Divergence through differential frequency: The grammaticalization of the Japanese connective *soredewa* 'now/then'

CrossMark

Koji Tanno

Arizona State University, School of International Letters & Cultures, Arizona State University PO Box 870202, Tempe, AZ 85287-0202, USA

HIGHLIGHTS

- Reduced and non-reduced variant forms of the Japanese connective *soredewa* experienced divergence.
- The reduced variants *dewa* and *jā* became specialized in pragmatic use.
- The non-reduced variants reversed their trends of change over time.
- The observed functional differentiations follow the Reducing and Autonomy Effects proposed by J. Bybee (2007).

ARTICLE INFO

ABSTRACT

Based on an extensive analysis of Early Modern and Modern Japanese texts, the present study illustrates how the Japanese connective *soredewa* and its variants underwent semantic-pragmatic changes over time. More specifically, the quantitative evidence provided in this study reveals that the reduced and non-reduced forms of *soredewa* progressively diverged. The reduced form became strongly associated with newer functions, while the non-reduced forms reverted to their previous uses after the reduced forms increased their presence in the language. The development of the reduced forms was found to follow the Reducing and Autonomy Effects of high token frequency proposed by Bybee (2007). These results shed new light on the functional relationships that develop between reduced and non-reduced forms during grammaticalization, a topic in need of more attention and evidence in historical pragmatics.

1. Introduction

The present study aims to illustrate how the Japanese connective *soredewa* 'now/then' underwent a process of divergence by quantitatively tracking the developments of its variant forms.[1] According to [33]; divergence is a type of semantic-pragmatic change in which a new lexical form splits off from the original lexical form, which may remain as an autonomous element. The resulting forms may continue developing on their own and coexist with the original form for several centuries. In the case of the English indefinite article 'a/an', for instance, the word 'an' in Old English, pronounced as the segment 'one' in the word 'stone', eventually split into two lexical forms, 'one' and 'a/an': the former is derived from the full form, while the latter is derived from a cliticized form. By tracking the semantic-pragmatic uses of all of the variants of the Japanese connective *soredewa* over three centuries, the present paper demonstrates that *soredewa* experienced a similar divergence, whereby the forms that retained the anaphoric component *sore* 'that' diverged in usage from those that lost it. More importantly, this study shows that, as far as the development of *soredewa* is concerned, this divergence process involved not only the acquisition of a stronger pragmatic force in the reduced forms but also the opposite change in the original full form, reversing its course of change by regaining the original semantic meaning.

The analysis of the historical development of the connective *(sore)dewa/ja* in this study adopts a usage-based approach [8,16,52]. This approach regards language as a complex adaptive system

---

E-mail address: ktanno@asu.edu.

[1] The present study adopts the term *connective*, an English translation of the Japanese grammatical category *setsuzokushi*, to refer to connective expressions that primarily serve to create coherence at the semantic and pragmatic levels, similar to *and*, *but*, and *so* in English. These Japanese expressions are free morphemes that typically appear at the sentence-initial position [48]. explains this Japanese grammatical category in detail. The term *discourse marker* will be used as a cover term to refer to that functional category of expressions whose primary function is to signal non-propositional aspects of communicative force such as discourse organization, cognitive process, modalization, and interactional management. This group of expressions covers a wide range of expressions, including connectives and other types of pragmatic markers such as modal particles (such as the German *doch*) and interactional markers (such as the English *say*).

shaped by the language user's experience with it, in contrast with the static self-contained system conceived by more traditional linguistic analysis. Rather than relying on speakers' grammaticality judgments, the usage-based approach considers the actual use of language in discourse as its primary data and methodology and emphasizes the importance of the frequency of occurrence in analyzing linguistic elements. Such an approach allows one to address the non-discrete nature of linguistic categories and capture gradual changes that arise as a consequence of real human behavior, thus providing an appropriate and effective way to examine grammaticalization and other semantic-pragmatic changes.

In this study, the proposed divergence change in the connective *soredewa* is revealed quantitatively based on frequency analysis, a method that yields a straightforward and accurate picture regarding its shift in dominance from semantic to pragmatic use. In addition, special attention is paid to how the effects of token frequency, namely, the Reducing and the Autonomy Effects, interact and affect the pragmatic strengthening of *(sore)dewa*. The Reducing Effect refers to the effect whereby frequently used words or phrases reduce their forms [7,10,32]. This effect is often, but not always, observable in the erosion and/or phonological reduction of words or phrases, as well as in word and morpheme fusion [33,56,57]. The Autonomy Effect refers to the effect whereby frequently used and morphologically opaque words or phrases become more autonomous, resulting in their tendency to be stored as whole units in the lexicon [7,10,11]. On the other hand, infrequent and morphosemantically transparent words or phrases are considered to be less autonomous and are derived by morphological rules. The degrees of form reduction and autonomy are expected to shift over time with token frequency. More specifically, as token frequency increases, increased erosion should occur, and, consequently, more eroded variants should appear. Because form erosion and/or phonological reduction renders the morphosemantic structure of the word more opaque, this increase in eroded variants also leads to increased autonomy. This prediction is tested in the present study by quantitatively tracking the semantic-pragmatic meanings expressed by *soredewa* and its variants in historical Japanese texts from the mid-17th to mid-20th centuries.

2. Background

The pattern of morphological structure and grammatical functions of *sore-de-wa* are typical of the majority of common connectives, such as *sore-de-mo* 'but' and *sore-da-kara* 'so', in present-day Japanese. Their full forms are internally an adverbial phrase, consisting of (1) the anaphoric term, (2) a copula, and (3) a particle, and all of these connectives have reduced counterparts without the anaphoric component. The first morpheme of the connective *sore-de-wa* is a mesial demonstrative pronoun that, when used anaphorically, refers to a prior discourse segment. The second morpheme *de* 'to be' is the gerundive copula, which connects to the following clause to form a bi-clausal sentence, serving a similar function as the English conjunction 'and'. The sequence *sore-de*, thus, conveys the meaning of 'being so, and'. The final morpheme *wa* is a topic-marking particle, but when combined with a copula, as in the sequence *de-wa*, it conveys the conditional meaning 'if' [1,18,44].[2] Thus, the entire sequence *sore-de-wa* serves as an adverbial clause, roughly meaning 'if being so, then'.

As shown in Table 1, reduction in a different segment of the

Table 1
Variant forms of *soredewa*.

	Increased reduction in *dewa*	
Increased reduction in *sore*	soredewa	sorejā[a]
	soidewa	soijā
	sondewa	sonjā
	ndewa	njā
	dewa	jā

[a] In this paper, no distinction in the length of the final vowel is made for *sorejā* and *jā*. The only exception is when referring to a particular token appearing in an excerpt.

connective results in a different variant form. The anaphoric term *sore* can be reduced to *soi/son* to *n* to zero [24]. The non-anaphoric component *dewa* can be reduced to *jā* [50].[3] The variant form chosen appears to be related to the formality or spontaneity of the conversation [24]. shows that the construction *soredewa* tends to be reduced in casual conversations with close friends to *soijā* or even shorter form *jā*, or alternatively, *sonjā* or *njā*.

[45] proposes that the original adverbial phrase *soredewa* is grammaticalized into a discourse marker. This process is accompanied by the appearance of reduced non-anaphoric variants such as *dewa* and *jā*. Recent studies also suggest that this is a typical pathway in the development of Japanese connectives [47]. proposes that *datte* 'because' stemmed from the adverbial construction *sore-dattemo* 'even so' [80]. and [31] demonstrate the presence of a gradual transition from *sore-dakara* 'because being so' to *dakara* 'because' [46]. maintains that *daga* 'but' and *dakedo* 'but' stemmed from *sore-daga* 'even so' and *sore-dakedo* 'even so', respectively.

Diachronic studies on the connective *soredewa* have shown that it was initially an anaphoric adverbial phrase that developed into a semantic connective that expressed a conditional relation, and subsequently became a pragmatic connective that signals relations between speech acts and discourse boundaries. Many studies note that *soredewa* and *dewa* began to appear during the Early Modern Japanese period (1603–1868) [38,39,45,54,80]. The former form first appeared in texts from the 17th to 18th centuries. As shown in (1) and (2), it behaved more like an anaphoric expression that refers back to a prior discourse segment than a lexical item with its own semantic meaning [38,39,45] and was often used to introduce a negative consequence [80,81].

(1)   *Suō-otoshi* (Published in 1642)

(1.1)   A: *Ojigo-sama-wa   Ise-e   ikō   demo   nashi,*

          Uncle-mister-TOP[4]   Ise-to   go.will   COP-FP   not.and

          iku-mai   dem   gozaran

---

[3] The contraction of the *dewa* to *jā* morphemes is not an isolated change to the connective *soredewa* in modern Japanese, and can be found in other copulaic constructions and nominalizations, as well as in some western dialects [65]. Historical texts show that the original copulaic form *dearu*, from which all the other copulaic forms evolved, became reduced to *dea*, and eventually *jā*, which started appearing in Late Middle Japanese (1200–1600) [18,51].

[4] ACC = accusative, CL = noun classifier, CN = connective, COP = copula, DAT = dative, FOC = focus particle, FP = sentence-final particle, GEN = genitive, GEP = general extender particle, IMP = imperative, HRT = hortative, HON = honorific, N = nominalizer, NOM = nominative, PRG = progressive, PSS = passive, PST = past tense, QT = quotative particle, RST = resultant, TOP = topic marker, VLT = volitional.