

A Historical and Morphosyntactic Analysis of Japanese Epistemic Markers (*Dearoo/Daroo* and *-Oo*)

AKITAKA YAMADA
Osaka University

1 Introduction

The epistemic modal marker in Early Middle Japanese (EMJ) is pronounced *-amu*, which is considered to be a suffix, for it is preceded by a be-support (*ar*-support) if it is not adjacent to a verb, as seen in (1)a (cf., Watanabe 2009, 39). Likewise, its historical descendent in Contemporary Japanese (CJ), *-oo*, also requires an *ar*-support, as seen in (1)b, and is also considered as a suffix.

- (1) a. [*atsu k*] *(*ar*-)*amu*. b. [*atsu k*] *(*ar*-)*oo*.
 hot PRED be-EPI hot PRED be-EPI
 ‘(It) will be hot.’ ‘(It) will be hot.’

CJ also has another epistemic modal expression *dearoo* (*daroo*), which is etymologically derived from three morphemes *de*, *ar*- and *-oo*, as in (2).

- (2) *atsu i* *de* *(*ar*-)*oo*. (3) *atsu i* *dearoo*.
 hot PRED.PRS DE be-EPI hot PRED.PRS EPI
 ‘(It) will be hot.’ ‘(It) will be hot.’

Japanese/Korean Linguistics 30.
Edited by Sara Williamson, Adeola Aminat Babayode-Lawal, Laurens Bosman, Nicole Chan, Sylvia Cho,
Ivan Fong, and Kaye Holubowsky.
Copyright © 2023, CSLI Publications.

When examining the gloss in (2), one may find this construction to be similar to (1)b in that *-oo* triggers an *ar*-support, thus proposing that *dearoo* is decomposed into three morphemes in the CJ grammar. However, this paper argues that the sequence of *de*, *ar*, and *-oo* underwent a diachronic reanalysis, and they serve as a lexicalized/unanalyzable unit in CJ, as indicated by the gloss in (3).

Although *dearoo* (*daroo*) has been actively discussed in the existing literature, previous studies in most cases concern the semantics and pragmatics (Hara 2018 amo.). A few syntactic studies discuss the classification of modal expressions/projections (e.g., genuine-modals and quasi-modals, or E-modals and U-modals; Inoue 2007; Ueda 2008; Haraguchi and Shuhama 2011), but they do not provide a finer-grained analysis as to how they interact with morphosyntactic operations, let alone the historical change.

This paper attempts to fill this gap. After reviewing the fundamentals of Japanese copular and epistemic modal constructions (Section 2), a morphosyntactic analysis is provided for epistemic modal constructions in both EMJ and CJ (Section 3). Then in Section 4, we discuss how the old system was replaced by the new system; we argue that the unification of the conclusive and adnominal form (i.e., the *syusi* and *reintai-kei*) facilitated the reanalysis, causing a domino effect in language change. This hypothesis is empirically supported by a survey of historical corpus data (Section 5).

2 Copular Sentences and Epistemic Modal Markers in Japanese

2.1 Distribution of Copular Markers

Japanese has two copular elements (Nishiyama 1997, 1999; Yamada 2023). Since the understanding of these elements is indispensable for the investigation of epistemic modals, let us first examine their distributions in CJ and EMJ.

Contemporary Japanese. Compare the small clauses in English (4) and CJ (5). Although the English small clause contains no overt copula, the Japanese sentence obligatorily pronounces one even in a non-tensed environment, and this boldface element is called the PREDICATIVE COPULA.

(4) Bernie considers [Alex smart].

(5) Contemporary Japanese (CJ)

- | | | | | | | |
|----|------------------|-----------------|----------------------------|------------------|---------------|-------------------------------------|
| a. | <i>Bernie-ga</i> | [<i>Alex-o</i> | { <i>gakusya/siawase</i> } | <i>ni</i> | <i>si-ta.</i> | |
| | Bernie-NOM | Alex-ACC | scholar/happy | | PRED do-PST | |
| | | | | | | ‘Bernie made Alex a scholar/happy.’ |
| | | | | | | <i>NP/NAP</i> |
| b. | <i>Bernie-ga</i> | [<i>Alex-o</i> | <i>utukusi</i> | <i>ku</i> | <i>si-ta.</i> | |
| | Bernie-NOM | Alex-ACC | beautiful | | PRED do-PST | |
| | | | | | | ‘Bernie made Alex beautiful.’ |
| | | | | | | <i>CAP</i> |

The predicative copula is pronounced either *ni* or *ku* in CJ. This choice is solely dependent on the category of the preceding element; a noun always takes *ni*, but adjectives are split into two types: (i) the one that takes *ni* is called the NOMINAL ADJECTIVE (NAP, e.g., *siawase* ‘happy’), (ii) while the one with *ku* is the CANONICAL ADJECTIVE (CA, e.g., *utukusi* ‘beautiful’). In the affirmative, tensed-environment, *ni* and *ku* are pronounced *de* and *ku*, as in (6), and they can be optionally contracted with the following element *ar-* (*at-* is its allomorph) to be pronounced *dat-* and *kat-*, respectively, as in (7).

- (6) a. *Alex-ga* {*gakusya/siawase*} ***de*** *at-ta*.
 Alex-NOM scholar/happy PRED be-PST
 ‘Alex was a scholar/happy.’ NP/NAP
- b. *Alex-ga* *utukusi* ***ku*** *at-ta*.
 Alex-NOM beautiful PRED be-PST
 ‘Alex was beautiful.’ CAP
- (7) a. *Alex-ga* {*gakusya/siawase*} ***dat-ta***.
 Alex-NOM scholar/happy PRED.be-PST
 ‘Alex was a scholar/happy.’ NP/NAP
- b. *Alex-ga* *utukusi* ***kat-ta***.
 Alex-NOM beautiful PRED.be-PST
 ‘Alex was beautiful.’ CAP

This contraction is only permitted when the two elements are adjacent. For example, when a particle is attached to the PredP, no contraction is triggered:

- (8) a. *Alex-ga* [{*gakusya/siawase*} ***de***]-*wa* *at-ta*.
 Alex-NOM scholar/happy PRED-FOC be-PST
 ‘It is true that Alex was a scholar/happy.’ NP/NAP
- b. *Alex-ga* [*utukusi* ***ku***]-*wa* *at-ta*.
 Alex-NOM beautiful PRED-FOC be-PST
 ‘It is true that Alex was beautiful.’ CAP

The second type is the DUMMY COPULA (*ar-* and its allomorph *at-*), and it is the element that precedes the past tense marker in (6) and (8), and the element that the predicative copula is fused with. This is akin to the English *do*-support: it appears when a suffix needs morphological support. The negation marker also triggers this dummy element (cf., *I *(do)-ed not run*):

- (9) a. *Hasit-ta*.
 run-PST
 ‘(S/he) ran.’
- b. *hasir-anak* ***(*at-*)***ta*.
 run-NEG be-PST
 ‘(S/he) did not run.’ VP

	English	CJ		EMJ	
		N/NA	CA	N/NA	CA
Predicative	∅	<i>ni/de</i>	<i>ku</i>	<i>ni</i>	<i>ku</i>
Dummy	<i>be</i>	<i>ar-</i>		<i>ar-</i>	

TABLE 1 Predicative and dummy copulas in English, CJ Japanese and EMJ Japanese

EMJ. The predicative and dummy copular is also observed in EMJ:

- (10) a. *Hasiri-keri*
run-PST
'(S/he) ran.'
- b. *hasir-az* **(ari-)**keri*.
run-NEG be-PST
'(S/he) did not run.'
- VP
- (11) a. [_{PredP} {*yama/apare*} *ni*]-*zo* **(ari-)**keru*.
mountain/amazing PRED-PRT be-PST.ADN
'(It) was a mountain/amazing.'
- NP/NAP
- b. [_{PredP} *asa* *ku*]-*zo* **(ari-)**keru*.
shallow PRED-PRT be-PST.ADN
'(It) was shallow.'
- CAP

It is clear that the *ar*-support is triggered when the past tense suffix is not immediately preceded by a verb. PRED and *be* can be contracted, as in (12).

- (12) a. {*yama/apare*} *nari-keri*.
mountain/amazing PRED.be-PST
'(It) was a mountain/amazing.'
- NP/NAP
- b. *asa* *kari-keri*.
shallow PRED.be-PST
'(It) will be shallow.'
- CAP

Table 1 summarizes the findings so far. Whether *-tal-keri* locally lowers to V/v (but not to A), or a verb (but not an adjective) head-moves to T, we can explain the distribution in the same way as the English data are explained.

2.2 Distribution of Epistemic Modal Markers

First, the distribution of *-amu* is the same as that of *-tal-keri*, as shown in (13) through (15): it involves an *ar*-support, unless it is adjacent to a verb.

- (13) a. *hasir-**amu***.
run-EPI
'(S/he) will run.'
- b. *hasir-az ar-**amu***.
run-NEG be-EPI
'(S/he) will not run.'
- VP
- (14) a. [_{PredP} {*yama/apare*} *ni*]-*ya* **(ar-)**amu*.
mountain/amazing PRED-PRT be-EPI
'Will (it) be a mountain/amazing?'
- NP/NAP

- b. [_{PredP} *asa* *ku*]-*ya* *(*ar*-)***amu***.
shallow PRED-PRT be-EPI
‘Will (it) be shallow?’ CAP
- (15) a. {*yama/apare*} ***nar-amu***.
mountain/amazing PRED.be-EPI
‘(It) will be a mountain/amazing.’ NP/NAP
- b. *asa* ***kar-amu***.
shallow PRED.be-EPI
‘(It) will be shallow.’ CAP

Second, being a historical descendent, *-oo* also shows a similar distributional pattern. Just like *-ta* and *-keri*, it is a suffix requiring *ar*-support:

- (16) a. **hasir-oo*. b. *hasir-anak ar-oo*.
run-EPI run-NEG be-EPI
‘(S/he) will run.’ (intended) ‘(S/he) will not run.’ VP
- (17) a. [_{PredP} {*yama/aware*} *de*]-*wa* *(*ar*-)***oo***.
mountain/pathetic PRED-PRT be-EPI
‘Will (it) be a mountain/pathetic?’ NP/NAP
- b. [_{PredP} *asa* *ku*]-*wa* *(*ar*-)***oo***.
shallow PRED-PRT be-EPI
‘Will (it) be shallow?’ CAP
- (18) a. {*yama/aware*} ***dar-oo***.
mountain/pathetic PRED.be-EPI
‘(It) will be a mountain/pathetic.’ NP/NAP
- b. *asa* ***kar-oo***.
shallow PRED.be-EPI
‘(It) will be shallow.’ CAP

But there is an important exception: the epistemic reading (e.g., **s/he will run*) is unavailable in (16)a: although the sequence itself is a possible Japanese sentence, it only has the volitional reading, unlike the other examples.

Finally, consider the distribution of *daroo* (*dearoo*). As shown below, it can be used with a VP, NP, NAP, and CAP, without an *ar*-support.

- (19) a. *hasir-u* {***daroo/dearoo***}.
run-PRS EPI
‘It may be the case that (s/he) runs.’ VP
- b. {*yama/aware*} (**ad*-) {***dearoo/daroo***}.
mountain/pathetic be- EPI
‘(It) will be a mountain/pathetic.’ NP/NAP

- c. *asa i (*ad-) {dearoo/daroo}*.
 shallow PRED.PRS be EPI
 ‘(It) will be shallow.’ CAP

3 Proposal

I argue that the similarities and differences of *-amu* and *-oo* are best explained, by inheriting the following commonly-accepted views from the literature:

- (20)
- a. Epistemic modality is represented by EpiP (not CP/TP), which appears in a position higher than Vol(ition)P (Cinque 1999).
 - b. When we fail to establish a head chain (head movement/lowering), do/be-support is triggered (Arregi and Pietraszko 2019).
 - c. There are postsyntactic morphological operations (vocabulary insertion, fusion, linearization etc.; Halle and Marantz 1993).
 - d. Head displacement is subject to historical change (Roberts 2007).
 - (i) V-to-T (T-Lowering): EMJ (✓), CJ (✓)
 - (ii) T-to-Epi: EMJ (✓), CJ (*)
 - (iii) A/N/Pred-to-T: EMJ (*), CJ (*)

3.1 The Morphosyntax of *-Amu* and *-Oo*

Consider the EMJ sentence in (13)a. In EMJ, the underlined heads in (21) form a head-complex via head movement (or lowering). The T-suffix is combined with V, so no *ar*-support is triggered; I use a dagger to indicate a suffix.

- (21) [EpiP [TP [VP ... V] ... T[†]] Epi[†]]
 |
hasir-amu

When V-to-T movement (or T-lowering) is hindered, the underlined head chain in (22) is now split (Arregi and Pietraszko 2019), and *ar-* is inserted to support the suffix *-amu*, as in (13)b and (14). For example, (22) represents how (13)b is derived. When Pred is adjacent to this *ar*-support element, a fusion is triggered to yield *nar* and *kar*, as in (15).

- (22) [EpiP [TP [NegP [VP ... V] ... Neg] T[†]] Epi[†]]
 | | |
hasir az ar-amu

The CJ sentences are derived in the same fashion save for (16)a. Unlike EMJ, CJ does not allow T-to-Epi movement (or Epi-to-T lowering). Thus, (16)a is illicit, for the suffix *-oo* remains stranded, as in (23)a. But a verb can move to the head of Vol(ition)P, as in (23)b; hence, the volitional reading.

- (23) a. $[_{\text{EpiP}} [_{\text{TP}} [_{\text{VP}} \dots \underline{\text{V}}] \text{T}^{\dagger}] \text{Epi}^{\dagger}]$ b. $[_{\text{TP}} [_{\text{VolP}} [_{\text{VP}} \dots \underline{\text{V}}] \text{Vol}^{\dagger}] \text{T}^{\dagger}]$
| | |
hasir **-oo* *hasir-oo*

Our analysis makes a good prediction about the contrast between EMJ and CJ in the past epistemic form. Since there is no T-to-Epi head movement, *-oo* cannot be immediately preceded by T, as in (24)a. So the sentence should require an *ar-* support. This prediction is borne out in (24)b, by inserting *ar-*, which is reduced to *r-* due to the sequence of the same vowels. The otherwise ill-formed sentence is now grammatical.

- (24) a. **hasit-ta -oo.* b. *hasit-ta r-oo.*
run-PST EPI run-PST be-EPI
‘It may be that (s/he) ran.’ (intended) ‘It may be that (s/he) ran.’

Compare this sentence with the EMJ counterpart. In EMJ, the epistemic inference about a past event is marked by a single morpheme *-kemu* ‘PST.EPI’ (not *keri-amu* ‘PST-EPI’). Since a fused form is considered possible iff two heads appear in single terminal node forming a head-complex, the existence of a fused form indicates the presence of T-to-Epi movement.

- (25) a. *hasiri-kemu.* b. *hasir-az ari-kemu.*
run-PST.EPI run-NEG be-PST.EPI
‘It may be that (s/he) ran.’ ‘It may be that (s/he) did not run.’

3.2 The Morphosyntax of *Dearoo* and *Daroo*

Now let us turn to the last epistemic modal marker, namely *daroo* (*dearoo*). One might have noticed that the sequence of *de*, *ar-* and *-oo* and its contracted form *daroo* are already seen in (17)a and (18)a. Therefore, one may propose that the other epistemic modal expression *dearoo* (*daroo*) is, in fact, not a single morpheme, but is decomposable into three independent morphemes.

This analysis, however, makes a number of erroneous predictions. First, it predicts that *daroo* (*dearoo*) can only be used with an NP or an NAP. However, this prediction is not borne out; it can also be used with a VP and a CAP:

- (26) a. $[_{\text{TP}} \text{hasir-}u]$ {*daroo/dearoo*}.
run-PRS EPI
‘It may be the case that (s/he) runs.’
b. $[_{\text{TP}} \text{asa} \quad \quad \quad i]$ {*daroo/dearoo*}.
shallow PRED.PRS EPI
‘It may be the case that (it) is shallow.’

Second, it predicts that the element preceding *daroo* (*dearoo*) must not be preceded by a tensed element. These predictions also fail:

- (27) a. [TP [{*yama/aware*} *de*]-*wa* *(*at-ta*)] {*daroo/dearoo*}.
 mountain/pathetic PRED-PRT be-PST EPI
 ‘It will be the case that (it) was a mountain/pathetic.’
 b. [TP [*asa* *ku*]-*wa* *(*at-ta*)] {*daroo/dearoo*}.
 shallow PRED-PRT be-PST EPI
 ‘It will be the case that (it) was shallow.’
 c. [TP *Hasit**(-*ta*)] {*daroo/dearoo*}.
 run-PST EPI
 ‘It may be the case that (s/he) ran.’
- (28) a. [TP {*yama/aware*} *dat-ta*] {*daroo/dearoo*}.
 mountain/pathetic PRED.be-PST EPI
 ‘It will be the case that (it) was a mountain/pathetic.’
 b. [TP *asa* *kat-ta*] {*daroo/dearoo*}.
 shallow PRED.be-PST EPI
 ‘It will be the case that (it) was shallow.’

Third, unlike the predicative copula, the *de* in *dearoo* can be coordinated:

- (29) a. *[*kaze-ga tsuyo i de*] *katsu* [*atsu i de*] *ar-oo*.
 window-NOM strong PRS *de* and hot PRS *de* be-EPI
 ‘(It) will be windy and hot.’ (intended)
 b. [*zyuudai de*] *katsu* [*konnan de*] *ar-oo*.
 serious PRED and difficult PRED be-EPI
 ‘(It) was serious and difficult.’

Given these observations, we must conclude that *daroo* and *dearoo* are distributed in a position higher than TP, and that the *de* in *dearoo* is distinct from the predicative copula. The entire expression *daroo* (*dearoo*) occupies the Epi position, the same as *-oo*. The difference between *daroo* (*dearoo*) and *-oo* should rather be attributed to their suffixal status: unlike *-am* or *-oo*, *daroo* (*dearoo*) is a free morph. Consider the derivation below.

- (30) CJ: derivation for (27)a
- | | | |
|----|-----------------------------------|----------------|
| a. | [EpiP [TP ... T [†]] | Epi] |
| b. | [EpiP [TP ... [T <i>be</i> T]] | Epi] |
| c. | [EpiP [TP ... [T <i>ar ta</i>]] | <i>daroo</i>] |

First, the only suffix-marked head is T (= (30)a). The *ar*-support is, thus, needed only for T (= (30)b). Second, the vocabulary items are plugged in each

terminal node (= (30)c). Third, after the hierarchical structure is linearized, phonological operations are applied to yield the sequence in (27).

4 Historical Changes in the Epistemic Modal Construction

If all the discussions so far are on the right track, we need to ask how the new forms emerged in a context where only *-oo* (and *-amu*) had been used. To answer this, one may wish to propose that people in the past could replace an NP (e.g., [_{NP} *konnan*] in (31)a) with a TP (e.g., [_{TP} *hasir-u*], as in (31)b). As a result of this substitution, the bold-face elements in (31)b are reanalyzed as being a single morpheme, encoding the epistemic modal meaning.

- (31) a. [_{PredP} [_{NP} *konnan*] *de*] *ar-oo*.
 difficult PRED be-EPI
 ‘(It) will be difficult.’
 b. [_{TP} *hasir-u*] ***de-ar-oo***.
 run-PRS PRED-be-EPI
 ‘(S/he) will run.’

This naïve replacement analysis, however, runs into problems. First, if an NP can be freely replaced by a TP, it predicts that (32)b is as grammatical as (31)b, which is contrary to fact. Second, if the sister node of Pred can be replaced by a TP, it is unclear why this does not hold with *ku*, as in (33).

- (32) a. [_{PredP} [_{NP} *konnan*] *de*] *at-ta*.
 difficult PRED be-PST
 ‘(It) was difficult.’
 b. * [_{TP} *hasir-u*] ***de-at-ta***.
 run-PRS PRED-be-PST
 ‘(S/he) ran.’
 (33) * [_{TP} *hasir-u*] ***ku-ar-oo***.
 run-PRS PRED-be-EPI
 ‘(S/he) will run.’

Despite these apparent challenges, this paper assumes that the basic insight of replacement analysis is essentially correct, and shows that these problems are solved when the details are fleshed out. To this end, we discuss how predicative copulas are derived first in Section 4.1, and then consider how the reanalysis proceeded during the transition from EMJ to CJ (Section 4.2).

4.1 Markedness in Pred

We saw that the distribution of *de* is much wider than that of *ku*: while *ku* is limited to a CA, *de* can be used with an NP and an NAP. This means that the

vocabulary insertion for Pred is sensitive to the category with which Pred is externally merged, and *de* is the unmarked, elsewhere vocabulary item. For these reasons, this paper proposes the following rules (cf., Yamada 2023):

- (34) a. Pred \leftrightarrow *de* (CJ), *ni* (EMJ)
 b. Pred_[Sel:⟨CA,1⟩] \leftrightarrow *ku*
 c. #Pred + *dummy* # \leftrightarrow *dar* (CJ), *nar* (EMJ)
 d. #Pred_[Sel:⟨CA,1⟩] + *dummy* # \leftrightarrow *kar*

To see how these rules work, consider the sentences in (8), and their derivations in (35) and (36). When vocabulary items are inserted, different items are selected on the basis of the category of the phrase with which Pred is externally merged. In (8)b, it is a CAP; hence the more specific rule in (34)b is chosen, as in (35). In (8)a, it is not a CAP; hence the more general rule in (34)a is utilized, as in (36).

- (35) a. [TP [PredP CAP Pred]-PRT T[†]].
 b. [TP [PredP CAP Pred]-PRT [T *be*-T]].
 | | | | |
 utukusi ***ku*** *wa* *ar ta*
- (36) [TP [PredP NP Pred]-PRT [T *be*-T]].
 | | | | |
 gakusya ***de*** *wa* *ar ta*

For the fused form, as in (7), the following derivations are assumed. Predicative copulas are fused with the dummy copula; *k-ar* is the marked form, because it is used only with a CAP, as shown below; n.b., the underline indicates that they form a head complex via head-movement (or lowering).

- (37) a. [TP [PredP CAP Pred_[Sel:⟨CA,1⟩]] T[†]].
 b. [TP [PredP CAP Pred_[Sel:⟨CA,1⟩]] [T *be* -T]].
 | |
 asa *kar ta*

4.2 Reanalysis

With the distinction of Preds in mind, let us consider the aforementioned questions, repeated below:

- (38) a. What licenses the reanalysis in (31)?
 b. Why is the reanalysis not triggered with a CAP (= (33))?
 c. Why is there no reanalyzed form for the past tense (= (32)b)?

Reanalysis in the Epistemic Modal Construction. When a CAP is used with a predicative copula, the EMJ grammar has two different strategies:

presence of a CAP. This is because the sisternode with which it merges is an NP, not a CAP; hence the elsewhere form is selected.

The reanalysis in question is now seen as the simplification of these complex heads (indicated by the two underlined segments in (42)a) into single terminal nodes, as shown in (42)a and (42)d.

(42)	a.	$[\text{EpiP}[\text{TP}[\text{PredP}[\text{NP}[\text{TP} [\text{CAP Pred}] \text{T}] \text{NMLZ}] \text{Pred}] \text{be-T}] \text{Epi}]$	
	b.	$\begin{array}{ccc} & & \\ \text{asa} & \text{ki} & \text{ni ar (nar) amu} \end{array}$	
	c.	$\begin{array}{ccc} & & \\ \text{asa} & \text{i} & \text{de ar (dar) oo} \end{array}$	
	d.	$[\text{EpiP}[\text{TP} \quad \quad \quad] \quad \quad \quad] \quad \quad \quad [\text{CAP Pred}] \text{T}] \quad \quad \quad \text{Epi}]$	

That is, in EMJ, the structure in (42)a generates the sequence in (42)b, and its historical descendants in (42)c. However, later generations exposed to a sequence of this kind abductively inferred that these sentences were generated from the structure in (42)d (cf., Roberts 2007), where *dearoo* and *daroo* were analyzed as the realization of the head of EpiP.

An important change that enabled the reanalysis which took place in the LMJ (Late Middle Japanese) was the loss of the distinction between the conclusive and adnominal form (Frellesvig 2010, 404, amo). That is, in EMJ, the adnominal form, for example, *ki* in (42)b, prevented the reanalysis, because the morphology clearly guarantees that the sequence of *asa ki* ‘shallow PRED.PRS.NMLZ’ is unambiguously a nominalized phrase. However, the unification of the adnominal and conclusive form, as in (42)c, made it possible for the sequence to be analyzed not only as an NP (under the interpretation that *i* is the adnominal form), but also as a TP (under the interpretation that *i* is the conclusive form), and this secondary interpretation allowed the bracketing structure in (42)d. Thinking this way, we can answer the questions posed in (38)a and (38)b. The trigger of the reanalysis is the loss of distinction in conjugation morphology, and the reason why *ku aroo* did not evolve into a full-fledged epistemic marker is that it involves a nominalization, and the predicative copula used for an NP, is *de*, not *ku*.

The Past-Tense Construction. Now let us turn to the last question in (38)c. If the reanalysis is, in this way, triggered in the epistemic modal construction, why does an equivalent change not happen with a past-tense construction (= (38)b)? Certainly, the nominalized construction is fine with the past tense:

(43)	<i>mizu-no</i>	<i>kokoro-no</i>	<i>asa</i>	<i>ki</i>	<i>nari-keri.</i>
	water-GEN	heart-GEN	shallow	PRED.NMLZ	PRED.be-PST
	‘The water in the river is not what I want it to be.’ (lit. The water’s sympathy for me is shallow; Tosa Nikki, 934)				

The derivation of this sentence is analyzed as follows, in much the same way as in (41). The only difference is that Epi is not present, so the head-complex of *nari keri* does not include Epi in its internal structure.

- (44)
- | | | | |
|------|---|---------------------------|-----|
| a. | [PredP CAP Pred] | | |
| b. | [TP [PredP CAP Pred] T] | | |
| c. | [NP[TP [PredP CAP Pred] T] NMLZ] | | |
| d. | [PredP[NP[TP [PredP CAP Pred] T] NMLZ] Pred] | | |
| e. | [TP[PredP[NP[TP [PredP CAP Pred] T] NMLZ] Pred] T [†]] | | |
| f. | [TP[PredP[NP[TP [PredP CAP Pred] T] NMLZ] Pred] <i>be-T</i>] | | |
| g. | [TP[PredP[NP[TP [PredP CAP Pred] T] NMLZ] Pred] <i>be-T</i>] | | |
| h. | | | |
| | <i>asa</i> <i>ki</i> | <i>ni ari (nari) keri</i> | |
| i. * | | | |
| | <i>asa</i> <i>i</i> | <i>de at (dat) ta</i> | |
| j. * | [TP | [TP [PredP CAP Pred] T] | T] |

Notice that the reanalysis from (44)g to (44)j is not as justified as before, because it does not make any sense to have a double TP structure. Arguably, there should be no language that redundantly uses two TPs one above the other. Hence, the reanalysis only emerged when there is an overt Epi suffix, which does not permit the sentence in (44)i.

5 Verifying the Hypothesis: A Corpus Survey

Our analysis is falsifiable, making several testable predictions. For example, by examining corpus data, we can empirically verify whether the emergence of new forms (*dearoo* and *daroo*) are preceded by the unification of the conclusive and adnominal form. To this end, this study conducted a survey using the Corpus of Historical Japanese (CHJ). With the search formulae below, instances of *-amu*, *dearoo* and *daroo* were extracted (last accessed Mar 2, 2023). Their distribution over time is shown in Figure 1. (Since instances with NP and NAP are ambiguous, we only examine the uses with CAP and VP.)

- (45) ***-amu-oo***: POS LIKE “(CA%|V%)” AND FOLLOWING WORDS: LEXEME *mu* ON 1 WORDS FROM KEY
- (46) ***dearoo***: POS LIKE “(CA%|V%)” AND FOLLOWING WORDS: WRITTEN FORM *de* ON 1 WORDS FROM KEY AND FOLLOWING WORDS: (LEXEME *aru* AND CONJUGATED FORM LIKE volitional/inferential form%) ON 2 WORDS FROM KEY
- (47) ***daroo***: POS LIKE “(CA%|V%)” AND FOLLOWING WORDS: (LEXEME *da* AND CONJUGATED FORM LIKE volitional/inferential form%) ON 1 WORDS FROM KEY

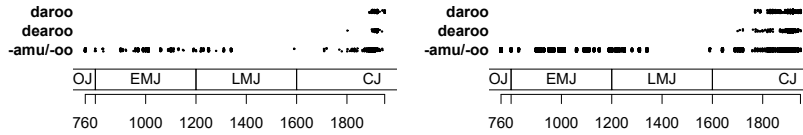


FIGURE 1 Historical change in the epistemic modal construction: CAP + EPI (Left) and VP + EPI (Right)

The unification of the conclusive and adnominal forms is known to have started around the LME period (Okimori 2010; Frellesvig 2010). If our analysis is on the right track, it is predicted that the emergence of *dearoo* (*daroo*) should be found only after the unification had been completed. As clearly shown in Figure 1, this prediction is borne out. The new forms came into use after the LMJ period was over.

Yet it must be acknowledged that the data in CHJ are limited, and mostly restricted to written texts. So, the initial examples of *dearoo* and *daroo* are likely to appear in much earlier days in colloquial registers. Nonetheless, even though they started being used 100 or 200 years earlier than the earliest examples in Figure 1, our prediction would not be seriously challenged.

6 Conclusion and Future Directions

The paper has proposed an analysis not only of the way the old and new grammars differ, but also of the transition from the former to the latter. First, as for the difference among the epistemic modal constructions, we have made the following claims:

- (48)
- a. The underlying structure of epistemic modal constructions is the same.
 - b. But *-amu* and its descendent *-oo* are both suffixes in the head of EpiP, while *dearoo* and *daroo* are free morphs.
 - c. The loss of T-to-Epi movement makes it impossible for VP+*-oo* to express the epistemic modal meaning, unlike VP+*-amu*.

Second, as for the transition, it is shown that the unification in the conjugation system caused the reanalysis, creating a new vocabulary insertion rule (Epi \leftrightarrow *dearoo/daroo*).

This change is also a change in the *do/be*-support system. In English, too, there is a great number of studies discussing the change in the status of *be* (Lightfoot 2006). It is, thus, fruitful to cross-linguistically compare the change in such dummy elements in future research to reveal commonalities over languages.

Acknowledgments

This research has been supported by The JSPS Grant-in-Aid for Scientific Research (C) (#22K00507) and The JSPS Core-to-Core Program (A. Advanced Research Networks “International Research Network for the Human Language Faculty” #JPJSCCAJ221702004).

References

- Arregi, K. and A. Pietraszko. 2019. Do-Support as Spellout of Split Head Chains. *Proceedings of the 49th Annual Meeting of the North East Linguistic Society*, 63–72.
- Cinque, G. 1999. *Adverbs and Functional Heads*. Oxford University Press: Oxford.
- Frellesvig, B. 2010. *A History of Japanese Language*. Cambridge: Cambridge University Press.
- Halle, M. and A. Marantz. 1993. Distributed Morphology and the Pieces of Inflection. *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, ed. K. Hale and S. J. Keyser, 111–176. Cambridge, MA: MIT press.
- Hara, Y. 2018. *Daroo* as an Entertain Modal: an Inquisitive Approach. *Japanese/Korean Linguistics* 25.
- Haraguchi, T. and Y. Shuhama. 2011. On the Cartography of Modality in Japanese. *Online Proceedings of Glow in Asia Workshop for Young Scholars 2011*, 102–110.
- Inoue, K. 2007. Nihongo no Modaru Saiko [Japanese Modals Revisited]. *Nihongo no Shubun Gensho [The Main Clause Phenomena in Japanese]*, ed. N. Hasegawa, 227–260. Tokyo: Hituzi Shoboo.
- Lightfoot, D. 2006. *How New Languages Emerge*. Cambridge: Cambridge University Press.
- Nishiyama, K. 1997. What does the Copula do? *University of Pennsylvania Working Papers in Linguistics* 4(2), 227–243.
- Nishiyama, K. 1999. Adjectives and the Copulas in Japanese. *Journal of East Asian Linguistics* 8, 183–222.
- Okimori, T. 2010. *Hazimete Yomu Nihongo no Rekisi [Introduction to the History of Japanese]*. Tokyo: Beret Publishing.
- Roberts, I. 2007. *Diachronic Syntax*. Oxford: Oxford University Press.
- Ueda, Y. 2008. Person Restriction and Syntactic Structure of Japanese Modals. *Scientific Approaches to Languages* 7, 211–150.
- Watanabe, A. 2009. *Seisei Bunpo [Generative Grammar]*. Tokyo: University of Tokyo Press.
- Yamada, A. 2023. Looking for Default Vocabulary Insertion Rules: Diachronic Morphosyntax of the Japanese Addressee-Honorification System. *Glossa* 8(1).