which was used in the formation of the nonpast form (e.g. *akuru* <= *aku-* + *ru*).

The monograde vowel stem verbs, UM and LM, had no such alternation between a basic and a derived stem, but used the basic stem in formation of all forms. In EMJ short (one-mora) UB verbs shifted to monograde conjugation, so that short *i*-stem verbs belonged to UM, and stems of more than two moras belonged to UB (8.1.2), but other than that the difference between bigrade and monograde vowel stem conjugation remained through LMJ. However, in the course of the first half of NJ, bigrade conjugation was lost in most dialects (and is today retained only in a very small number of dialects) and all LB verbs became LM, and UB verbs became UM.

As part of this change the shape of the nonpast changed so that it too came to be formed on the basic stem, see (2a) below. With the changes in the formation of the volitional (15.1.1), the levelling of the vowel base conjugations resulted in greater transparency in the morphophonology of the vowel base verbs, with direct manifestation of the basic stem in all inflected forms. The change of the LB verbs to LM took place first with short verbs and at the end of LMJ a few short *-e* base verbs were in the process of shifting and had variant nonpast forms, e.g. *fe-* 'pass': *furu* ~ *feru*.

15.1.2.2 Merger of n*-irregular and quadrigrade verbs*

The (irregular) *n-* base verb *sin-* lost its irregular conjugation and became a regular QD verb. As with the levelling of the bigrade verbs to monograde, this change mainly affected the formation of the nonpast, (2b):

(2)		LMJ	NJ
a.	*ake-* 'open'	*akuru*	*akeru*
	ne- 'sleep'	*nuru*	*neru*
	oki- 'arise'	*okuru*	*okiru*
b.	*sin- 'die'*	*sinuru*	*sinu*

15.1.3 Summary of verbal conjugation classes

The general changes among verbal conjugation classes from OJ to NJ are summarized in Table 15.2.[1] The main change between OJ and EMJ (see 8.1.2)

[1] In Japanese school grammar, the regular consonant base verbs in cNJ are called *godan* ('quinquigrade') instead of *yodan*, reflecting that the *katsuyōkei* system posits an extra stem, ending in *-o*, whose only function is to be used to form the volitional by attaching *-o*. Thus, in the *katsuyōkei* system *kak-* 'write' has the *suiryōkei* ('conjectural stem') *kako-* to which *-o* is attached to give the volitional *kakoo*. The sole motivation for this analysis is to fit the formation of all forms into the *katsuyōkei* system.

Table 15.2 *Changes among verbal conjugation classed from OJ to NJ*

	OJ	EMJ	LMJ	NJ
QD	*CVC-*	*CVC-*	*CVC-*; *ar-*	*CVC-*, *ar-*, *sin-*, *ker-*
r-irr	*ar-*	*ar-*		
n-irr	*sin-*	*sin-*	*sin-*	
LM		*kwe-*	*ke-*	*Ce-*, *CVCe-*
LB	*CVCe-*, *Ce-*	*Ce-*, *CVCe-*	*Ce-*, *CVCe-*	
UM	*Ci-*, *Cwi-*	*Ci-*	*Ci-*	*Ci-*, *CVCi-*
UB	*CVCwi-*, *Cwi-*	*CVCi-*	*CVCi-*	
k-irr	*ko-*	*ko-*	*ko-*	*ko-*
s-irr	*se-*	*se-*	*se-*	*se-*

was the emergence of the LM class with EMJ *kwe-* > LMJ *ke-* as its only member, and the migration of the short (one-mora) UB verbs to UM, so that in EMJ/LMJ short *-i* base verbs were UM and more than two -mora *-i* base verbs were UB. In the course of LMJ, the *r*-irr class merged into the QD verbs as a result of the concl/adn merger (12.1.2). Already at the end of LMJ a few short LB verbs were in the process of migrating to LM, but in NJ the bigrade conjugation was lost and the UB verbs merged with the UM verbs to become UM, and the LB verbs became LM (15.1.2.1); however, the sole EMJ/LMJ LM verb *ke-* 'kick' migrated to QD and is now *ker-*. In addition to these general changes, a number of individual verbs have migrated between conjugation classes through the history of the language.

15.2 Adjectives and copula

The paradigms of both the adjectival and the regular copula show a simplification compared with LMJ. The main cNJ forms of the adjectival copula are shown in Table 15.3. Comparing the LMJ (12.2.1) and NJ paradigms, the main differences are the establishment of an inflected past tense for the adjectival copula as with the verbs and the shift from expressing negation by an auxiliary *-n-* (*-karan-*) to an analytic construction with *na-* (or polite *arimasen*) following the infinitive (*-ku na-*, *-ku arimasen*); see further 16.3. The LMJ concessive, *-keredomo* was lost (replaced by *-kutemo*), as was the LMJ conditional, *-kuwa* ~ *-kunba* (replaced by *-kattara*, which reflects the secondary adjective conjugation, supplemented with *-kutewa*). Both *-kutemo* and *-kutewa* conform to the same pattern as the verbs of using the gerund as a stem for further formations. The conjectural and past conjectural are no longer generally used productively in standard or Kansai cNJ, but have been replaced by

Table 15.3 *cNJ adjectival copula forms*

Finite	
Nonpast	*i*
Past	*katta*
(Conjectural)	*(karoo)*
(Past conjectural)	*(kattaroo)*
Non-finite	
Infinitive	*ku*
Gerund	*kute*
Conditional	*kattara ~ kutewa*
Provisional	*kereba*
Concessive	*kutemo*
Negative	*ku na-*

	taka- 'tall'	*utsukushi-* 'beautiful'
Finite		
Nonpast	*takai*	*utsukushii*
Past	*takakatta*	*utsukushikatta*
(Conjectural)	*(takakaroo)*	*(utsukushikaroo)*
(Past conjectural)	*(takakattaroo)*	*(utsukushikattaroo)*
Non-finite		
Infinitive	*takaku*	*utsukushiku*
Gerund	*takakute*	*utsukushikute*
Conditional	*takakattara ~ takakutewa*	*utsukushikattara ~ utsukushikutewa*
Provisional	*takakereba*	*utsukushikereba*
Concessive	*takakutemo*	*utsukushikutemo*
Negative	*takaku na-*	*utsukushiku na-*

attaching the conjectural form of the copula to the nonpast or past: *takai daroo* (*yaroo*), *takakatta daroo* (*yaroo*). in standard cNJ the infinitive is *-ku* (*takaku*, *utsukushiku*) and the *onbin*-variant, *-u*, is not used, except in the super-polite formation (cf. 12.3.3), but it is used in the Kansai dialects (*takoo*, *utsukushuu*, etc.), where it is also used in the forms built on the infinitive (gerund *takoote*, negative *takoo na-*, sometimes abbreviated, e.g. *tako na-*).

The main inflected copula forms are as in Table 15.4. *Dewa* is often abbreviated to *ja*. As with the adjectives, the past conjectural is very rare today. The finite forms are built on a stem *da* which ultimately derives from *de aru* (see 12.2.2), probably through *de aru* > *daru > *da*. *Da* is almost absent from LMJ,

Table 15.4 *cNJ copula forms*

	Plain	Polite
Nonpast	*da*	*desu*
Attributive	*na ~ no*	
Past	*datta*	*deshita*
Conjectural	*daroo*	*deshō*
(Past conjectural)	*(dattaroo)*	–
Infinitive	*ni ~ to*	
Gerund	*de*	
Conditonal	*dewa*	
Provisional	*nara(ba)*	
Concessive	*demo*	
Negative	*dewa na-*	

which instead has *dya* (< *dyaru* < *de aru*), and is thought to be an eastern Japanese feature which found its way into the common language used in Edo and from there into standard cNJ. It is in that case not an isolated dialect feature of very long standing, for *de aru*, which is the origin of both *da* and *dya*, is an LMJ formation. Note, however, that the difference in the pattern of fusion, *de aru* > Kansai (western) *dyaru* > *dya* :: cNJ (eastern) *daru > *da*, is similar to the difference in the OJ statives (3.1.4.7.2, 5.1): *saki-ar- > (western) OJ *sakyer-* :: EOJ *sakar-*. Some parts of Kansai use *ja*, the direct reflex of LMJ *dya*, but today Kyoto and Osaka mostly use nonpast *ya*, conjectural *yaroo*, past *yatta*, built on a stem *ya*, which is usually thought to be a further reduction of LMJ *dya*, but which as mentioned above may instead originate in a reanalysis of the EMJ particle *yarɔɔ* (cf. 12.1.4). The cNJ polite copula is thought to have originated in further reductions of LMJ *de sɔɔ*, copula gerund *de* + *sɔɔ*, a severely reduced form of the polite existential and auxiliary verb *sɔɔrɔɔ* (i.e., a polite form of *de ar-*), cf. 12.7.2.

15.3 Other new modern Japanese grammatical forms

In addition to the changes in verb and adjective inflection described above, a few other grammatical forms are an important part of cNJ and should be mentioned here: The conjunctional particle ***shi*** 'and (moreover)' is thought to be from the adjectival conclusive ending *-si* which disappeared as an inflectional ending with the concl/adn merger (12.2.1). It was used first with nouns and adjectival infinitives towards the end of LMJ and came to be used with verbs in NJ. The evidential extension ***rasi-*** 'seems, appears' developed earlier in Kyoto NJ than in Edo; it is not a direct continuation of the OJ extension

rasi- (which went out of use in EMJ, cf. 8.5), but is thought to have developed in early NJ from the derivational morpheme *-rasi-* which appeared in the second half of LMJ and which derives adjectives meaning '-like, typical of, suitable for' (e.g. (from *Vocabulario*) *afôraxij* /afɔɔ-rasii/ 'foolish', *afɔɔ* 'fool' > NJ *aho*); derivational *-rasi-* still coexists with the extension *rasi-*. Very recently, a similar course of development has given the extension ***ppoi*** (*iku ppoi* 'seems to be about to go') which is more or less synonymous with *rasii* and which until recently was only used as a derivational morpheme (*mizu-ppoi* 'watery'). Finally, the 'hearsay' extension ***soo-da*** originates in the LMJ evidential auxiliary *-(i)soo-na* (12.1.5.3) which continues to be used as an auxiliary today.

16 Eastern dialect features of the standard language

Although most features of standard NJ derive straightforwardly from the language reflected in the sources from the end of LMJ, there are a number of features of standard NJ which reflect influence of eastern Japanese dialects on the Edo version of the common language which came to form the basis for the standard language. It is interesting to consider what Rodrigues says in *Arte* about the language of Kanto (*Quantô* /kwantoo/, described as the area stretching from Mikawa (the southern part of present-day Aichi-ken) eastwards), that is to say, about eastern Japanese before the establishment of Edo as a political centre (*Arte*, pp. 612–13). He first introduces Kanto Japanese:

> Generally, the manner of speaking is rough and sharp. Many syllables are swallowed and not pronounced. Further, there are many coarse words particular to this region which can only be understood among the people from there.

He then goes on to list a number of features typical of Kanto Japanese. Other than the use of the particle *bei* (< *be-i* the necessitive extension + nonpast adjectival copula), these features are described contrastively, summarized in (1).

(1)		Kyoto	Kanto
a.	/se/ =>	[ʃe]	[se]
b.	Allative particle	*e*	*sa*
c.	Shape of intentional	*-ɔ-ɔze-*	*-a-nze-*
d.	Adjectival copula infinitive	*-u*	*-ku*
e.	*Onbin* of *-w* base verbs	*-u*	*-Q*
f.	Negative auxiliary	*-nu*	*-nai*

(a) describes a sound change that took place earlier in Kanto than in Kansai (loss of palatalization before /e/). The allative particle *sa* is not used in standard NJ. (c)–(e) are *onbin* forms, where the western forms are more vocalic and the eastern forms more consonantal; the intentional is no longer used, but the distribution of adjectival and verbal forms is like that today, with the Kanto

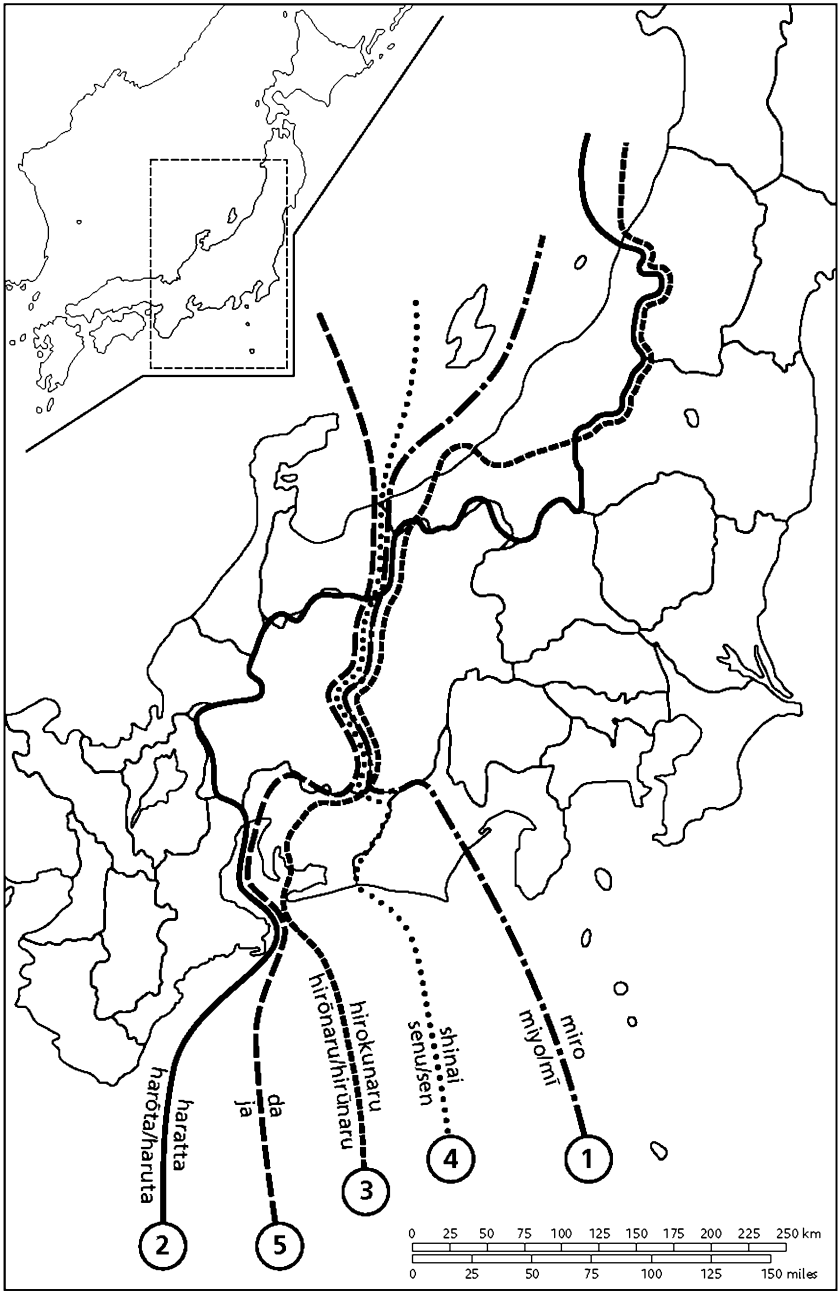
forms being used in the standard language (see 16.1). Finally, it is also the Kanto negative auxiliary which is used in standard NJ (see 16.3).

The following features of standard Japanese reflect influence from eastern Japanese dialects: In phonology, the main difference between standard (Tokyo) cNJ and Kyoto cNJ is the prosodic system (see 7.4). In this case it is clear that the cNJ standard (Tokyo) prosodic system does not reflect the Kyoto based common language, but instead an independent eastern Japanese system whose split from the system reflected in EMJ (Kyoto) and in NJ Kyoto predates the written records noting pitch (late EMJ). This is the main instance of an old eastern Japanese feature which became part of Edo language and subsequently of the variant of the common language used in Tokyo which became the standard language.

In grammar, the use of *ka* as a general sentence final interrogative particle which is now widespread in all of Japan is quite different from the LMJ pattern of distribution of *ka* and *zo* exemplified in 12.6.1.2.2 above and is thought to be an eastern dialect feature. Other eastern features of standard cNJ include *onbin*-forms (16.1), the copula *da* (16.2), the negative auxiliary *-(a)na-* (16.3) and the imperative form of vowel base verbs (16.4). In fact, these are the main morphological isoglosses usually listed as separating western and eastern Japanese, cf. Map 16.1.

16.1 *Onbin* forms

The use of verbal and adjectival *onbin* forms is a feature that separates eastern and western Japanese, and also standard (Tokyo, eastern) NJ and Kansai (Kyoto, western) NJ. As mentioned above (15.2), the standard language does not use *-u*, the *onbin* variant of the adjectival copula infinitive, whereas Kansai (including Kyoto) and Kyushu do. With the verbal *onbin* stems the situation is more complex. In cNJ, the *onbin* stem is used in the formation of the past (*-(I)ta*), gerund (*-(I)te*), conditional (*-(I)tara*), and representative (*-(I)tari*). The forms in (2a) are mostly shared between western and eastern cNJ. Where LMJ exhibited variation between vocalic and consonantal forms (for *-w*, *-b*, *-m* base verbs), they are now divided as in (2b), which also includes *-s* bases where eastern cNJ uses a form identical with the infinitive whereas western cNJ retains use of a distinct *onbin* stem. The very rough basic pattern is that eastern and standard cNJ uses consonantal and western cNJ uses vocalic variants. However, there is today a great deal of variation in the shapes and use of verbal *onbin* stems within western Japan. Thus for *-b* and *-m* base verbs, most of Kansai, including Kyoto, have come to use consonantal, standard language, forms (*yonde*), whereas far western Honshu and Kyushu use vocalic forms (*yōde*). For *-s* base verbs, most of Kansai uses the vocalic *onbin* form (*daite*) but Kyoto uses the standard non-*onbin* form (*dashite*). For *-w* base verbs



Map 16.1 Major east–west isoglosses (from Shibatani 1990: 197)

western cNJ generally uses the vocalic (*kōte*) and standard cNJ the consonantal (*katte*) form. For an overview of the distribution of *onbin* forms in cNJ, see maps 92–105 (verbal forms) and 137–9 (adjectival forms) in the *Grammar atlas of Japanese dialects* (Kokuritsu Kokugo Kenkyūjo 1989–2006). In addition, there are slight irregularities for individual verbs; for example, in standard cNJ, the *onbin* stem of *ik-* 'go' is *iQ-* (past *itta*), not expected **iI-* (**iita*), or *tow-* 'ask' is *toV-* (*tōta*), not **toQ-* (**totta*).

(2)	Basic stem	*Onbin*-stem		Gerund	
		Western	Eastern	Western	Eastern
a.	*mot-* 'hold'	*moQ-*		*motte*	
	tor- 'take'	*toQ-*		*totte*	
	kak- 'write'	*kaI-*		*kaite*	
	kog- 'row'	*koI-*		*koide*	
	shin- 'die'	*shiN-*		*shinde*	
b.	*kaw-* 'buy'	*koV-*	*kaQ-*	*kōte*	*katte*
	yob- 'call'	*yoV-*	*yoN-*	*yōde*	*yonde*
	yom- 'read'	*yoV-*	*yoN-*	*yōde*	*yonde*
	das- 'put out'	*daI-*	*dashi-*	*daite*	*dashite*

It should be noted that there is more variability in this respect in the LMJ sources than in either of standard and Kyoto Japanese and that all of the forms used in both standard and Kyoto cNJ are represented in the LMJ sources. However, it should also be noted that Rodrigues, as mentioned above, describes the use of the standard NJ adjectival and verbal *onbin* as typical of Kanto, and that the Christian sources from the very end of LMJ use forms that correspond to Kansai NJ. Thus, while we cannot say that the NJ distribution necessarily is of a very long standing, it seems to have been well established from the end of LMJ.

16.2 Copula *da*

The paradigm of the standard cNJ plain copula *da* in Table 15.4 above directly reflects a number of LMJ forms, but also includes new forms in *da(-)* which were either absent or extremely rare in LMJ (12.2.2). These forms are like the equivalent LMJ forms thought to reflect contraction of *de ar-* such as *de aru* > *dea* > *da*, rather than the usual LMJ outcome *dya*. This, too, is thought to be an eastern dialect development, and as mentioned above (15.2), this development is similar to the EOJ pattern of phonological fusion in the morphological stative: pre-OJ *saki ar- > (western) OJ *sakyer-* :: EOJ *sakar-*, contra.

16.3 The negative auxiliary

In standard cNJ the negative auxiliary is *-(a)na-* which inflects like an adjective and has the main forms in (3). The conjectural is not used productively today, but was in use until the Meiji period.

(3)

Finite			
Nonpast	*nai*	*yomanai*	*tabenai*
Past	*nakatta*	*yomanakatta*	*tabenakatta*
(Conjectural)	*(nakarō)*	*(yomanakarō)*	*(tabenakarō)*
Non-finite			
Infinitive	*naku*	*yomanaku*	*tabenaku*
Gerund	*nakute*	*yomanakute*	*tabenakute*
Conditional	*nakattara*	*yomanakattara*	*tabenakattara*
Provisional	*nakereba*	*yomanakereba*	*tabenakereba*
Concessive	*nakutemo*	*yomanakutemo*	*tabenakutemo*

This formation has in cNJ entirely replaced the LMJ negative auxiliary (12.1.5.1), which is, however, reflected in Kyoto NJ and in many other dialects. Rodrigues mentions negative *-(a)na-* as a feature of Kanto Japanese, giving examples such as *aquenai* /agenai/ 'doesn't raise', *yomanai* 'doesn't read' and *narauanai* /narawanai/ 'doesn't learn' (*Arte*, p. 612). This shows that this negative formation was in use already at the end of LMJ and was not a form which emerged in Edo. It is sometimes speculated that the negative *-(a)na-* is related to the EOJ negative *-(a)nap-* or *-(a)nani* (see Chapter 5), but it is more likely that it reflects the adjective *na-* grammaticalized from a free (grammatical) word to an auxiliary *-(a)na-*. A possible source for that is a construction such as (4a.ii) which would be the negative counterpart of (4a.i), which has the existential verb *ar-* used in a light verb type construction in order to topicalize the lexical verb; (4a.ii) uses the adjective *na-* as a suppletive negative form of the existential verb *ar-*. The constructions in (4a) are not attested, and the LMJ common language equivalent would be *yomi wa se-*, using *se-* 'do' as the light verb. However, EMJ or LMJ eastern dialects are not attested in writing, and it is quite possible that constructions such as (4a) were in use there. In late LMJ *na-* had become the regular suppletive negated form of *ar-* and in the late LMJ common language sources there are several analytic forms and constructions with *ar-* whose negative counterparts use *na-*, e.g. (4b) analytic adjectival copula forms (12.2.1), (4c) regular copula forms (12.2.2), (4d) the respect construction (12.7.1.2), or (4e) the polite existential (12.7.2). The existence of such constructions may have reinforced the reinterpretation of *na-* as an auxiliary. The attachment of this auxiliary to the *a-* stem could in part have been motivated by the conventional attachment of negative forms to the *a-* stem

serving as an analogical model, but also here the proposed origin in *yomi wa na- fits well and would involve reduction of /i wa/ to /a/, reinterpreted as forming the *a-* stem: *yomi wa na- > *yomya na- > *yoma-na-*.

(4)	(i)	(ii)
a.	*yomi wa ar-	*yomi wa na- (>*yomya na- > *yoma-na-*)
b.	*taka-(k)u ar-*	*taka-(k)u na-*
c.	*de(wa) ar-*	*de(wa)-na-*
d.	*o-yomi-ar-*	*o-yomi-na-*
e.	*goza-ar-*	*goza-na-*

Note that the Kansai negative, in addition to the negative nonpast *-n* (*yoman* 'doesn't read') which reflects the LMJ negative auxiliary *-(a)n-*, has developed a new negative, *-(a)hen(-)* which attaches to the *a-* stem of consonant base verbs, e.g. *yoma-hen* 'doesn't read' (*yom-* 'read'). This new Kansai negative is thought to have developed in a way that is similar to that which we propose as the origin of the Kanto negative, namely through a light verb construction: thus *yomahen* derives from *yomi wa sen*, the negative form of the usual light verb construction with *se-* 'do', through (a) reduction of /i wa/ to /a/, reanalysed as forming the *a-* stem: *yomi wa* > *yomya > *yoma-*; and (b), weakening of /s/ to /h/, *sen* > *hen*, and reinterpretation of *hen* as a simple suffix. This innovative negative has in some varieties of Kansai Japanese acquired vowel harmony variants, using /hen/ with consonant and *-e* base verbs, but /hin/ with *-i* base verbs: *yoma-hen*, *tabe-hen* 'doesn't eat', but *oki-hin* 'doesn't arise'. Finally, note that both *-n* and *-hen* add *-katta* to form the past tense (*yomankatta*, *yomahenkatta* 'if someone didn't read') and *-kattara* to form the conditional (*yomankattara*, *yomahenkattara* 'if someone doesn't read'), both extracted from the paradigm of the adjectival copula.

16.4 The imperative

While both the choice of *onbin* stem and the replacement of the LMJ negative auxiliary in the standard language were influenced by eastern dialect features, the shape of the imperative is the only feature of standard cNJ which directly reflects an attested dialect feature of long standing. Thus, the shape of the imperative in vowel base verbs in standard cNJ corresponds to the EOJ vowel base imperative (cf. 5.2). The present-day Kyoto forms are usually thought to be the infinitive (usually elongated).

(5)		OJ	Kyoto cNJ	EOJ	Standard cNJ
	ake-	*ake(yo)*	*(ake:)*	*akero*	*akero*
	oki-	*oki(yo)*	*(oki:)*	*okiro*	*okiro*

17 The westernization of Japanese: Loanwords and other borrowings

Particularly since the Meiji period (1867–1912), Japanese has taken in a large number of loanwords from European languages, especially from English. However, Japan's initial contact with European languages is reflected in borrowings from as early as the sixteenth century. The ongoing impact on the Japanese lexicon is of a magnitude equal to the impact Chinese had on the language centuries earlier (see Chapter 9), not simply providing new words for new things and technology, but making available a whole intellectual and philosophical conceptual world which was unknown in Japan before the Meiji period. The intake of loanwords may be divided into three main waves: (a) pre-Meiji, (b) from Meiji to the end of WWII, and (c) from the end of the war, of which it is the latter two phases which have exerted profound influence on the language.

17.1 Vocabulary layers and hybrid words

Often the existence of three or four different 'vocabulary layers' is posited in descriptions of Japanese: native Japanese words (*wago* 和語, or *yamato-kotoba*), Sino-Japanese words (*kango* 漢語), (recent) loanwords (*gairaigo* 外来語), and sometime also mimetics (*giseigo* 擬声語, *gitaigo* 擬態語). The first three, the main lexical layers, overtly refer to historical origin, but this classification is in addition thought by many scholars to represent some synchronic linguistic reality for speakers of the language, and other defining criteria than etymology are invoked, for example phonological or socio-linguistic ones. However, an etymological classification of words or morphemes is highly relevant for the history of the language, but there is no one-to-one relationship between the etymologically defined layers and layers defined by other criteria. There are certainly phonological features which are exclusive to some recent loanwords (14.6), but they are not shared by all recent loanwords. And there are socio-linguistic values (17.1.1) which are characteristic of recent and not well-integrated loanwords, but this is a short-lived quality which changes if and when the words become better integrated; more generally this applies to much new vocabulary, regardless of its origin. On the

whole it seems that the notion of vocabulary layers, other than those defined strictly etymologically, is not useful for a book such as this one.

So-called 'hybrid' words, i.e. words composed of elements from more than one etymological layer, abound. We already saw examples of combinations of borrowed SJ vocabulary with preexisting material in 9.2.3.2, and a simple further example is the name of Yanase Takashi's popular cartoon character *anpanman*, lit. 'bean-jam bun man', which was coined by attaching the male superhero suffix *-man* borrowed from English to the compound *anpan* 'bean-jam bun', which was coined in the Meiji period from SJ *an* 'bean-jam' (餡) and *pan* 'bread' (from Portuguese). *Anpanman* is thus hybrid in combining three different layers of borrowed material.

17.1.1 Gairaigo

The Japanese word usually translated into English as 'loanword' is *gairaigo* (外来語), but *gairaigo* is different from *loanword* in several respects. First of all, *gairaigo* is used about recent, transparent borrowings, primarily from European languages, but does not refer to SJ vocabulary, which also is borrowed, or to old naturalized loanwords, such as those mentioned in Chapter 4. The word *gairaigo* was coined during the period of modernization of the Japanese lexicon which took place in the Meiji period (17.3) and was used first by the linguist Ueda Kazutoshi in 1895 (after returning from his studies in Germany in 1890), and it seems clear that *gairaigo* corresponds to and is intended to render the German *fremdwort* which literally means 'alien, foreign word'. This word, and Japanese *gairaigo*, captures better than Japanese *shakuyōgo* (借用語), German *lehnwort*, or English *loanword* the various and quite diverse socio-linguistic nuances associated with the use of loanwords from European languages (such as 'learned, trendy, exotic' and others), at least in the initial phase of their use. Words considered *gairaigo* are usually written in *katakana* and they are sometimes referred to as *katakana-go* '*katakana* words'.

17.2 Pre-Meiji; from the end of Late Middle Japanese to the middle of the nineteenth century

During the so-called 'Christian century', from the arrival of the missionaries in the 1540s to the ban on Christianity and expulsion of the missionaries in the 1630s, loanwords were mainly taken in from Portuguese. Later on, during the period from the 1640s to the opening of Japan in the 1850s, when contact with Europeans was almost limited to the Dutch or took place through the Dutch settlement on Dejima off Nagasaki, loanwords were taken in from Dutch. As we saw above, the Jesuits freely used Portuguese (and Latin) words

in their publications in Japanese, either spelling them as in the source language in the alphabet texts, or adapting them into *kana* in the texts in Japanese script (10.2.2), but many of these words never became part of the Japanese language. The following is a short list of loanwords, or foreign words, used in Japanese, listed by Thunberg (1792: 268–9).[1] Certainly, the list comprises words which must have been and remained foreign words, used for convenience in conversation with the few interpreters and others who had contact with the Dutch in Nagasaki, but it is also striking that a fair number of words in the list have survived into and are still used in cNJ (shown in boldface). The list gives an interesting glimpse of words which a visitor at that time such as Thunberg was exposed to. For some of these words, Thunberg cites a source language, here noted as 'h(olland)' for Dutch' and 'p(ortugal)' for Portuguese, and it is noteworthy that of the words surviving into cNJ, more than half are from Portuguese. Several of the words listed involve both borrowed and Japanese elements (the latter here written in plain type).

(1) ***tabaco*** (p) 'tobacco' (cNJ *tabako*, mainly used about cigarettes), ***bir*** (h) 'beer' (cNJ *biiru*), *Portugal* abra 'olive oil' (*abura* 'oil'), *unicorn* 'unicorn's horn', *boter* (h) 'butter', *boter* no *kas* 'cheese' (*no* genitive), ***karta*** (p) 'playing cards' (cNJ *karuta*), *kapitein* (h) 'captain', *lancetta* 'lancet', *krokodil* 'crocodile', ***pang*** (p) 'bread' (cNJ *pan*), *pokk* (p) 'venereal disease', ***savon*** (p) 'soap' (cNJ *shabon*, now not used much, but generally replaced by SJ *sekken* 石鹸), fige *savon* 'shaving soap' (*fige* > cNJ *hige* 'beard'), *flasco* (p) 'flask', *banco* (p) 'bench', *diamant* 'diamond', *faka* 'knife', *saffran* 'saphron', *biduro* (p) 'glass', *baso* 'dish', ***kopp*** (h) 'cup' (cNJ *koppu*), *ducaton* 'Dutch coin', *theriak* 'theriac', *gans* (h) 'goose', ***vein*** (h) 'wine' (cNJ *wain*), kakami *biduro* 'looking glass' (*kagami* 'mirror'), ***rassia***, **orassia** 'cloth' (*o-* Honorific prefix; cNJ *rasha*; Thunberg does not note it, but this word is from Portuguese *raxa*), *aderlaten* (h) 'blood-letting', *skerbekken* (h) 'basin for shaving', ***fork*** (h) 'fork' (cNJ *hooku, fooku*), *kananor* isi 'hematite' (Cannanore (Kannur) in India + *isi* 'stone'), ***tabaco* ire** 'tobacco pouch' (*ire* 'receptacle'), *tinta* 'Spanish wine'.

Other Portuguese loanwords which have survived into cNJ include *bateren* 'priest, padre' (< *padre*), *kasutera* 'sponge cake' (< (*paõ de*) *Castella* 'bread

[1] The list is introduced by: 'Sunt præterea res, in Japoniam ab Exteris allatæ, quæ nominibus vel Portugallicis vel Hollandicis insigniri solent, ut:' ('there are various things, brought to Japan from outside, which are usually designated by Portuguese or Dutch names, such as:').

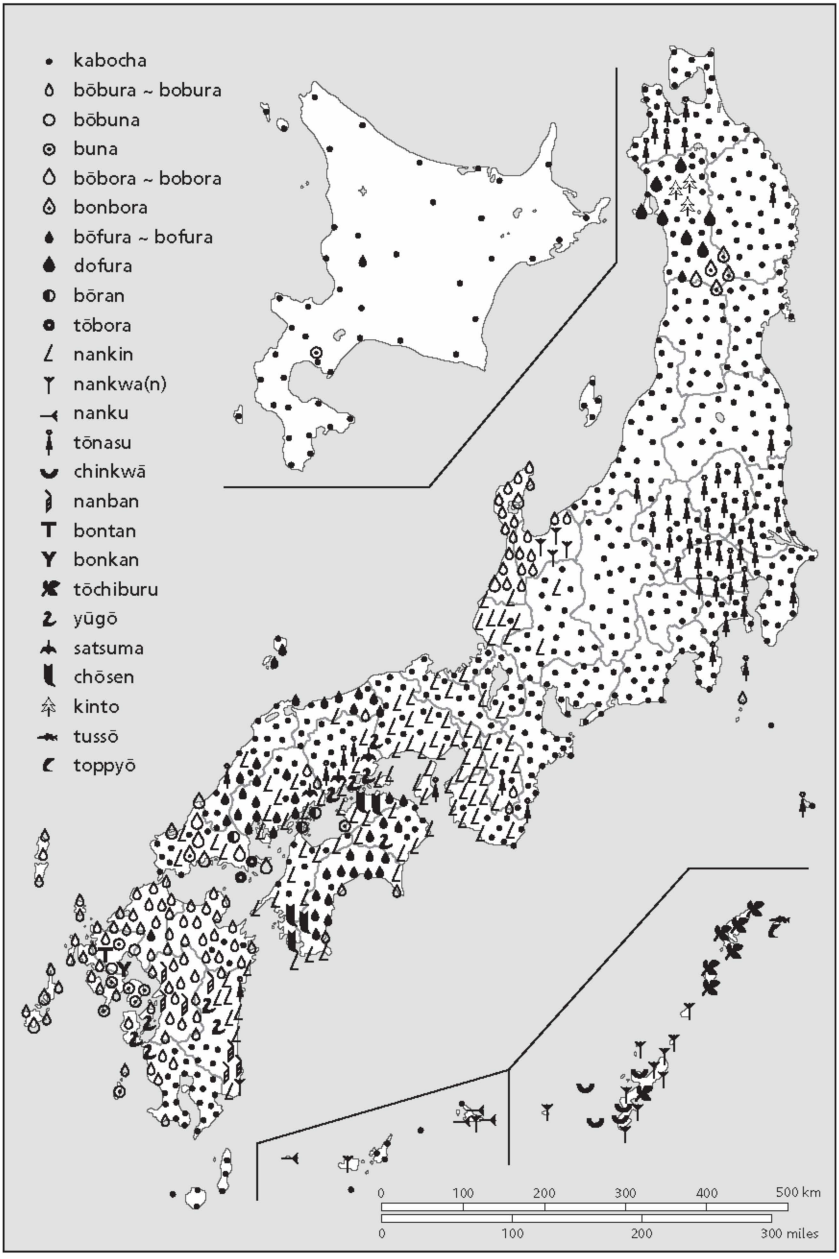
from Castilla'), and *tenpura* 'tempura; deep fried fish and vegetables in batter' (< *tempero* 'cooking'). A final example is the Japanese words for 'pumpkin', which were brought to Japan by the Portuguese in the late sixteenth century. The distribution of words throughout Japan is shown in Map 17.1. The most common word, used also in standard Japanese, is *kabocha* which is used in most mainland dialects, but another widely used word is *boobora* (which has a number of variants: *bobora*, *bo(o)bura*, *boobuna*, etc.). These two word(group)s reflect adaptations of the two parts of what the Portuguese called the pumpkin when they brought it to Japan: *Cambodia abóbora* 'Cambodian pumpkin' (Port. *abóbora* 'calabash, gourd, pumpkin').[2] The variant shapes and distribution of the words reflecting *abóbora* also illustrate the phonological adaptations and changes which may take place in the course of borrowing into Japanese, as well as the fact that Japanese dialects may treat loanwords differently.

From the middle of the seventeenth to the middle of the nineteenth century the Dutch were the only Europeans allowed in Japan – and Thunberg for example had to learn Dutch in order to pass himself off as a Dutchman to be allowed to stay in Japan, and to interact with the local translators. Contact with technology and ideas from Europe, in which through that time there was a small but steadily growing interest, took place exclusively through Dutch which was studied and learnt by many intellectuals, and the study of European medicine, technology, science, etc. came to be known as 'Dutch learning' (*ran-gaku* 蘭学, 'Dutch' (< *oranda* 'Holland') + 'learning'). In the course of this, a large number of Dutch words came to be used by those engaged in 'Dutch learning'. Most such words remained foreign and never gained general currency, but some entered the language and a few are still in use today. Dutch loanwords still in use include *mesu* 'surgical scalpel' (< *mes* 'knife') and *handon* 'half day off', < *han-* 'half' + *don(taku)* (< *zontag* 'Sunday').

17.3 From the beginning of Meiji to the end of WWII (1945)

Although the Dutch officially until the middle of the nineteenth century were the only Europeans allowed in Japan, there was some contact with other Europeans and their languages from the beginning of the nineteenth century, but it was not until the enforced opening of Japan in the 1850s that loanwords from other European languages than Portuguese and Dutch were taken in to any noteworthy extent. From the middle of the nineteenth century Japan was inundated with technology, science, ideas and material culture from Europe

[2] There are also words for 'pumpkin' which have different origin, such as place names presumably reflecting putative origins (*nankin* after Nanjing in southern China, or *chōsen* 'Korea') or descriptions (*tō-nasu* 'Chinese eggplant').



Map 17.1 Distribution of words for 'pumpkin' (from Satō 2002: 169)

Table 17.1 *Loanwords in use during the Taishō period (from Shibatani 1990: 149, table 7.5)*

Origin	Number of words	%
English	84	51.9
Dutch	45	27.8
Portuguese	23	14.2
French	6	3.7
German	2	1.2
Spanish	1	0.6
Others	1	0.6

and the US, which was eagerly adopted in the course of the effort to modernize and catch up with the 'western' countries. This resulted in a great amount of borrowed vocabulary, which was taken in either as direct loans (17.3.1), or as loan translations into SJ (17.3.2).

17.3.1 Loanwords from European languages

English dominated as a donor language of loanwords from the very beginning of the period of modernization, but it did not attain the virtual monopoly it has today until after the end of WWII. It is often observed that loanwords from the minor donor languages exhibit some degree of semantic specialization, correlating with areas of material and cultural contribution: e.g. German (*karute* 'medical patient record' < *karte*, *(aru)baito* 'part-time student job' < *arbeit)*, French (*bifuteki* 'steak' < *bifteck*, *konsome* < *consommé*), Italian (*opera*, *sonata*). Table 17.1 shows the origins of 162 loanwords from different European languages which were in common use around the middle of the Taishō period (1912–26). It is noteworthy that words borrowed from Dutch and Portuguese still at this time were quite prominent.

17.3.2 Loan translations; Sino-Japanese coinages

In addition to direct loans, a strategy was adopted of coining SJ words for the new notions, institutions and things to be named, that is, loan translation into (Sino-)Japanese. This coinage often consisted of finding some word, or *kanji* combination, in one of the Chinese Classics, which could be drafted in to write the new word, which was vocalized by using the SJ reading, or alternatively of reviving or adapting SJ words from earlier or specialized usage, but freer coinage from Sino-Japanese was also frequently used.

Some of the new SJ words originate in *kanji* writings intended as a logographic representation of a direct loanword, accompanied by reading glosses in *katakana*, subsequently reinvented as a SJ word. An example which illustrates some of the processes involved is cNJ *shokudō* 食堂 which is now the general word for most types of '(large, public) dining room, dining hall'. In the 1870s English *dining room* was rendered or borrowed as *dainingurūmu*, written as 会食堂 (ダイニングルーム), i.e. logographically in *kanji* with the pronunciation shown in *katakana*. Earlier 食堂 *zikidoo* < *zikidaũ* (which is *go-on*), was a specialized word for dining rooms in Buddhist temples, but following on from the writing of *dainingurūmu* as shown above, 食堂 was drafted in as a SJ translation equivalent and was, accompanying its new use, given a different pronunciation (reading) using *kan-on* for 食, but keeping *go-on* for 堂 (whose *kan-on* is *tō*). Incidentally, *dainingurūmu* is still occasionally written 食堂 (ダイニングルーム).

SJ words deliberately coined, revived or adapted in this way during the Meiji period make up the great majority of SJ words in use today – and the overwhelming majority of academic, political and intellectual vocabulary used in Japanese today – and some of these words were later adopted in China and Korea, with Chinese or Sino-Korean vocalization. Many of these new words are attributable to individuals who also otherwise played prominent roles in the process of modernization in the nineteenth and early twentieth century, such as Nishi Amane, Fukuzawa Yukichi and Inoue Tetsujirō, whose translation of William Fleming's *The vocabulary of philosophy, mental, moral and metaphysical* (1857), *Tetsugaku jii* (哲学字彙, 1881) provides more than 2,500 SJ words relevant to philosophy. Of the many new SJ words coined in this way in the late nineteenth century, a few examples are given in (2):

(2) *shakai* 社会 'society', *minshu(shugi)* 民主(主義) 'democracy', *jiyū* 自由 'liberty', *shinri(gaku)* 心理(学) 'psyche (psychology)', *tetsugaku* 哲学 'philosophy', *kaisha* 会社 'firm, company', *kaikeigaku* 会計学 'accounting', *rōdō* 労働 'work' (which includes 働 which was made in Japan, see 9.1.5), *yūbin* 郵便 'post', *jidōsha* 自動車 'automobile', *tetsudō* 鉄道 'railway', *denwa* 電話 'telephone', *enzetsu* 演説 'speech', *tōron* 討論 'discussion', *bungaku* 文学 'literature', *shōsetsu* 小説 'novel', *kokka* 国家 'nation', *kokumin* 国民 'people'.

As the short list suggests, new words for abstract concepts and institutions were usually coined in this way, rather than by direct loans. It is highly unlikely that the lexical, terminological modernization of the Meiji period would have been as successful as it was, if it had had to rely on direct loans, rather than

SJ coinages, which are short and in some cases provide educated readers with semantic clues. In addition, SJ coinages have the important function of giving the impression that these words were part of and belonged in an intellectual tradition.

17.3.3 Influence from European languages in grammar and usage

European languages, and especially translation practices from European languages not unlike the *kanbun-kundoku* practices described in 9.1 – i.e. 'reading' a Dutch or English text 'in Japanese' – have exerted some influence on Japanese, though not to the extent Chinese did earlier. For example, the use of the 3rd person pronouns *kare* 'he' and *kanojo* 'she' (< *ka-no-jo* 'DISTAL-GEN-woman'), which became widely used from the late nineteenth century in literary writing and eventually passed over into general language, originates in translation-inspired imitations of European literature; the imitation went so far as to use *kanojo* to refer to ships. In general, under this influence from Dutch and English, pronouns came to be used far more, especially in writing from the early twentieth century onwards, than they had been before the Meiji period.

Another example which is often mentioned is changes in the use of the passive. Because of the relatively free word order, a semantic patient can in Japanese be topicalized or fronted without the need for passivization, and the passive is traditionally said not to have been used much solely for this purpose, but usually to express some kind of affectedness on a human, or animate, patient, by a sentient or personified agent. However, since the Meiji period, more direct uses of the passive have become current, with both inanimate agents and patients, and this is usually ascribed to influence from or imitation of translation from Dutch and English languages. Kinsui (1997) has shown that marking of the agent in direct passives by *ni-yotte* originates in translation practices from Dutch, in which *door* 'through', which can be used both about path, means and way, and about agents in passive constructions, was conventionally translated by *ni-yorite* or *ni-yotte*, which originally meant 'because', also when used about passive agents. This usage was eventually generalized and made its way into general language. It is interesting that the title of Yamada Yoshio's ground-breaking study of the influence of *kanbun-kundoku* on Japanese (*Kanbun no kundoku ni yorite tsutaeraretaru gohō* from 1935), uses the direct passive with an inanimate agent and *ni-yorite* to mark the agent, both of which originate in similar influences from similar practices centuries after *kanbun-kundoku* made its mark on Japanese.

Finally, the obligatory marking of all core arguments (subjects and objects) by case particles in written Japanese today is a written language feature which was introduced in the establishment of the new normative standard written

Table 17.2 *Loanwords used in magazines (from Shibatani 1990: 148, table 7.4)*

Origin	Number of words	%
English	2,395	80.8
French	166	5.6
German	99	3.3
Italian	44	1.5
Dutch	40	1.3
Russian	25	0.8
Chinese	22	0.7
Portuguese	21	0.7
Spanish	21	0.7
Others	131	5.3

language. It was not a feature of written Japanese before the *genbun'itchi* reforms, nor was it ever, or is today, a feature of spoken Japanese, where omission of case particles (case drop) has always been frequent.[3] Its introduction into standard written Japanese is usually ascribed to a desire to have a normative, regular grammar for written language, as the European languages did, but it should also be noted that *kuntengo* (9.1.6), which is one formal genre of written Japanese, generally did not have case drop.

17.4 Post-WWII

The period after the end of WWII has seen an intake of loanwords from English on a much larger scale than in the period before the war and English now has a near monopoly as a donor of loanwords to Japanese, as shown in Table 17.2 which shows a clear growth in the proportion of loanwords from English in the post-war period compared to before the war, cf. Table 17.1. This has accompanied improvements in English education and widespread command of English (i.e., more (limited) bilingualism) and far greater contact with and access to English, in particular American English.

It is today difficult to distinguish between integrated loanwords from English and occasional use of phonologically more or less adapted English words (which may later be more widely adopted to become well-integrated loanwords). Conventions exist for spelling English words in *katakana*, and it is therefore easy to insert English words in written Japanese for stylistic effect.

[3] That is the reason a trick question like *pantsukutta koto aru?* (courtesy of Miss Saitō Sachiko) works. It can either be parsed as (i) *pan (o) tsukutta koto aru?* 'Have you ever made bread?' or as (ii) *pantsu (o) kutta koto aru?* 'Have you ever eaten pants?'

Needless to say, different speakers have different repertoires of integrated loanwords. In this respect the socio-linguistic situation in Japan is much the same as in most other developed countries, with quite different levels of command of English between different segments of the population and widespread use of English words by some speakers, especially educated and young urban people. Buzzwords often derive from English, for example *sekuhara* 'sexual harassment', abbreviated from *sekushuaru harasumento*, the phonologically adapted rendition of the English word; and inventive metaphorical usages can be surprising and amusing, as in *bākodo* 'comb-over' which is from *bar code*. English words and morphemes are in Japan often used creatively in combinations that are not usual in first language English speech communities, giving rise to 'loanwords' which are not used in English, for example *ōeru* 'female office worker' < *OL* (short for *office* + *lady*), *batontatchi* 'baton pass' or as a verb *batontatchi-se-* 'pass the baton to one's successor' < *baton* + *touch*, *sukinshippu* 'physical contact (especially between baby and mother)' < *skin* + *-ship*, or *makaroni-wesutan* 'spaghetti western'. *Camcorder* (Japanese *kamukōdā*) is an example of an English word coined in Japan and now used worldwide. These examples are all from the post-war period, but many examples are found also before the war, for example *ōrudomisu* 'old maid' < *old* + *miss*, or *afureko* 'dubbing' < *after* + *recording*.

REFERENCES

Westernization, general: Miller 1967: 235ff., Shibatani 1990: 142ff. Loanwords: *Kadokawa gairaigo jiten* (Arawaka 1977). Grammatical borrowing: Kinsui 1997.

APPENDIX Summary of the main regular phonemic changes between OJ and conservative NJ

This table exemplifies the main regular phonemic changes which have taken place between OJ and conservative NJ (see Chapter 14). It does not include the *onbin* sound changes (see 7.1.4), which were not regular sound changes, nor phonetic changes which have not resulted in phonemic change. The first row includes cross-references to the main text. In the first column is given a tag gloss, not reflecting changes in word meaning which may have occurred over time. Most words are given in their OJ shape in the second column, but SJ vocabulary or *kanji* readings are entered at around 950 in the shape they would have had then, although not all are attested from that time.

Gloss		Loss of *kō–otsu* distinctions (7.3.2.1)		Merger of /.ye/ and /.e/ (7.3.2.2)	Merger of medial /p/ and /w/ (7.3.1)	Loss of syllable initial /w/ before /o, i, e/ (7.3.2.3; 11.2)			Merger of /Ĩ, Ũ/ and /I, U/ (11.1.2)
	OJ	< 800 Cye > Ce Cwi > Ci	< 950 Cwo > Co	*c.* 950 .ye > .e	950–1000 /-p-/ > /-w-/ _ /i, e, a, o/ /-p-/ > Ø / _/u/	*c.* 1000 .w > Ø _/o/	*c.* 1100 medial position -.w > Ø _/i, e/	*c.* 1300 initial position #.w > Ø _/i, e/	early LMJ /Ĩ/ > /I/ /Ũ/ > /U/
‘front’	mapye	mape			mawe		mae		
‘today’	kyepu	kepu			keu				
‘love’	kwopwi	kwopi	kopi		kowi		koi		
‘shell’	kapi				kawi		kai		
‘indigo’	awi						ai		
‘voice’	kowe						koe		
‘river’	kapa				kawa				
‘branch’	yeda			eda					
‘thrush’	nuye			nue					
‘hill’	woka					oka			
‘blue’	awo					ao			
‘face’	kapo				kawo	kao			
‘well’	wi							i	
‘smile’	wemi							emi	
‘field’	para								
‘plain’	pira								
‘boat’	pune								
‘morning sun’	asapi								
‘the road to Ki’	kidi								
‘classic’ (経)				keĩ					kei
‘sutra’ (経)				kyaũ					kyau
‘attachment’ (執)				sipu	siu				
‘need’ (要)				eu					
‘public’ (公)				koũ					kou
‘fruit; cookie’ (菓子)				kwasi					
‘first year in a dynastic period’ (元年)				gwannen					

Fricativization of /p/ (11.3)	Monophthongization of /VU/ (11.5)	Merger of /d/ and /z/ before /i, u/ (14.1)	Merger of /ɔo/ and /oo/ (14.2)	Delabialization of /f/ (14.3)	Loss of /w/ after /k, g/ (14.4)	
early? LMJ /p/ > /f/	mid? LMJ /iU/ > /yuu/ /eU/ > /yoo/ /oU/ > /uu/ /aU/ > /ɔɔ/	17th century d > z / _{i, u}	17th century /ɔɔ/ > /oo/	?1700 /f/ > /h/	late 19th century /kwa, gwa/ > /ka, ga/	Conservative NJ
						mae
	kyoo					kyō
						koi
						kai
						ai
						koe
						kawa
						eda
						nue
						oka
						ao
						kao
						i
						emi
fara				hara		hara
fira				hira		hira
fune				hune		fune
asafi				asahi		asahi
		kizi				kiji
						kei
	kyɔɔ		kyoo			kyō
	syuu					syū
	yoo					yō
	koo					kō
					kasi	kashi
					gannen	gannen

References

Arakawa, Sōbē (ed.). 1977. *Kadokawa garaigo jiten*, 2nd edn. Tokyo: Kadokawa Shoten.

Bentley, John R. 2000. A new look at Paekche and Korean data from Nihon shoki. *Language Research* 36/2, 417–43.

2001a. *A descriptive grammar of Early Old Japanese prose*. Leiden: Brill.

2001b. The origin of man'yōgana. *Bulletin of the School of Oriental and African Studies* 64/1, 59–73.

Collado, Diego. 1632a. *Ars grammaticae iaponicae linguae*. Rome.

1632b. *Dictionarium sive thesauri linguæ Iaponicæ compendium*. Rome.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

de Boer, Elisabeth M. forthcoming. *The historical development of Japanese tone*; volume I: *The tone systems of the modern dialects*; volume II: *The introduction and adaptation of the Middle Chinese tones in Japan* (Veröffentlichungen des Ostasien-Instituts der Ruhr-Universität Bochum). Wiēsbaden: Harrassowitz.

DeFrancis, John. 1984. *The Chinese language: Fact and fantasy*. Honolulu: University of Hawai'i Press.

Doi, Tadao (trans.). 1955. *Nihon daibunten* [A Japanese translation of Rodrigues 1604–8]. Tokyo: Sanseidō.

Doi, Tadao, Morita Takeshi and Chōnan Minoru (eds. and trans.). 1980. *(Hōyaku) Nippo jisho* [A Japanese translation with introduction and postscript of Rodrigues 1603–4]. Tokyo: Iwanami Shoten.

Durst-Andersen, Per. 1996. Towards a multifunctional grammar: 'Language, reality and mind' in a grammatical description. *Hermes, Journal of Linguistics* 17, 75–102.

Elliott, Ralph W. V. 1959. *Runes: An introduction*. Manchester: Manchester University Press.

Endō, Orie. 2006. *A cultural history of Japanese women's language*. Ann Arbor: Center for Japanese Studies, the University of Michigan.

Engberg, Jonas. 1990. *Observationes in linguam japonicam: Observations on the Japanese language* [Introduction to and translation of Thunberg 1792]. Stockholm: Center for Pacific Asia Studies, University of Stockholm.

Farris, William Wayne. 1998. *Sacred texts and buried treasures*. Honolulu: University of Hawai'i Press.

Fleming, William. 1857. *The vocabulary of philosophy, mental, moral and metaphysical*. Glasgow.

Frellesvig, Bjarke. 1995. *A case study in diachronic phonology: the Japanese onbin sound changes*. Aarhus: Aarhus University Press.

1998. Accent, word tone and metatony in Central Japanese. *On'in Kenkyū/Phonological Studies* (published by the Phonological Society of Japan) 1, 195–202.

2001. A common Korean and Japanese copula. *Journal of East Asian Linguistics* 10, 1–35.

2007. On the verb morphology of Old Japanese. In *Current issues in the history and Structure of Japanese*, ed. Bjarke Frellesvig, Masayoshi Shibatani and John Charles Smith, 219–252. Tokyo: Kurosio Publishers.

2008. On reconstruction of proto-Japanese and pre-Old Japanese verb inflection. In *Proto-Japanese: Issues and prospects*, ed. Bjarke Frellesvig and John Whitman, 175–92. Amsterdam: John Benjamins.

Forthcoming. Chronological layering in proto-Japanese and pre-Old Japanese verb inflection. In *Studies in Japanese and Korean Linguistics*, ed. Bjarke Frellesvig, Jieun Kiaer and Janich Wrona. London: Lincom.

Frellesvig, Bjarke, Jieun Kiaer and Janick Wrona (eds.). Forthcoming. *Studies in Japanese and Korean linguistics*. London: Lincom.

Frellesvig, Bjarke, Masayoshi Shibatani and John Charles Smith (eds). 2007. *Current issues in the history and structure of Japanese*. Tokyo: Kurosio Publishers.

Frellesvig, Bjarke and John Whitman (eds.). 2008a. *Proto-Japanese: Issues and prospects*. Amsterdam: John Benjamins.

2008b. Evidence for seven vowels in proto-Japanese. In *Proto-Japanese: Issues and prospects*, ed. Bjarke Frellesvig and John Whitman, 15–41. Amsterdam: John Benjamins.

Forthcoming. Prenominal complementizers and the derivation of complex NPs in Japanese and Korean. *Japanese/Korean Linguistics* 18.

Fukuda, Yoshisuke. 1965. *Nara jidai Azuma hōgen no kenkyū*. Tokyo: Kazama Shoin.

Hamada, Atsushi. 1970. *Chōsen-shiryō ni yoru Nihongo kenkyū*. Tokyo: Iwanami Shoten.

Harbsmeier, Christoph. 1981. *Aspects of classical Chinese syntax*. London: Curzon Press.

Hashimoto, Shirō. 1966. Kodaigo no shiji-taikei: Jōdai o chūsin ni. *Kokugo Kokubun* 3516, 329–41.

1982. Shijigo no shiteki tenkai. In *Kōza nihongogaku*, vol. II: *Bunpōshi*, ed. Kenji Morioka *et al.*, 217–40. Tokyo: Meiji Shoin.

Hattori, Shirō. 1978–9. Nihonsogo ni tsuite 1–22. *Gekkan Gengo* 7/1–7/3, 7/6–8/12.

1979. Ōno Susumu 'Dōshi katsuyōkei no kigen' ni tsuite. In *Nihon no gengogaku*, vol. IV (*Bunpō* II), ed. Hattori Shirō, Kawamoto Shigeo and Shibata Takeshi, 658–63. Tokyo: Taishūkan Shoten.

Hayata, Teruhiro. 1998. Jōdai nihongo no onsetsu kōzō to o-retsu kō-otsu no betsu (The syllable and the distinction between o_1 and o_2 in Old Japanese). *Onsei Kenkyū/Journal of the Phonetic Society of Japan* 2, 25–33.

2000. The liquid and stem-final vowel alternations of verbs in ancient Japanese. *Gengo Kenkyū* 118, 5–27.

Hida, Yoshifumi, Endō Yoshihide, Katō Masanobu, Satō Takeyoshi, Hachiya Kiyoto, Maeda Tomiyoshi (eds.). 2007. *Nihongogaku kenkyūjiten (The research encyclopedia of Japanese linguistics)*. Tokyo: Meiji Shoin.

Horrocks, Geoffrey. 1997. *Greek: a history of the language and its speakers*. London: Longman.
Imaizumi, Tadayoshi. 1971. *Nippo jisho no kenkyū*. Tokyo: Ōfūsha.
Inoue, Tetsujirō. 1881. *Tetsugaku jii*. Tokyo: Tōkyō Daigaku Sangakubu.
Iwai, Yoshio. 1970. *Nihon gohō-shi*, vol. I: *Nara, Heian-jidai-hen*. Tokyo: Kasama Shoin.
1971. *Nihon gohō-shi*, vol. II: *Kamakura-jidai-hen*. Tokyo: Kasama Shoin.
1973. *Nihon gohō-shi*, vol. III: *Muromachi-jidai-hen*. Tokyo: Kasama Shoin.
1974. *Nihon gohō-shi*, vol. IV: *Edo-jidai-hen*. Tokyo: Kasama Shoin.
Iwanami (ed.). 1957–69. *Nihon koten bungaku taikei* (100 vols. + 2 index vols.). Tokyo: Iwanami Shoten.
Kageyama, Taro. 1999. Word formation. In *The handbook of Japanese linguistics*, ed. Tsujimura Natsuko, 297–325. Oxford: Blackwell Publishers.
Kamei, Takashi. 1954. On the authenticity of *on*-readings in Sino-Japanese. *The Annals of the Hitotsubashi Academy* 5, 97–105.
Kanno, Hiroomi. 1978. Chōsengo kara no shakuyōgo. *Gekkan Gengo* 7/2, 36–42.
Kariya, Bōshi (compiler), Yamada Yoshio, and Katori Hozuma (eds.). 1912. *Kokyō ibun*. Tokyo: Hōbunkan.
Kasuga, Masaji. 1985. *Saidaiji-bon Konkōmyō saishōō kyō no koten no kokugogakuteki kenkyū*. Tokyo: Benseisha.
Kawabata, Yoshiaki. 1976. *Yōgen*. In *Iwanami kōza nihongo*, ed. Ohno Susumu and Shibata Takeshi, vol. VI (*Bunpō* I), 169–217. Tokyo: Iwanami Shoten.
Kindaichi, Haruhiko, Hayashi Ōki, and Shibata Takeshi (eds.). 1988. *Nihongo hyakka daijiten (An encyclopedia of the Japanese language)*. Tokyo: Taishūkan Shoten.
Kindaichi, Haruhiko and Wada Minoru. 1980. Kokugo akusento ruibetsu goi-hyō. In Kokugogakkai 1980, 7–10.
King, J. Ross and John Whitman (eds.). Forthcoming. Special issue of *Sino-Platonic Papers*.
Kinsui, Satoshi. 1995. Nihongoshi kara mita joshi. *Gengo Kenkyū* 24/11, 78–84.
1997. The influence of translation on the historical development of the Japanese passive construction. *Journal of Pragmatics* 28/6, 759–79.
2006. *Nihongo sonzai hyōgen no rekishi*. Hitsuji Shobo.
2007. The interaction of argument and non-argument in the diachronic syntax of Japanese. In *Current issues in the history and structure of Japanese*, ed. Bjarke Frellesvig, Masayoshi Shibatani and John Charles Smith, 253–61. Tokyo: Kurosio Publishers.
Kitagawa, Kazuhide. 1982. *Shoku nihongi senmyō: Kōhon, sōsakuin*. Tokyo: Yoshikawa Kōbunkan.
Kiyose, Gisaburō. 1985. Heian-chō ha-gyō shiin P-onron (Evidence that modern Japanese h- was p- in the Heian period). *Onsei no kenkyū* 21, 73–87.
Kobayashi, Akemi. 1981. Ennin no kijutsu suru sansukuritto onsetsu 'ca' no onka: kyū-seiki nihongo suitei no kokoromi. *Ōsaka Gaikokugo Daigaku Gakuhō* 52, 63–80.
Kokugogakkai (eds.). 1955. *Kokugogaku jiten*. Tokyo: Tōkyōdō Shuppan.
1980. *Kokugogaku daijiten*. Tokyo: Tōkyōdō Shuppan.
Kokuritsu Kokugo Kenkyūjo (ed.). 1989–2006. *Hōgen bunpō zenkoku chizu* (Grammar atlas of Japanese dialects), 6 vols. Tokyo: Ōkurashō Insatsukyoku.

Komatsu, Hideo. 1956. Nihon jion ni okeru shinnai-nisshō-inbi no sokuonka to zetsunainisshō e no gōryū-katei. *Kokugogaku* 25.
1981. *Nihongo no on'in* (Nihongo no sekai 7). Tokyo: Chūō Kōronsha.
Kōno, Rokurō. 1969. The Chinese writing and its influences on the scripts of the neighbouring peoples: With special reference to Korea and Japan. *Memoirs of the Research Department of the Toyo Bunko* 27, 83–140.
Kotani, Hiroyasu. 1986. *Mokkan to senmyō no kokugogakuteki kenkyū*. Osaka: Izumi Shoin.
Kubozono, Haruo (ed.). 2008. *Asymmetries in phonology: An East-Asian perspective*. Tokyo: Kurosio Publishers.
Kuroda, S.-Y. 2007. On the syntax of Old Japanese. In *Current issues in the history and structure of Japanese*, ed. Bjarke Frellesvig, Masayoshi Shibatani and John Charles Smith, 263–317. Tokyo: Kurosio Publishers.
Lange, Roland A. 1973. *The phonology of eighth-century Japanese*. Tokyo: Sophia University.
Levy, Ian Hideo. 1981. *Man'yōshū: A translation of Japan's premier anthology of classical poetry*. Tokyo: Tokyo University Press.
Lewin, Bruno. 1976. *Der koreanische Anteil am Werden Japans*. Opladen: Westdeutscher Verlag.
Mabuchi, Kazuo. 1971. *Kokugo on'inron*. Tokyo: Kasama Shoin.
1993: *Gojūonzu no hanashi*. Tokyo: Taishūkan.
Martin, Samuel E. 1968. On the forms of Japanese adjectives. *Glossa* 2, 46–69.
1975. *A reference grammar of Japanese*. New Haven: Yale University Press.
1987. *The Japanese language through time*. New Haven: Yale University Press.
1990. Morphological clues to the relationships between Japanese and Korean. In *Linguistic change and reconstruction methodology*, ed. Phili Baldi, 483–509. Berlin and New York: Mouton de Gruyter.
1991. Recent research on the relationships of Japanese and Korean. In *Sprung from some common source: Investigations into the prehistory of languages*, ed. Sydney M. Lamb and E. Douglas Mitchell, 269–92. Stanford: Stanford University Press.
Matsumori, Akiko. 2008. On the reconstruction of the proto-accentual system of Japanese. In *Proto-Japanese: Issues and prospects*, ed. Bjarke Frellesvig and John Whitman, 103–24. Amsterdam: John Benjamins.
Matsumoto, Katsumi. 1984. Gengoshi no saiken to gengofuhen. Gengo Kenkyū 86, 5–32.
Matsumura, Akira (ed.). 1971. *Nihon bunpō daijiten*. Tokyo: Meiji Shoin.
1998. *(Zōho) Edogo Tōkyōgo no kenkyū*. Tokyo: Tōkyōdō Shuppan.
Miller, Roy A. 1967. *The Japanese language*. Chicago: University of Chicago Press.
1971. *Japanese and the other Altaic languages*. Chicago: University of Chicago Press.
1989. Old Japanese *i*. In *Bruno Lewin zu Ehren: Festschrift aus Anlass seines 65. Geburtstages*, ed. Irmela Hijiya-Kirschnereit and Jürgen Stalph, vol. I, 251–91. Bochum: Brockmeyer.
Miyake, Marc. 1997. Pre-Sino-Korean and pre-Sino-Japanese: Reexamining an old problem from a modern perspective. *Japanese/Korean Linguistics* 6, 179–211.
2003a. *Old Japanese: a phonetic reconstruction*. London: RoutledgeCurzon.
2003b. Philological evidence for *e and *o in Pre-Old Japanese. *Diachronica* 20/1, 83–137.

Mizushima, Yoshiharu. 1984. *Man'yōshū azuma-uta honbun kenkyū narabi ni sōsakuin*. Tokyo: Kasama Shoin.

2003. *Man'yōshū sakimori uta zenchūshaku*. Tokyo: Kasama Shoin.

Morita, Takeshi. 1993. *Nippo jisho teiyō*. Osaka: Seibundō Shuppan.

Murayama, Shichirō. 1957. Vergleichende Betrachtung der Kasus-Suffixe im Altjapanischen. *Studia Altaica: Festschrift fur Nikolaus Poppe*, 126–31. Wiesbaden: Otto Harrassowitz.

Murayama, Shichirō and Roy A. Miller. 1979. The Inariyama tumulus sword inscription. *Journal of Japanese Studies* 5, 405–38.

Nakada, Norio. 1969. *Tōdaiji fujumonkō no kokugogakuteki kenkyū*. Tokyo: Kazama Shobō.

Nakada, Norio and Hayashi Chikafumi. 2000. *Nihon no kanji*. Tokyo: Chūōkōron Shinsha.

Nelson, Andrew Nathaniel. 1974. *The modern reader's Japanese-English character dictionary*, second revised edition. Rutland, VT: Tuttle.

Nishizaki, Tōru. 1992. *'Hokke mongu' koten no kokugogaku-teki kenkyū: Honbun-hen*. Tokyo: Ōfūsha.

(ed.). 1995. *Nihon kojisho o manabu hito no tame*. Kyoto: Sekai Shisōsha.

Nomura, Takeshi. 1989. Jōdaigo no *tu* to *nu* ni tsuite. *Kokugogaku* 158, 1–14.

Numoto, Katsuaki. 1986. *Nihon kanjion no rekishi*. Tokyo: Tōkyōdō Shuppan.

Ogura, Hajime. 1998. Sa-gyō shiin no rekishi. *Kokugogaku* 195, 42–55.

Ohno, Susumu. 1953. Nihongo no dōshi no katsuyōkei no kigen ni tsuite. *Kokugo to Kokubungaku* 350, 47–56.

1977. Kanazukai no rekishi. In *Iwanami kōza nihongo*, ed. Ohno Susumu and Shibata Takeshi, vol. VIII (*Moji*), 301–39. Tokyo: Iwanami Shoten.

1978a. Dōshi katsuyōkei no kigen. In *Nihongo bunpō o kangaeru*, Ohno Susumu, 1979–222. Tokyo: Iwanami Shoten. (Reprinted 1979 in Hattori Shirō, Ōno Susumu 'Dōshi katsuyōkei no kigen' nit tsuite. In *Nihon no gengogaku*, vol. IV (*Bunpō* II), ed. Hattori Shirō, Kawamoto Shigeo and Shibata Takeshi, 658–63. Tokyo: Taishūkan Shoten.)

1978b. *Nihongo no bunpō o kangaeru*. Tokyo: Iwanami Shoten.

1980. *Nihongo no seiritsu* (Nihongo no sekai 1). Tokyo: Chūō Kōronsha.

1993. *Kakarimusubi no kenkyū*. Iwanami Shoten.

Ohno, Susumu *et al.* (eds.). 1990. *Iwanami kogo jiten*, 2nd, enlarged edn. Tokyo: Iwanami Shoten.

Omodaka, Hisataka *et al.* (eds.). 1967. *Jidai-betsu kokugo daijiten: Jōdaihen*. Tokyo: Sanseido.

Ōtsuka, Mitsunobu and Kita Takashi. 1999. *Esopo no haburasu: Honbun to sōsakuin*. Osaka: Seibundō Shuppan.

Philippi, Donald L. 1968. *Kojiki: Translated with an introduction and notes*. Tokyo: Tokyo University Press.

1990. *Norito: A translation of the ancient Japanese ritual prayers* (with a new preface by Joseph M. Kitagawa). Princeton : Princeton University Press.

Pulleyblank, Edwin G. 1984. *Middle Chinese: A study in historical phonology*. Vancouver: UBC Press.

1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: UBC Press.

Quinn, Charles Joseph. 1987. A functional grammar of predication in Classical Japanese. Doctoral dissertation, University of Michigan.

Ramsey, S. Robert. 1979a. Language change in Japan and the odyssey of a teisetsu. *Journal of Japanese Studies* 8/1, 97–131.

1979b. The Old Kyōto dialect and the historical development of Japanese accent. *Harvard Journal of Asiatic Studies* 39/1, 157–75.

1991. The polysemy of the term *kokugo*. In *Schriftfestschrift: Essays on writing in honor of John DeFrancis on his eightieth birthday* (Sino-Platonic Papers 27), ed. Victor H. Mair, 37–47. Philadelphia: University of Pennsylvania, Department of Oriental Studies.

Rodrigues, Ioão. 1603–4. *Vocabulario da lingoa de Iapam*. Nagasaki.

1604–8. *Arte da lingoa de Iapam*. Nagasaki.

1620. *Arte breve da lingoa Iapoa*. Macao.

Russell, Kerri L. 2006. A reconstruction and morphophonemic analysis of proto-Japonic verbal morphology. Doctoral dissertation, University of Hawai'i at Manoa.

Forthcoming. *A grammar of Eastern Old Japanese*. Folkestone: Global Oriental.

Saeki, Umetomo. 1959. *Nara jidai no kokugo*. Tokyo: Sanseidō.

Sandness, Karen. 1986. The pronunciation of the s-initial syllables in the Nara period: a critical look at Arisaka's and Mabuchi's theories. *Papers in East Asian Languages* 3, 1–16.

1999. *The evolution of the Japanese past and perfective suffixes* (Michigan Monograph Series in Japanese Studies, 26). Ann Arbor: Center for Japanese Studies, University of Michigan.

Satō, Kiyoji (ed.). 1977. *Kokugogaku kenkyūjiten*. Tokyo: Meiji Shoin.

(ed.). 1982. *Kodai no goi*. Tokyo: Meiji Shoin.

Satō, Ryōichi (ed.). 2002. *O-kuni kotoba o shiru: Hōgen no chizuchō*. Tokyo: Shōgakukan.

Satō, Takeyoshi. 1982. Kodai goi no gaisetsu. In K. Satō 1982, 45–77.

Seeley, Christopher. 1991. *A history of writing in Japan*. Leiden: E. J. Brill.

Shibata, Takesi. 1998/1965. The rise and fall of dialects. In *Takesi Sibata: Sociolinguistics in a Japanese context*, ed. Kunihiro Tetsuya, Inoue Fumio, and Daniel Long, 183–206. Berlin and New York: Mouton de Gruyter.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

2009. Elements of complex structures, where relativization isn't: The case of relativization. In *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution*, ed. Talmy Givón and Masayoshi Shibatani, 163–98. Amsterdam: John Benjamins.

Shimabukuro, Moriyo. 2007. *The accentual history of the Japanese and Ryukyuan language: A reconstruction*. Folkestone: Global Oriental.

Shinzato, Rumiko. 1991. Where do temporality, evidentiality and epistemicity meet? A comparison of Old Japanese *-ki* and *-keri* with Turkish *-di* and *-miş*. *Gengo Kenkyū* 99, 25–57.

Shōgakukan (ed.). 2000–2. *Nihon kokugo daijiten*, 2nd edn. Tokyo: Shōgakukan.

Siebold, Philipp Franz von. 1826. *Epitome linguae japonicae*. Batavia.

Sohn, Ho-min. 1999. *The Korean language*. Cambridge: Cambridge University Press.

Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76, 859–90.

Spear, Richard L. 1975. *Diego Collado's grammar of the Japanese language: Edited and translated by Richard L. Spear*. Kansas: Center for East Asian Studies, The University of Kansas.

Sugahara, Norio. 2000. *(Kirishitan shiryō o shiten to suru) Chūsei kokugo no kenkyū*. Tokyo: Musashino Shoin.

Sugimoto, Tsutomu. 1985. *Edo no onnakotoba: Asobase to arinsu to*. Tokyo: Sōtakusha.

1988. *Tōkyōgo no rekishi*. Tokyo: Chūō Kōronsha.

1989. *Seiyōjin no nihongo hakken: Gaikokujin no nihongo kenkyūshi (The discovery of the Japanese language by western people: A historical survey of learning and study of Japanese)*. Tokyo: Sōtakusha.

Suzuki, Isao. 1978. Bongokigen no Nihongoi. *Gekkan Gengo* 7/2, 30–5.

Takeuchi, Lone. 1987. *A study of Classical Japanese tense and aspect*. Copenhagen: Akademisk Forlag.

1999. *The structure and history of Japanese: From Yamatokotoba to Nihongo*. London: Longman.

Thunberg, Carl Peter. 1792. Observationes in linguam japonicam. *Nova Acta Regiae Societatis Scientiarium Upsaliensis* 2/5, 258–73.

Tsukishima, Hiroshi. 1969. *Heian-jidaigo shinron*. Tokyo: Tōkyō Daigaku Shuppankai.

1977. *Kokugo no rekishi*. Tokyo: Tōkyō Daigaku Shuppankai.

1986. *Heian jidai kunten-bon ronkō: okototen-zu kanajitai-hyō*. Tokyo: Kyūko Shoin.

1987. *Heian jidai no kokugo*. Tokyo: Tōkyōdō Shuppan.

(ed.). 1993. *Studies in Sino-Japanese* (Acta Asiatica 65). Tokyo: Tōhō Gakkai.

Tsunoda, Ryūsaku and L. Carrington Goodrich. 1951. *Japan in the Chinese dynastic histories: Later Han through Ming dynasties* (Perkins Asiatic monographs 2). South Pasadena: P. D. and I. Perkins.

Twine, Nanette. 1991. *Language and the modern state: The reform of written Japanese*. London: Routledge.

Umegaki, Minoru. 1978. *Nihon gairaigo no kenkyū*. Tokyo: Kenkyusha.

Unger, J. Marshall. 1977/1993. Studies in early Japanese morphophonemics: Yale University dissertation, reproduced by IULC.

1988. Middle Chinese syllable-final stops in Sino-Japanese: A critique of Vance's theory. *Journal of the American Oriental Society* 108/4, 627–31.

2001. Layers of words and volcanic ash in Japan and Korea. *Journal of Japanese Studies* 27/1, 81–111.

2003. Substratum and adstratum in prehistoric Japanese. In *Language contacts in prehistory: Studies in stratigraphy*, ed. Henning Andersen, 241–58. John Benjamins.

2007. Dating the collapse of *koo-otu* distinctions and the status of /mo/. In *Current issues in the history and structure of Japanese*, ed. Bjarke Frellesvig, Masayoshi Shibatani and John Charles Smith, 339–50. Tokyo: Kurosio Publishers.

Unger, J. Marshall and Yōkō Itō Tomita. 1983. The classification of Old Japanese adjectives. *Papers in East Asian Languages and Literatures* 1, 52–65.

Uwano, Zendō. 1977. Nihongo no akusento. In *Iwanami kōza nihongo*, ed. Ohno Susumu and Shibata Takeshi, vol. V (*On'in*), 283–321. Tokyo: Iwanami Shoten.

Vance, Timothy J. 1986. The development of Chinese syllable-final stops in Sino-Japanese. *Papers in East Asian Languages* 3, 48–62.

1987. *An introduction to Japanese phonology*. Albany: SUNY Press.

2007. Have we learned anything about *rendaku* that Lyman didn't already know? In *Current issues in the history and structure of Japanese*, ed. Bjarke Frellesvig, Masayoshi Shibatani and John Charles Smith, 153–70. Tokyo: Kurosio Publishers.

Vovin, Alexander. 1993. Long vowels in proto-Japanese. *Journal of East Asian Linguistics* 2, 125–34.

1994. Genetic affiliation of Japanese and methodology of linguistic comparison. *Journal de la société finno-ougrienne* 85, 241–56.

1997. On the syntactic typology of Old Japanese. *Journal of East Asian Linguistics* 6, 273–90.

2003. *A reference grammar of Classical Japanese prose*. London: RoutledgeCurzon.

2005. *A descriptive and comparative grammar of western Old Japanese*, part 1: *Phonology, script, lexicon and nominals*. Folkestone: Global Oriental.

2008. *A descriptive and comparative grammar of western Old Japanese*, part 2: *Adjectives, verbs, adverbs, conjunctions, particles, postpositions*. Folkestone: Global Oriental.

2009. *Man'yōshū to Fudoki ni mirareru fushigi-na kotoba to jōdai nihon-rettō ni okeru ainugo no bunpu (Strange words in the Man'yōshu and the Fudoki and the distribution of the Ainu language in the Japanese islands in prehistory)*. Kyoto: Kokusai Nihonbunka Kenkyūsentā.

Washio, Ryūichi. 2002. Jōdai nihongo ni okeru jodōshisentaku no mondai: seiōshogo to no hikaku kara miete kuru mono (Auxiliary selection in Old Japanese: a comparative perspective). *Nihongo Bunpō/Journal of Japanese Grammar* 2/1, 109–31.

Watanabe, Akira. 2002. The loss of overt Wh-movement in Old Japanese. In *Syntactic effects of morphological change*, ed. David Lightfoot, 179–95. Oxford: Oxford University Press.

Wenck, Günther. 1959. *Japanische Phonetik*, vol. IV. Wiesbaden: Otto Harrassowitz.

Whitman, John. 1985. The phonological basis for the comparison of Japanese and Korean. Doctoral dissertation, Harvard University.

1997. *Kakarimusubi* from a comparative perspective. In *Japanese/Korean Linguistics* 6, ed. Ho-min Sohn and John Haig, 161–78.

1999. Personal pronoun shift in Japanese: A case study in lexical change and point of view. In *Function and structure*, ed. Akio Kamio and Ken-Ichi Takami, 357–86. Amsterdam: John Benjamins.

2008. The source of the bigrade conjugation and stem shape in pre-Old Japanese. In *Proto-Japanese: Issues and prospects*, ed. Bjarke Frellesvig and John Whitman, 159–73. Amsterdam: John Benjamins.

Wrona, Janick. 2008. *The Old Japanese complement system.: A synchronic and diachronic study*. Folkestone: Global Oriental.

Forthcoming. The early history of *no* as a nominalizer. In *Studies in Japanese and Korean linguistics*, ed. Bjarke Frellesvig, Jieun Kiaer and Janick Wrona, forthcoming. London: Lincom.

Wrona, Janick and Bjarke Frellesvig. 2010. The Old Japanese case system: The function of *wo*. *Japanese/Korean Linguistics* 17, 565–79.

Yamada, Yoshio. 1913/1952. *Heian-chō bunpōshi*. Tokyo: Hōbunkan.

1913/1954. *Nara-chō Bunpō-shi*. Tokyo: Hōbunkan.

1935. *Kanbun no kundoku ni yorite tsutaeraretaru gohō*. Tokyo: Hōbunkan.

Yamaguchi, Akiho and Akimoto Morihide (eds.). 2001. *Nihongo bunpō daijiten*. Tokyo: Meiji Shoin.

Yamamoto, Matsuo. 1955. Keiyōshi ku-katsoyō shiku-katsuyō no imijō no sōi ni tsuite. *Kokugogaku* 23, 71–5.
Yamauchi, Yōichirō. 2003. *Katsuyō to katsuyōkei no tsūjiteki kenkyū*. Osaka: Seibundō Shuppan.
Yamazaki, Kaoru. 1992. *Keiyōshi jodōshi no kenkyū*. Osaka: Izumi Shoin.
Yanagida, Yuko and John Whitman 2009a. Alignment and word order in Old Japanese. *Journal of East Asian Linguistics* 18, 101–44.
Yanagida, Yuko and John Whitman 2009b. Ninshō to katsukaku ruikei: Jōdai nihongo no daimeishi taikei no shiten kara.
Yoshida, Kanehiko, Tsukishima Hiroshi, Ishizuka Harumichi and Tsukimoto Masayuki (eds.). 2001. *Kuntengo jiten*. Tōkyōdō Shuppan.
Yuzawa, Kōkichirō. 1928. Amakusa-bon Heike monogatari no gohō. *Kyōiku* (Special extraordinary issue), 55–126.
1929/1970. *Muromachi jidai gengo no kenkyū:* shōmono *no gohō*. Tokyo: Kazama Shobō.
1957. *Edo-kotoba no kenkyū*. Tokyo: Meiji Shoin.
1970. *Tokugawa-jidai gengo no kenkyū*. Tokyo: Kazama Shobō.
Yuzawa, Tadayuki. 1996. *Nihon kanjion-shi ronkō*. Tokyo: Benseisha.

Index of main grammatical forms