Constructions 18 (2026)
doi 10.24338/cons-756

Yakuwarigo 'Role Language' in Translated Japanese: Representing Interpersonal Relationships in Constructions

Masaru Kanetani[1]
[1] University of Tsukuba

Abstract

When speeches delivered in foreign languages are translated into Japanese and broadcast on television, the resulting Japanese often sounds unnatural. Drawing on a small-scale exploratory case study of Japanese copula usage, this study examines such unnatural translated speech as an instance of *yakuwarigo* 'role language' (Kinsui 2003), which indexes non-fluent Japanese speakers. To support this claim, this article argues that fluent Japanese speakers store constructional knowledge that encodes appropriate interpersonal relationships, and that this knowledge can be described within the framework of Relational Morphology (Jackendoff and Audring, 2020). Furthermore, this study highlights the importance of the speaker's visual image in shaping the perception of translated speech as *yakuwarigo* and proposes that these translations are multimodal constructs composed of video and textual elements.

1 Introduction

When speeches in foreign languages are broadcast on Japanese TV programs, they are typically translated into Japanese, dubbed, and subtitled, as shown in Figure 1.

Figure 1: A (modified) screenshot from the TV program "Shirabetemitara", originally broadcast on Fuji Television (https://www.youtube.com/watch?v=b_3P7jGXvjE)[1]

[1] The video clip is no longer accessible as of September 4, 2025. It was viewable on the broadcaster's official YouTube channel (https://www.youtube.com/@FNNnewsCH), but was either made private or removed after a certain period following the broadcast.

Corresponding author(s): Masaru Kanetani, University of Tsukuba, kanetani.masaru.gb@u.tsukuba.ac.jp

During an interview conducted in English (Fig. 1), a man from Australia talks about what brought him to Japan and his experiences there. This is accompanied by Japanese translation provided through subtitles and dubbing. The subtitle at the bottom of the screen is transcribed as in (1):

(1) 1998 年から 2017 年まで住んでいたんだ
1998-nen kara 2017-nen made sundeita-nda.
1998-year from 2017-year until live.PST-it.is.that[2]
'I lived here from 1998 to 2017.'[3]

At least to the author of the present paper, the underlined part in the Japanese sentences in (1) sounds unnatural. A more idiomatic version would arguably be (2).

(2) 1998-nen kara 2017-nen made sundeita-ndesu-yo.
1998-year from 2017-year until live.PST- it.is.that.POL.-SFP
'I lived here from 1998 to 2017.'

Japanese predicates generally display alternations among their plain, polite, super-polite, honorific, and humble forms, which formally index the interpersonal relationship between interlocutors (see Section 3 for details). For example, the copula appears as plain *da*, polite *desu*, and super-polite *degozaimasu*. Information status and interpersonal stance are also indexed by sentence-final particles (SFPs) such as *yo* 'I tell you', *ne* 'you know', and their combinations. Replacing the plain form of the predicate *sundeita-nda* in the original translation in (1) with the polite form *sundeita-ndesu-yo*, as in (2), yields an utterance that is more idiomatic for this street-interview context with an unfamiliar interviewer. The subsequent lines in the clip are likewise rendered in the plain style, as transcribed in (3):

(3) ヒカリエはもう建っていたと思うけどたくさん変わったねぇ
Hikarie-wa moo tatteita-to omou-kedo
Hikarie-TOP already built.be.PST-COMP think-but
takusan kawatta-nee.
a.lot change.PST-SFP
'I believe that Hikarie was already there, but things have changed a lot, haven't they?'

Once again, I would use the polite forms *omou-ndesu* 'believe [with the explanatory modality *n*]' and *kawarimashita* 'changed' instead of the plain form *omou* and *kawatta*, respectively.

[2] The abbreviations used in the glosses are as follows: 1 = first person; ACC = accusative marker; ATT = attributive suffix; COMP = complementizer; COND = conditional conjunctive particle; COP = copula; DAT = dative marker; GEN = genitive marker; NOM = nominative marker; POL = polite; PST= past form; SG = singular; SFP = sentence-final particle; TOP = topic marker. Transcriptions of interviews are accompanied by Japanese texts.

[3] This English sentence is not the original speech. Since the Japanese translation is dubbed over the original utterance, the original utterance is not audible and the English sentence is my translation of the translated speech in Japanese.

This exploratory case study asks the question why broadcast translations sometimes sound less idiomatic to Japanese ears and argues that, in interview contexts, viewers often arrive at a *yakuwarigo* reading: visual cues evoke an INTERVIEW frame, and when the subtitled/dubbed line diverges from the expected register, the incongruity is coerced into *yakuwarigo* 'role language' (Kinsui 2003) — leading the viewers to interpret it as "this is how such a (perceived) non-fluent speaker would talk" (see Section 2 for details of *yakuwarigo*). Specifically, despite the unnaturalness, viewers infer that the speaker would likely use such unnatural Japanese based on the speaker's perceived identity as a non-native speaker of Japanese.

At the same time, this article subscribes to Construction Grammar (CxG) and contends that the accurate representation of constructional knowledge is essential. It is argued that since indexing interpersonal relationship is inseparable from situation report in language use in Japanese (Hirose 2013, 2015, 2024), such information about interpersonal relationships needs to be represented in constructions and stored as part of constructional knowledge. Non-fluent Japanese speakers who lack this knowledge may fail to select appropriate constructions. This observation supports treating unnatural utterances like (1) as *yakuwarigo*. More specifically, it is argued that *yakuwarigo*, as conceived in the present study, should be treated as a construction with social meaning, rather than a stylistic phenomenon.

The remaining sections are organized as follows. Section 2 proposes a view of Japanese used in these translated speeches as *yakuwarigo* 'role language' (Kinsui 2003). That is, although Japanese lexical items and syntax are used, making the utterance intelligible, the translated speech is not natural but rather virtual Japanese—intentionally used to index a non-fluent speaker of Japanese. Section 3 argues that in ordinary Japanese speech, interpersonal relationship needs to be linguistically encoded (Hirose 2013, 2015, 2024), and hence this knowledge of interpersonal relations is stored as part of constructional knowledge. Section 4 describes constructional knowledge by using Jackendoff and Audring's (2020) Relational Morphology templates. It discusses how the choice of plain or polite forms, the presence or absence of SFPs, and their combinations are distinguished across different interlocutors and contexts. Finally, Section 5 proposes a view of *yakuwarigo* in translated speeches broadcast on TV as video/image-text multimodal constructs, emphasizing the importance of the visual information alongside the textual information. Section 6 concludes the discussion.

2 Translated speech as *yakuwarigo*

Unnatural utterances in translated speech can be explained in terms of *yakuwarigo* 'role language", proposed by Kinsui (2003) and defined as follows:[4]

[4] A reviewer noted that "the definition seems to imply that *yakuwarigo* is not restricted to fictional data but in fact describes relational work choices". While *yakuwarigo* was originally proposed to capture conventionalized mappings between linguistic forms and character personae in fiction (Kinsui 2003), the definition explicitly allows for two-way mappings—the style-to-persona and persona-to-style mappings. This study extends the term in the latter sense to non-fictional broadcast translation as well, where viewers infer "how someone would speak" from the on-screen persona. This does not equate *yakuwarigo* with relational work: whereas relational work addresses the broader "negotiation" of interpersonal relations (Locher and Watts 2005: 11; see also

> When a particular style of speech (vocabulary, grammar, phrasing, intonation, etc.) evokes a specific image of a person (age, gender, occupation, class, era, appearance, personality, etc.), or when a specific image of a person allows one to imagine the style of speech that the person would likely use, that style of speech is called *yakuwarigo*. (Kinsui 2003:205, my translation)

For example, the simple utterance "yes, I know" can be translated in various ways as listed in (4a–h), each of which evokes a specific image of a speaker, as indicated in the parentheses after the example.

(4)	a.	Soo-yo, atashi-ga shitte-ru-wa	(a young girl)
	b.	Soo-ja, washi-ga shitte-oru	(an old scientist)
	c.	Soya, wate-ga shitt-oru-dee	(a man from Osaka)
	d.	Soo-ja, sessha-ga zonjite-oru	(a samurai)
	e.	Soo-desu-wa-yo, watakushi-ga zonjite-ori-masu-wa-yo	(a lady)
	f.	Soo-dayo, boku-ga shitte-ru-no-sa	(a young boy)
	g.	Nda, ora shitteru-da	(a countryman)
			(Kinsui 2003:v)

The response *yes* is translated as *Soo-yo* in (4a), *Soo-ja* in (4c), and *Nda* in (4g); the predicate *know* is translated *shitte-oru* in (4b), *shitt-oru-dee* in (4c), and *zonjite-ori-masu-wa-yo* in (4d). Sentence-final expressions such as *-yo*, *-ja*, and *-desu-wa-yo*, and first-person expressions, such as *atahi*, *watashi*, *wate*, and *sessha*, are typical elements of *yakuwarigo*.[5] These expressions all evoke a specific type of speaker.[6] However, what the speakers evoked by these *yakuwarigo* represent nothing but linguistic stereotypes (Kinsui 2003:35). For example, an old scientist may utter the line in (4b) in comics, dramas, or plays, but scientists in reality would not speak in that way.

Note that Kinsui's (2003:205) definition of *yakuwarigo*, cited above, involves two directions of matching between a speech style and a speaker: (i) *from speech style to person*, where a particular style of speech evokes a specific image of a

---

Halliday 1978), *yakuwarigo* refers to stereotypical form-persona associations that can be activated in context (Ota 2009, 2017; Yoda 2011). As discussed in Section 4, students use polite form when addressing teachers, while super-polite forms are typically employed in customer service contexts. Crucially, however, the roles are not assigned by the expressions they use, nor does specific expressions evoke their roles. Rather, because of the pre-existing social roles, they are engaged in the relational work and use certain discerning styles.

[5] I use the term "first-person expressions" rather than "first-person pronoun", because the expressions are distinguished from personal pronouns. Hirose (2000:1630) points out that "in Japanese a variety of words are used to represent the public self, depending on who is talking to whom." In contrast, the pronoun *I* is a special word for public self, and personal pronouns "are defined primarily as public expressions" (ibid.).

[6] Because of the rich person-referring expressions and sentence-final expressions, Japanese displays many *yakuwarigo*. Yamaguchi (2007) observes that English, a language poor in these expressions, also has some, if not many, *yakuwarigo* as well, as exemplified in (i):

(i) "<u>Las'</u> time I saw you, <u>you was</u> only a baby," said the giant. "<u>Yeh</u> look a lot like a like <u>yer</u> dad, but <u>yeh</u>'ve got <u>yer</u> mum's eyes."(*Harry Potter and the Philosopher's Stone*, as cited in Yamaguchi 2007:12, underlining added)

The quotations are utterances by Hagrid, a giant gamekeeper and groundskeeper at Hogwarts wizarding school in *Harry Potter and the Philosopher's Stone*. Yamaguchi (2007) points out that the underlined elements in (i) make Hagrid sound West Country accented.

speaker; and (ii) *from person to speech style*, where a specific image of a speaker enables one to infer the style of speech that the person would likely use. Those in (4a–h) instantiate *yakuwarigo* in the sense of definition (i). When dealing with translated speech through the lens of *yakuwarigo*, definition (ii) is relevant; that is, we have a specific image of the original speaker, from which we can infer the speech style they are likely to use. This article restricts the application of *yakuwarigo* to this sense. Thus, it can be understood in sociolinguistic terms as an enregistered linkage between linguistic forms and personae—as expressions capable of indexing the speakers' stereotypical characteristics (Agha 2007: 55). A reviewer also correctly suggests that Kinsui's second definition of *yakuwarigo* (the mapping from person to speech-style) amounts to Silverstein's (2003) third-order indexicality, where stereotyped personae are invoked to license form selection and interpretation. In other words, *yakuwarigo* readings are not intrinsic properties of forms, but rather, higher-order, metapragmatically guided construals (see also footnote 4).

2.1 Ota's (2009, 2017) studies on interviews with athletes

Analyzing translated speech in terms of *yakuwarigo* in this way is not entirely new. For example, Ota (2009) investigated translations in TV interviews with athletes during the 2008 Beijing Olympics and observes that utterances by athletes tend to be translated into Japanese containing *yakuwarigo* allied to certain characters, as summarized in Table 1 (see also Ota 2017).

athletes/contexts	characters represented by *yakuwarigo*
Male superstars (e.g., Usain Bolt)	strong and tough
"Queen of X" (e.g., Yelena Isinbayeva, a.k.a. Queen of pole-vaulting)	strong and tough
Athletes with easily imitable characters (e.g., Dextler St. Louis)	frank and friendly
Disadvantaged athletes (e.g., Natalie du Toit)	sincere
Messages to rivals (e.g., Teddy Riner to Satoshi Ishii)	aggressive and persisting messages

Table 1: Athletes and the characters represented by *yakuwarigo*

Consider, for example, the following translated speech by Usain Bolt, a Jamaican sprinter:

(5) 俺がナンバーワンだ
Ore-ga nanbaawan-da
1.SG.-NOM number.1-COP
'I am number 1!' (Ota 2009:57)

The nominative first-person expression, *ore-ga*, and the plain form copula *da* are characteristics of this translation. Following Kinsui's (2003, 2007) observation that in postwar boys' comics, a smart hero pursuing his ideals, full of wisdom, courage, and a sense of justice tends to refer to himself as *boku*, while *ore* is typically used by a strong and wild hero, Ota (2009:63) contrasts the first-person expressions *ore* and *boku* and highlights the different images of the character

evoked by each expression. That is, because of his showy and confident pauses and performances before and after races (Ota 2009:63), the typical image of his character as a charismatic superstar with a strong personality overlaps with the comic book hero archetype associated with the use of *ore*. Note that these first-person expressions existed long before they were tied to the stereotypical images of the characters, but the repeated uses in comics or other literary works as *yakuwarigo* might enhance the images of the speakers tied to the words. *Shogakukan's Seisenban Nihon Kokugo Daijiten* (a Japanese dictionary) describes the shift in meaning of *ore* as follows: "a self-reference term, which was widely used in the past when *addressing both superiors and inferiors, regardless of social status or gender*. Today, it is used by men in informal settings when addressing peers or those of lower status" (italics added). In contrast, the dictionary defines *boku* as "a self-reference term *for men*" (italics added) and explains the meaning shift as follows. It was originally used in classical Chinese literature, as a term referring to men of great humility since ancient times. From the Edo period onward, *boku* was used by young men and students as a euphemism for referring to peers or those of lower status. Since then, its use has become widespread, and today it is commonly used to refer specifically to young boys.

Ota (2009) also analyzes the polite copula *desu*, as opposed to the plain copula *da*, as an example of *yakuwarigo* in translation. As we discuss in Sections 3 and 4, the plain form *da* is rarely used in normal conversation, and this fact makes the utterance in (1) unnatural. Nevertheless, Ota's observation indicates that the plain form *da* is dominant in translated speech in sports television programs, thus making the use of the polite form *desu* rather marked in such programs.[7] Ota then argues that the use of *desu* adds a particular nuance to the utterance. For example, an interview with Natalie du Toit, a South African swimmer who lost her left leg due to a traffic accident in 2001, is translated using the polite form *desu*, as shown in (6):

(6) 障害があっても努力をすれば夢は実現するのです

Shoogai-ga	atte-mo	doryoku-o	sure-ba	yume-wa
disability-NOM	exist-even.if	effort-ACC	do-COND	dream-TOP

<u>jitsugensuru-nodesu</u>
realize-it.is.that.<u>POL</u>

'Even with disability, I can realize my dream by working hard.'
(Ota 2009:67, underlining added)

In (6), the polite sentence-final modality *nodesu*, as opposed to the plain form *noda*, is used. Ota claims that the polite forms, which indicate weakness, are used when disadvantaged athletes who have faced hardship appeal or speak of their thought from the heart. This effectively allows viewers to imagine how they have confronted those difficulties and strived with all their effort, while enhancing their adverse circumstances.

[7] Ota (2017:28) reports that plain forms account for 66%, while the polite forms account for 10% of the overall translated speech in interviews during the 2016 Rio de Janeiro Olympics. Since Ota also examines interviews with non-Japanese speakers, the predominance of the plain form could reflect a general tendency in translated speech within interview contexts. However, further investigation of cross-genre interviews involving different types of speakers is necessary.

It should be noted that the first-person pronoun in Bolt's utterance is not always translated as *ore*. Ota (2017) reports that in 7 out of 10 broadcast translations, when Bolt speaks about his final Olympic participation with some emotional feeling, his speech is translated containing *watashi*, a formal first-person expression with a calm tone, as shown in (7):[8]

(7) これが私の最後のオリンピックだ

Kore-ga	watashi-no	saigo-no	orinpikku-da
this-NOM	1.SG.-GEN	last-GEN	Olympics-COP

'This is the last Olympics to me.' (Ota 2017:29, underlining added)

This suggests that the distinction between *ore* and *watashi* as *yakuwarigo* is attributed not only to specific speaker's identity but also to the contents and context of the utterance. This is also true to the daily use of person-indicating expression in relational work. There are many person-indicating expressions in Japanese, and the choice "depends on such factors as age, sex, and social status of the conversation participants; the social relationships that hold between them; the degree of intimacy or formality of the conversation; and combinations of these factors" (Fillmore 1975:77). That is, a single individual uses various expressions to refer to oneself and whether one refers to oneself as *watashi*, *ore*, or *boku* largely depends on the context.

Crucially, in translations of athletes' speeches, "the athletes actually do not speak Japanese but the translators choose linguistic expressions according to the athletes' characters" (Ota 2009:57, my translation). Thus, the translated speech containing *yakuwarigo* reflects the translators' perspectives on the original speakers. By listening to the athletes' translated speech while watching their figures, the viewers may unify their perspectives with those of the translators and believe that the athletes have these characters, which actually represent just made-up stereotypes imposed on the viewers. In other words, one can manipulate a person's impressions by using *yakuwarigo*. We may have certain stereotypical images toward famous athletes and people in showbusiness, who are repeatedly shown on a TV screen, and *yakuwarigo* that appears to accord with their stereotypical image may help amplify such images. However, the man in Figure 1 is a layperson, who came to see his friends in Japan from Australia and happened to respond to a Japanese TV interview; the viewers have no prior knowledge about him. If the unnatural speech in (1) is analyzed as *yakuwarigo* in the sense of definition (ii)—similar to the translated speech of athletes—what image do the viewers form of this man and what role is the man assigned?

2.2 Yoda's (2007, 2011) research on 'Westerner's language'

Relevant insights are also presented in Yoda's (2007, 2011) study of what she calls *Seeyojin-go* 'the Westerner's language', a speech style that many Japanese people think Westerners would use. Note that despite its label "Westerner's language", the lexical items and grammatical structures are Japanese (e.g., *Oo, Romio, Romio* (a Japanese translation of "O Romeo, Romeo [wherefore art thou Romeo?]" from Shakespeare's *Romeo and Juliet*). Yoda (2007) explains how the Westerner's language is diachronically established and shared in the Japanese

[8] Ota (2017:32) also suggests that the polite form of copula *desu* is compatible with this specific speech.

community as follows. First, the Japanese who hear literal translations of Western plays found them awkward, because they contain elements that are not commonly used in natural spoken Japanese. Then, the sense of awkwardness, arising from intercultural contact, left strong impression on Japanese speakers, who (mis)interpreted the translated utterances as typical speech styles of Westerners.

In her subsequent research on the Westerner's language, Yoda (2011) identifies the following characteristics of Westerner's language: (i) the wrong use of the SFP *ne* 'you know' (e.g., *Ichido kekkon shippai shita-ne* / once marriage fail did-SFP/ 'I once failed at marriage' (p.214)); (ii) the addition/deletion of a mora (e.g., *Kono jiken waakarimashita-kaa*? [cf. the standard speech: *wakarimashita-ka*] / this case understand.PST.POL-Q / 'Did you understand this case?' (p.215); and (iii) the (random) insertion of the speaker's mother tongue (e.g., *Nihongo berii muzukashii-deesu* / Japanese very difficult-COP.POL / 'Japanese is very difficult' (p.217).[9] In addition to Yoda's observations, the absence of functional words indicating semantic roles in these examples should not be overlooked. While sematic roles are basically marked by word order in English, they are marked by particles in Japanese, which has a relatively free word order (cf. *John-ga Mary-ni tegami-o okutta* [=*Mary-ni John-ga tegami-o okutta* / *Mary-ni tegami-o John-ga okutta*] 'John sent Mary a letter' vs. *Mary-ga John-ni tegami-o okutta* 'Mary sent John a letter'). Thus, the absence of such functional words, as well as the elements listed in (i)–(iii), jointly makes the utterances awkward, and indexes the non-fluent Japanese speaker. Adding the dative marker *ni* and the topic marker *wa* to the examples would make them a little more idiomatic: *kekkon-ni shippai-shita-ne*; *jiken-wa wakarimasshita-ka*; and *Nihongo-wa berii muzukashii-desu*.

Yoda (2011:233) argues that "awkward" Japanese expressions characterized by these elements are related to "the sociable and likable characters" often attributed to Westerners. In fact, these elements are observed in the interview video clip in Figure 1. Following the utterances transcribed in (1) and (3), the man describes his experience at a Japanese summer festival, which is translated as in (8):

(8) 日本の夏祭りは行けるならぜひ行ってほしいね。とても日本らしいからね。たくさんの美しい浴衣とおいしい食べ物。ハッピーな人たち。いい時間だよ。

Nippon-no	natsumatsuri-wa	ikeru-nara	zehi
Japan-DAT	summer.festival-TOP	can.go-COND	by.all.means

ittehoshii-ne.	Totemo	Nihonrashii-kara-ne.	Takusan-no
go.want-SFP	very	Japanese-because-SFP	a.lot-GEN.

utsukushii	yukata	to	oishii	tabemono.	Happii-na
beautiful	yukata	and	delicious	food	happy-ATT

hitotachi.	Ii	jikan-da-yo.
people	good	time-COP-SFP

'If you have the opportunity to go to a Japanese summer festival, I highly recommend it. They are so quintessentially Japanese. There are beautiful *yukata*, delicious food and happy people everywhere. It's a great time!'

[9] Note that Japanese is a mora-timed language. Note also that the word *berii* (ベリー) is a transcription of *very* within the Japanese phonetic and orthographic system.

The underlined parts in (8) indicate the unnatural use of the SFP *ne* and the use of the English word *happy* (transcribed as *happii* (ハッピー) in Japanese; see footnote 9), characteristic of the Westerner's language in Yoda's (2011) terms. However, it is not clear why and how these elements indicate "the sociable and likable characters" of Westerners. The man in the video clip does look sociable and friendly, but it is just an impression and not related to the linguistic expressions. As Yoda (2007) points out, the repeated use of these elements with sociable and likable Western characters in comics, dramas, plays, and translations over time has gradually built such a stereotypical view. That said, there is no direct, motivated correlation between the linguistic expressions and "the sociable and likable characters".

In this section, assuming that the translated speech broadcast in the TV program in Figure 1 — parts of which are transcribed in (1), (3), and (8) — can be analyzed as *yakuwarigo*, two previous studies were reviewed: Ota's (2009, 2017) analysis of athletes' translated speech as *yakuwarigo* and Yoda's (2009, 2011) research on the Westerner's language used in plays, comics, dramas, and other media. Kinsui (2003:205) defines *yakuwarigo* either as (i) a particular style of speech that evokes a specific image of a speaker, or as (ii) a specific image of a speaker that enables one to infer the style of speech that the person would likely use. It is in the sense of definition (ii) that is relevant when analyzing unnatural Japanese used in the translated speeches of famous athletes and the traveler depicted in Figure 1. In contrast, the Westerner's language is more or less similar to those listed in (4a–g), i.e., *yakuwarigo* as characterized by definition (i).

Unlike the original speakers of translated speech, the speakers of the Westerner's language and the *yakuwarigo* listed in (4a–g) are fictional characters created by the scriptwriters and comic artists. These forms of *yakuwarigo* are used to imbue unreal personas with the creators' intended characters, thereby allowing those intended characters to be shared with the viewers. The Australian man's speech, translated as in (8), contains certain elements typical of the Westerner's language—such as the unnatural use of *ne* and the insertion of the English word *happy*. However, it cannot be viewed as the Westerner's language in Yoda's (2009, 2011) terms. Although Yoda (2011) argues that the "awkward" Japanese evokes a "sociable and likable" Westerner, there is no clear or direct relation between them. Rather than attributing awkwardness or unnaturalness to specific characters based on impressions, the present study claims that such features should be more directly linked to the non-fluency of the language, because a non-fluent speaker of a language cannot speak that language idiomatically. Based on our knowledge of famous athletes, we may perceive, for example, Usain Bolt as a superstar, and Natlie du Toit as a disadvantaged athlete. However, as noted earlier, viewers have no prior knowledge about the man in Figure 1. Due to his appearance as a typical non-Japanese person, many viewers may unconsciously assume that he does not speak Japanese fluently. Although he stayed in Japan for 19 years (from 1998 to 2017) and he may have a good command of the Japanese language, this is not crucial.

What matters is the specific image that viewers evoke, not from his speech style but from his appearance. In this regard, the figure of the man is essential to analyze the utterances in (1), (3), and (8) as *yakuwarigo*. The interaction of the visual and text information will be discussed in Section 5. Before proceeding, to account for why (1) reads as less idiomatic, Sections 3 and 4 set aside *yaku-*

warigo and examine how interpersonal relationship is typically indexed in Japanese interaction—especially through predicate morphology and sentence-final particles.

3 Indexation of interpersonal relationships in Japanese

In this section I will outline the gramatico-pragmatic theory known as "the three-tier model of language use", developed through a series of studies by Yukio Hirose (e.g., Hirose 2013, 2015, 2024) to show the significance of the appropriate indication of interpersonal relationships in language use in Japanese (Section 3.1). Afterwards, I will report on a small survey conducted on predicate usage and copula choice (Section 3.2) which aims to show that in Japanese interpersonal relationships need to be indexed.

3.1 Hirose's three-tier model

Prior to the development of his three-tier model, Hirose (2000) proposed that the speaker is deconstructed into the *public self* (the subject of public expression) and the *private self* (the subject of private expression). Public expression and private expression correspond to the communicative function and the thought-expressing function of language, respectively (Hirose 2000:1624). Thus, the private self construes a certain situation and forms their thought in their private expression, while the public self, facing an addressee, reports the construed situation to them, using public expressions. Since the public self interacts with others in the actual events of language use, interpersonal relationships, such as who they are speaking to, also need to be concerned. Hence, Hirose (2015) argues that language use composes the following three tiers: the situation construal tier, in which "the speaker as private self construes a situation, forming a thought about it" (p.123); the situation report tier, in which "the speaker as public self reports or communicates his construed situation to the addressee" (p.123); and the interpersonal relationship tier, in which "the speaker as public self construes and considers his interpersonal relationship with the addressee" (p.123). Hirose then claims that languages differ as to which speaker serves as the deictic center and how the three tiers of language use are unified. For example, English is characterized as a public-self-centered language, in which the public self in the situation report tier serves as the deictic center and situation report is unified with situation construe, as shown in Figure 2. In contrast, Japanese is considered a private-self-centered language, in which the private self in the situation construal tier is the deictic center and situation report is unified with the interpersonal relationship, as shown in Figure 3.

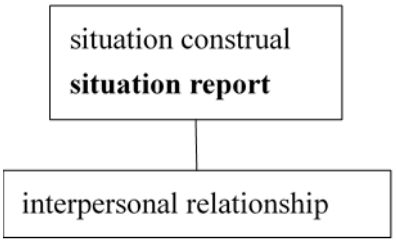


Figure 2: English as a public-self-centered language (based on Hirose 2015:124)

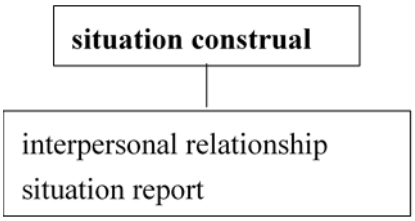


Figure 3: Japanese as a private-self-centered language (based on Hirose 2015:125)

In Figures 2 and 3, the tiers printed in boldface represent the location of the default deictic center in each language.[10]

Crucially, as shown in Figure 3, the interpersonal relationship tier is unified with the situation report tier in Japanese. Thus, when reporting a situation to someone in Japanese, "the speaker must always construe and consider his relationship with the addressee, defining himself and the addressee in terms of that relationship" (Hirose 2015:124–125). This is not the case in English, because situation report is independent of the interpersonal relationship tier in English. Consequently, a plain sentence with interpersonal relationship unmarked is used as a public expression in English, but not in Japanese. Therefore, Hirose suggests that while sentence (9a) has a performative function, as shown in (9b), the Japanese plain sentence in (10a) cannot have a similar performative function, as shown in (10b).[11] Instead, when talking to someone, appropriate marking of the speaker's "social or sociopsychological relationship with the addressee" (p. 122) is necessary, which is done by adding a SFP like *yo* 'I tell you' or turning the plain copula *da* to its polite or super-polite form (*desu* or *degozaimasu*), as indicated in (10c).

(9) a. Today is Saturday. (Hirose 2015:128)
b. I SAY TO YOU Today is Saturday. (Hirose 2015:128)

(10) a. Kyoo-wa doyoobi-da
today-TOP Saturday-COP
'Today is Saturday.' (Hirose 2015:128)
b. # I SAY TO YOU Kyoo-wa doyoobi-da (Hirose 2015:128)
c. Kyoo-wa doyoobi-dayo/desu/degozaimasu
today-TOP Saturday-COP.SFP/COP.POL/COP.Super POL.
'Today is Saturday.' (Hirose 2015:122)

Note in passing that the existence of an unmarked mode of situation report in English does not entail that the interpersonal tier does not exist in the language. Interpersonal relations can be optionally signaled, for instance by adding (optional) terms of address to an otherwise unmarked utterance (e.g., *Today is Saturday, Jane/madam/honey* (Hirose 2015:130)). As illustrated in Figure 2, however, the interpersonal relationship tier is only optionally added to the situation report unified with the situation construal tier in English. This has been discussed

[10] These are simplified forms of Hirose's (2015:124–125) original figures. For their full descriptions, see Hirose (2015).

[11] Hirose (2015:122) suggests that the plain form in (10a) can "rarely" be used except when one talks to oneself or to someone very close to one.

in terms of linguistic strategies, whereby linguistic selections index interpersonal stance and relational positioning, rather than encoding fixed politeness values (Locher and Graham 2010; Locher and Watts 2005). Goffman's (1967, 1974) classical studies stress the notion of frames, which cue expectations about what is going on (1974: 8) and the modulation of one's "ceremonial behavior" (1967: 77), i.e., *demeanor*. In more recent works, Locher and Watts (2005), for instance, regard (im)politeness as a matter of participants' evaluations negotiated against locally salient norms, i.e., relational work. Spencer-Oatey's (2005) rapport-management theory models how face, sociality rights and obligations, and interactional goals condition linguistic choices and their uptake across contexts and cultures. Crucially, these modulations are strategic options for building better relationship to others; hence, as Hirose (2015) claims, in English, the unmarked or plain sentence can often function as a default to address anyone unless local norms render relational work particularly salient.

With the characteristic of Japanese in mind, let us consider the example in (1), repeated here as (11):

(11) 1998 年から 2017 年まで住んでいたんだ (= (1))
1998-nen kara 2017-nen made <u>sundeita-nda</u>.
1998-year from 2017-year until live.PST-it.is.that
'I lived here from 1998 to 2017.'

The predicate *sundeita-nda* 'lived' is used in the plain forms. As suggested in (2), I would use the polite form *sundeita-ndessu* instead.

3.2 Survey on predicate-marking and copula choice

To confirm my intuition, I asked 43 Japanese undergraduate students (17 males and 26 females) what they would reply to a TV interviewer asking where their hometown is. The survey was conducted in May 2024 using Google Forms. The data were collected anonymously from the participants recruited in undergraduate courses I taught on a volunteer basis. Students were asked to select one of the following forms *da* (plain), *da-yo* (plain + SFP = friendly), *desu* (polite), *desu-yo* (polite + SFP), *su* (polite yet casual/frank), *degozaimasu* (super-polite), and other. Of the 43 respondents, 42 selected *desu*, the polite form, and one *su*, the polite yet casual/frank form. This result suggests that we perceive a TV interviewer as someone who is neither personally close nor so distant that deference is required.

One might ascribe the use of the plain form in (1) [= (11)] to the nature of translation and argue that the predicates need not to be in polite form because what is important in translation is to convey the correct propositional content. This may be true, and in fact a lot of non-native speeches are translated in plain forms (often with a somewhat literary tone) and subtitled. However, the translation in (1) is not only subtitled in text but dubbed over the original speech, incorporating paralinguistic features such as tones and laughter, as if the interviewee were speaking in Japanese. Besides, certain interpersonal characteristics, such as the expression of modality *n(o)da* 'it is that' and the SFPs *nee* 'you know' and *yo* 'I tell you' are used in this and subsequent utterances. The modality expression *nda*, the contracted form of *noda*, is considered to be categorized as "interpersonal mood that bears the relational function" in the sense of Noda

(1997), adding an exclamatory nuance.[12] In the video clip, following the narrator's introduction saying (in Japanese), "Phil, from Australia, came to visit his fried living in Japan after six years", he explains what brought him to Japan, beginning with the utterance in (1). That is, the modality expression in the translated speech is considered to be employed to signal that the explanatory utterance is directed to the interviewer (and viewers of the program). If the translation were intended to genuinely convey what is said, those expressions indicating how it is said would not be necessary, either. Thus, the coexistence of these interpersonal elements and the plain forms makes the utterance non-native. In other words, the unnaturalness of the translated speech in (1) results from the coexistence of the expressions indicating communicative intention *-nda* (as well as SFPs) and the plain copula *da*, the predicate for non-communicative intention.

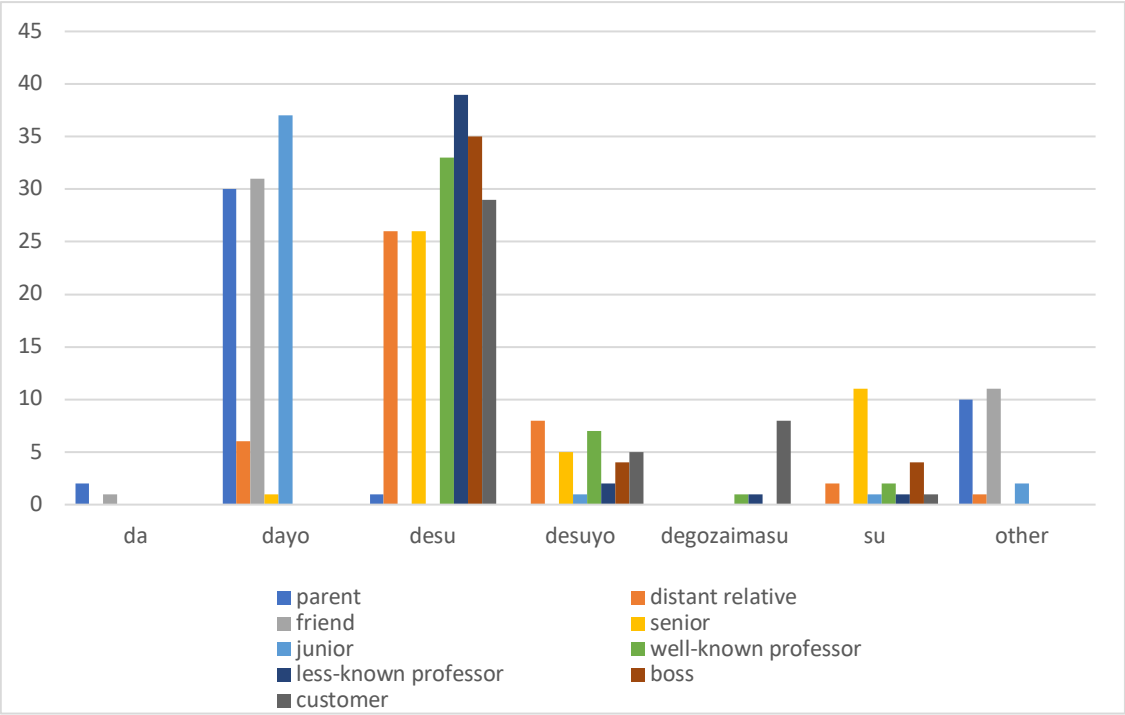
It has just been argued that predicates are rarely used in plain forms in Japanese due to the unification of situation report with interpersonal relationship (Hirose 2013, 2015). The polite form is the most appropriate form to use when talking to a TV interviewer (as 42 out of 43 respondents selected it). This indicates that predicates require appropriate marking that reflects the specific interpersonal relationship in the context.

In the survey, the respondents were also asked how they would respond if they were asked what day it was by each of the following individuals: (i) a parent; (ii) an uncle they meet after a long time, (iii) a friend, (iv) a senior member in their club, (v) a junior member in their club, (vi) a professor who knows them well, (vii) a professor who does not know them well, (viii) boss at their part-time job, (ix) a customer at their part-time job.[13] The respondents were asked to imagine replying that it was Saturday and to select one of the copula forms listed in the right-most column of Table 2 to form a predicate, like *Doyoobi-da* 'it is Saturday [plain form]'.[14]

[12] Noda (1997) observes that *noda* represents the "interpersonal" and "propositional" moods, each of which is further subcategorized into the one that bears the function of relating utterances or situations and the one that does not. It is beyond the scope of this article to describe their differences in detail.

[13] The terms "junior" and "senior" are used here not to mean the third year and fourth year of university, but to mean elder and younger, respectively. The Japanese society is known to emphasize a vertical hierarchy based on seniority rules; therefore, the younger are basically expected to speak politely and respectfully to the elder. Generally, this hierarchy is observed particularly rigidly in athletic teams at universities and schools.

[14] Simply to observe the general tendencies of relations between each form and interpersonal relationship, the results are presented without performing statistical analysis.



	parent	distant relative	friend	senior	junior	well-known professor	less-known professor	boss	customer
da (plain)	2	0	1	0	0	0	0	0	0
dayo (plain + SFP [friendly])	30	6	31	1	37	0	0	0	0
desu (polite)	1	26	0	26	0	33	39	35	29
desuyo (polite + SFP)	0	8	0	5	1	7	2	4	5
degozaimasu (superpolite)	0	0	0	0	0	1	1	0	8
su (polite yet frank/casual)	0	2	0	11	1	2	1	4	1
other	10	1	11	0	2	0	0	0	0

Table 2: Results of the survey on the linguistic forms used when speaking to specific interlocutors

The results suggest that each copula form exhibits certain characteristics. First, as expected, the plain form *da* is not used virtually to any interlocutors in communication. Second, most responses clustered around *dayo* (the plain form followed by the SFP *yo*) and *desu* (the polite form). The former is generally considered the most appropriate form to use when speaking to close people, while the latter is preferred when addressing distant people. Note in passing that the super-polite form *degozaimasu* is not commonly used.[15] As far as the types of addressees examined in the present survey are concerned, the use of the super-polite form *degozaimasu* basically seems to be restricted to situations involving customer service or other commercial interactions. The choice between the super-polite form and the polite form is not solely characterized by the degree of deference, but by the sociopsychological distance between the interlocutors.

Third, as a combination of the polite form *desu* and the SFP *yo* 'I tell you', it is not surprising that the form *desuyo* is mostly directed to the individuals to whom the polite form *desu* is used—only with one exception in (v) a junior member in their club. How, then, should *desu* be distinguished from *desuyo*? From the perspective of the "territory of information", Kamio (1995) argues that *yo* is used when the information belongs to the speaker's territory, whereas *ne* 'you know' is used when the information falls in the addressee's territory. According to Inoue (1997), the SFP *yo* "describes the situational context surrounding the speaker and the listener by referring to the truth of the proposition" (p.65, my translation), while connoting "impositional attitude" (ibid.:67, n.4). Thus, compared with *desu*, *desuyo* may make the utterance sound more like a revelation or warning. In fact, Inoue observes that the use of *yo* makes the utterance odd in contexts where the speaker expects the addressee simply listen to and understand the content, as in (12):

(12) [context: the speaker is simply explaining the procedure of saving a file]
… mazu "fairu" to-iu tokoro-o kurikkusimasu (??yo).
first file COMP-say place-ACC click.POL (I.tell.you)
'First, please click on the "file" tab.' (Inoue 1997:65)

In the context used in the survey, the respondents were merely expected to inform the individuals of the day. Therefore, using *desuyo* would go beyond the expectation, indicating that the individuals should have known that it was Saturday.

Fourth, the polite yet frank/casual form *su* is typically used when speaking to senior members in their club (= (iv) in Table 2).[16] Although its meaning is indicated as "polite yet frank/casual", its precise meaning is difficult to define. The

[15] Although one respondent—a student who took one of my courses—claimed to use the super-polite form *degozaimasu* when addressing a professor, I have neither heard it directed to me nor used it myself when I was a student. It could be used in the context, but presumably, despite the speaker's superficial civility, the utterance would sound insolent.

[16] The *su*-form may be (stereo)typically associated with speech by young males directed toward senior students (in schools and universities) or senior colleagues (in workplaces). In fact, all 11 respondents who reported using *su* when speaking to senior members in their clubs were male students aged 18–20. However, Nakamura (2020) observes that *su* is also used in female speech as a marker of liberation from traditional female identity, which has often been understood in terms of male hegemony (p. 207).

form has been analyzed as a contracted variant of the polite form *desu* (Nakamura 2020:16). Nakamura investigates a thirty-minute discourse between male college students (two junior members and one senior member in a university *kendo* club). She observes that *su*, as well as *desu*, is used exclusively by the junior members when speaking to the senior member, but neither form is used in the senior member's speech directed to the juniors or in interactions between the junior members (p.55). Furthermore, Nakamura identifies certain functions of *su* distinguishing it from *desu*. The *su* form can convey intimacy without sacrificing the politeness toward the addressee, thus serving an affiliative function (pp.56–65).

Finally, there are a considerable number of responses in the "other" category. Most of them (23 out of 24 responses) consisted only of the noun *doyoo(bi)* 'Satur(day)', containing no copula.[17] Such responses may seem to completely lack interpersonal markers, but they are in fact fragmentary answers, used specifically to respond to a particular question. Fragmentary answers are also "very often" observed in "responses to questions" (Huddleston and Pullum 2002:1542) in English, a public-self-centered language (e.g., *What day is it today? — Saturday.*). The result also suggests that these fragmentary answers are restricted to those who are close to the speakers such as parents and friends.[18]

4 Interpersonal Relationships as part of constructional knowledge

In the previous section, we have observed that each of the six copula forms studied exhibits unique interpersonal restrictions in use. Although they are truth-conditionally equivalent, they differ in terms of interpersonal relationship; their differences should be linked directly to their phonological forms. The knowledge is described using Jackendoff and Audring's (2020) Relational Morphology templates, with the interpersonal relationship component (abbreviated as IR). Relational Morphology enables us to simultaneously represent the word-internal relations between the components via "interface links" and word-to-word relations via "relational links" using the same indices.

	da	***dayo***	***desu***
semantics	predication$_1$	predication$_1$ + new info$_2$	predication$_1$
morphosyntax	copula$_1$	copula$_1$-SFP$_2$	copula$_1$
phonology	/da/$_1$	/da$_1$ yo$_2$/$_3$	/desu/$_4$
IR		close$_3$	polite$_4$

[17] Another response in this category was *doyoobi-desho!* 'Saturday, huh?'. I do not analyze it in detail, but it should be noted that this can also be used only among close relationship.

[18] In this regard, Hirose (2013:24) observes that as shown in (ib), subject omission in English "does not go well with respectful forms of address such as *sir* and *Professor Brown*".

(i) a. I hope you like it, {sir/Professor Brown}. (Hirose 2013:24)
b. ? Hope you like it, {sir/Professor Brown}. (ibid.)

Omission phenomena in general may be cross-linguistically linked to the degree of interpersonal closeness between interlocutors. However, more detailed investigations are necessary to compare different phenomena in different languages.

	desuyo	***degozaimasu***	***su***
semantics	predication$_1$ + new info$_2$	predication$_1$	predication$_1$
morpho-syntax	copula$_1$-SFP$_2$	copula$_1$	copula$_1$
phonology	/desu$_4$ yo$_2$/$_5$	/degozaimasu/$_6$	/su/$_7$
IR	[polite$_4$ + impositional]$_5$	super-polite$_6$	[polite + casual]$_7$

The interface links notated by coindexes 3, 4, 5, 6, and 7 between the phonological and IR components in each word indicate that the different phonological forms are only differentiated in terms of interpersonal relationship. Given the basic representation of the plain copula *da* using the interface links notated by index 1, the relational links coindexed as 1 across the semantics and morphosyntax of these words show that they belong to the same grammatical category and share the same truth-conditional meaning as the plain form *da*—bearing the basic predicative function as the copula.

The SFP *yo* is categorially distinguished from the copula and has its own function of presenting new information to the listener, or the information belonging to "the speaker's territory" (Kamio 1994). The phonological and morphosyntactic form of *dayo* can be analyzed as the composition of *da* and *yo*, corresponding to the compositional semantic meaning. However, the entire expression also carries the irreducible IR function, whereby it can be used with close individuals, but not with distant ones. This is notated by index 3. Since the unmarked copula *da* is typically used as private expression (see footnote 11) but since the polite form *desu* inevitably indicates distance, *dayo* is the appropriate form to indicate both publicness and intimacy. The IR of *desuyo*, unlike that of *dayo*, can be compositionally computed based on its components *desu* and *yo*, distinguishing *desuyo* from *dayo* and *desu*. The "impositional" connotation, derived from the presence of *yo* (Inoue 1997), is concerned with the speaker's attitude toward the listener, and hence, represented in the IR component.

Relational Morphology conforms to the usage-based model (e.g., Tomasello 2003, Langacker 2009, Diessel 2019), in which constructions (including words) are assumed to be acquired based on actual usage events. Jackendoff and Audring (2020:219) note that "learning a word requires building the structures of each component and establishing the interface links between them [… and] components of a lexical item do not have to be learned at once". That is, they may be learned independently and then are connected to each other via interface links. Thus, the IR information of individual words is acquired based on usage events through actual interactions with others. As illustrated in Figure 3, in Japanese, the unification of the situation report tier and interpersonal relationship tier is independent of the situation construal tier. Given that the plain variant *da* is typically used at the situation construal level (Hirose 2015), it is natural for the IR information to remain unmarked. In contrast, other variants are used for situation report, which is unified with interpersonal relationship. Therefore, it follows that each variant is learned in association with specific interpersonal relationships through actual language use. Although replacing copulas and deleting or adding SPFs do not affect truth-conditional meaning (as notated by coindex 1 in semantics), fluent speakers distinguish among them depending on their interlocutors and the context. To put it the other way around, non-fluent speakers of Japanese do not have good IR information in their constructional representation.

The impositional connotation of *desuyo* can be understood as resulting from the addition of the SFP *yo* to the polite form *desu*, which per se functions as a marker of public expression. In this regard, Ikarashi's (2013) discussion of how speakers mark informational superiority in English is insightful. He argues that the performative expression *I tell you* in English indicates the speaker's informational superiority. Although utterances with the performative clause *I tell you* are generally unnatural in English, as shown in (13), they may be accepted in certain contexts, as demonstrated in (14).

(13) *I tell you that it is so.
(Brown and Levinson 1987:190, as cited in Ikarashi 2013:112)

(14) Lightyear: They are a trillium-carbonic alloy, and I can fly.
Woody: No, you can't.
Lightyear: Yes, I can.
Woody: You can't.
Lightyear: Can.
Woody: Can't. Can't. Can't!
Lightyear: *I tell you, I could fly around this room with my eyes closed!*
Woody: Okay, then, Mr. Light Beer, prove it.
(*Toy Story*, as cited in Ikarashi 2018:113, italics in original)

From the perspective of the three-tier model, communication in English is regarded as the "reciprocal information sharing" (Hirose 2013:25) based on the symmetrical relationship between the interlocutors. However, "the use of *I tell you* changes [the] linguistically symmetrical relationship in default case to a linguistically asymmetrical one" (Ikarashi 2013:118), and changes the mode of communication to "one-sided information giving" (ibid.). Ikarashi (2013:119) observes that in (14), by saying *I tell you*, Lightyear unilaterally gives the information that he can fly and attempts to impose it on Woody, as the communicative goal of the target utterance is to make Woody accept the content in question. Recall that in English, a plain sentence is the default choice when the interlocutors reciprocally share information (Hirose 2013). In communication, therefore, without any specific context, the addition of *I tell you* to a plain sentence disrupts the default symmetrical speaker-listener relationship (e.g., (13)). Likewise, *desuyo*, which combines the markers of public expression (*desu* and *yo*), conveys the speaker's attitude of "I know—and I can tell—what you do not know", which may sound imposing on the listener in the survey context.

The super-polite form *degozaimasu* does convey politeness, but as discussed earlier in this section, it is distinguished from the polite form *desu* not solely on the degree of deference toward the addressee; the distinction is based on not only social but also psychological distancing to the addressees. As a result, the use of *degozaimasu* is more restricted than that of *desu*. In other words, the "politeness" conveyed by *degozaimasu* should be distinguished from that of *desu*, as indicated by their different indices, 4 and 6, respectively, in their IR representations. The *su* form also carries its own politeness with a casual tone, as notated by coindex 7[19]

[19] If the *su* form is a contracted form of *desu* (Nakamura 2020:16), it may inherit its politeness from *desu* and thus be assigned coindex 4. At the same time, the phonological representations of /desu/$_4$ and /su/$_7$ should be specified as /*de*su/$_4$ and /* *su/$_{4,7}$, respectively, where the stars indicate that the "two phonological strings are the same except for the materials between [the] stars" (Jackendoff and Audring 2020:xiii). Furthermore, *desu* is further analyzable into *de-* (the

In summary, various forms of the Japanese copula reflect specific interpersonal relationships, despite sharing the same basic meaning. Therefore, IR information must be learned and represented along with the phonological and morphosyntactic form and the semantic meaning in each lexical item. Their forms, meanings, as well as IR information, are represented integrally as constructions in Relational Morphology terms (Jackendoff and Audring 2020).

5 *Yakuwarigo* translation as multimodal constructs

As proposed in Section 4, fluent Japanese speakers store various predicate forms, SFPs, and their combinations, corresponding to one semantic meaning but differing in terms of interpersonal relationships. Nevertheless, forms that do not reflect the appropriate interpersonal relationship are often used for translated speech. Taking *register* as a system of socially recognized signs with established pragmatic values that index roles and relationships in interaction (Agha 2007: 80), we can identify register clashes in the translated speeches discussed in Sections 1 and 2. In Section 2, such register-clash expressions were characterized as inviting a coerced interpretation as if they were uttered by a (perceived) non-fluent speaker of Japanese—i.e., a *yakuwarigo* reading. It was also pointed out that the figure of a non-Japanese man plays a significant role in viewing the utterance in (1) as *yakuwarigo*. In this section, it is argued that translations with *yakuwarigo* seen in TV programs, as in Figure 1, can be considered video-text multimodal constructs, the form of video and text information paired with certain meanings.

Although studies on multimodal construct(ion)s typically focuses on grammatical constructions used together with kinesic information such as gesture, facial expressions, posture (e.g., Zima and Bergs 2017, Hoffmann 2021), Dancygier and Vandelanotte (2017) analyze image macro internet memes as image-text multimodal constructions. They argue that the notion of frame metonymy plays a crucial role in the comprehension of image macro memes. That is, limited formal means used in memes, such as iconic image and quotation marks, serving as characteristic parts of a frame, metonymically call up whole frames in detail (Dancygier and Vandelanotte 2017:565–567). These accessible formal cues allow one to understand the detailed but unexpressed information of the entire frame, against which internet memes are comprehended.

Analog to image macro memes, translations in TV programs can be viewed as multimodal constructs. In Figure 1, the unnatural Japanese utterance in (1) is dubbed over the lip-sync video of a typical non-Japanese person; then, the viewers may mistakenly think that the speaker is speaking in Japanese. Thus, the entire information that one can get from the interview video clip in Figure 1 is represented in (15):

(15) The lip-sync video of a non-Japanese man responding to a street interview with dubbed audio makes it look like he's saying in Japanese, "1998-nen kara 2017-nen made sundeita-nda (I lived here from 1998 to 2017)", using the plain form of the copula.

---

continuative form of *da*) and *su* (an auxiliary verb expressing politeness); *degozaimasu* into *de-* (the continuative form of *da*) and *gozaimasu* (the super-polite form of the verb *aru* 'exist'). However, such finer-grained representations are not necessary for the purposes of the present study. The representations adopted here are therefore simplified to avoid unnecessary complexity.

Upon seeing the video, one may invoke a frame, which may be called the "interview frame", in which one (= the interviewee) is interviewed by a TV crew (= the interviewer) while being video recorded and interacting. The term "frame" is used here in the sense of frame semantics (e.g., Fillmore 1982) and is understood as a "script-like conceptual structure" (Ruppenhoffer et al. 2010: 1). Frames are generally *evoked* by "lexical and grammatical material observable in the text" (Fillmore 1982: 124) and *invoked* in the translator to "assign coherence to a text" (ibid.:124). In the case of the interview video clip in Figure 1, it is the scene in the video itself—as a constituent of the multimodal construct—that directly evokes the interview frame, rather than any specific linguistic elements used therein. As shown in Section 2, the majority of fluent Japanese speakers (42 out of 43 respondents) claims to use the polite form when responding to a TV interview, which suggests that an utterance like (2) is more or less expected in this frame. The use of the polite form (or other interactive forms) in interaction between the interviewer and interviewee composites part of the entire interview frame, which one can metonymically call up from the scene of the interview, as suggested by Dancygier and Vandelanotte (2017). Once the interview frame is invoked, proficient Japanese viewers typically expect the interviewee to use the polite form. However, they are met with the unexpected use of the plain form, resulting in a noticeable clash of register. The overall mechanism of frame evocation/invocation is illustrated in Figure 4.

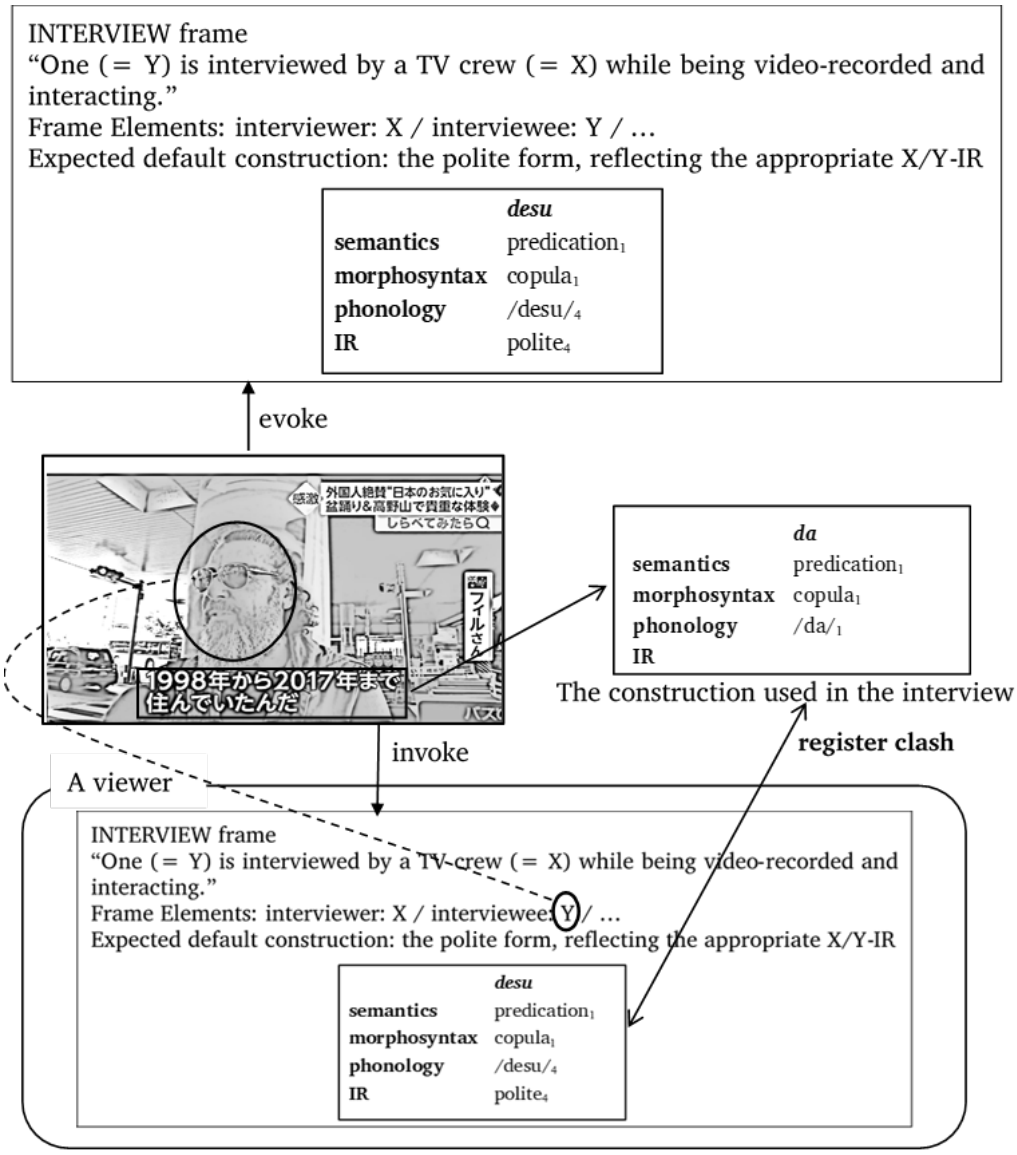


Figure 4: Evocation/invocation of the interview frame and the emergence of register clash

The solid arrows labeled "evoke" and "invoke" indicate that the scene in the video clip directly evokes the interview frame, which is simultaneously invoked by the viewer. As mentioned earlier, the frame designates as the expected default speech style the polite form, which is assumed to reflect the appropriate interpersonal relationship between the interviewer and interviewee. The broken line indicates the mapping of the image of the interviewee in the video clip onto the frame. The knowledge of the frame allows the viewers to notice that the registers clash in the frame and the actual utterance in the video clip, as indicated by the bidirectional arrow labeled "register clash".

As discussed in Section 2, the noticeable register clash is resolved by coercing the utterance into *yakuwarigo*. Recall that Kinsui's (2003:205) definition of *yakuwarigo* involves two directions of matching between a speech style and a speaker. It is the second direction of matching that is relevant here; namely, from person to speech style. That is, due to his *apparent* lack of fluency in Japanese,

he would *likely* make such a deviation from the expected norm, thus allowing the viewers to interpret the register clash as *yakuwarigo*.

In frame semantic terms, by mapping the visual image of the non-native Japanese speaker onto the interviewee role of the invoked interview frame (as indicated by the broken line in Figure 4), the default expectation of using the polite form may be overridden. Note that this mapping process involves the conceptual blending of the interview frame and the interview video clip. It is this blending that allows for the coerced interpretation of the register clash construction as *yakuwarigo*, thus resolving the mismatch in register. The essence of conceptual blending is "to constructing a partial match between two input mental spaces, to project selectively from those inputs into a novel 'blended' mental space, which then dynamically develops emergent structure" (Fauconnier and Turner 2002:2). By blending the interview frame (input space 1) and the figure of a non-native Japanese speaker (input space 2), a modified version of the interview frame is created (blended space), in which the interviewee role and the construction he uses are inherited from input space 2. Then, the non-native Japanese speaker is mapped onto the interviewee role and the expected default construction is overridden. The blended space thus projected is represented in Figure 5, where the elements inherited from input space 2 are underlined and the emergent structures are boldfaced. Specifically, the apparent register-clash expression acquires the *yakuwarigo* reading due to the blending of the interview frame and the figure of a non-native Japanese speaker. Thus, added to the constructional representation of the plain copula *da* is the social meaning tier.[20]

INTERVIEW frame (modified)
"One (= Y) is interviewed by a TV crew (= X) while being video-recorded and interacting."
Frame Elements: interviewer: a Japanese TV crew / interviewee: a non-native Japanese speaker / ...
Construction: the plain form, not reflecting the appropriate X/Y-IR

	da
semantics	predication$_1$
morphosyntax	copula$_1$
phonology	/da/$_{1,8}$
IR	____
social meaning	***yakuwarigo***$_8$

Interpretation of the register-clash construction:
The non-native speaker of Japanese — due to his *apparent* lack of fluency in Japanese — would likely use it = *yakuwarigo*

Figure 5: *Yakuwarigo* construction as a multimodal blend of the interview frame and the perceived non-fluent interviewee

[20] The term *social meaning* is understood here in Weissman's (2025) *reverse* sense: "relevant pre-existing beliefs about a speaker's beliefs should influence how a hearer interprets that speaker's utterances" (p. 142). See also Hansen and Terkourafi (2023) for a relevant discussion on what they call "Hearer's Meaning". For a general overview of social meaning, see Beltrama (2020).

The mismatch in register is thus resolved by the viewer's interpreting the construction in such a way that the incongruity is neutralized within the context. Specifically, the extended Relational Morphology representation shown in Figure 5 introduces a social meaning tier, which allows yakuwarigo to be modeled as an integral component of the construction, rather than merely as an interpretive effect. This integrated representation establishes the status of yakuwarigo within the linguistic system: although the copula da used in the translated speech in (1) is semantically, morphosyntactically and phonologically identical to the plain copula da in Japanese, as notated by coindex 1, the two are distinguished in terms of social meaning, as notated by index 8.

This neutralization is essentially explained by the override principle proposed by Michaelish (2004: 25), according to which a lexical item that is semantically incompatible with its morphosyntactic context conforms to the meaning of the structure in which it is embedded. As the semantic value of a lexical item conforms to the meaning of the structure, when combined with the video clip in Figure 1, the non-native speaker of Japanese is mapped onto the interviewee role of the frame, and the plain form used in the frame may be interpreted in accordance with this modified version of the interview frame. Thus, in this blending, the visual image of the interviewee clearly plays a significant role in shaping the viewers' stereotypical perception of them. The viewers reframe the interview scene so that the construction used in the interview frame no longer appears to contain a register clash; instead, it may be interpreted as the construction that the interviewee—though not known to the viewers—would likely use and is thus coerced into *yakuwarigo*.

6 Concluding remarks

In this article, the source of unnaturally-sounding Japanese utterances like (1) and (3) was identified as their *yakuwarigo* reading (Kinsui 2003). Previous studies have attributed *yakuwarigo* either to translated speech by famous athletes with well-established and widely recognized personalities, such as superstars or disadvantaged athletes (Ota 2009, 2017), or to the words spoken by fictional characters in plays, dramas, comics or other media (Yoda 2007, 2011). Unlike such speakers with established or fictional identities, viewers typically have no prior knowledge of ordinary people who appear on screen, such as individuals who happen to be interviewed on the street. This study argued that, based on the on-screen identity of non-native speakers of Japanese, viewers may infer that they would likely use the construction with inappropriate indication of interpersonal relationship regardless of the context. To support this analysis, it was first discussed that fluent Japanese speakers possess knowledge that non-fluent speakers often lack; namely, the appropriate representation of interpersonal relationship. In Japanese, such representations are inseparable from situation report (Hirose 2013, 2015, 2024) and are stored as part of constructional knowledge. This was described integrally using the framework of Relational Morphology (Jackendoff and Audring 2020). Relational Morphology templates not only allow for the description of different interpersonal relationships encoded in individual (word) constructions, but also enable us to capture the relations across the constructions that differ in phonological forms yet share morphosyntactic specification and truth conditional meaning. Furthermore, the relations among these components (i.e., interface links) are assumed not to be necessarily acquired all

at once but acquired independently and related gradually through repeated exposure to actual usage events (Jackendoff and Audiring 2020:219).

Particularly, interpersonal relationship is learned through socialization and enculturation (for language socialization and enculturation, see, for example, Schieffelin and Ochs 1987). Kinsui (2003:v–vi) reports that "almost 100% of native speakers of Japanese who grew up in Japan would never return incorrect answers [in the matching task of those *yakuwarigo* listed in (4a–h) and their corresponding speakers]" (my translation), while noting that the same task may be more difficult for learners of Japanese as a foreign language. The contrast between native speakers and learners suggests that competence in *yakuwarigo* is not innate but gradually acquired through exposure to the Japanese speech community. This interpretation is consistent with usage-based constructionist approaches.

The significant role that the visual information from the video plays in understanding unnatural Japanese superimposed on the laypeople's utterances as *yakuwarigo* was also emphasized. A survey revealed that fluent Japanese speakers typically use polite forms when responding to a TV interview, which constitutes the interview frame. By invoking this frame, viewers activate the expected default construction as well via frame metonymy (cf. Dancygier and Vandelanotte. 2017). The apparent mismatch between the construction expected in the frame (the polite form) and the actual utterance (the plain form) was claimed to be resolved by coercing the interpretation of the use of the plain form into *yakuwarigo* in accordance with the context where it is used. Specifically, by blending the interview frame and the interview video clip, the non-native Japanese speaker is mapped onto the interviewee role and the utterance in unnatural Japanese overrides the default construction expected in the frame. The frame is thus modified so that the mismatch can be resolved. In accordance with the modified frame, the viewers are coerced to interpret the unnatural Japanese as *yakuwarigo*. Therefore, the present study claimed that the *yakwuarigo* (in the sense of definition (ii)) is best characterized as a multimodal construction that integrates image/video and text.

To conclude, this study has argued that *yakuwarigo* can be systematically modeled as a social-meaning component within multimodal constructional representations, integrating insights from enregisterment (Agha 2007) and indexicality (Silverstein 2003), rather than treating it as a purely interpretive or stylistic phenomenon. Through this analysis, the kind of sociolinguistic knowledge that should be integrated into constructional templates was elucidated. The interpersonal relationship information is inseparably represented for Japanese (Hirose 2013, 2015, 2024). The additional social meaning tier is also necessary for interpreting an (unidiomatic) utterance as *yakuwarigo*. Translated speeches broadcast on TV may sound Japanese, using Japanese lexical items and grammar. However, since they are *virtual*, as opposed to *natural*, language (Kinsui 2003), the appropriate representation of the social meaning within the constructional temperate is necessary.

That said, this study was based on a small set of illustrative broadcast cases and focused on a limited range copula forms and interview video clips in Japanese. Taking this as an exploratory case study, the present analysis opens up a broader research agenda for investigating how social meaning is integrated into constructional representations across languages and genres. Intriguing questions left open include how interpersonal relationship is represented in other constructions sensitive to interpersonal relationship—such as various first- and second-

person expressions (e.g., *watashi*, *ore*, *boku*, etc., for *I*; *kimi*, *anata*, *omae*, etc., for *you*), and different verb forms (e.g., the plain verb *iu* 'say' appears as *iimasu* (polite), *iwareru* (systematic subject honorific), *ossharu* (idiomatic subject honorific), *moosu* (humble))—and how these multiple constructions interact with each other in a single utterance. For example, the polite form of copula *desu* appears to align more naturally with the first-person expression *watashi* (having a formal and calm tone) than with *ore* (typically used by boys and men with an informal tone) (see footnote 8).[21]

Moreover, the private-/public-self-centeredness of other languages and their translations need to be examined. Although European languages other than English are generally considered public-self-centered, Wada (2008) identifies the degree of public-self-centeredness as forming the inclination in the order of English > Swedish > Spanish > Dutch > French > German, where English is more public-self-centered than German. Kim (2022) points out that Korean, like Japanese, is a private-self-centered language. Investigating Modalpartikel (modal particles) in German through the lens of the three-tier model of language use, Oya (2022) and Okamoto (2022) suggests that although German and English are both public-self-centered, German is less strongly so than English. Although a detailed review of these studies are not provided here, what is important is that the public-/private-self-centeredness forms a continuum, on which English and Japanese are located at opposite ends on the spectrum. Korean is close to Japanese. Although German is not as close as Korean, it is relatively closer to Japanese compared with English in terms of interpersonal relationship encoding. A question then arises as to whether speeches by Korean and German ordinary people are translated into Japanese differently from those by English speakers. If so, what do the translations sound like, and are they also conceived of as *yakuwarigo*? Such questions are also left open for future research.

Acknowledgments

I thank two anonymous reviewers and the editors for their constructive criticism and helpful comments. Needless to say, any remaining errors and shortcomings are mine alone.

Conflict of interest statement

The author declares none.

[21] I conducted an advanced Google search on October 10, 2025 within the x.com domain limiting the language to Japanese for the exact phrases *watashi-wa totemo ureshii-desu* and *ore-wa totemo ureshii-desu*, both of which are translated as "I'm very happy". The words *watashi*, *ore*, and *ureshii* are written in *kanji* characters, as 私はとても嬉しいです and 俺はとても嬉しいです, respectively. The only difference between them lies in the first-person expressions *watashi* (私) and *ore* (俺). The former returns 2310 results, while the latter returns only 8.

License

This work is licensed under a "CC BY 4.0" license.

References

Agha, Asif. 2007. *Language and Social Relations*. Cambridge: Cambridge University Press.

Beltrama, Andrea. 2020. Social meaning in semantics and pragmatics. *Language and Linguistics Compass* 14.9. https://doi.org/10.1111/lnc3.12398

Brown, Penelope and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Dancygier, Barbara and Lieven Vandelanotte. 2017. Internet memes as multimodal constructions. *Cognitive Linguistics* 28.3: 565–598. https://doi.org/10.1515/cog-2017-0074

Diessell, Holger. 2019. Usage-based construction grammar. In Ewa Dąbrowska and Dagmar Divjak (eds.), *Cognitive Linguistics: A Survey of Linguistic Subfields*, 50–80, Berlin/Boston: De Gruyter Mouton.

Fauconnier, Gilles and Mark Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Fillmore, Charles J. 1975. *Santa Cruz Lectures on Deixis 1971*. Bloomington: Indiana University Linguistics Club.

Fillmore, Charles J. 1982. Frame Semantics. In the Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, 111–137, Seoul: Hanshin.

Goffman, Erving. 1967. *Interactional Ritual: Essays on Face-to-face Behavior*. Garden City: Anchor Books.

Goffman, Erving (1974). *Frame Analysis. An Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.

Hansen, Maj-Britt Mosegaard and Marina Terkourafi. 2023. We need to talk about Hearer's Meaning! *Journal of Pragmatics* 208: 99–114. https://doi.org/10.1016/j.pragma.2023.02.015.

Halliday, Michael A. K. 1978. *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Hirose, Yukio. 2000. Public and private self as two aspects of the speaker: A contrastive study of Japanese and English. *Journal of Pragmatics* 32.11: 1623–1656. https://doi.org/10.1016/S0378-2166(99)00111-3

Hirose, Yukio. 2013. Deconstruction of the speaker and the three-tier model of language use. *Tsukuba English Studies* 32: 1–28.

Hirose, Yukio. 2015. An overview of the three-tier model of language use. *English Linguistics* 32.1: 120–138.

Hirose, Yukio. 2024. *Jibun-no Gengogaku: Kotoba-no Shutaisee-no Arika*. Tokyo: Kaitakusha.

Hoffmann, Thomas. 2021. Multimodal construction grammar: From multimodal constructs to multimodal constructions. In Xu Wen and John R. Taylor (eds.), *The Routledge Handbook of Cognitive Linguistics*, 78–92, New York/London: Tylor and Francis.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Ikarashi, Keita. 2013. The performative clause *I tell you*, interpersonal relationship, and informational superiority. *Tsukuba English Studies* 32: 111–126.
Inoue, Masaru. 1997. Moshi moshi, kippu o otosare-mashita-yo. *Gekkan Gengo* 26.2: 62–67.
Jackendoff, Ray and Jenny Audring. 2020. *The Texture of the Lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.
Kamio, Akiko. 1995. Territory of information: The case of Japanese. *Journal of Pragmatics* 21.1: 67–100. https://doi.org/10.1016/0378-2166(94)90047-7
Kim, Sunmi. 2022. Kan-Nichi shuujoshi to shuuketsugobi no joohoo-kanri-kinoo to hanashite/kikite tono sookansei, soshite sansoo-moderu-no shiza. In Yukio Hirose, Masaharu Shimada, Naoaki Wada, and Akiko Nagano (eds.) *Hikaku/Taishoo Gengogaku no Aratana Tenkai: Sansoo-moderu-niyoru Hirogari to Fukamari*, 190–210, Tokyo: Kaitakusha.
Kinsui, Satoshi. 2003. *Vaacharu Nihongo: Yakuwarigo-no Nazo*. Tokyo: Iwanami Shoten.
Kinsui, Satoshi. 2007. Kindai manga-no gengo. In Satoshi Kinsui (ed.), *Yakuwarigo-no Chihei*, 97–108, Tokyo: Kurosio.
Langacker, Ronald W. 2009. A dynamic view of usage and language acquisition. *Cognitive Linguistics* 20, 627-640. https://doi.org/10.1515/cogl.2009.027
Locher, Miriam A. and Richard J. Watts 2005. Politeness theory and relational work. *Journal of Politeness Research* 1: 9–33.
Michaelis, Laura. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15.1: 1–67. https://doi.org/ 0.1515/cogl.2004.001
Nakamura, Momoko. 2020. *Shin-Keigo "Mazi Yabai-ssu": Shakaigengogaku-no Kanten-kara*. Tokyo: Hakutakusha.
Noda, Harumi 1997. *No(da) no Kinoo*. Tokyo: Kurosio.
Okamoto, Junji. 2022. Sansoo-moderu-de shintaishi-no shiyoo-o setsumei-suru: Soogo-kooi-soo-no doonyuu. In Yukio Hirose, Masaharu Shimada, Naoaki Wada, and Akiko Nagano (eds.) *Hikaku/Taishoo Gengogaku no Aratana Tenkai: Sansoo-moderu-niyoru Hirogari to Fukamari*, 165–189, Tokyo: Kaitakusha.
Ota, Makie. 2009. Usaion Boruto no "I"-wa, naze "ore"-to yakusareru-noka: Supootsu hoosoo-no "yakuwarigo". *Hoosoo Kenkyuu to Choosa* 59(3): 59–73.
Ota, Makie. 2017. Saikoo oinpikku hoosoo-no "yakuwarigo": "Nihonjin senshu-o shujinkoo-to-shita'monogatari'" to iu shiten-kara. *Hoosoo Kenkyuu to Choosa* 67(3): 26–45.
Oya, Toshiaki. 2022. Taiken-wahoo + shintaishi *ja* = shin-nai hatsuwa: Doitsu-go niokeru kooteki-jiko to shiteki-jiko. In Yukio Hirose, Masaharu Shimada, Naoaki Wada, and Akiko Nagano (eds.) *Hikaku/Taishoo Gengogaku no Aratana Tenkai: Sansoo-moderu-niyoru Hirogari to Fukamari*, 140–164, Tokyo: Kaitakusha.
Ruppenhoffer Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. 2010. *FrameNet II: Extended Theory and Practice*. Berkeley: International Computer Science Institute.
Schieffelin, Bambi B. and Ochs, Elinor. 1987. *Language Socialization across Cultures*. Cambridge: Cambridge University Press,
Shogakukan's Seisenban Nihon Kokugo Daijiten. 2006. Tokyo: Shougakukan.
Silverstein, Michael. 2003. Indexical Order and the Dialectics of Sociolinguistic Life. *Language and Communication* 23.3–4: 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2.

Spencer-Oatey, Helen. 2005. (Im)Politeness, face and perceptions of rapport: Unpackaging their bases and interrelationships. *Journal of Politeness Research-language Behaviour Culture* 1: 95-119. https://doi.org/10.1515/jplr.2005.1.1.95

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Wada, Naoaki. 2008. Kootekijiko-chuushinsei-no doai to seeooshogo-no hoo/jisei genshoo-no sooi. In Yuichi Mori, Yoshiki Nishimura, Susumu Yamada, and Mitsuaki Yoneyama (eds.) *Kotoba no Dainamizumu*, 277–294, Tokyo: Kurosio.

Weissman, Benjamin. 2025. Social meaning in reverse: Expectations of English role noun use based on speaker identity. *Journal of Pragmatics* 242, 141–155. https://doi.org/10.1016/j.pragma.2025.04.008.

Yamaguchi, Haruhiko. 2007. Yakuwarigo-no kobetsusei to fuhensei: Nichi-Ei-no taishoo-o tooshite. In Satoshi Kinsui (ed.), *Yakuwarigo-no Chihei*, 9–26, Tokyo: Kurosio.

Yoda, Megumi. 2007. <Seeyoojin-go> "Oo, Romio!" no bunkee: Sono kakuritsu to fukyuu. In Satoshi Kinsui (ed.), *Yakuwarigo-no Chihei*, 159–178, Tokyo: Kurosio.

Yoda, Megumi. 2011. Yakuwarigo toshiteno katakoto Nihongo: Seeyoojin kyarakuta-o chuushin-ni. In Satoshi Kinsui (ed.), *Yakuwarigo Kenkyuu-no Hatten*, 213–248, Tokyo: Kurosio.

Zima, Elizabetta and Alexander Bergs (Eds.). 2017. Towards a multimodal construction grammar. *Linguistics Vanguard* (Special Issue), 3.