Rules and Similarity in Linguistic Generalization: Inflectional Verb Forms in L1 and L2 Japanese

Yoko Nakano and Kenta Kishimoto (Kwansei Gakuin University)
y-k.nakano@kwansei.ac.jp

Two major views of morphological production have been proposed. One assumes that inflectional forms are produced through an operation directed by a rule while the other assumes that they are produced through similarity to the existing verb forms (Pinker, 1999). This study investigated the two views of morphological generalization using two inflectional forms (past tense and volitional) in two Japanese verb groups (Appendix 1). In Study 1, verb-form production tasks (tasks to type a conjugated form of a verb presented on a PC monitor) were administered to elicit past-tense verb forms (Experiment 1) and volitional forms (Experiment 2) in L1 and L2 Japanese. In Study 2, computational learning was implemented using a software program (Minimal Generalization Learner [MGL]) (Albright & Hayes, 2003), which can generate multiple default rules, and the output was compared with the data from human production.

Study 1: Participants. Twenty-one native speakers (NSs) of Japanese, 24 advanced Japanese learners of German, and 20 Japanese learners of Chinese (Exp. 1), and 23 NSs of Japanese, 18 advanced Japanese learners of German, and 19 Japanese learners of Chinese (Exp. 2) participated in the experiments. German participants were not taught conjugation patterns as explicit rules. Materials. Similarity-based generalization is sensitive to the similarity of novel verbs to real verbs, but rule-based generalization is not. Therefore, novel verbs with differing degrees of similarity to real verbs were constructed. In Japanese, Group I (Gr-I) verbs are the most frequent, and Group II (Gr-II) verbs are less frequent than Gr-I verbs. Japanese is a moraic-language, with one mora typically comprising a consonant (C) and a vowel (V), and the two verb groups can be distinguished by the last two moras (CVCV), particularly by the penultimate V and the final C (underlined). The Vs of typical Gr-I verbs are /a/, /o/, or /ɯ/, and their C is any consonant other than /ɾ/. The Vs of atypical Gr-I verbs are /i/ or /e/, and their C is /ɾ/. By combining typical and atypical Vs and Cs, four categories of novel verbs were constructed (very similar: *SS*; somewhat similar: *SD* and *DS*; very dissimilar: *DD*). The right *S* and *D* stand for the *S*imilar and *D*issimilar Vs and the left ones for *S*imilar and *D*issimilar Cs. There were 24 novel verbs in Exp. I and 64 in Exp. 2. Samples are presented in Table 1. Prediction. Similarity-based generalization predicts that the proportion of Gr-I verb conjugation will vary along the similarity, i.e., will show a gradient similarity effect (*SS* > *SD*, *DS* > *DD*) while rule-based generalization predicts an invariant proportion between the four categories.

Study 2: The Balanced Corpus of Contemporary Written Japanese contains 903 real, non-compound verbs whose frequency is over 10 per million. Of these, 785 verbs were given to MGL for training. After checking the plausibility of the outputs with 118 real verbs, we gave MGL 48 novel verbs used in Study 1 in the test phase.

Comparison with Results for Studies 1 and 2:

The mean percentages of Group I conjugation in past-tense and volitional forms produced by Human and MGL are in *Figures 1* and *2*. When the participants did not adopt Group I conjugation pattern for a particular item, they adopted Group II conjugation. On the one hand MGL is capable of generating multiple default rules, and on the other hand, NSs' results show a lack of gradient similarity effect across conditions, suggesting a possibility that they adopted a single conjugation rule, rather than a multiple rule. Significant differences between NSs and MGL are observed in the *DD-eru* category of past-tense forms, and the *DD-iru* category of past-tense and volitional forms, and the *DS* category of volitional form. As for the learners, no difference was found between German and Chinese learners across conditions. The learners' results were different from NSs in the *DD-eru* and *DD-iru* categories in both past-tense and volitional forms. However, the results did not indicate any significant difference from MGL in the *DD-iru* category of both past-tense and volitional forms except the German group in past-tense forms, suggesting a possibility that, unlike NSs, the learners have two default rules, one for the verbs in the *SS, SD and DS* categories and the other for those for the *DD* categories. In sum, NSs possibly applied a single default rule to any types of novel verbs, whereas, the learners applied two default rules to different types of novel verbs.

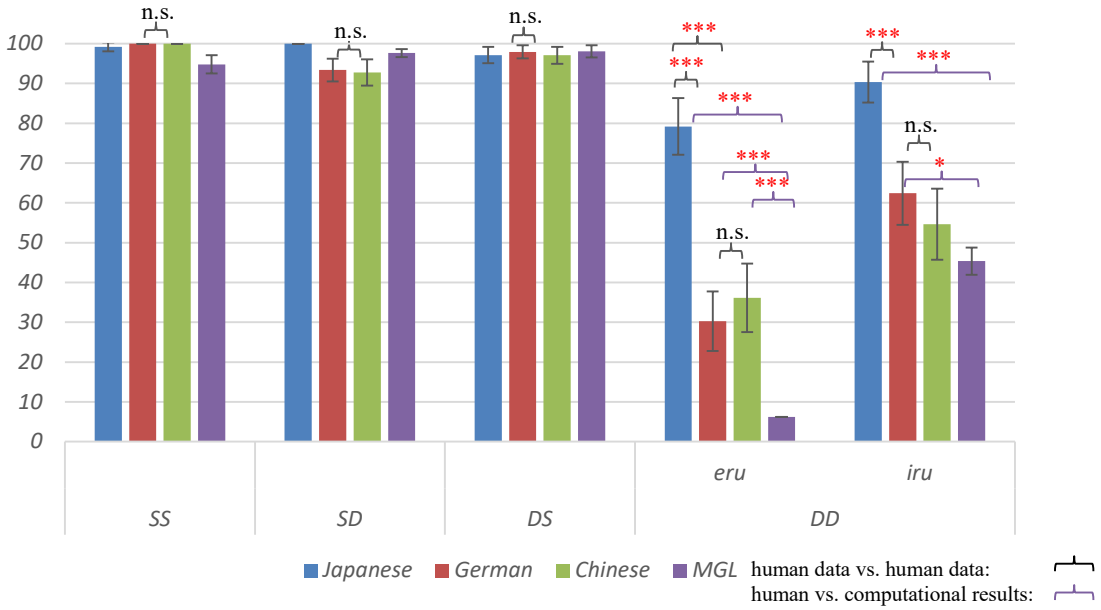

Figure 1. *Mean percentage of past-tense forms that matched Group I verb conjugation patterns in human data and computational output*

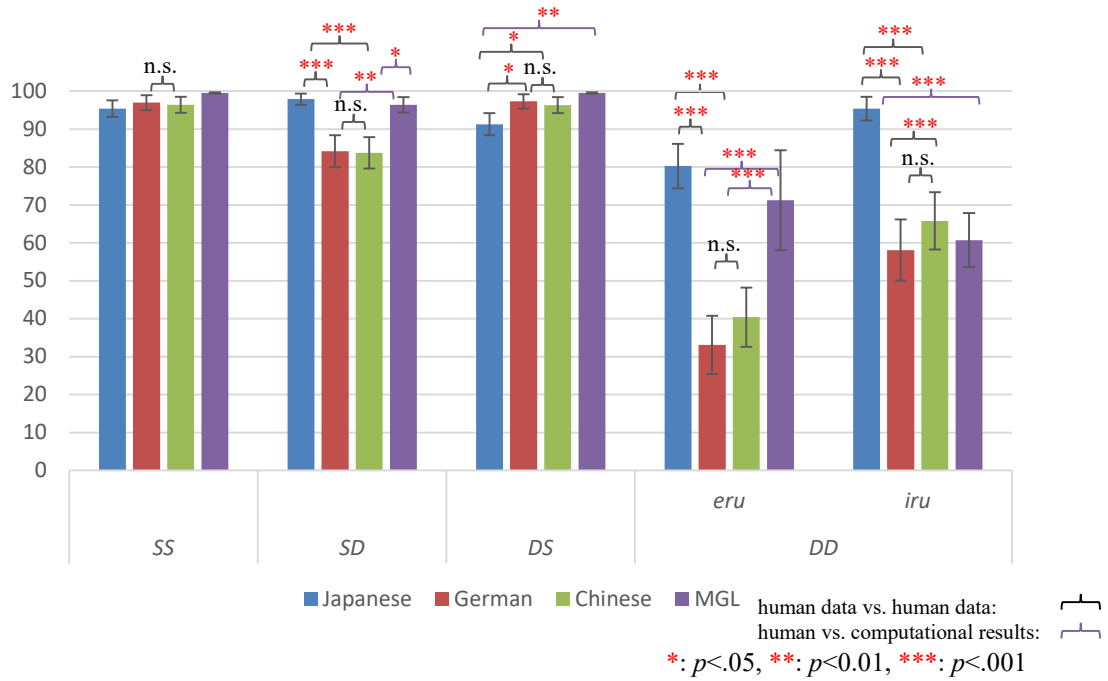

Figure 2. *Mean percentage of volitional forms that matched Group I verb conjugation patterns in human data and computational output*

References.

Pinker, S. (1999). *Words and rules: The ingredients of language*. Basic Books.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*(2), 119–161.

Fushimi, T. et al. (2007). Inflecting Japanese verbs. In T. Fushimi (Ed.), *Research report of grant-in-aid for scientific research*, *the mechanism of inflecting Japanese verbs: Cognitive, neuropsychological, and computational approach* (pp. 53–77). Tokyo.

Appendix 1

Japanese Verb Groups and Past-Tense and Volitional Forms

Japanese verbs can be categorized into three groups: Group I or consonant-stem verbs, Group II or vowel-stem verbs and Group III or irregular verbs. Group-I verbs (67%) are the largest of the three, and Group-II verbs are the second (31 %) and Group III has two verbs.

Past-tense forms of Group-I verbs are produced by adding the past-tense morpheme *ta* to the verb stem. When the past-tense suffix *ta* is added to a verb stem, phonological change occurs. For instance, the dictionary form of verb "write" *kaku* consists of a stem *kak* and the non-past suffix *u*. When the past-tense suffix *ta* is added to the stem *kak*, the stem-end consonant /*k/* is changed into /i/, and the past-tense form becomes *kaita*. Nine different consonants /k, g, t, m, n, ɾ, w, s/ appear at the stem end, and each of them is phonologically changed when *ta* is added. Group-II verbs end with *iru* or *eru*, and past-tense forms of Group-II verbs end are produced by adding *ta* to the verb stem but no phonological change occurs. For instance, the dictionary form of Group-II verb *taberu "eat"* consists of the stem *tabe* and the non-past suffix *ru*. Its past-tense form *tabeta* consists of the stem *tabe* and suffix *ta.*

Volitional forms of Group-I verbs consist of a verb stem and the volitional suffx oo /ɔ:/, e.g., *kaku "write" >> kakoo "let's write",* and volitional forms of Group-II verbs consist of a verb stem and the volitional suffix yoo /jɔ:/, e.g., *taberu "eat" >> tabeyoo "let's eat".* No phonological change occurs in volitional forms.

Although Group-I and Group-II verbs can be distinguished with the last two moras, as described in the first page, some Group-I verbs share the phonetic features with Group-II verbs. The dictionary forms of all Group-II verbs end with either *eru* (*neru, "sleep"*) or *iru* (*kiru*, *"wear*"), and atypical Group-I verbs also end with *eru* (*keru, "kick"*) or *iru* (*kiru*, *"cut*") (the *DD* category)*,* and less typical Group-I verbs end with *ru* ending (the *DS* category). Real Group-I verbs and Group-II verbs with *eru* ending accounts for 4.5% (58 words) vs. 95.5% (1222 words), and those with *iru* ending for 51.4% (146 words) vs. 48.6% (138 words) (Fushimi et al. 2007). Therefore, it is conceivable that the verbs in the *DD* category have memorable size to the learners as well as NSs but they need to be memorized in order to generate correct inflectional forms, and the novel verbs in the *DD* and *SD* categories may be challenging to MGL.

Table 1. *Classification of real verbs by the similarity to typical Group-I verbs, and their past-tense and volitional forms of Japanese verbs*

Vowel	***Consonant***	***Code***	***Dictionary Form of Real Verb***	***Past-Tense Form***	***Volitional Form***	***Sample Novel Verb***
*S*imilar: /a/, /o/, or /ɯ/	*S*imilar: non-/ɾ/	*SS*	*kak-u "write"*	*kai-ta*	*kak-oo*	*baku*
	*D*issimilar: /ɾ/	*SD*	*kar-u "mow"*	*kat-ta*	*kar-oo*	*baru*
*D*issimilar: /i/ or /e/	*S*imilar: non-/ɾ/	*DS*	*kik-u "listen"*	*kii-ta*	*kik-oo*	*beku*
	*D*issimilar: non-/ɾ/	*DD*	*kir-u "cut"*	*kit-ta*	*kir-oo*	*biru*

Note. 1. -: a boundary between a stem and an affix

2. The columns *Vowel* and *Consonant* indicate the penultimate vowel and the last consonant in the phonetic structure of the last two moras in the dictionary form of Japanese verbs (/CVCV/, underlined).
3. *S*imilar and *D*issimilar: V and C that are dis-/similar to the V and C of typical Group-I verbs.