L1/L2 acquisition of Japanese high-applicative expressions

Akitaka Yamada*
Osaka University
a.yamada.hmt@osaka-u.ac.jp

1 Introduction

Long after the idea of construction networks was introduced by Berkeley linguists in the 1990s, analytical details keep being updated and developed in new and important ways (Hilpert 2014; Diessel 2023, for a recent review). One important change is the proposal of 'quantified edges' (Yamada, 2019; Diessel, 2023). By analyzing L1 and L2 learners' acquisition of Japanese high-applicative (HA) constructions, this current study advances this theoretical movement, by proposing language acquisition is a learning process of weights (of edges) in a given construction network.

More specifically, by using an elaborate statistical method (i.e., the State-Space Model, SSM, Hagiwara 2021; Yamada 2022a, 2023a,c), this study demonstrates how two HA constructions—*-te kure* and *-te moraw*—are acquired by L1 and L2 learners. It shows that one construction is much harder to learn than the other, despite the two having almost the same meaning. It is claimed that the difference arises from whether a competing or confounding construction scheme is already entrenched in a given network. It is argued that the presence of such a misleading construction scheme makes it difficult for a beginner to establish an appropriate or intended association within the network.

2 Previous literature

2.1 High applicatives (HAs) in Japanese: commonality

When expressing an optional event participant, English relies on an adjunct phrase. For example, consider the sentences in (1). The verb *run* cannot introduce a participant who receives a benefit from the depicted event. But when the preposition *for* is used, a beneficiary is now felicitously expressed.

(1) a. *Taro ran Hanako.
b. Taro ran [for Hanako].

Likewise, Japanese utilizes the adjunct phrase *-no tame-ni*, which roughly corresponds to the English phrase *for the sake of* to achieve the same goal:

(2) a. **Taro-ga Hanako-ni hasit-ta.*
Taro-NOM Hanako-DAT run-PST
'Taro ran for Hanako (intended).'

b. *Taro-ga* [*Hanako-**no tame-ni***] *hasit-ta.*
Taro-NOM Hanako-GEN sake-for run-PST
'Taro ran for (the sake of) Hanako.'

An adjunct phrase is, however, not the only linguistic device to introduce an optional argument. Some languages have developed special morphemes to manipulate the valency already determined by the predicate, and such morphemes are called high-applicatives (Kuno, 1973; McGinnis, 2001, 2005; McGinnis and Gerdts, 2003; Jeong, 2006; Pylkkänen, 2002, 2008). Japanese is a language

*This research has been supported by the JSPS Grant-in-Aid for Scientific Research (C) (#22K00507).

with many high applicative expressions and is, therefore, often cited in discussions of applicatives in syntax (Pylkkänen, 2008; Nishigauchi, 2014; Hasegawa, 2018; Aoyagi, 2020; Ikawa, 2022), semantics (Kuno and Kaburaki, 1977; Kubota and Uegaki, 2009; Bosse et al., 2012; Tomioka and Kim, 2017), pragmatics (Yamada, 2020, 2022b), and historical linguistics (Shiina, 2021; Yamada, 2023b). Consider the example below. Here, *-te moraw* '-CV HA' (n.b., CV is a converb suffix) introduces an applied beneficiary argument (here, *Hanako*) and the sentence now includes two event participants.

(3) a. *Taro-ga hasit-ta.*
Taro-NOM run-PST
'Taro ran.'

b. *Hanako-ga Taro-ni hasir-**te morat**-ta.*
Hanako-NOM Taro-DAT run-CV HA-PST
'Taro ran, which benefactively affected Hanako.'

Japanese also possesses a semantically synonymous benefactive expression *-te kure*. Since the logical form $P \wedge \neg Q$ becomes a contradiction if and only if P and Q are semantically equivalent ($P = Q$), we can conclude from the data in (4) that the core meanings of *-te moraw* and *-te kure* are truth-conditionally identical. Note that Japanese has a third option, *-te age (yar)*, but due to limited space, we do not discuss it in this paper.

(4) a.*[$_P$ *Taro-ga hasir-**te kure**-ta*] *ga,* [$_Q$ *Taro-ni hasir-**te moraw***]*-anak at-ta.*
Taro-NOM run-CV HA-PST but Taro-TOP run-CV HA-NEG COP-PST
'*Taro ran, from which I benefited, but he did not run, although I would have benefited from his running.'

b.*[$_P$ *Taro-ni hasit-**te morat**-ta*] *ga,* [$_Q$ *Taro-wa hasir-**te kure***]*-nak at-ta*
Taro-DAT run-CV HA-PST but Taro-TOP run-CV HA-NEG COP-PST
'*Taro ran, from which I benefited, but he did not run, although I would have benefited from his running.'

2.2 High applicatives (HAs) in Japanese: differences

While being synonymous, *-te kure* and *-te moraw* have two major differences (Kuno, 1978; Hasegawa, 2018). First, they are different in case alignment (Yamada and Nagano, 2023, to appear). Unlike in the baseline example in (5), AG(ent) in (6)a receives the dative case, with BEN(eficiary) being marked with the nominative case. Conversely, in (6)b, AG remains nominative-marked (Yamada and Nagano, 2023, to appear). Note that, while implying the presence of a beneficiary, *-te kure* never overtly introduces an applied benefactive argument. Compare (6)b with (7): if one wishes to overtly pronounce the beneficiary, one must exploit the adjunct phrase strategy, as seen in (2)b.

(5) *Taro-**ga*** [AG] *hasit-ta.*
Taro-NOM run-PST
'Taro ran.'

(6) a. *Hanako-**ga*** [BEN] *Taro-**ni*** [AG] *hasit-**temorat**-ta.*
Hanako-NOM Taro-DAT run-HA-PST
'Taro ran, from which Hanako benefited.'

b. *Taro-**ga*** [AG] *hasit-**tekure**-ta.*
Taro-NOM run-HA-PST
'Taro ran, from which [I] benefited.'

(7) *Watasi-*{**ni/notameni*}[BEN] *Taro-**ga*** [AG] *hasit-**tekure**-ta.*
I-DAT/for Taro-NOM run-HA-PST
'Taro ran for me.'

Second, the two constructions differ in viewpoint restriction (Kuno 1987; see Kurihara and Nakahama 2010, 142–144 for a review). Compare (6) with (8). For *-te kure*, the referent of the

subject cannot be the speaker him/herself, whereas for *-te moraw* it is the dative marked object that cannot be the first person. Kuno (1978) argues that the *-te moraw* construction has a viewpoint rule such that the speaker is more sympathetic to the subject than to the object. Since the speaker cannot be more sympathetic to *Hanako* than *watasi* 'I,' (8)a is judged infelicitous. Reversely, as in (8)b, *-te kure*, it must be the case that the speaker feels less sympathy to the subject than the beneficiary. (8)b contradticts this construction specific constraint.

(8) a. *?*Hanako-**ga*** [BEN] *watasi-**ni*** [AG] *hasit-**temorat**-ta.*
Hanako-NOM I-DAT run-HA-PST
'I ran, from which Hanako benefited (intended).'

b. **Watasi-**ga** [AG] *hasit-**tekure**-ta.*
I-NOM run-HA-PST
'I ran, from which [I] benefited (intended).'

2.3 Acquisition of applicatives

Researchers working on the acquisition of Japanese HAs have made the following observations (Horiguchi 1979; Clancy 1985; Uyeno et al. 1978; Ishiguro 1985; Tashiro 1995; Okabe 2005; Nakahama and Kurihara 2007; Takemura 2010; Okabe 2011; Ohba and Deen 2020, 2022, amo). First, adult Japanese L1 speakers are known to consistently pick a single viewpoint during a narrative (unless the scene has substantially changed; Kuno 1978; Nakahama and Kurihara 2007), whereas L2 learners do not maintain the viewpoint as often (Tashiro, 1995; Takemura, 2010).

Second, young Japanese L1 learners are known to acquire such a viewpoint (or, empathy) restriction in an incremental fashion, and not simultaneously. Ohba and Deen (2020, 2022), for example, reveal that Japanese children as young as Age 4 showed adult-like knowledge of *-te moraw*, while the empathy-encoding of *(-te) kure* is learned at around Age 6. Note, however, that this finding holds only when arguments are unpronounced. When arguments are overtly pronounced, it is known that *(-te) kure* is known to be acquired in a much earlier stage than *(-te) moraw* (Uyeno et al., 1978; Ishiguro, 1985; Okabe, 2005).

Despite the progress in research, there remain some issues to be invesitigated. First, previous L1 studies were all concerned with preschool children. Given that even 6-year old children have some difficulty acquiring these constructions, it is likely that elementary school children may still have some trouble. When do they become competent in using them as adults? Second, less attention has been paid to the comparison between L1 and L2 acquisition. The explanation proposed to account for the difficulty—for example, the view that the effort to learn *-te moraw* is attributed to the irregular semantic-role of the *ni*-marked NP—seems to apply to L2 learners as well. Is this prediction borne out? Can we see a clear delay in acquiring *-te moraw* in the L2 grammar as well?

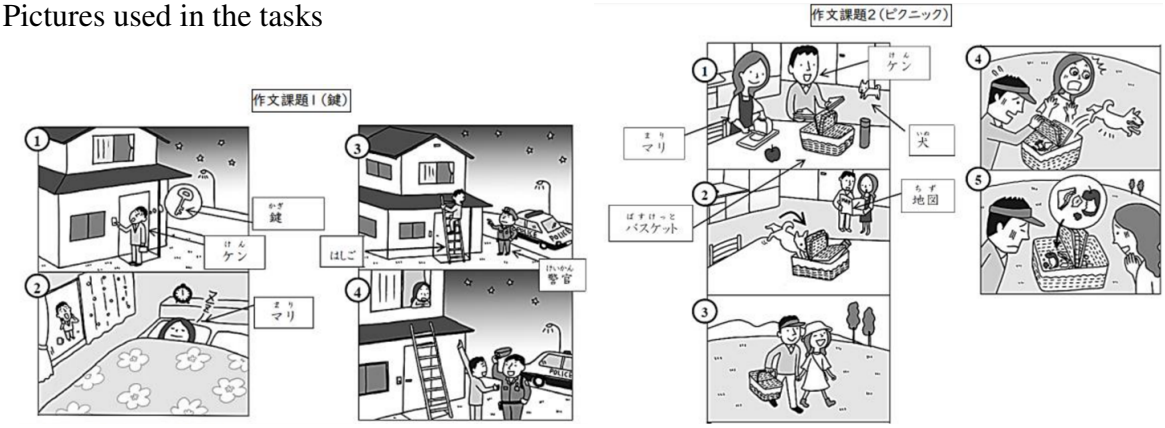
3 Data

3.1 Picture description task

In the 1970s, it was discovered that not only L1 (Brown, 1973; Pinker, 1984) but also L2 learners (Krashen, 1982, 1985) acquire morphosyntactic patterns in certain orders. Such early research findings are, however, often criticized for their small/unbalanced samples or inappropriate uses of statistics. Recently, however, I-JAS (Sakoda, 2020) and JASWRIC (Ishikawa et al., 2023)—two well-designed corpora of L2 and L1 learners of Japanese—were released, enabling us to conduct cross sectional corpus research in a well-controlled circumstance. JASWRIC is a collection of picture description tasks of L1 learners (from the first grade of elementary school to the first grade of college), while I-JAS contains the L2 learners' results for the same tasks. The pictures used in these corpora are shown in (9) (Ishikawa et al. 2023, 397-398). Since the same pictures are used for each group, we

can compare the narratives of L1 and L2 learners in a controlled setting.

(9) Pictures used in the tasks



Below are a few representative examples from L1 learners: (10a) (first-year elementary school), (10b) (first-year high school), (11a) (second-year junior high), and (11b) (third year high school).

(10) *-te kure*

a. *keisatukan-wa yurusi-**te** **kure**-ta.*
policeman-TOP forgive-CV HA-PST
'The policeman forgave (him), from which Ken benefitted.' [Task 1]

b. *Ken-wa* [*Mari-ga tuku-**te** **kure**-te i-ru*] *tyuusyoku-o basuketto-ni tume-masi-ta.*
Ken-TOP Mary-NOM make-CV HA-CV PRF-PRS lunch-ACC basket-into put-AH-PST
'Ken put the lunch [Mary made into the basket, from which Ken benefitted].' [Task 2]

(11) *-te moraw*

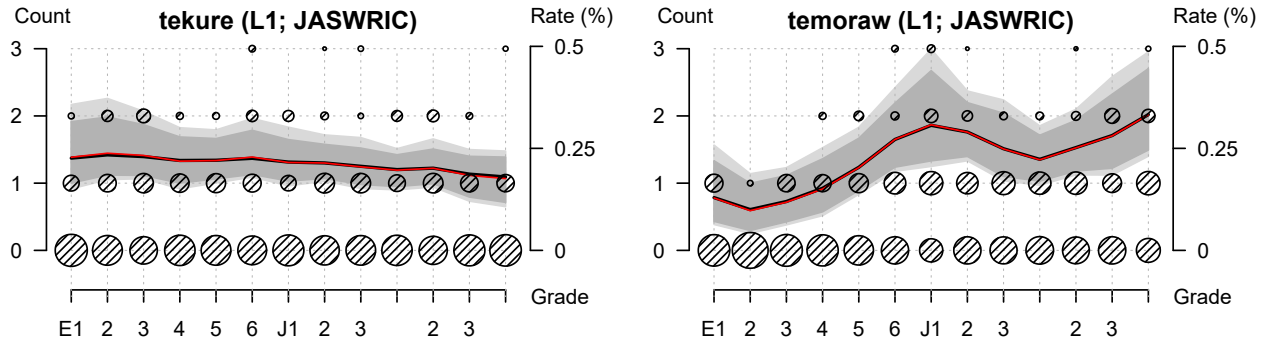
a. [*Mari-ni ake-**te** **mora**-oo-to*] *omoi-masi-ta-ga,* [...] [Task 1]
Mari-DAT open-CV HA-will-COMP think-AH-PST-although
'[Ken] decided to have Mary open [the door], from which he would have benefitted but ...'

b. *Hutari-tomo* [*kyuuzitu na node*], *petto-no inu-ni orusuban-o si-**te** **morai**,*
two-both holiday COP since pet-GEN dog-DAT staying at home-ACC do-CV HA,
yukkuri pikunikku-o tanosimu yoo des-u.
relaxed picnic-ACC enjoy seem COP-PRS
'Since it is a holiday, they both seem to have the dog—which is their pet—stay at home, from which they would have benefitted, and enjoy the picnic in a relaxed manner.' [Task 2]

3.2 Statistical model: the State-Space Model

Participants were asked to describe the pictures in their own words; some were imaginative, while others just provided an objective recounting of the images. If we were to look at the individual level, we would be overwhelmed by the variation among participants with the same language proficiency. However, at the group level (that is, by averaging the results), we would expect participants' idiosyncrasies to level out, and hence we could easily detect how the overall tendency grows and changes as language acquisition proceeds. Arguably, it is predicted that when one is young or at the beginner phase, the production rate starts at almost 0%. That rate, however, grows substantially to converge into a percentage by which the average competent language user produces constructions when describing the pictures. In what follows, we will test this hypothesis by estimating an average production rate per each grade (for L1 learners) and proficiency level (for L2 learners).

Certainly, one could just count how often each construction is produced at each stage. However, such a descriptive approach runs into a problem. While it reveals what the sample (i.e., the corpus



(a) The production of *-te kure*

(b) The production of *-te moraw*

Figure 1: Growth of high-applicatives in L1 grammar

data) is like, it does not say anything about what the population is like. To resolve this shortcoming, we used a sophisticated model suitable for time series data, the State-Space Model. Using Bayesian statistics, we estimated the posterior distributions of the parameters for each grade/proficiency, as well as discussing how confident we were about the production rate for each stage.

More specifically, in order to analyze how the frequency of their uses of V-*te kure* and V-*te moraw* changed over grade (L1 learners) and proficiency (L2), we built the model in (12).

(12)
$$y_i^{(t)} \sim \text{Poisson}(\lambda_i^{(t)})$$
$$\frac{\lambda_i^{(t)}}{a_i} = \exp(\eta_i^{(t)})$$
$$\eta_i^{(t)} = \beta_0^{(t)} \sim N(\beta_0^{(t-1)}, \sigma^2)$$

Note that $y_i^{(t)}, \lambda_i^{(t)}, a_i^{(t)}$ are the frequency of HAs, the parameter of Poisson Distribution, and the offset of the *i*-th learner at time *t*. In our context, this t refers to the grade (for L1 learners) and proficiency level (for L2 learners). $\beta_0^{(t)}$ is designed to follow a random walk with Variance σ^2.

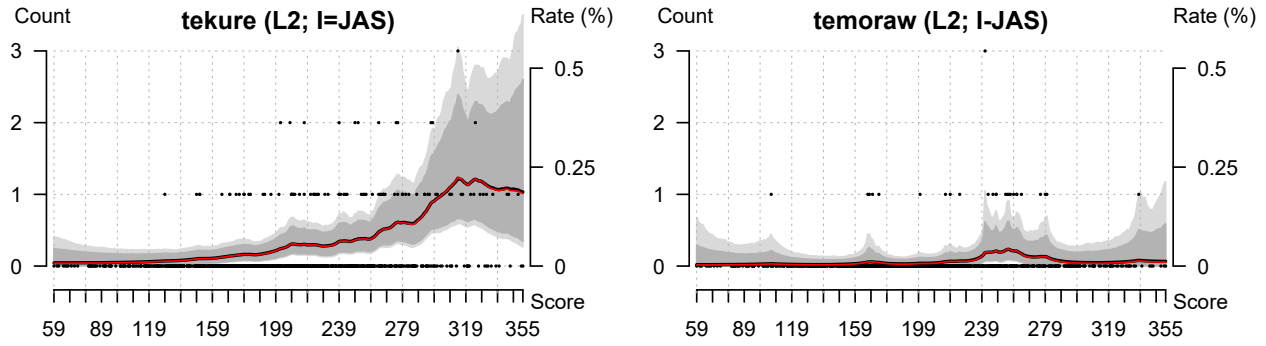
4 Results

The posterior distributions were estimated based on the Hamiltonian Monte Carlo Simulation, using Stan on R (Hagiwara 2021; Yamada 2022a, 2023a,c; R Core Team 2023; Stan Development Team 2023). Table 1 summarizes the values of the parameters set up for this numerical computation. All the R-hat values were checked to be lower than 1.01 (Vehtari et al., 2021), which we take to be an indication of convergence (Gelman et al., 2013).

setting	value
warm-up periord	4,000
iteration	15,000
thinning	2
adapt_delta	0.995
chain	4

Table 1: Values specified when running Stan on R

The estimated posterior distributions of $\lambda^{(t)}$'s are shown in Figures 1 through 2. They are plotted according to the following conventions:



(a) The production of *-te kure*

(b) The production of *-te moraw*

Figure 2: Growth of high-applicatives in L2 grammar

- The horizontal axis is the grade of the L1 learners (JASWRIC) and the proficiency level of the L2 learners measured by the J-CAT score (I-JAS).
- The y-axis (on the left) represents how many HAs were used during the tasks. [descriptive stats]
  - The bubbles (Figure 1) and dots (Figure 2) represent how often participants used HAs. For example, Figure 1(a) shows that most of the first year students did not use *-tekure*, but some used it just once, and a few used it twice in their writing tasks.
- The y-axis (on the right) represents estimated $\lambda^{(t)}$ (the production rate) [inferential stats]
  - The solid line is the posterior mean for each grade/score.
  - The grayed areas represent the 95% and 99% posterior Credible Intervals.

4.1 L1 learners

Consider the results from L1 learners (Figure 1). The production rate of U1 (first year university students) is reasonably seen as representing the rate at which a competent adult native speaker uses an HA to describe the aforementioned pictures. Figure 1(a) shows that even the first year elementary students (E1) have already acquired the grammar of *-te kure* at much the same rate as adults. In contrast, Figure 1(b) suggests that the E1 learners' rate of *-te moraw* is lower than that of the baseline adult speakers, and the acquisition of *-te moraw* takes time to complete over the elementary school days. Recall from Section 2.3 that, among the two differences (case-alignment and viewpoint), preschool children have already become accustomed to the viewpoint restrictions (Ohba and Deen, 2020, 2022). So the asymmetry shown here is considered to be associated with the acquisition of case alignment, as we saw in Section 2.2 (Uyeno et al., 1978; Ishiguro, 1985; Okabe, 2005, 2019). The case alignment of *-te moraw* is more difficult to acquire.

4.2 L2 learners

4.2.1 Overall tendency

Figure 2(a) demonstrates that L2 learners gradually acquire the use of *-te kure* in their intermediate stages (around the 200s in the J-CAT score). In contrast, it is clearly seen that even the advanced L2 learners (students in the 300s) avoid using *-te moraw* (Figure 2[b]). Just as with L1 learners, L2 learners also have difficulty using this construction.

4.2.2 Error analysis

There is an asymmetry in the error pattern between the two constructions. First, out of the 29 instances of *-te moraw* of L2 learners, 11 examples were incorrectly used, which amounts to 38.0% of the *-te moraw* instances. In nine of these cases, *-te moraw* is used when *-te kure* is expected, as in (13).

(13) *Yatto, keisatu-san-wa rikaisi-**temorai**-masi-ta.* (intended: *rikaisi-**tekure**-masi-ta*)
finally policeman-Mr.-TOP undertand-HA-AH-PST understand-HA-AH-PST
'The policeman understood [that], from which [Ken] benefited.' (I-JAS, IID11-SW2)

Second, and conversely, *-te kure* had 112 instances but only 7 examples were erroneous (e.g., (14)) (6.2%). Although no clear error trend was found, it is safe to conclude that L2 learners make mistakes much less frequently.[1]

(14) *Tuma-ga doa-o ake-**tekure**-tehosii* (intended: *Tuma-ni doa-o ake-**tehosii***)
wife-NOM door-ACC open-HA-want wife-DAT door-ACC open-want
'[Ken] wants his wife to open the door for him.' (I-JAS, CCH50-SW2)

5 Discussion

The results of this study consistently show that *-te moraw* is more difficult to acquire both for L1 and L2 learners. It is thus important for us to ask why it is so difficult.

5.1 Usage-based view on language acquisition

In construction grammar, grammar is seen as a dynamic system of constructions—it is made up of conventionalized form-function (syn-sem) units. Learners acquire the usage of a construction in interaction with the context where it used, based on the domain-general cognitive abilities unique to human beings (Tomasello, 1999). However, constructions are not stored in our grammar in a haphazard, random way. Rather, *a structured inventory of linguistic units* (Langacker, 1987, 73) is analyzed as being represented as schematic, hierarchical networks, in which constructions are interlinked (Goldberg 1995; Hilpert 2014; see Diessel 2019, 2023 for relations beyond a hierarchical association).

When children are extremely young (e.g., around two years-old), they do not exhibit adult-like productivity/creativity. They merely repeat or slightly modify what is available from the previous discourse (Lieven et al., 2003; Hilpert, 2014). However, learners in this present study were much older, at least 6 years of age. It is, therefore, safe to say that they are learning their language as adults through categorization and schematicization, for example (Brooks and Tomasello, 1999). This suggests that the aforementioned error patterns reflect their process of acquiring the target grammar via their cognitive abilities. But how are these errors and abilities related?

5.2 Language acquisition as an estimation of weights in a construction network

To answer this question, it helps to examine a newer trend in construction grammar. In the classic model of construction grammar, originally developed in the 1990s, the researchers' primary concerns were the 'nodes,' that is, the information of each construction scheme. For example, a subject-relative clause and an object-relative clause (child nodes) are assumed to have the same property inherited from a higher-level construction scheme (parent nodes) for a finite relative clause.

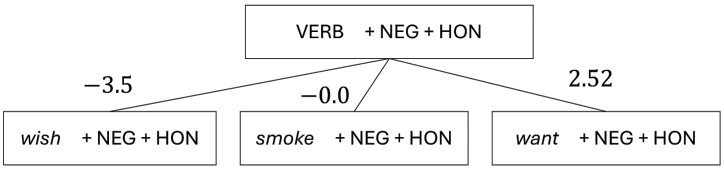
Of course, the classic model, too, paid attention to 'edges.' For instance, types of inheritance links were extensively discussed, with the aim to reveal the nature of a given constructional network (i.e., polysemy, subpart, instance, and metaphorical links; Goldberg 1995). However, after the 2000s, in tandem with the developments of corpora and experimental methods, edges were analyzed as an 'association,' rather than merely being treated as a 'link,' in line with many scientific disciplines beyond linguistics (Elman et al., 1996; Buchanan, 2002; Diessel, 2023). While a wide range of research possibilities was opened up by considering networks as associations—see Diessel (2023) for details—this present study would like to pick a single important aspect: by framing an edge as an association, we will discuss the strengths of two nodes. Associations are 'quantifiable' and more

[1] **Bouletic Modality**. The error in (14) is indicative of the fact that, semantically, applicatives have some modal meanings. See Yamada (to appear) for a detailed semantic analysis.

amenable to linguistic frameworks that investigate probabilistic relations (Bybee and Hopper, 2001; Bod, 2009; Suethanapornkul, 2020).

Consier one example. In conducting a corpus study of constructional alternation between polite negative forms in Japanese, Yamada (2019) provided an explicit, hence falsifiable, statistical model that estimates the association strengths among a construction scheme (a parent and a child node). As illustrated in (15), an elaborate machine learning method estimates a quantified score for each edge of a construction network, which indicates what variant each realization prefers, and to what extent (ibid.:127).[2]

(15)



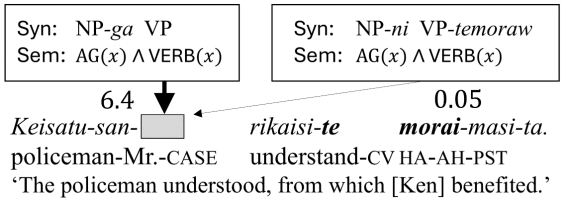
Our intuition on language is subtle, gradual, and delicate. Weights within a constructional network of this kind well-capture this otherwise ineffable, elusive, and ephemeral aspect of the grammar. In Yamada's study, it is the machine that estimates the weight values based on the corpus data. But if we replace the machine with a human, it is interpreted as a model of human language acquisition process.

5.3 Weight estimation in multiple inheritances

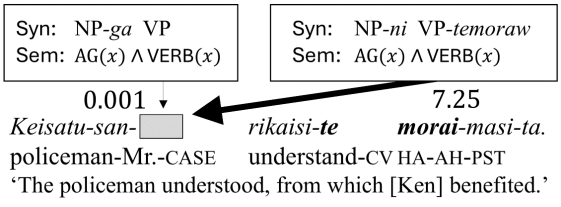
If we put forth this idea, it naturally follows that becoming a competent language user—or acquiring construction networks of a given language—is to converge to estimate the weights of all relevant association scores. With this in mind, we can now explain why *-te moraw* is difficult to acquire, and how it nevertheless ends up being acquired.

Consider the error in (13). We saw in Section 2.2 that the baseline sentence (NP-*ga* VP) and the *-te moraw* construction (NP-*ni* VP) differ in how AG receives case marking. In the baseline configuration, AG receives the nominative marker *-ga* (= (5)), whereas in the *-te moraw* construction, AG must take the dative marking *-ni* (= (6)a). Language acquisition is assumed to proceed from a frequent, prototypical, and unmarked construction. So, at an early stage, learners rely more heavily on the baseline construction scheme when referring to AG even in the *-te moraw* construction, as illustrated in (16)a.

(16) a. Earlier stage



b. Later stage



This is because the learning has not yet converged. Consequently, temporal values are used as an estimated value for the association weight of the two nodes. Tentative as it is, the weight for the baseline construction scheme is much bigger than that for the *-te moraw* construction, so the gray-shaded blank slot in (16)a is realized as *-ga*, yielding the 'wrong' case marking.[3] However, in later

[2] **Convention.** In Yamada's (2019) study, he is concerned with a variable selection, as is discussed in Variation Theory (Cedergren and Sankoff, 1974). To this end, a positive value in (15) is used to indicate that the instantiation (e.g., VERB → *want*) prefers the new variant, while a negative value prefers the old, prescriptive form (e.g., VERB → *wish*) A value of 0 therefore shows that the instantiation (e.g., VERB → *smoke*) is neutral in construction choice. In contrast, the weights in (16) show how tightly the construction schemes are associated with the produced example. Despite this notational difference, both conventions share the important insight that language learning is a weight-identification process. It is left to future study to decide which approach is better in modeling human language acquisition processes.

[3] **Values in (16)**. The values used in (16) (i.e., 6.4, 0.05, etc.) are used for illustrative purposes only, and learners need

stages, as in (16)b, learners incrementally update the weight values so that they end up having a higher weight for the *-te moraw* construction. Now that the association strength between the instance and the *-te moraw* construction scheme is much tighter than that for the baseline, learners successfully use *-ni*.

The analysis so far gives a plausible answer to the question introduced at the beginning of this section: why is *-temoraw* difficult for learners regardless of the L1 and L2 distinction? This is because, regardless of the L1/L2 difference, it is only *-te moraw* that has the irregular case marking pattern. Since the *-te kure* construction maintains the baseline case alignment pattern, it is easier to use, causing an ordering in acquisition between the two HA constructions. By the same token, L2 learners tend to avoid using *-te moraw* as much as possible so as not to make a mistake, as indicated by Figure 2(b).

6 Conclusion and future directions

In the study of acquring HAs, it was not clear how elementary school children gradually acquire a construction scheme, and how it is different from, or similar to, that of L2 learners. As this study demonstrated, a detailed discussion would not be easy—or even possible—to make without well-controlled learner corpora, and elaborate statistical modeling. The results show a clear delay in the acquisition of the *-te moraw* construction, and this tendency is seen for both L1 and L2 learners.

Within the framework of construction grammar, it is claimed that the difficulty of *-te moraw* is due to the conflict of two competing constructional schemas. On the one hand, having been exposed to language used by care-givers, teachers, and/or other competent speakers, learners have discovered that the nominative-marked NP is typically an Agent. Frequency-wise, this mapping scheme is considered so entrenched in the grammar of language learners that beginners easily use it when producing a language. On the other hand, the *-te moraw* construction requires the Agent to be marked with a dative case marker. This scheme is less frequent. If there is a competing scheme of much heavier weight, this constructional scheme is easily overwhelmed, yielding the wrong case alignment. Hence, it is more difficult for both L1 and L2 learners to acquire.

To promote this idea, this paper proposes that a language learner is modeled as keeping estimating weights within the network from ambient language data. This theory predicts that during the trial-and-error phase (e.g., (16)a), learners with immature weight values are expected to produce 'incorrect' sentences as their 'best' outcome. This prediction is in agreement with the well-known U-shaped development (Marcus et al., 1992). Future research is expected to further reanalyze such well-known phenomena through the lens of weight-learning analysis, to better understand and, most importantly, model the human language acquisition processes.

References

Aoyagi, Hiroshi. 2020. "How high is high applicative in Japanese and Korean?" In *Japanese/Korean Linguistics*, Volume 28. 1–14.

Bod, Rens. 2009. "From Exemplar to grammar: a probabilistic analogy-based model of language learning." *Cognitive Science* 33 752–793.

Bosse, Solveig, Benjamin Bruening, and Masahiro Yamada. 2012. "Affected experiencers." *Natural Language and Linguistic Theory* 30 (4): 1185–1230.

Brooks, Patricia, and Michael Tomasello. 1999. "Young children learn to produce passives with nonce verbs." *Developmental Psychology* 35 29–44.

Brown, Roger. 1973. *A first language*. Cambridge, Mass: HUP.

to figure out the best set of parameter values, based on the data (ambient language use), just as machines do when they conduct a numerical computation, in which the values are gradually updated until they converge to the final values.

Buchanan, Mark. 2002. *Nexus: small worlds and the groundbreaking science of networks*. New York: W.W.Norton & Companys.

Bybee, Joan, and Paul Hopper. 2001. "Probabilistic relations between words: evidence from reduction in lexical production." In *Frequency and the emergence of linguistic structure*, 229–253, Amsterdam: John Benjamins.

Cedergren, Henrietta J., and David Sankoff. 1974. "Variable Rules: Performance as a Statistical Reflection of Competence." *Language* 50 (2): 333–355.

Clancy, Patricia M. 1985. "The acquisition of Japanese." *The crosslinguistic study of language acquisition* 1 373–524.

Diessel, Holger. 2019. *The grammar network: how linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.

Diessel, Holger. 2023. *The Constructicon: Taxonomies and Networks*. Cambridge: Cambridge University Press.

Elman, Jeffrey, Elizabeth Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett. 1996. *Rethinking innateness: a Connectionist perspective on development*. Cambridge: The MIT Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian data analysis [3rd edition]*. London: CRC Press.

Goldberg, Adele E. 1995. *Constructions: a construction grammar approach to argument structure*. University of Chicago Press: Chicago.

Hagiwara, Junichiro. 2021. *Time Series Analysis for the State-Space model with R/Stan*. Singapore: Springer.

Hasegawa, Nobuko. 2018. "Benefactives." In *The Cambridge Handbook of Japanese Linguistics*, edited by Hasegawa, Yoko 509–529, Cambridge: CUP.

Hilpert, Martin. 2014. *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.

Horiguchi, Sumiko. 1979. "Nenshooji no jukyuu hyoogen [Japanese preschoolers' giving- and receiving-expressions]." In *Kotoba no hattatu [Development of language]*, 51–76, Hiroshima: Bunka Hyoron Publishing.

Ikawa, Shiori. 2022. *On Agree Feeding Interpretation: Honorification Empathy, and Switch reference*. Ph.D. dissertation, Rutgers University.

Ishiguro, Hiroaki. 1985. "Developmental study of the comprehension of giving and receiving sentences in Japanese children: case and point of view." *The Japanese Journal of Psychology* 56 (4): 192–199.

Ishikawa, Shin'ichiro, Tatsuya Tomonaga, Ryohei Onishi, Toshiaki Okamoto, Naoki Katsube, Hisayo Kawashima, Tatsuya Kishimoto, and Reiko Muranaka. 2023. "Syootyuukoodaisei ni yoru nihongo ebyoosya sutoorii raitengu koopasu [Japanese Students' L1 Story Writing Corpus (JASWRIC): A New Dataset for Analysis of L1/L2 Japanese]." In *Proceedings of Language Resources Workshop*, 393–416.

Jeong, Youngmi. 2006. *The landscape of applicatives*. Ph.D. dissertation, The University of Maryland.

Krashen, Stephen D. 1982. *Principles and practices in Second Languaeg Acquisition*. Pregamon: OUP.

Krashen, Stephen D. 1985. *The Input Hypothesis: issues and implications*. New York: Longman.

Kubota, Yusuke, and Wataru Uegaki. 2009. "Continuation-based semantics for conventional implicatures: the case of Japanese benefactives." In *Proceedings of SALT 31*, 306–323.

Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, Mass.: The MIT Press.

Kuno, Susumu. 1978. *Danwa no bunpo [Grammar in discourse]*. Tokyo: Taishukan.

Kuno, Susumu. 1987. *Functional syntax: anaphora, discourse and emphathy*. Chicago: University of Chicago Press.

Kuno, Susumu, and Etsuko Kaburaki. 1977. "Empathy and syntax." *Linguistic Inquiry* 8 627–672.

Kurihara, Yuka, and Yuko Nakahama. 2010. "Sutoorii kootiku ni okeru siten [Perspectives in narratives]." In *Gengogaku to nihongo kyooiku VI [Linguistics and Japanese language education VI]*, 141–156, Tokyo: Kuroshio Publishers.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Vol.1 Theoretical Prerequisites*. Stanford: Stanford University Press.

Lieven, Elena, Heike Behrens, Jenny Speares, and Michael Tomasello. 2003. "Early syntactic creativity: a usage-based approach." *Journal of Child Language* 30 333–370.

Marcus, Gary F., Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, and Fei Xu. 1992. *Overregularization in language acquisition*. University of Chicago Press.

McGinnis, Martha. 2001. "Phases and the syntax of applicatives." In *Proceedings of NELS 31*, edited by Mchombo, Sam 183–199, Amherst, Mass: GLSA.

McGinnis, Martha. 2005. "UTAH at Merge: evidence from multiple applicatives." *MIT Working Papers in Linguistics* 49 183–200.

McGinnis, Martha, and Donna B. Gerdts. 2003. "A phase-theoretic analysis of Kinyarwanda multiple applicatives." In *Proceedings of the 2003 Annual Conference of the Canadian Linguistic Association*, 154–165, Department de linguistique et de didactique des langues, Úniversité à Québec à Montréal.

Nakahama, Yuko, and Yuka Kurihara. 2007. "Viewpoint setting in L1 and L2 Japanese narratives." In *Papers from the Sixth Annual Conference of the Japanese Society for Language Sciences*, 141–156.

Nishigauchi, Taisuke. 2014. "Reflexive binding: awareness and empathy from a syntactic point of view." *Journal of East Asian Linguistics* 23 157–206.

Ohba, Akari, and Kamil Ud Deen. 2020. "Acquisition of perspective and empathy verbs in Japanese." In *Proceedings of the 44th Boston University Conference on Language Development [BUCLD 44]*, 430–443.

Ohba, Akari, and Kamil Ud Deen. 2022. "Acquisition of emphathy in child Japanese." *Language Acquisition* 29 (3): 260–295.

Okabe, Reiko. 2005. "Children's acquisition of benefactives and passives in Japanese.." In *Proceedings of the 29th annual Boston University Conference on Language Development*, 437–447, Somerville, MA: Cascadilla Press.

Okabe, Reiko. 2011. "Two types of benefactives in child Japanese: A preliminary experimental study." In *The proceedings of the 12th Tokyo Conference on Psycholinguistics*, 191–208.

Okabe, Reiko. 2019. "Juyo doshi kobun no kakutoku saiho: kopasubunseki [Revisiting the acquisition of applicative constructions: a corpus analysis]." *The journal of the Senshu University Research Society [Senshu Jinbun Ronshu]* 105 219–245.

Pinker, Steven. 1984. *Language leranability and language development*. Cambridge, Mass: HUP.

Pylkkänen, Liina. 2002. *Introducing arguments*. Ph.D. dissertation, MIT.

Pylkkänen, Liina. 2008. *Introducing Arguments*. Cambridge, MA.: MIT press.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, `https://www.R-project.org/`.

Sakoda, Kumiko. 2020. "I-JAS tanjo no keii. [The background of compilation of I-JAS]." In *Phrase Structure and the Lexicon*, edited by Sakoda, Kumiko, Shinichiro Ishikawa, and Jaeho Lee 2–13, Kuroshio Publishers.

Shiina, Michi. 2021. *"Saseteitadaku" no goyooron: hito wa naze tukaitakunarunoka [The pragmatics of "saseteitadaku": why are people eager to use it?]*. Tokyo: Hitsuji Shobo.

Stan Development Team. 2023. "RStan: the R interface to Stan." `https://mc-stan.org/`, R package version 2.32.3.

Suethanapornkul, Sakol. 2020. *Statistical learning of predictive dependencies in the tense-aspect system of a miniature language by English and Thai first language adults*. Ph.D. dissertation, Georgetown University.

Takemura, Miwa. 2010. "[The use of viewpoint expressions in the personal telling task and the picture description Task by native Japanese speakers and Chinese learners of Japanese]]." *Hiroshimadaigaku Daigakuin Kyoikugaku Kenkyuka Kiyo* 59 289–298.

Tashiro, Hitomi. 1995. "Tyuzyokyu nihongo gakusyusya no bunsyo hyogen no mondaiten." *Nihongo Kyoiku* 85 25–37.

Tomasello, Michael. 1999. *The cultural origins of human cognition. An essay.*. Cambridge: Harvard University Press.

Tomioka, Satoshi, and Lan Kim. 2017. "The *give*-type benefactive constructions in Korean and Japanese." *Journal of East Asian Linguistics* 26 233–257.

Uyeno, Tazuko, S. I. Harada, Hideo Hayashibe, and Hiroshi Yamada. 1978. "Comprehension of sentences with giving and receiving verbs in Japanese children." *Annual bulletin of Research Institute of Logopedics and Phoniatrics* 12 167–185.

Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. "Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC (with Discussion)." *Bayesian Analysis* 16 (2): 667–718.

Yamada, Akitaka. 2019. "A Quantitative Approach to Addressee-Honorific Markers:Identification of Crucial Independent Variables and Prototypes." *Mathematical Linguistics* 32 (3): 117–132.

Yamada, Akitaka. 2020. "An OT-driven dynamic pragmatics: high-applicatives, subject-honorific markers and imperatives in Japanese." In *New Frontiers in Artificial Intelligence. JSAI-isAI 2019. Lecture Notes in Computer Science*, Volume 12331. 354–369, Cham: Springer.

Yamada, Akitaka. 2022a. "Diachronic constructional alternation of high-applicative forms: a Time-Series analysis using a State-Space Model for the choice among sasete itadak, sasete moraw, sasete kudasar, and sasete kure [tekiyokei no tuziteki kobun kotai: "saseteitadaku" "sasetemorau" "sasetekudasaru" "sasetekureru" no sentaku nitaisuru zyotai kukan moderu o motiita zikeiretubunseki]." In *Proceedings of the 66th conference of Mathematical Linguistics Society of Japan*, 12–17.

Yamada, Akitaka. 2022b. "Subject-honorific markings in imperatives: an OT-driven Dynamic Pragmatics." In *Proceedings of the 164th Meeting of the Linguistic Society of Japan*, 163–169.

Yamada, Akitaka. 2023a. "Constructionalization of the Japanese addressee-honorification system." In *Proceedings of the 23rd Meeting of the Japanese Cognitive Linguistics Association*, 171–180.

Yamada, Akitaka. 2023b. "Historical Pragmatics using State-Space Models." In *Proceedings of the 24th Conference of the Pragmatics Society of Japan*.

Yamada, Akitaka. 2023c. "Looking for default vocabulary insertion rules: diachronic morphosyntax of the Japanese addressee-honorification system." *Glossa: a journal of general linguistics* 8 (1): 1–47.

Yamada, Akitaka, and Taika Nagano. 2023. "The 'passivized' high applicative construction in Japanese." In *Proceedings of the 166th Meeting of the Linguistic Society of Japan*, 246–252.

Yamada, Akitaka, and Taika Nagano. to appear. "Not always introducing arguments: the syntax of high-applicative constructions in Japanese." In *Proceedings of the 17th Meeting of the WAFL*.

Yamada, Akitaka. to appear. "A modal approach to the Japanese high-applicative expression -*temoraw*." In *Proceedings of the 17th Meeting of the WAFL*.