

Article

Nicole C. De Los Reyes* and Ute Römer-Barron

A collostructional approach to Japanese noun-modifying clause construction use and acquisition: a learner corpus study

<https://doi.org/10.1515/cllt-2024-0020>

Received February 23, 2024; accepted March 1, 2024; published online March 25, 2024

Abstract: Japanese features a general noun-modifying clause construction (NMCC) with a more versatile range of semantic and pragmatic interpretations than equivalent constructions in other languages. Motivated by the learning challenge NMCCs pose to Japanese as a foreign language (JFL) learners, this article examines speech data from the International Corpus of Japanese as a Second Language (I-JAS) to compare learner use of NMCCs against a large L1 Japanese corpus. Instances of the construction from both corpora were analyzed to identify high-frequency part-of-speech categories and subcategories in the modifying clause predicate and head noun slots. A simple collexeme analysis was then employed to identify strongly attracted and repelled lexical items among those identified in realizations of the construction. Taken together, findings from these analyses revealed an important connection between the semantic weight of head nouns in NMCCs and the idiomacity of the construction, with learner productions demonstrating a tendency toward heavy head nouns. This study lays the groundwork for future research seeking to explore the NMCC at different levels of granularity and to improve its treatment in JFL pedagogical materials.

Keywords: collostructional analysis; Japanese learner corpus; noun-modifying clause construction

***Corresponding author: Nicole C. De Los Reyes**, Applied Linguistics and ESL, Georgia State University, 25 Park Place, 15th Floor, Atlanta, 30303, GA, USA, E-mail: ndelosreyes1@gsu.edu. <https://orcid.org/0009-0004-6892-7723>

Ute Römer-Barron, Applied Linguistics and ESL, Georgia State University, Atlanta, USA. <https://orcid.org/0000-0002-0896-3165>

1 Introduction

Cross-linguistic investigations into noun-modifying clause constructions (NMCCs), a commonly observed linguistic feature in which a modifying clause combines with a noun, have been a particularly productive area of research at the intersection of linguistic typology (e.g., Comrie 1996, 2002, 2007; Keenan and Comrie 1977), first language (L1) acquisition (e.g., Kidd 2011), and second language (L2) acquisition (e.g., Giacalone Ramat 2002). Research exploring nodes in the network of NMCCs has largely focused on testing the acquisition of relative clause (RC) constructions and the various syntactic relationships between their constituents against the noun phrase accessibility hierarchy (NPAH) of Keenan and Comrie's (1977) foundational study. Discussions on the grammar and typology of Japanese NMCCs (e.g., Matsumoto 1997, 2017, 2018), however, suggest that Japanese has only one general noun-modifying clause construction, which subsumes a range of semantic and pragmatic meanings (e.g., RC constructions, noun complement clause constructions) that in other languages would be represented by multiple distinct clause types. This characteristic of Japanese NMCCs has been identified as creating difficulties for not only the typological analysis of the construction (e.g., Matsumoto 1989) but also for its acquisition by Japanese as a foreign language (JFL) learners (e.g., Yabuki-Soh 2007, 2013).

Motivated by the learning challenge NMCCs pose to JFL learners, the study reported on using a collostructional approach to identify high-frequency NMCCs and their lexical associations in JFL learner and L1 Japanese speaker production, highlighting similarities and differences in usage patterns between the two data types. Previous research on Japanese NMCCs for an English-speaking readership has sought to contribute to the paucity of literature on the acquisition of prenominal RC languages by exploring the universality of the NPAH, namely, its ability to predict the acquisition of prenominal RC constructions using corpus (Ozeki and Shirai 2007) and experimental data (Yabuki-Soh 2007). Pedagogically motivated inquiries into Japanese clausal modification which take the broader NMCC as their starting point, such as Yabuki-Soh's (2013) examination of noun-modifying clause types used in JFL textbooks, remain limited. Building on and going beyond the literature in this area, our descriptive exploratory study adopts a corpus-based approach to examine the NMCC repertoire of L2 learners, including frequent construction types and dominant lexical associations, with reference to an L1 baseline.

The present study uses learner data from a large corpus, the International Corpus of Japanese as a Second Language (I-JAS; Sakoda et al. 2016, 2020), to investigate the structural features of NMCCs produced by JFL learners and to compare them to those found in L1 Japanese usage. The main goal of the study is to gain a better understanding of JFL learner knowledge of NMCCs and how it relates to L1 language use, both in terms of dominant subtypes and lexical associations of the construction.

2 Literature review

2.1 Overview of Japanese clausal noun modification

The Japanese NMCC (hereafter NMCC) has been discussed in the literature (e.g., Matsumoto 2017, 2018) as a single construction with two main constituents: a modifying clause and a head noun (HN), see Figure 1.

The construction is described as having a “wide range of interpretations” (Matsumoto 2018: 464) due to relations between the HN and the modifying clause that are grammatically and semantically broad (Matsumoto 2017). Structurally, the modifying clause predicate is a well-formed independent clause (Matsumoto 2018: 465) that consists of a noun combined with a copula, a verb, or an adjective in the finite or adnominal form. Three simplified examples of typical varieties of NMCCs are provided in Table 1.

As can be seen in the examples in Table 1, the modifying clause predicate can share the same verb (e.g., *kat-ta*, ‘buy’) in the past prenominal/finite form and contain unexpressed arguments (e.g., subject, direct object) and adjuncts, which are interpreted as having an association with the HN. However, the array of

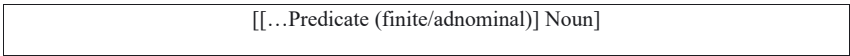


Figure 1: The NMCC construction (adapted from Matsumoto 2018: 464).

Table 1: Simplified typical varieties of the NMCC.

Argument structure	Examples				Relativization type
Argument internal to clause	[<i>hon</i>	<i>o</i>	<i>kat-ta</i>]	<i>gakusei</i>	Subject
	Book	ACC	buy-PST	student	
	‘the student [who bought the book]’				Direct object
	[<i>gakusei</i>	<i>ga</i>	<i>kat-ta</i>]	<i>hon</i>	
	Student	NOM	buy-PST	book	
	‘the book [that the student bought]’				Oblique
Adjunct of the modifying verb	[<i>gakusei</i>	<i>ga</i>	<i>kat-ta</i>]	<i>mise</i>	
	Student	NOM	buy-PST	store	
	‘the store [where the student bought (x)]’				

Note. Adapted from Matsumoto (2018: 465).

Table 2: Different interpretations of the same NMCC.

Example				Potential interpretations
[<i>hon</i>	<i>o</i>	<i>kat-ta</i>]	<i>gakusei</i>	'the student [who bought the book]' 'the student [from whom (x) bought a book]' 'the student [for whom (x) bought a book]' etc.
book	ACC	buy-PST	student	

Note. Adapted from Matsumoto (2018: 465).

arguments in the modifying clause will result in different interpretations of the NMCC (e.g., argument NMCC, adjunct NMCC). For example, the HN can play the role of the subject or direct object of the verb as seen in the instances where the argument is internal to the clause. The HN can also indicate the location where an action or event occurred, as in the second argument structure in Table 1, where the HN is the adjunct of the modifying verb. Matsumoto (2018) notes that language users' acceptability judgements concerning interpretations of the construction are not only constrained by syntactic, semantic, and pragmatic knowledge but also shared knowledge of the world. An example demonstrating a selection of possible interpretations for a single instance of the NMCC as determined by the interaction between linguistic and contextual factors is provided in Table 2.

It is important to note that, while it is possible for Japanese NMCCs to have a modifying clause that is identical in structure to an independent clause in terms of expressed arguments, there are instances where the HN cannot be associated with a missing argument in the modifying clause. There are also instances where the HN cannot be interpreted as an argument or adjunct of the modifying clause (see Table 3 for examples of these instances). Matsumoto (2017) observed that these characteristics of Japanese NMCCs can lead to the creation of instances with potentially ambiguous interpretations.

Table 3: Examples of potentially ambiguous NMCCs.

Examples				
(1)	[<i>atama</i>	<i>no</i>	<i>yoku-naru</i>]	<i>hon</i>
	head	GEN	good-become	book
	'the book [(by reading) which (one's) head gets better (i.e., one becomes more intelligent)]'			
(2)	[<i>hon</i>	<i>o</i>	<i>kat-ta</i>]	<i>uwasa</i>
	book	ACC	buy-PST	rumor
	'the rumor [that (x) bought a book]'			

Note. Adapted from Matsumoto (2018: 466).

2.2 Descriptive approaches to analyzing Japanese NMCCs

The descriptive approach to analyzing NMCCs can be attributed to the in-depth analyses of the construction in early typological research, most notably Teramura (1969). Teramura (1969) investigated the syntactic and semantic relationships between the HN and the modifying clause predicate within an NMCC. The author identified inner- and outer-relationship constructions as two broad types of clausal noun modification. Inner-relationship constructions (*uchi no kankei*) embed into the modifying clause predicate a HN with one of the five grammatical functions: Subject (SU), Direct Object (DO), Indirect Object (IO), Oblique (OBL), and Genitive (GEN). Conversely, outer-relationship constructions (*soto no kankei*) have HNs that cannot be embedded into the modifying clause predicate due to syntactic restrictions related to the semantic properties of the HNs. Teramura's (1969) bipartite classification of NMCCs based on the relationship between the semantically loaded HN and the modifying clause predicate has informed the design of subsequent typological studies and empirical investigations into the acquisition of the construction by L1 Japanese speakers and JFL learners, the latter of which will be the focus of Section 2.3.

2.3 Empirical studies of NMCC acquisition by L2 Japanese learners

We are only aware of a small number of empirical studies written for an English-speaking readership that have investigated the acquisition of NMCCs by JFL learners (Ozeki and Shirai 2007; Yabuki-Soh 2007). Ozeki and Shirai's (2007) corpus study tested the applicability of Keenan and Comrie's (1977) noun phrase accessibility hierarchy (NPAH). The authors hypothesized that JFL learners are capable of producing more marked NMCCs even at lower levels of proficiency. Using the KY (Kamada Yamaguchi) Corpus (Kamada 1999, 2006), a learner corpus of transcribed ACTFL Oral Proficiency Interviews, instances of L1 and L2 NMCCs were analyzed. NMCCs were operationalized as nouns modified by verbs, adjectives with complements, and adjectives in the past tense (see Appendix A for examples of each of these NMCC types). The study found that with the exception of L1 Korean-speaking learners, NMCC use was contingent on the animacy status of the HN rather than its markedness as predicted by the NPAH. JFL learner productions from the KY Corpus revealed an initial mapping of SU relatives with animate-head NPs and DO and OBL relatives with inanimate-head NPs. Similar animacy effects were found in the analysis of the L1 Japanese data.

The quasi-experimental study by Yabuki-Soh (2007) explored whether JFL learners' acquisition of a marked NMCC, particularly the OBL relativization type, would facilitate learning NMCC types higher in the NPAH. Adult learners of first-year Japanese were randomly assigned to a form-based, meaning-based, or form-plus meaning-based instructional group and were exposed to only the OBL-type NMCC in three treatment sessions. Comprehension and sentence completion tasks were used to measure participants' progress across the three groups. Highly explicit instruction was found to play an important role in participants' ability to recognize and use NMCCs. Moreover, instruction on only the OBL-type was found to facilitate participants' ability to generalize its structure to other NMCC types. The study demonstrated that the acquisition of Japanese NMCCs is a complex and nonlinear process, which cannot be encapsulated by the NPAH predictions correlating less marked constructions and learner accessibility, suggesting a need to examine NMCCs using alternative theoretical frameworks and analytical approaches.

Departing from research testing the NPAH, Yabuki-Soh (2013) analyzed NMCCs in seven postsecondary JFL textbooks to determine the extent to which instructional materials reflect the characteristics of Japanese clausal noun modification. The author found that textbooks tend to focus on inner-relationship NMCCs, namely, the SU and DO types, and neglect outer-relationship NMCCs. The author makes several suggestions to improve the existing coverage of NMCCs in JFL textbooks, notably, the provision of grammar explanations that extend the use of NMCCs from providing additional information about people and things (i.e., inner-relationship NMCCs) to expressing conceptual content (i.e., outer-relationship NMCCs) and the inclusion of more varied HNs in discussions of outer-relationship NMCCs.

Taken together, the above studies show changes in the trajectory of research on NMCCs, which shifted from a focus on the learnability of RC constructions by JFL learners to the quality of JFL textbooks regarding their inclusion and treatment of noun-modifying clauses of different types. There remains a need, however, to better understand JFL learners' knowledge of NMCCs as reflected in their use of the construction. Our study seeks to address this research gap by drawing on learner corpus data to analyze NMCCs in JFL learner and L1 Japanese speaker production and to identify potential discrepancies in their usage patterns.

2.4 The present study

According to Mori et al. (2020), the development of quality instructional materials for JFL teaching is contingent on “the integration of fine-grained qualitative linguistic

analysis and a systematic analysis of a large database enabled by technology” (109). Answering this call for more corpus-based research investigating L2 Japanese development, our study uses quantitative methods to complement existing qualitative linguistic analyses of the structure of and lexical items in NMCCs. The study aims to provide insights that can inform the treatment of NMCCs in pedagogical materials and in classroom instruction. Three research questions guide this inquiry:

1. What NMCC types are part of JFL learners’ spoken repertoire?
2. How does JFL learners’ use of NMCCs in spoken Japanese compare to that of L1 Japanese speakers?
3. Are there any significant associations between NMCCs and specific head nouns in L1 and L2 usage?

The remaining sections will outline the methodology of the study, present findings on the three research questions, and conclude with a discussion of results, limitations, and future directions.

3 Methods

3.1 The International Corpus of Japanese as a Second Language (I-JAS)

The International Corpus of Japanese as a Second Language (I-JAS; Sakoda et al. 2016, 2020) is a large-scale (8,076,969 words) learner corpus in the National Institute for Japanese Language and Linguistics (NINJAL) family of corpora. The corpus is composed of written and spoken language data from three groups of adult learners ($n = 1,000$) ranging from 17 to 46 years in age: (1) learners of Japanese in a foreign language context ($n = 850$); (2) learners of Japanese in a classroom context in Japan ($n = 100$); and (3) learners of Japanese in naturalistic contexts ($n = 50$) (Sakoda et al. 2016, 2020). It also includes comparison data from L1 Japanese speakers ($n = 50$) ranging from 20 to 50 years in age. A key characteristic of the I-JAS is its inclusion of rich metadata pertaining to learners’ backgrounds and learning environments, which was collected through a 20-item questionnaire administered online in the participants’ respective L1 (see Appendix B for a breakdown of the metadata collected in the questionnaire). Additionally, the corpus includes information on various aspects of learners’ Japanese language proficiency, such as their linguistic knowledge and automatic processing ability, as measured by the Japanese Computerized Adaptive Test (Imai et al. 2009) and the Simple Performance-Oriented Test, respectively.

3.2 Data collection and analysis

The present study draws from the dialogue task (*taiwa tasuku*; ‘dialogue task’) in the I-JAS, a 5,952,601-word subcorpus of spoken interview data, in which the participant data isolated from interviewer data totals 3,168,198 words. The task is a 30-min semi-structured conversation facilitated by a L1 Japanese-speaking interlocutor on 15 predetermined topics largely pertaining to daily life. Appendix C presents the overall structure of the dialogue task and provides sample topics and questions. Instances of NMCCs produced by 850 learners of Japanese in a foreign language context, ranging in proficiency from beginner to advanced,¹ with a majority of intermediate level learners (see Appendix D), and 50 Japanese L1 speakers, referred to as JJJ by the corpus compilers, were extracted in two separate key-word-in-context (KWIC) searches using the online corpus concordance system Chunagon (version 2.7.0; National Institute for Japanese Language and Linguistics n.d.). The graphical user interface-based query builder uses dependency parsing to retrieve mono- and multi-morphemic lexical items, more commonly referred to as short-unit words (SUWs) and long-unit words (Koiso et al. 2020; Sakoda et al. 2016, 2020), from individual NINJAL corpora based on annotated morphological information. Adopting an SUW search for lexemes tagged as nouns preceded by lexemes in the attributive form, a total of 25,738 KWIC concordance lines were retrieved from the corpus of L2 Japanese learner utterances (2,466,125 words), hereafter referred to as the Japanese as a Foreign Language (JFL) corpus. See Table 4 for an overview of the composition of the JFL corpus. A total of 4,441 KWIC concordance lines with 661 SUW HN lemma types were retrieved from the corpus of L1 Japanese speaker utterances or JJJ corpus (228,818 words).

A combination of Microsoft Excel and R scripts was used to prepare and process the data from the two corpora for analysis. The raw word counts were normalized to frequencies per 1,000,000 words to allow for comparisons between speaker groups. A simple collxeme analysis (Stefanowitsch and Gries 2003) was performed, focusing on HNs in the JFL and L1 Japanese corpora.

3.3 Collostructional analysis

Collostructional analysis (CA) is described as a “family of methods” (Gries 2019: 388) – collxeme analysis, distinctive collxeme analysis, and co-varying collxeme

¹ While it would have been possible to analyze the stratified learner corpus data separately, the JFL learner texts were grouped and analyzed together, given disparities in the number of participants across the seven proficiency levels identified in the corpus metadata (see Appendix D).

Table 4: Overview of the JFL corpus.

Native language	Participants	Subcorpus size
Chinese (C)	200	574,764
Korean (K)	100	347,595
English (E)	100	276,778
German (G)	50	159,177
Russian (R)	50	158,031
Indonesia (I)	50	144,142
Thai (TTH)	50	143,970
Hungarian (H)	50	142,819
Turkish (TTH)	50	138,571
Vietnamese (V)	50	135,076
French (F)	50	124,943
Spanish (S)	50	120,259
Total	850	2,466,125

analysis – developed by Stefanowitsch and Gries (2003) that measures the strength of attraction or repulsion of lexical items (collexemes) to slots in constructions (Goldberg 1995: 1–6, 2013). Advantages of CAs include increased precision in grammatical descriptions over traditional collocation analysis, which have practical implications for fields such as language pedagogy wherein knowing the attraction between collexemes and constructions can inform materials design and classroom practices.

The collostructional analysis selected for the present study was a simple collexeme analysis (e.g., Stefanowitsch and Gries 2003), which focused on the HN slot of the NMCC. Chunagon was first used to extract all instances of nouns preceded by attributive forms in the JJJ and the JFL corpora. Using the downloadable search results, pivot tables were then generated in Microsoft Excel to identify the types and the tokens of nouns present in the corpora. To determine the observed frequency of each noun, all instances of nouns in the JJJ and the JFL corpora were also extracted. Pivot tables were again created to obtain all noun types and their frequencies in the corpora. Due to the larger size of the JFL corpus and limitations of the concordancer regarding downloadable search results (i.e., a maximum of 100,000 concordance lines), nouns in the JFL corpus had to be retrieved separately for each L1 group and were later combined into a single spreadsheet, mirroring that of the JJJ corpus. The resulting information was compiled into two data frames containing three columns – types of nouns in the HN slots, their frequency in the NMCC, and their frequency in the corpus – which served as input for the simple collexeme analysis performed using R (version 4.2.2) (R Core Team 2022). The package {collostructions}

(Flach 2021) was used to compute the collostructional strength, measured as log-likelihood (Dunning 1993), for the HNs in the corpora, and a Bonferroni correction was applied. The alpha level was set at $p < 0.05$. In the calculation, the size of the complete JJJ corpus (in words) was used for the “corpsize” argument of the function.

4 Results

The distribution of L1 and L2 NMCCs was analyzed using the JFL and the JJJ corpora described in the previous section. A total of 30,161 instances of NMCCs were extracted from the I-JAS dialogue task data. Of these instances 25,750 (10,441.48 per million words [pmw]) were from the JFL corpus and 4,411 (19,277.33 pmw) were from the JJJ corpus.

4.1 Distribution of NMCCs in JFL learner speech

Two searches – nouns preceded by attributive forms and attributive forms followed by nouns – were performed to isolate SUWs in the HN and the modifying clause predicate slots of the NMCC in the JFL corpus. Analyses of the modifying predicate slot revealed that JFL learners use a variety of NMCC types categorizable by POS. Four POS categories, in particular, were identified (see left side of Table 5):² auxiliary verbs (*jodōshi*; 10,068 instances), verbs (*dōshi*; 9,130 instances), adjectives (*keiyōshi*; 6,506 instances), and suffixes (*setsubiji*; 34 instances). With the exception of auxiliary verbs, these categories are further subdivided as indicated in Table 5. Analyses of the HN slot in the NMCC revealed four categories of nouns (see left side of Table 6). The most frequent category was common nouns (*meishi-fūtsūmeishi*; 25,456 instances) with six subtypes as detailed in Table 6.³ The second most frequent noun types were proper nouns (*meishi-koyūmeishi*; 196 instances) of six subtypes (see Table 6). Less frequent noun types were auxiliary nouns (*meishi-jodōshigokan*; 60 instances) and numeral nouns (*meishi-sūshi*; 38 instances).

4.2 Contrastive analysis of NMCCs in L1 and L2 Japanese

Comparisons of the NMCC predicates in the L1 and L2 data showed complete overlap in the POS categories and subcategories of SUW directly to the left of the HN (see

² See <https://clrd.ninjal.ac.jp/unidic/glossary.html> and <https://www.sketchengine.eu/tagset-jp-mecab/> for details on the POS tagset used in the I-JAS annotation.

³ Glosses were retrieved from Tono et al. (2013).

Table 5: NMCC predicates and top 3 most frequent lexemes within POS categories.

POS type	POS subtype	JFL			JJJ		
		Tokens	Percentage	Types	Tokens	Percentage	Types
<i>jodōshi</i> 'auxiliary verb'		10,068	39.12 %	<i>-da</i> 'copula'; <i>-ta</i> 'past'; <i>-nai</i> 'not'	1,985	45.01 %	<i>-ta</i> 'past'; <i>-da</i> 'copula'; <i>-te iru</i> , <i>-teru</i> 'continuation'
<i>dōshi</i> 'verb'	<i>dōshi-ippan</i> 'general verb'	9,130 5,698	35.47 % 22.14 %	<i>iu, yū</i> 'say, speak, talk'; <i>hanasu</i> 'talk, tell, speak'; <i>hairu</i> 'enter, come in, go in'	1,684 1,234	38.19 % 27.98 %	<i>iu, yū</i> 'say, speak, talk'; <i>chi-gau</i> 'be different, be wrong'; <i>omou</i> 'think, believe, feel, expect'; <i>asobu</i> 'play'
	<i>dōshi-hijir-itsukanō</i> 'bound verbs'	3,432	13.33 %	<i>suru</i> 'do, make'; <i>iru</i> 'be, exist, stay'; <i>iku</i> , <i>yuku</i> 'go, come'	450	10.20 %	<i>suru</i> 'do, make'; <i>iru</i> 'be, exist, stay'; <i>aru</i> 'be (existence), have (possession), happen, occur'
<i>keiyōshi</i> 'adjective'	<i>keiyōshi-ippan</i> 'general <i>i</i> -adjective'	6,506 4,929	25.28 % 19.15 %	<i>chīsai</i> 'small, little, tiny'; <i>warui</i> 'bad'; <i>ōki</i> 'big, large, great'	732 547	16.60 % 12.40 %	<i>sugoi</i> 'fantastic, wonderful, terrible'; <i>chīsai</i> 'small, little, tiny'; <i>kowai</i> 'frightening, scary, terrified'; <i>warui</i> 'bad'
	<i>keiyōshi-hijir-itsukanō</i> 'bound <i>i</i> -adjective'	1,577	6.13 %	<i>ii, yoi</i> 'good'; <i>nai</i> 'there is no...', no...'; <i>hoshi</i> 'want, desire'	185	4.20 %	<i>ii, yoi</i> 'good'; <i>nai</i> 'there is no...', no...'; <i>hoshi</i> 'want, desire'
<i>setsubiji</i> 'suffix'		33 29	0.13 % 0.11 %	<i>-rashii</i> 'seem,	9 8	0.20 % 0.18 %	<i>-yasui</i> 'easy to...'; <i>-poi</i> 'ish,

Table 5: (continued)

POS type	POS subtype	JFL			JJJ		
		Tokens	Percentage	Types	Tokens	Percentage	Types
	<i>setubiji-keijō-shiteki</i> 'adjective suffix'			look'; -yasui 'easy'; -poi 'ish, -like'			-like'; -rashii 'seem, look'
	<i>setubiji-dōshiteki</i> 'verbal suffix'	4	0.02 %	-garu 'show signs of being, feel, think'	1	0.02 %	-garu 'show signs of being, feel, think'
Total		25,737	100.00 %		4,410	100.00 %	

Table 5 for an overview). The most frequent predicate type in both corpora were auxiliary verbs (39.12 % of all NMCC tokens in JFL and 45 % in JJJ data), followed by verbs (35.47 % in JFL; 27.98 % in JJJ) and adjectives (25.38 % in JFL; 16.59 % in JJJ). Suffixes were the least frequent type of predicate in NMCCs in both corpora (0.13 % in JFL; 0.20 % in JJJ).

An examination of the lexemes in the HN slots of L1 and L2 NMCCs also revealed overlap in POS categories. In the NMCCs produced by both JFL learners and L1 Japanese speakers, the HN slot of the construction was filled predominantly by common nouns (98.86 % in JFL; 99.27 % in JJJ) of the same five subcategories: general common nouns, adverbial nouns, verbal nouns, counter nouns, and adjectival nouns (see Table 6). Similarities were also observed in JFL learners' and L1 Japanese speakers' use of other noun subcategories, such as country names, auxiliary nouns, counter nouns, general places, verbal adjectival nouns, general names, and first names. NMCCs with general proper nouns and last names in the HN slot of the construction are the only point of difference between the two corpora, occurring exclusively in the JFL data while being the only noun subcategories absent from L1 Japanese speaker productions. These findings can be attributed to self-referential utterances typical of lower-proficiency speakers of a language, which would include proper nouns in the form of country of origin and general place and location names relating demographically heterogeneous JFL speakers. Table 6 summarizes the HN tokens and the top three lexemes within each subcategory.

The next step in the analysis involved identifying and comparing realizations of the construction within the three most frequent modifying clause predicate types (auxiliary verbs, verbs, adjectives). Suffixes were omitted given their low frequency.

Table 6: NMCC HN tokens and top 3 most frequent lexemes within noun categories.

POS type	POS subtype	JFL		JJ	
		Tokens	Percentage	Tokens	Percentage
<i>meishi-fūtūmeishi</i> 'common noun'	<i>Ippan</i> 'general noun'	25,456 16,297	98.86 % 63.29 %	4,379 2,689	99.27 % 60.96 %
	<i>fukushikanō</i> 'adverbial noun'	5,606	21.77 %	871	19.75 %
	<i>Sahenkanō</i> 'verbal noun'	2,947	11.44 %	587	13.31 %
	<i>jōsūshikanō</i> 'counter nouns'	450	1.75 %	125	2.83 %
	<i>keijiyōshikanō</i> 'adjectival noun'	150	0.58 %	105	2.38 %
	<i>sahenkeijiyōshikanō</i> 'verbal adjectival noun'	6	0.02 %	2	0.05 %
<i>meishi-koyūmeishi</i> 'proper noun'	<i>chimei-koku</i> 'country'	196 145	0.76 % 0.56 %	17 3	0.39 % 0.07 %
	<i>chimei-ippan</i> 'general places'	28	0.11 %	9	0.20 %
	<i>Ippan</i> 'general places'	11	0.04 %	0	0.00 %

Table 6: (continued)

POS type	POS subtype	JFL		JJI	
		Tokens	Percentage	Types	Percentage
<i>meishi-jodō-shigokan</i> 'auxiliary noun'	<i>jinmei-ippan</i> 'names of people'	5	0.02 %	<i>Ri</i> 'Lee'; <i>Kirisuto</i> 'Christ'	2 0.05 %
	<i>jinmei-sei</i> 'last names of people'	5	0.02 %	<i>Tokoro</i> 'Tokoro'; <i>Abe</i> 'Abe'; <i>Natsume</i> 'Natsume'; <i>Sakurai</i> 'Sakurai'	0 0.00 %
	<i>Jinmei-na</i> 'first names of people'	2	0.01 %	<i>Chiaki</i> 'Chiaki'; <i>Nashika</i> 'Nausicaä'	3 0.07 %
		60	0.23 %	<i>sō</i> 'people say that, it is said that, I hear that'	7 0.16 %
		38	0.15 %	<i>ichi</i> 'one'; <i>san</i> 'three'; <i>go</i> 'five'	8 0.18 %
Total		25,750	100.00 %	4,411	100.00 %

In the auxiliary_HN realization, the top three most frequently used auxiliary lexemes in the JJJ corpus were *-ta*, *-da*, and *-te iru/teru* (continuation), which contrasted with the lexemes *-da*, *-ta*, and *-nai* found in the JFL corpus. The general nouns *koto* and *toki* were the two most frequent HNs in NMCCs in both corpora. The third most frequent HN, however, differed between the L1 and the L2 Japanese speakers with the former demonstrating a preference for the general noun *kanji* over the general noun *hito*, which was preferred by the latter. In verb_HN NMCC realizations, both JFL learners and L1 Japanese speakers showed a strong preference for the general verb *iu/yū* as the modifying predicate. This lexeme, however, was more frequent in the JJJ corpus (73.72 %) than in the JFL corpus (48.53 %). Realizations of *i-Adjective_{general}_HN* and *i-Adjective_{bound}_HN* NMCCs types were more frequent in the JJJ corpus (3,198.91 instances pmw) than in the JFL corpus (2,638.15 instances pmw). There was minimal overlap in the top three modifying clause types in the *i-Adjective_{general}_HN* NMCC. The most frequent types in the JJJ corpus were *sugoi* (463.23 pmw), *chīsai* (463.23 pmw), and *kowai* (135.47 pmw), whereas *chīsai* (352.27 pmw), *warui* (137.05 pmw), and *ōki* (114.35 pmw) were most frequent in the JFL corpus. As for the types of HNs in the *i-Adjective_{general}_HN* NMCC, there was little variation in lexical preferences between speaker groups. The adverbial nouns *koro* (231.62 pmw) and *toki* (201.02 pmw) followed by the general noun *koto* (113.62 pmw) were the three most frequent HNs used by L1 Japanese speakers in this construction. In contrast, the top three HNs in the JFL corpus were *koto* (195.85 pmw), *toki* (169.09 pmw), and *hito* (123.68 pmw). In *i-Adjective_{bound}_HN* NMCC realizations, the modifying clause predicates *ii*, *nai*, and *hoshi* were the top three most frequently used across both the JJJ (493.82, 244.73, and 69.92 instances pmw) and the JFL corpora (508.08, 111.11, and 20.27 instances pmw). The HN types were also similar. The top three most frequent HNs were *mono* (91.77 pmw), *sensei* (61.18 pmw), and *koto* (52.44 pmw) in the JJJ corpus and *sensei* (63.66 pmw), *hito* (62.85 pmw), and *koto* (49.47 pmw) in the JFL corpus.

4.3 Simple colllexeme analysis of head nouns in NMCCs

The collostructional strength measured as log-likelihood (G^2) was computed for each colllexeme in the HN slot of the NMCC. In the JJJ corpus, 101 out of the 650 colllexemes used by L1 Japanese speakers were found to be significantly attracted to the construction after the Bonferroni correction. Of the 1,246 colllexemes used by JFL learners in the JFL corpus, 222 noun colllexemes were found to be significantly attracted to the construction. Tables 7 and 8 present the top 20 colllexemes in each corpus organized by collostructional strength, that is, their level of attraction to the NMCC. In the L2 NMCC productions, the top 20 significantly attracted colllexemes in descending order were *koto*, *toki*, *tokoro*, *mono*, *hito*, *kanji*, *sensei*, *machi*, *tame*, *kata*,

Table 7: Top 20 significantly attracted collexemes in the HN slot of the NMCC – JJJ corpus.

	Collexeme	Freq.Corp.	Freq.CX (Obs.)	Freq.CX (Exp.)	G ²	P-value
1	<i>koto</i> ‘thing’	854	656	16.50	4,364.10	0.00E+00
2	<i>tokoro</i> ‘place’	518	376	10.00	2,398.85	0.00E+00
3	<i>mono</i> ‘thing’	387	275	7.50	1,727.73	0.00E+00
4	<i>kanji</i> ‘feeling, impression, atmosphere’	459	283	8.80	1,648.79	0.00E+00
5	<i>toki</i> ‘time’	706	275	13.60	1,261.08	2.11E-273
6	<i>wake</i> ‘reason, cause’	141	123	2.70	867.90	6.05E-188
7	<i>hito</i> ‘person, people, human being’	374	128	7.20	543.40	2.23E-117
8	<i>fū</i> ‘style, type, way, like’	85	75	1.60	532.46	5.35E-115
9	<i>kata</i> ‘person, man’	447	127	8.60	485.31	9.71E-105
10	<i>sensei</i> ‘teacher’	417	110	8.00	402.01	1.30E-86
11	<i>ki</i> ‘mind, heart’	155	69	3.00	336.35	2.58E-72
12	<i>koro</i> ‘time, about, when’	154	64	3.00	300.80	1.43E-64
13	<i>ko</i> ‘child’	127	54	2.40	256.77	5.62E-55
14	<i>hanashi</i> ‘story, talk’	196	62	3.80	251.08	9.81E-54
15	<i>yatsu</i> ‘guy, fellow’	60	39	1.20	231.50	1.82E-49
16	<i>omoide</i> ‘memory, reminiscence’	51	33	1.00	195.37	1.39E-41
17	<i>imi</i> ‘meaning, sense’	54	54	1.00	189.54	2.61E-40
18	<i>basho</i> ‘place, spot, position’	56	30	1.10	160.82	4.87E-34
19	<i>tame</i> ‘for’	46	28	0.90	160.45	5.85E-34
20	<i>kioku</i> ‘memory’	67	27	1.30	124.63	3.99E-26

ato, *basho*, *koro*, *tabemono*, *tsumori*, *ko*, *kodomo*, *shigoto*, *hanashi*, and *keiken*. The top 20 significantly attracted collexemes in the L1 Japanese NMCC productions were *koto*, *tokoro*, *mono*, *kanji*, *toki*, *wake*, *hito*, *fū*, *kata*, *sensei*, *ki*, *koro*, *hanashi*, *yatsu*, *omoide*, *imi*, *basho*, *tame*, *kioku*. Comparison between the corpora indicated substantial overlap (70 %) among these top-20 HNs. Both speaker groups show strong preferences for general common nouns that broadly refer to concrete or abstract entities. Notable differences include the HNs *kodomo*, *machi*, *tsumori*, *tabemono*, *ato*, *shigoto*, and *keiken* which are among the most significant collexemes in the JFL but not in the JJJ corpus, and the HNs *hazu*, *omoide*, *ki*, *wake*, *fū*, *imi*, and *kioku* for which the opposite is the case. These differences do not pertain to POS subcategorization but rather the semantic weight of the HNs.

The simple collexeme analysis also identified HNs that are significantly repelled by the NMCC. Only those in the JFL corpus, however, reached statistical significance. Table 9 displays the most significantly repelled collexemes. Of the 131 repelled

Table 8: Top 20 significantly attracted collexemes in the HN slot of the NMCC – JFL subcorpus.

Collexeme	Freq.Corp.	Freq.CX (Obs.)	Freq.CX (Exp.)	G ²	P-value
1 <i>koto</i> ‘thing’	8,912	5,747	92.80	42,302.93	0.00E+00
2 <i>toki</i> ‘time’	11,180	2,640	116.40	12,318.88	0.00E+00
3 <i>tokoro</i> ‘place’	2,655	1,431	27.60	9,504.22	0.00E+00
4 <i>mono</i> ‘thing’	2,029	1,186	21.10	8,144.73	0.00E+00
5 <i>hito</i> ‘person, people, human being’	5,299	1,498	55.20	7,528.03	0.00E+00
6 <i>kanji</i> ‘feeling, impression, atmosphere’	962	562	10.00	3,844.92	0.00E+00
7 <i>sensei</i> ‘teacher’	6,329	829	65.90	2,792.03	0.00E+00
8 <i>machi</i> ‘town, city’	1,629	445	17.00	2,184.12	0.00E+00
9 <i>tame</i> ‘for’	1,123	365	11.70	1,936.68	0.00E+00
10 <i>kata</i> ‘person, man’	2,101	378	21.90	1,511.72	0.00E+00
11 <i>ato</i> ‘after, later’	3,070	406	32.00	1,369.30	1.23E-296
12 <i>basho</i> ‘place, plot, position’	547	204	5.70	1,148.52	1.17E-248
13 <i>koro</i> ‘time, about, when’	903	238	9.40	1,147.14	2.34E-248
14 <i>tabemono</i> ‘food’	1,186	258	12.40	1,134.78	1.14E-245
15 <i>tsumori</i> ‘intention’	230	154	2.40	1,116.55	1.04E-241
16 <i>ko</i> ‘child’	510	192	5.30	1,085.30	6.44E-235
17 <i>kodomo</i> ‘child’	2,791	302	29.10	898.90	2.12E-194
18 <i>shigoto</i> ‘work, job’	2,119	263	22.10	852.72	2.32E-184
19 <i>hanashi</i> ‘story, talk’	995	201	10.40	851.73	3.82E-184
20 <i>keiken</i> ‘experience’	407	144	4.20	791.98	3.72E-171

collexemes in the JFL corpus, 15 reached statistical significance, namely, *-go*, *ni*, *jū*, *kane*, *haha*, *jibun*, *issho*, *mina/minna*, *ichi*, *toshi*, *go*, *yon/shi*, *ima*, *chichi*, *kinō*, *tsuki*. The analysis also revealed 16 repelled collexemes in the JJJ corpus after the Bonferroni correction, only one of which came close to reaching statistical significance: *ima* ‘now.’

5 Discussion

Our first research question examined the NMCC repertoire of L2 learners of Japanese. To answer this question, transcribed spoken interview data from 850 adult JFL learners were sampled. The analysis revealed that JFL learners produce varied NMCCs with four broad types of modifying clause predicates: auxiliary verbs, verbs, adjectives, and suffixes. Of these four, NMCCs containing auxiliary verbs, specifically

Table 9: Significantly repelled collexemes in the HN slot of the NMCC – JFL corpus.

	Collexeme	Freq.Corp.	Freq.CX (Obs.)	Freq.CX (Exp.)	G ²	P-value
1	<i>go</i> ‘language’	7,203	4	75.00	119.47	1.03E-24
2	<i>ni</i> ‘two’	3,374	1	35.10	61.55	5.39E-12
3	<i>jū</i> ‘ten’	3,175	1	33.10	57.50	4.22E-11
4	<i>kane</i> ‘money’	3,895	5	40.60	50.56	1.44E-09
5	<i>haha</i> ‘mother’	2,358	1	24.60	40.97	1.93E-07
6	<i>jibun</i> ‘oneself’	2,603	3	27.10	35.25	3.61E-06
7	<i>issho</i> ‘together, with’	2,745	4	28.60	33.68	8.08E-06
8	<i>mina</i> ‘everyone, everything’	2,011	2	20.90	28.68	1.06E-04
9	<i>ichi</i> ‘one’	5,171	20	53.80	28.35	1.26E-04
10	<i>nen/toshi</i> ‘year, age’	3,754	12	39.10	26.07	4.11E-04
11	<i>go</i> ‘five’	1,806	2	18.8	24.82	7.85E-04
12	<i>yon/shi</i> ‘four’	1,527	1	15.9	24.43	9.63E-04
13	<i>ima</i> ‘now’	5,541	26	57.70	22.17	3.11E-03
14	<i>chichi</i> ‘father’	1,272	1	13.2	19.45	1.29E-02
15	<i>kinō</i> ‘yesterday’	1,482	2	15.4	18.82	1.79E-02
16	<i>tsuki</i> ‘month, moon’	1,417	2	14.8	17.64	3.33E-02

the copula *-da*, past tense marker *-ta*, and the negative marker *-nai*, in the slot to the immediate left of the HN made up most instances, closely followed by those containing general and bound verbs. A simple descriptive analysis of items that frequently occupy the HN slot of the NMCC revealed that JFL learners demonstrated a strong preference for common nouns of the general type, particularly *koto* (‘thing’).

The second research question compared JFL learners’ use of NMCCs in spoken Japanese to that of L1 Japanese speakers. Our analysis showed that while auxiliary verbs were the most frequent predicate type in both corpora, the most frequently used auxiliary lexeme differed by speaker group. JFL learners showed a preference for the copula to the left of the HN, whereas L1 speakers preferred the past tense marker *-ta*. A likely explanation for this finding is the prevalence of auxiliary_{da}-HN NMCCs in grammar descriptions and NMCC examples provided in JFL textbooks (Yabuki-Soh 2013). Similarly, JFL learners’ preference of head nouns representing concrete or abstract entities can likely be attributed to exercises prompting the identification of people and things in picture description tasks (Yabuki-Soh 2013). Systematic analyses of JFL teaching materials would be helpful to determine dominant patterns in the language input learners receive in JFL instructional settings. The similarities in NMCC use that exist between the speaker groups, especially pertaining

to the use of abstract general common nouns, can be interpreted in light of Takara's (2012) finding that specific head nouns play roles in high frequency adverbial clauses or fixed expressions.

To answer our third research question, we performed a simple collexeme analysis to determine associations between the NMCC and nouns occurring in the HN slot of the construction in L1 and L2 Japanese speaker productions. This analysis allowed us to go beyond the simple frequency-based analyses we carried out to address RQs 1 and 2. It also allowed us to expand on the scope of previous NMCC studies that examined the acquisition of the construction by JFL learners using experimental (Ozeki and Shirai 2007) and interventional designs (e.g., Yabuki-Soh 2007). Specifically, the collexeme analysis enabled us to make inferences about L1 and L2 NMCCs use beyond those available through the examination of frequency measures alone, pertaining to the attraction of HNs to the construction.

The simple collexeme analysis revealed overlap in 13 of the top-20 HNs attracted to the NMCC in the JFL and JJJ corpora. For the remaining seven HN selections where the two groups differed, the distinction lies in the probability of using light versus heavy HNs in terms of semantic weight, as categorized according to the parameters in Takara (2012). These most strongly associated HNs in the NMCC repertoire of JJJ speakers included six light HNs (viz. *yatsu*, *ki*, *wake*, *fū*, *imi*, *kioku*) and one heavy HN (viz. *omoide*). This strong attraction of light over heavy HNs to the NMCC in JJJ speaker production is consistent with the finding in Ozeki and Shirai (2007) that L1 conversational Japanese contained a 70–20 split of light to heavy HNs by frequency percentage, with the remaining 10 % occupied by “gray” nouns of ambiguous semantic weight. Conversely, the HNs most strongly attracted to NMCCs in JFL learner production included four heavy HNs (viz. *machi*, *tabemono*, *shigoto*, *keiken*), two light HNs (viz. *tsumori*, *ato*), and one gray HN (viz. *kodomo*). This revealed a significantly stronger attraction of heavy HNs to the NMCC for JFL learners than for JJJ speakers, a finding that is likely the result of input effects of pedagogical materials (Yabuki-Soh 2013).

Koto was by far the most significantly attracted collexeme in both corpora with the highest G^2 value. This may not be surprising given the various functions of *koto* in Japanese, including its use as a formal noun, nominalizer, and constituent of grammaticalized NMCCs.⁴ We found that other collexemes attracted to the HN slot of the construction were also formal nouns belonging to grammaticalized NMCCs (e.g., *toki*; {V/Adj} (i) inf *toki* (ni) ‘at the time when/when’; *wake*; {V/Adj} (i) inf *wake* (da)

4 We acknowledge that the use of a tupleization approach (Gries 2019) over G^2 , a one-dimensional measure of collexeme strength, would likely enable a more nuanced understanding of *koto* in the JFL and JJJ corpora. Such an approach is recommended for future research on the topic.

that's why'), semantically 'light' HNs (e.g., *koto*, *tokoro*, *hito*), or HNs in a semantic gray zone (*kanji*, *ko*, *basho*). A possible explanation for JFL learners' use of *koto* is its use in high frequency grammatical constructions that are often taught early to first- and second-year Japanese students to build their knowledge of basic Japanese grammar (Makino and Tsutsui 2011 [1989]). In comparison, L1 Japanese speakers' use of *koto* might be explained by Takara's (2012) observation that HNs in NMCCs do not always require 'heavy' nouns if interlocutors can determine their referents from the same or previous turns.

The results of the simple collexeme analysis also revealed 16 significantly repelled HNs in the JFL corpus which belonged to the subcategories of numeral nouns (*ni*, *jū*, *ichi*, *go*, *yon/shi*), general nouns (*go*, *kane*, *haha*, *jibun*, *chichi*), adverbial nouns (*mina/minna*, *nen/toshi*, *ima*, *kinō*), counter nouns (*tsuki*), and verbal nouns (*issho*). The presence of repelled HNs exclusively in the L2 data could indicate that JFL learners have an insufficient understanding of the potential slot fillers of prototypical NMCCs (perhaps missing a number of nouns that are idiomatic choices in NMCCs), which are hypothesized in Takara (2012) to "contain 'light' heads rather than 'heavy' heads" (35). In other words, JFL learners may not be fully aware of the properties shared by strongly attracted HNs and the pragmatic motivations behind their selection in conversation.

Taken together, the simple collexeme analysis and the semantic analysis it facilitated suggest an important connection between the choice of HNs of a given semantic weight and the HN slot of the NMCC: light nouns in Japanese frequently perform grammaticalized functions that require NMCCs. To remove these light nouns from their context in NMCCs would be insufficient for idiomatic production, which highlights a problem with JFL learners' textbook treatment of NMCCs: it is primarily limited to their function as descriptive predicates to a HN, while grammaticalized light HNs are presented as isolated grammar points with no explicit relation to other NMCCs. One might infer that the L2 learners who contributed to the JFL corpus tended toward heavy HNs more so than the JJJ speakers, not simply for lack of vocabulary, but for lack of the abstracted understanding of NMCCs necessary to incorporate the vocabulary grammatically.

6 Conclusions

The aim of the present study was to build on and extend the largely qualitatively-descriptive literature on Japanese NMCCs by analyzing L1 and L2 productions of the construction taken from the I-JAS, a large-scale learner corpus, and employing a

simple collexeme analysis to identify strongly attracted and repelled lexical items in the HN slot of the construction. This aim was motivated by a desire to better understand JFL learners' use of NMCCs against an L1 Japanese baseline and to stimulate discussions on the treatment of NMCCs in existing Japanese textbooks, which has been identified by Yabuki-Soh (2013) as an area for improvement. While the work of Matsumoto (2018) provided valuable insights into NMCCs as a single construction with the potential for multiple interpretations, no known study for an English-speaking readership so far had identified (1) the POS categories and sub-categories that are most frequent in the modifying clause predicate and (2) the types of nouns that are not only used frequently in, but are also strongly attracted to the HN slot of the construction. Our study completed both tasks. The information that this study provides offers valuable insights for the treatment of NMCCs in pedagogical grammars, such as which realizations of NMCCs are most frequent and hence worth teaching. Moreover, our study suggests the need to treat NMCCs as a network of constructions with a wide range of potential slot fillers, each with their own level of attraction to or repulsion of the construction.

The main limitation of the present study is its scope in terms of both breadth and depth. The breadth was restricted to the collexeme analysis of only the HN slot of the NMCC; examinations of the modifying clause predicate beyond the rightmost SUW could provide additional important insights into JFL learners' NMCC use. Future extensions of this research could include further investigations using quantitative methods such as collostructional analysis into what slot fillers are typical in the modifying clause predicate of the construction. The depth of the NMCC analysis did not exceed the level of granularity necessary for the current inquiry; the richness of the NMCC data suggests the potential for more detailed and systematic syntactic–semantic analyses in future studies. Further, JFL learners were treated as a homogeneous group, obscuring the potential effects of variation in L1 backgrounds and levels of proficiency on L2 learners' construction repertoire and its relation to L1 usage.

To conclude, JFL learners' use of NMCCs, identified as a learning challenge in second language acquisition studies, had been examined insufficiently, mostly in the context of experimental studies. By adopting a corpus approach and targeting the NMCC as a collostruction, we were able to provide useful insights into the structure of the NMCC, its subtypes, and typical lexical associations in L1 and L2 spoken Japanese. The findings presented here arguably lay the groundwork for future studies seeking to explore realizations of the construction at different levels of granularity and to improve their treatment in JFL pedagogical materials.

Appendix A: Classification of RC types.

Classification of RC types		
RC type	Relativized structure	Original sentence
SU	<i>[PC-o katta] gakusei</i> PC-ACC bought student 'the student who bought a PC'	<i>gakusei-ga PC-o katta</i> student-NOM PC-ACC bought 'A student bought a PC.'
DO	<i>[Ken-ga tabeta] piza</i> Ken-NOM ate pizza 'the pizza that Ken ate'	<i>Ken-ga piza-o tabeta</i> Ken-NOM pizza-ACC ate 'Ken ate a pizza.'
IO	<i>[Ken-ga purezento-o ageta] gakusei</i> Ken-NOM present-ACC gave student 'the student whom Ken gave a present'	<i>Ken-ga gakusei-ni purezento-o ageta</i> Ken-NOM student-to present-ACC gave 'Ken gave a student a present.'
OBL	<i>[Ken-ga tomatta] hoteru</i> Ken-NOM stayed hotel 'The hotel that Ken stayed at.'	<i>Ken ga hoteru-ni tomatta</i> Ken-NOM hotel-LOC stayed 'Ken stayed at the hotel.'
GEN	<i>[bōjifurendo-ga nakunatta] jōsei</i> boyfriend-NOM passed:away woman 'The woman whose boyfriend passed away'	<i>zyoseee-no bōjifurendo-ga nakunatta</i> woman-GEN boyfriend-NOM passed:away 'The woman's boyfriend passed away.'

Note. Adapted from Ozeki and Shirai (2007: 178).

Appendix B: Metadata collected in 20-item questionnaire.

1	調査地 <i>chosachi</i> 'study region'
2	性別 <i>seibetu</i> 'sex'
3	身分 <i>mibun</i> '(social) status'
4	職業経験 <i>shokugyō keiken</i> 'work experience'
5	出身国 <i>shushinkoku</i> 'country of origin'
6	年齢 <i>nenrei</i> 'age'

(continued)

-
- 7 現在住んでいる国
genzai sundeiru kuni
 ‘country of residence’
 住んでいる年数
sundeiru nensū
 ‘years of residence’
- 8 母語（一番強い言語）
bogo (ichiban tsuyoi gengo)
 ‘mother tongue (strongest language)’
- 9 家族の母語
kazoku no bogo
 ‘family’s mother tongue’
- 10 住んでいるところで日常的に日本語が話されているか。
sundeiru tokoro de nichijōteki ni Nihongo ga hanasareteiru ka.
 ‘Is Japanese routinely used where you live?’
 「はい」の場合は、誰が話していますか。
[Hai] no bāi wa, dare ga hanashiteimasu ka.
 ‘If [yes] who speaks it?’
- 11 親しい友達に日本語母語話者はいますか。
shitashii tomodachi ni Nihongobogowasha wa imasu ka.
 ‘Do you have a close friend who is a native Japanese speaker?’
- 12 日本語の授業以外で、どのような時に日本語を使いますか。
Nihongo no jūgyō igai de, dono yō na toki ni Nihongo otsukaimasu ka.
 ‘Outside of Japanese class, when do you use Japanese?’
- 13 母語以外に日常的に使える言語はありますか。
bogo igai ni nichijōteki ni tsukaeru gengo wa arimasu ka.
 ‘Other than your mother tongue, is there a language you use regularly?’
 「はい」の場合は、その言語はなんですか。
[Hai] no bāi wa, sono gengo wa nan desu ka.
 ‘If [yes], what language is it?’
- 14 外国語の授業以外で、授業中に先生が話す時に使用していた言語を教えてください。
gaikokugo no jūgyō igai de, jūgyōchū ni sensei ga hanasu toki ni shiyō shiteita gengo o oshiete kudassai.
 ‘Outside of your foreign language class(es), what language do your instructors use in class?’
- 15 日本語を学習し始めたきっかけはなんですか。
Nihongo ogakushūshi hajimeta kikkake wa nan desu ka.
 ‘What was your impetus for studying Japanese?’
- 16 現在、どのように日本語を学んでいますか。
genzai, dono yō ni Nihongo omanandeimasu ka.
 ‘How do you currently study Japanese?’
- 17 日本語では次のどの生活をしていますか。普段しているものを選んでください。
Nihongo dewa tsugi no dono seikatsu oshiteimasu ka. fudan shiteiru mono o erande kudasai.
 ‘What kind of activities do you do in Japanese? Choose all that apply.’
- 18 現在とこれまでに教育機関で日本語を勉強したことがありますか。
genzai to kore made ni kyōikukikan de Nihongo o benkyōshita koto ga arimasuka.
 ‘Have you ever studied Japanese at an educational institution?’

(continued)

	「はい」の場合、期間と時期を教えてください。
	<i>[Hai] no bāi, kikan to jiki o oshiete kudasai.</i>
	‘If [yes], please indicate the length of time.’
19	これまでに勉強した日本語の教科書を分かる範囲で教えてください。
	<i>Koremade ni benkyōshita Nihongo no kyōkasho o waku hani de oshiete kudasai.</i>
	‘What Japanese textbooks have you used up to this point to be best of your knowledge?’
20	《日本以外》日本に行ったことがありますか。
	<i><Nihon igai> Nihon ni itta koto ga arimasu ka.</i>
	‘<Non-Japanese Residents> Have you been to Japan?’
	《日本在住者》以前にも日本に来たことがありますか。
	<i><Nihon zaijūsha> izen nimo Nihon ni kita koto ga arimasu ka.</i>
	‘<Japanese Residents> Have you come to Japan before (becoming a resident)?’
	「はい」の場合、期間・目的を教えてください。
	<i>[Hai] no bāi, kikan to mokuteki o oshiete kudasai.</i>
	‘If [yes], please indicate the length of time and reason for coming.’

Note. Adapted from Sakoda et al. (2016, 2020).

Appendix C: Structure of I-JAS dialogue task.

Stage	Focus	Sample topics/questions
1	Warm-up	N/A
2	Questions about past experiences	Incentive to learn Japanese; favorite books, dramas, etc.; hometown landscape, sightseeing spots, etc.
3	Questions about the future	Dream(s) for the future
4	Opinion statements	Do you want to live in the city or the countryside? Do you think money or time is more important? etc.
5	Cool down	N/A

Note. Adapted from Sakoda et al. (2020: 38).

Appendix D: JFL learners’ proficiency – J-CAT scores.

Level	Participants	Percentage
Beginner	37	3.70
Intermediate low	145	14.50
Intermediate mid	318	31.80
Intermediate high	298	29.80
Advanced low	160	16.00
Advanced mid	39	3.90
Near native	3	3.00
Total	1,000	

Note. Adapted from Sakoda et al. (2020: 85).

References

- Comrie, Bernard. 1996. The unity of noun-modifying clauses in Asian languages. *Pan-Asiatic Linguistics: Fourth International Symposium on Languages & Linguistics* 3. 1077–1088.
- Comrie, Bernard. 2002. Typology and language acquisition: The case of relative clauses. In Anna Giacalone Ramat (ed.), *Typology and second language acquisition*, 19–37. Berlin & New York: De Gruyter Mouton.
- Comrie, Bernard. 2007. The acquisition of relative clauses in relation to language typology. *Studies in Second Language Acquisition* 29(2). 301–309.
- Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19(1). 61–74.
- Flach, Susanne. 2021. *Collostructions: An R implementation for the family of collostructional methods*. Package version v.0.2.0. <https://sfla.ch/collostructions/>.
- Giacalone Ramat, Anna. (ed.). 2002. *Typology and second language acquisition*. Berlin & New York: De Gruyter Mouton.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2013. Constructionist approaches. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, 14–31. Oxford, UK & New York: Oxford University Press.
- Gries, Stefan Th. 2019. 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics* 24(3). 385–412.
- Imai, Shingo, Sukero Ito, Yoichi Nakamura, Kenichi Kikuchi, Yayoi Akagi, Hiromi Nakasono, Akiko Honda & Takekatsu Hiramura. 2009. Features of J-CAT Japanese computerized adaptive test. In David J. Weiss (ed.), *Proceedings of the 2009 GMAC conference on computerized adaptive testing*. <http://www.iacat.org/sites/default/files/biblio/cat09imai.pdf> (accessed 15 February 2024).
- Kamada, Osamu. 1999. KY-kōpasu to dainigengo toshi no Nihongo no shūtoku kenkyū [KY corpus and acquisition research of Japanese as a second language]. In *Heisei 8~10nendo kagaku kenkyūhi hoztokin kenkyū seika hōkokushyo, Kiban Kenkyū: Dainigengo toshi no Nihongo shūtoku Kenkyū ni kansuru aōgō kenkyū* [The report for grant-in-aid for scientific research, 1996–1998: Comprehensive research on the acquisition of Japanese as a second language], 335–350. Nagoya: Nagoya University.
- Kamada, Osamu. 2006. KY-kōpasu to Nihongo kyōiku kenkyū [KY corpus and Japanese language education research]. *Nihongo Kyoiku* [Journal of Japanese Language Teaching] 130. 42–51.
- Kidd, Evan (ed.). 2011. *The acquisition of relative clauses: Processing, typology, and function*. Amsterdam: John Benjamins.
- Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1). 63–99.
- Koiso, Hanae, Masayuki Asahara, Salvatore Carlino, Ken'ya Nishikawa, Kazuki Aoyama, Yuichi Ishimoto, Aya Wakasa, Michiko Watanabe, Yoshimi Yoshikawa, Nobuko Kibe & Kikuo Maekawa. 2020. Speech corpora in NINJAL, Japan demonstration of corpus concordance systems: Chunagon and Kotonoha. *Proceedings of the 3rd International Symposium on Linguistic Patterns in Spontaneous Speech (LPSS 2019)*, 8–12. Institute of Linguistics, Academia Sinica, Taipei. <https://doi.org/10.15084/00003051>
- Matsumoto, Yoshiko. 1989. Japanese-style noun modification...in English. *Berkeley Linguistics Society* (BLS) 15. 226–237.
- Matsumoto, Yoshiko. 1997. *Noun-modifying constructions in Japanese: A frame-semantic approach*. Amsterdam: John Benjamins.

- Matsumoto, Yoshiko. 2017. General noun-modifying clause constructions in Japanese. In Yoshiko Matsumoto, Comrie Bernard & Sells Peter (eds.), *Noun-modifying clause constructions in languages of Eurasia: Rethinking theoretical and geographical boundaries*, vol. 116, 23–43. Amsterdam: John Benjamins.
- Matsumoto, Yoshiko. 2018. Clausal noun modification. In Yoko Hasegawa (ed.), *The Cambridge handbook of Japanese linguistics*, 463–484. Cambridge: Cambridge University Press.
- Makino, Seiichi & Michio Tsutsui. 2011 [1989]. *A dictionary of basic Japanese grammar*. Tokyo: The Japan Times.
- Mori, Yoshiko, Atsushi Hasegawa & Junko Mori. 2020. Trends and developments of L2 Japanese research in the 2010s. *Language Teaching* 54(1). 90–127.
- National Institute for Japanese Language and Linguistics. n.d. *Chunagon: Kōpasu kensaku apurikēshon*. Chunagon: Corpus Search Application. <https://chunagon.ninjal.ac.jp>.
- Ozeki, Hiromi & Yasuhiro Shirai. 2007. Does the noun phrase accessibility hierarchy predict the difficulty order in the acquisition of Japanese relative clauses. *Studies in Second Language Acquisition* 29(2). 169–196.
- R Core Team. 2022. *R: A language and environment for statistical computing* (version 4.2.2) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Sakoda, Kumiko, Shinichiro Ishikawa & Jeho Lee (eds.). 2020. *Nihongo gakushūsha kōpasu I-JAS nyūmon: Kenkyū kyōiku ni do tsukaika* [Introduction to the I-JAS: application for research and teaching]. Tokyo: Kurosio Publishers.
- Sakoda, Kumiko, Madoka Konishi, Aiko Sasaki, Wakako Suga & Yoko Hosoi. 2016. Tagengo bogo no Nihongo gakushūsha Ōdan Kōpasu [International corpus of Japanese as a second language]. *Kokugoken Purojekuto Rebyū [NINJAL Project Review]* 6(3). 93–110.
- Takara, Nobutaka. 2012. The weight of head nouns in noun-modifying constructions in conversational Japanese. *Studies in Language* 36(1). 33–72.
- Teramura, Hideo. 1969. The syntax of noun modification in Japanese. *Journal-Newsletter of the Association of Teachers of Japanese* 6(1). 64–74.
- Tono, Yukio, Makoto Yamazaki & Kikuo Maekawa. 2013. *A frequency dictionary of Japanese*. London: Routledge.
- Yabuki-Soh, Noriko. 2007. Teaching relative clauses in Japanese: Exploring alternative types of instruction and the projection effect. *Studies in Second Language Acquisition* 29(2). 219–252.
- Yabuki-Soh, Noriko. 2013. Types of Japanese noun-modifying clauses used in JFL textbooks. *Japanese Language & Literature* 47(1). 59–92.