

Inverting and Modeling Morphological Inflection

Yohei Oseki

Faculty of Science & Engineering
Waseda University
oseki@aoni.waseda.jp

Yasutada Sudo

Department of Linguistics
University College London
y.sudo@ucl.ac.uk

Hiromu Sakai

Faculty of Science & Engineering
Waseda University
hsakai@waseda.jp

Alec Marantz

Department of Linguistics & Psychology
New York University
marantz@nyu.edu

Abstract

Previous “wug” tests (Berko, 1958) on Japanese verbal inflection have demonstrated that Japanese speakers, both adults and children, cannot inflect novel present tense forms to “correct” past tense forms predicted by rules of existent verbs (de Chene, 1982; Vance, 1987, 1991; Klafehn, 2003, 2013), indicating that Japanese verbs are merely stored in the mental lexicon. However, the implicit assumption that present tense forms are bases for verbal inflection should not be blindly extended to morphologically rich languages like Japanese in which both present and past tense forms are morphologically complex without inherent direction (Albright, 2002). Interestingly, there are also independent observations in the acquisition literature to suggest that past tense forms may be bases for verbal inflection in Japanese (Klafehn, 2003; Murasugi et al., 2010; Hirose, 2017; Tatsumi et al., 2018). In this paper, we computationally simulate two directions of verbal inflection in Japanese, Present \mapsto Past and Past \mapsto Present, with the rule-based computational model called Minimal Generalization Learner (MGL; Albright and Hayes, 2003) and experimentally evaluate the model with the bidirectional “wug” test where humans inflect novel verbs in two opposite directions. We conclude that Japanese verbs can be computed online via some generalizations and those generalizations do depend on the direction of morphological inflection.

1 Introduction

In her seminal “wug” test, Berko (1958) demonstrated that English speakers, both adults and children, can inflect novel nouns (e.g. *wug*) and novel verbs (e.g. *rick*) to “correct” plural forms (e.g. *wugs*) and “correct” past tense forms (e.g. *ricked*),

respectively. This demonstration strongly suggests that, since the novel words cannot be experienced before by the experimental participants and thus accessed from the mental lexicon, the inflected forms must have been produced online via some productive generalizations, whose nature has been actively debated in the literature (O’Donnell, 2015; Yang, 2016).

Nevertheless, “wug” tests might be too easy to “pass” in morphologically sparse languages like English in which present tense forms are unaffixed and homonymous (except 3rd person singular) with infinitival forms, and past tense forms are generated via simple affixation. In fact, previous “wug” tests on Japanese verbal inflection have demonstrated that Japanese speakers, both adults and children, cannot inflect novel present tense forms to “correct” past tense forms predicted by rules of existent verbs (de Chene, 1982; Vance, 1987, 1991; Klafehn, 2003, 2013). The results of these previous “wug” tests have been taken to indicate that Japanese verbs are merely stored in the mental lexicon, not produced online via productive generalizations.

However, the implicit assumption that present tense forms are bases for verbal inflection should not be blindly extended to morphologically rich languages like Japanese. As pointed out by Albright (2002), various factors conspire to determine which cell of the paradigm should be identified as the base and, consequently, in which direction morphological inflection should be implemented. In particular, unlike English whose present and past tense forms are asymmetrically complex, both present and past tense forms are suffixed in Japanese and thus morphologically complex without inherent direction, as in Table 1:

Final segment	Meaning	Root	Present	Past	
V-final	/e/	‘eat’	$\sqrt{\text{tabe}}$	tabe-ru	tabe-ta
	/i/	‘wear’	$\sqrt{\text{ki}}$	ki-ru	ki-ta
C-final	/r/	‘mow’	$\sqrt{\text{kar}}$	kar-u	kat-ta
	/t/	‘win’	$\sqrt{\text{kat}}$	kat-u	kat-ta
	/w/	‘buy’	$\sqrt{\text{kaw}}$	ka-u	kat-ta
	/m/	‘read’	$\sqrt{\text{yom}}$	yom-u	yon-da
	/b/	‘call’	$\sqrt{\text{yob}}$	yob-u	yon-da
	/n/	‘die’	$\sqrt{\text{sin}}$	sin-u	sin-da
	/k/	‘draw’	$\sqrt{\text{kak}}$	kak-u	ka-ita
	/s/	‘lend’	$\sqrt{\text{tas}}$	kas-u	kas-ita
/g/	‘sniff’	$\sqrt{\text{kag}}$	kag-u	ka-ida	

Table 1: Japanese verbal inflection (McCawley, 1968)

Interestingly, there are independent observations in the acquisition literature to suggest that past tense forms may be bases for verbal inflection in Japanese. First, Japanese children acquire past tense forms around age 2 before present tense forms (Clancy, 1985; Klafehn, 2003), unlike English children who acquire present tense or infinitival forms first (Brown, 1973). Second, Japanese children exclusively produce past tense forms as Root Infinitive analogues before age 2 (Murasugi et al., 2010), unlike bona fide Root Infinitives in English (Wexler, 1994). Third, Japanese children overregularize present tense forms (e.g. *sim-u* ‘die’; Hirose, 2017), but not past tense forms like English (e.g. *go-ed*; Klafehn, 2003). Finally, Japanese children seem to have an inductive bias to prefer past tense forms to present tense forms (Tatsumi et al., 2018). Those observations converge on the hitherto unexplored hypothesis that past tense forms are bases in Japanese.

In this paper, inspired by the acquisition literature (Klafehn, 2003; Murasugi et al., 2010; Hirose, 2017; Tatsumi et al., 2018), we computationally simulate two directions of verbal inflection in Japanese, Present \mapsto Past and Past \mapsto Present, with the rule-based computational model called Minimal Generalization Learner (MGL; Albright and Hayes, 2003) and experimentally evaluate the model with the bidirectional “wug” test where humans inflect novel verbs in two opposite directions. The following questions will be addressed:

1. Which direction is computationally less complex for the model?
2. Which direction is experimentally more accurate for humans?
3. In which direction do the model and humans correlate more strongly?

Model complexity will be measured via three evaluation metrics derived from the confidence scores of rules induced based on the lexicon of existent verb pairs (Albright, 2002). In addition, human accuracy and the correlation between the model and humans are explicitly evaluated against the model, avoiding impressionistic interpretations as in previous “wug” tests, where “the meaning of the word ‘pass’ is a 60% or better score” (Klafehn, 2013, p.182).

The organization of this paper is as follows. Section 2 describes the methodological details of the Minimal Generalization Learner, the bidirectional “wug” test, and the statistical analyses to compare the two. Section 3 presents the results of model complexity, human accuracy, and the correlation between the model and humans, corresponding to the three questions above. Section 4 summarizes the results and discuss theoretical implications. Section 5 concludes the paper.

2 Methods

2.1 Minimal Generalization Learner

Training: The rule-based computational model called *Minimal Generalization Learner* (MGL; Albright and Hayes, 2002, 2003) was employed from the literature. The MGL was trained on the lexicon of 1269 existent verb pairs (Suski, 1942) in two directions (Present \mapsto Past and Past \mapsto Present), with V-V compounds and light verb constructions removed in order to avoid inflation of the number of particular inflections. Then, rules were induced through minimal generalization for each direction. See Albright and Hayes (2002, 2003) for the rule induction algorithm.

Testing: Novel verbs were then fed into the trained MGL as input and the inflected forms of those verbs were produced as output with the reliability and confidence scores defined below.

Reliability score: The reliability score of a rule, \hat{p} , is defined as Equation 1:

$$\hat{p} = \frac{\# \text{ forms correctly derived (= hits)}}{\# \text{ forms potentially derived (= scope)}} \quad (1)$$

Confidence score: Since weak rules supported by smaller data should be penalized (Mikheev, 1997), the reliability score of a rule is transformed into the confidence score, π , defined as Equation 2:

$$\pi = \hat{p}^* - z(1 - \alpha)/2 \times \sqrt{\frac{\hat{p}^* \times (1 - \hat{p}^*)}{n}} \quad (2)$$

where \hat{p}^* is the smoothed reliability $\hat{p}^* = \frac{hits+0.5}{scope+1}$, $\sqrt{\frac{\hat{p}^* \times (1-\hat{p}^*)}{n}}$ is the estimated variance, and α is the free parameter called confidence value (the higher α , the greater penalty for weak rules) assumed here as $\alpha = .75$.

Mean confidence: Three evaluation metrics can be derived from the confidence score (cf. Albright, 2002). First, the mean confidence score of winning outputs is defined as Equation 3:

$$\bar{\pi} = \frac{1}{n} \sum_{i \in I} \max_{o \in O_i} \pi_o \quad (3)$$

where I is the set of all inputs, O_i is the set of all outputs derived from the input i , and π_o is the confidence score of the output o . The higher mean confidence, the more efficient grammar.

Mean margin: Second, the mean confidence margin of winning outputs is defined as Equation 4:

$$\bar{\Delta} = \frac{1}{n} \sum_{i \in I} [\max_{o \in O_i} \pi_o - \max_{s \in S_i} (\pi_s, 0)] \quad (4)$$

where S is the subset of all outputs derived from the input i without winning outputs. The higher mean margin, the more efficient grammar.

Mean entropy: Finally, the mean confidence entropy of possible outputs is defined as Equation 5:

$$\bar{H} = \frac{1}{n} \sum_{i \in I} \sum_{o \in O_i} \pi_o \log_2 \pi_o \quad (5)$$

where the confidence scores of all outputs derived from the input, O , are normalized, such that the scores sum up to 1. Unlike the metrics above, the lower mean entropy, the more efficient grammar.

2.2 Bidirectional ‘‘Wug’’ Test

Participants: The bidirectional ‘‘wug’’ test was conducted with the within-participants design. The experimental participants were 45 undergraduate students at Waseda University in Japan. They were given a ¥500 book coupon for their participation. The 6 participants who were non-native speakers of Japanese or misinterpreted the instructions were excluded from statistical analyses, resulting in 39 participants in total.

Stimuli: Bisyllabic novel roots (‘‘wug’’ roots) were created by randomly combining two open syllables (CV). The open syllables were extracted from the lexicon of 1269 existent verb stems (Suski, 1942), with V-V compounds and light verb constructions removed, and only those with token

frequency ≥ 20 were included. The ‘‘wug’’ roots that resemble actual Japanese words or contain repetitions of the same segments were excluded. Since the native Japanese words cannot generally start with voiced obstruents (Tanaka and Yashima, 2013) or /r/ (Labrune, 2014), the ‘‘wug’’ roots starting with those segments were also excluded. Then, 32 past tense forms of the ‘‘wug’’ roots were created as target stimuli in the Past \mapsto Present direction, by attaching (i) the V-final past tense endings /ta/ to 16 ‘‘wug’’ roots (i.e. 8 ending with /e/ and 8 ending with /i/) and (ii) the 4 C-final past tense endings (i.e. /tta/, /nda/, /ita/, and /ida/) to 4 ‘‘wug’’ roots each, hence 16 V-final and 16 C-final past tense forms. In the same vein, 32 present tense forms of the ‘‘wug’’ roots were created as target stimuli in the Present \mapsto Past direction, by attaching (i) the V-final present tense ending /ru/ to 16 ‘‘wug’’ roots (i.e. 8 ending with /e/ and 8 ending with /i/) and (ii) the 8 C-final present tense endings (i.e. /tu/, /lu/, /mu/, /bu/, /nu/, /ku/, /su/, and /gu/) to 2 ‘‘wug’’ roots each, hence 16 V-final and 16 C-final present tense forms. Note that the 4 V-final past tense forms whose roots end with /hi/ and 2 C-final present tense forms ending with /nu/ turned out to be not attested in the training data and thus excluded from the statistical analyses.

The 4 frames were also created in which the target stimuli are presented. Each frame consisted of two sentences A and B. In the Present \mapsto Past direction, A sentences include present tense forms, while B sentences contain a blank and elicit past tense forms. In the Past \mapsto Present direction, A sentences include past tense forms, while B sentences contain a blank and elicit present tense forms. In order to make sure that the participants produce target forms in B sentences, temporal adverbs are placed at the sentence initial position to maximally contrast sentences A and B. Specifically, A sentences constitute ‘‘Temporal Adverb + Proper Noun + **Verb** + Evidential’’, whereas B sentences ‘‘Temporal Adverb + Proper Noun + _____ + Sentence Final Particle’’, where the participants are asked to inflect the **Verb**.

Procedure: The task was written production ‘‘wug’’ test in the form of the questionnaire. At the top of the questionnaire were some biographical questions such as (i) birthplaces of participants and their parents, (ii) whether participants were born and grew up in Japan, and (iii) whether parents spoke Japanese to participants at home.

The English translation of the original Japanese instructions is reproduced below:

“This experiment examines your intuition about Japanese. There are 2 blocks and 48 questions in each block, where both actual and novel verbs in Japanese appear. For each question, please change bold and underlined **verbs** of A sentences to appropriate forms and complete the underlined portion of B sentences. Since there are no correct or incorrect answers, please answer the questions based on your intuition without too much reflection.”

The experiment was divided into two blocks, corresponding to Present \mapsto Past and Past \mapsto Present directions, and the order of the two directions was counterbalanced across participants. At the beginning of each block were two example questions with answers in B sentences completed, one example with an actual verb and another with a novel verb. The order of stimulus presentation was randomized across participants by creating 2 random orders for each direction, hence 4 different versions of randomization. The 16 actual verbs, 8 V-final and 8 C-final, were interspersed as fillers in each block, on the condition that no more than 3 target stimuli were presented in sequence.

2.3 Statistical Analyses

Human accuracy: In contrast with the previous “wug” tests, the human accuracies were explicitly evaluated against the MGL. The winning outputs with highest confidence scores were defined as “correct” among possible outputs derived from each input by the MGL. For example, suppose that the MGL produced three possible outputs X, Y, and Z for an input, among which X was the winning output with the highest confidence score. If the output X was produced by 30 participants, the human accuracy would be $30/39 = .769$.

Correlation between model and humans: The correlation between model confidence scores and human production probabilities were also analyzed (Albright and Hayes, 2003). The human production probabilities can be simply computed by dividing the frequencies of produced outputs by the number of participants. Given the distribution of human production probabilities being bimodal, nonparametric Kendall’s rank correlation analyses

were performed in R between model confidence scores and human production probabilities. The outputs not produced by both model and humans were not included in the correlation analyses.

3 Results

3.1 Model Complexity

The result of model complexity is summarized in Table 2, where three model complexity metrics, mean confidence score ($\bar{\pi}$), mean confidence margin ($\bar{\Delta}$), and mean confidence entropy (\bar{H}), are shown for each direction of verbal inflection:

	Present \mapsto Past	Past \mapsto Present
Mean confidence ($\bar{\pi}$)	.904	.959
Mean margin ($\bar{\Delta}$)	.724	.849
Mean entropy (\bar{H})	.387	.294

Table 2: Result of model complexity

The three evaluation metrics all converge on the conclusion that the Past \mapsto Present direction is computationally less complex than the Present \mapsto Past direction: the mean confidence and margin were higher, while the mean entropy was lower. On closer inspection, the confidence scores of possible outputs for /ru/-final present tense forms were almost a tie in the Present \mapsto Past direction, which increased the mean entropy.

3.2 Human Accuracy

The result of human accuracy is summarized in Table 3, where the accuracies of the current experiment are shown for each direction of verbal inflection and compared with six previous “wug” tests (de Chene, 1982; Vance, 1987, 1991; Klafehn, 2003, 2013):

Reference	Modality	Task	Accuracy
de Chene (1982)	oral	production	46%
Vance (1987)	written	choice	51%
Vance (1991)	written	choice	63%
Klafehn (2003)	written	choice	53%
Klafehn (2013)	oral	production	32%
Present \mapsto Past	written	production	48%
Past \mapsto Present	written	production	72%

Table 3: Result of human accuracy

First, the accuracy of the Present \mapsto Past direction (48%) is generally comparable to the literature, especially the oral production experiment by (46%; de Chene, 1982), despite different participants and stimuli between the experiments. Second, and more importantly, the accuracy of the

Past \mapsto Present direction (72%) was significantly higher than the literature, even the forced choice experiments (Vance, 1987, 1991; Klafehn, 2003), which would be regarded as “pass” on the assumption that “the meaning of the word ‘pass’ is a 60% or better score” (Klafehn, 2013).

3.3 Correlation Between Model and Humans

The result of correlation between the model and humans is shown in Figure 1, where the x -axis is the model confidence score computed by the MGL (“Model confidence score”), whereas the y -axis is the human production probability of the forms produced by the experimental participants (“Human production probability”). Color represents the Past \mapsto Present (red) and Present \mapsto Past (blue) directions. Shape of the data points represents C-final (circle; ●) and V-final (triangle; ▲) roots. The lines are the fitted linear models.

There were 1248 forms in total produced for each direction (39 participants * 32 stimuli = 1248 forms). In the Past \mapsto Present direction, 1225 forms were present tense forms, while 23 forms were errors (i.e. non-present tense forms). Out of the 1225 present tense forms, 956 forms were also produced by the MGL. In the Present \mapsto Past direction, 1225 forms were past tense forms, while 23 forms were errors (i.e. non-past tense forms). Out of the 1225 past tense forms, 821 forms were also produced by the MGL. Kendall’s rank correlation analyses revealed that the model confidence scores and human production probabilities were strongly correlated in the Past \mapsto Present direction ($z = 5.0618$, $\tau = 0.534$, $p < 0.001^{***}$), but only weakly in the Present \mapsto Past direction ($z = 2.3058$, $\tau = 0.2448$, $p < 0.05^*$).

4 Discussion

In summary, the results demonstrated that (i) the Past \mapsto Present direction was computationally less complex than the Present \mapsto Past direction, (ii) the Past \mapsto Present direction was experimentally more accurate than the Present \mapsto Past direction, and (iii) the model and humans were correlated strongly in the Past \mapsto Present direction, but only weakly in the Present \mapsto Past direction. The present work should be regarded as the computational psycholinguistic approach to the Paradigm Cell Filling Problem (Ackerman et al., 2009) and, importantly, the result of model complexity harmonizes well with the Low Entropy

Conjecture (Ackerman and Malouf, 2013), which may in turn provides an insight into SIGMORPHON Shared Task on morphological reinflection (Cotterell et al., 2018). In the following, two theoretical implications will be discussed: the past tense debate and language learning.

4.1 The Past Tense Debate

In the context of the past tense debate on rule vs. analogy (Pinker and Ullman, 2002), three logically possible models have been proposed in the literature: the single route rule-based model (Yang, 2002), the single route analogy-based model (Rumelhart and McClelland, 1986), and the dual route model (Pinker and Prince, 1988). The results above at least indicate that Japanese verbs can be computed online via some generalizations and those generalizations do depend on the direction of morphological inflection, contrary to the conclusion of previous “wug” tests that Japanese verbs are merely stored in the mental lexicon (de Chene, 1982; Vance, 1987, 1991; Klafehn, 2003, 2013). However, although the MGL is “rule-based”, the nature of those generalizations is still an open question to be addressed via the systematic comparison with contemporary analogy-based models such as Recurrent Neural Networks (RNN: Kirov and Cotterell, 2018) and Naive Discriminative Learning (NDL: Baayen et al., 2011) couched in Word and Paradigm models of morphology (Stump, 2001; Blevins, 2006).

In addition, given the different strength of correlation with the rule-based computational model in two opposite directions, we can hypothesize that the Past \mapsto Present direction is rule-based, while other directions including the Present \mapsto Past direction is analogy-based. Then, following the electroencephalography (EEG) experiment by Kobayashi et al. (2012) who demonstrated that rule and analogy are indexed by event-related potential (ERP) components called LAN and N400, respectively, we predict that the Past \mapsto Present direction is reflected by LAN, whereas the Present \mapsto Past direction by N400. This prediction is summarized in Table 4 and left for future research.

Direction	Wug	Model	Route	ERP
Past \mapsto Present	“pass”	Symbolic	Rule	LAN
Present \mapsto Past	“fail”	Neural	Analogy	N400

Table 4: Prediction of ERP components

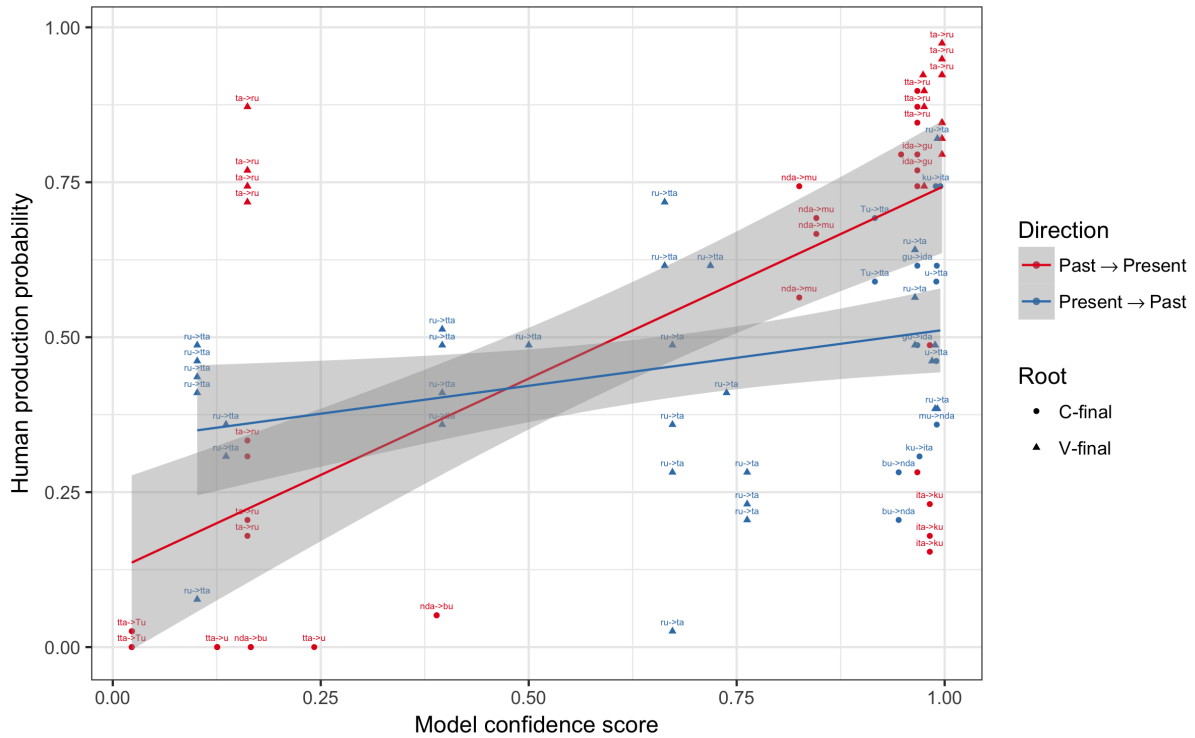


Figure 1: Result of correlation between the model and humans

4.2 Human and Machine Language Learning

The results above further suggest that human language learning can provide insights into machine language learning. Interestingly, the Past \mapsto Present direction was motivated by not only the language acquisition literature (Klaffehn, 2003; Murasugi et al., 2010; Hirose, 2017; Tatsumi et al., 2018) but also artificial language learning by Yin and White (2018) who show that humans have an inductive bias against neutralization and homophony. In other words, the Past \mapsto Present direction is preferred to the Present \mapsto Past direction because the present tense forms of /r/, /t/, /w/-final “wug” roots and /m/, /b/, /n/-final “wug” roots are neutralized into the homonymous past tense forms ending with /tta/ and /nda/, respectively.

Nevertheless, there were several limitations with the current experiment. First, the assumption that the training data is fed to the model in pairs organized as paradigms is unrealistic due to Zipf’s law (Zipf, 1949), where paradigms are almost always incomplete in human language learning (Yang, 2017; Blevins et al., 2017). In order to simulate realistic language learning, child-directed speech (CDS) should be employed as the training data, especially given that relative frequencies of present and past tense forms seem

to be diverse (Tatsumi et al., 2018). Second, the MGL is “rule-based” but learns only product-oriented generalizations (Becker and Gouskova, 2016) in that present and past tense forms are mapped to each other without underlying roots. In order to test source-oriented generalizations over underlying roots, morphological decomposition must be implemented to retrieve the roots from which present and past tense forms are derived (Taft, 1979, 2004). Finally, different “wug” roots were employed in two directions of verbal inflection (though created similarly) so that different results might be attributed to different roots, which remains to be controlled in future.

5 Conclusion

In this paper, inspired by the acquisition literature (Klaffehn, 2003; Murasugi et al., 2010; Hirose, 2017; Tatsumi et al., 2018), we computationally simulated two directions of verbal inflection, Present \mapsto Past and Past \mapsto Present, with the rule-based computational model called Minimal Generalization Learner (MGL; Albright and Hayes, 2003) and experimentally evaluated the model with the bidirectional “wug” test where humans inflected novel verbs in two opposite directions, addressing the following questions:

1. Which direction is computationally less complex for the model?
2. Which direction is experimentally more accurate for humans?
3. In which direction do the model and humans correlate more strongly?

The results revealed that (i) the Past \mapsto Present direction was computationally less complex than the Present \mapsto Past direction, (ii) the Past \mapsto Present direction was experimentally more accurate than the Present \mapsto Past direction, and (iii) the model and humans were correlated strongly in the Past \mapsto Present direction, but only weakly in the Present \mapsto Past direction. We conclude that Japanese verbs can be computed online via some generalizations (pace de Chene, 1982; Vance, 1987, 1991; Klafehn, 2003, 2013) and those generalizations do depend of the direction of morphological inflection.

Acknowledgments

We would like to thank Takane Ito, Ryo Ootoguro, Yoko Sugioka, and SIGMORPHON anonymous reviewers for valuable suggestions. This work was supported by JSPS KAKENHI Grant Number JP18H05589.

References

- Farrell Ackerman, James Blevins, and Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James Blevins and Juliette Blevins, editors, *Analogy in Grammar: Form and Acquisition*, pages 54–82. Oxford University Press, New York.
- Farrell Ackerman and Robert Malouf. 2013. Morphological organization: The low entropy conjecture. *Language*, 89:429–464.
- Adam Albright. 2002. *The Identification of Bases in Morphological Paradigms*. Ph.D. thesis, UCLA.
- Adam Albright and Bruce Hayes. 2002. Modeling English past tense intuitions with Minimal Generalization. *Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology*.
- Adam Albright and Bruce Hayes. 2003. Rule vs. analogy in English past tenses: a computational/experimental study. *Cognition*, 90:119–161.
- Harald Baayen, Petar Milin, Dusica Filipovic Durdevic, Peter Hendrix, and Marco Marelli. 2011. An Amorphous Model for Morphological Processing in Visual Comprehension Based on Naive Discriminative Learning. *Psychological Review*, 118:438–481.
- Michael Becker and Maria Gouskova. 2016. Source-Oriented Generalizations as Grammar Inference in Russian Vowel Deletion. *Linguistic Inquiry*, 47:391–425.
- Jean Berko. 1958. The Child’s Learning of English Morphology. *Word*, 14:150–177.
- James Blevins. 2006. Word-based morphology. *Journal of Linguistics*, 42:531–573.
- James Blevins, Petar Milin, and Michael Ramscar. 2017. The Zipfian Paradigm Cell Filling Problem. In Ferenc Kiefer, James Blevins, and Huba Bartos, editors, *Morphological paradigms and functions*, pages 141–158. Brill, Leiden.
- Roger Brown. 1973. *A first language*. Harvard University Press, Cambridge, MA.
- Brent de Chene. 1982. The segmentation of Japanese verbs: Experimental evidence. *Papers in Japanese Linguistics*, 8:29–61.
- Patricia Clancy. 1985. The acquisition of Japanese. In Dan Slobin, editor, *The crosslinguistic study of language acquisition*, pages 373–524. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Arya D. McCarthy, Katharina Kann, Sebastian Mielke, Garrett Nicolai, Miikka Silfverberg, David Yarowsky, Jason Eisner, and Mans Hulden. 2018. The CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection. *CoNLL*.
- Yuki Hirose. 2017. *The Adventure of Little Linguists*. Iwanami Publisher, Tokyo.
- Christo Kirov and Ryan Cotterell. 2018. Recurrent Neural Networks in Linguistic Theory: Revisiting Pinker and Prince (1988) and the Past Tense Debate. *Transactions of the Association for Computational Linguistics*, pages 651–665.
- Terry Klafehn. 2003. *Emergent properties of Japanese verbal inflection*. Ph.D. thesis, University of Hawaii.
- Terry Klafehn. 2013. Myth of the wug test. *BLS*, 37:170–184.
- Yuki Kobayashi, Yoko Sugioka, and Takane Ito. 2012. ERP Responses to Violations in Japanese Verb Conjugation Patterns. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, pages 611–616.
- Laurence Labrune. 2014. The phonology of Japanese /r/: A panchronic account. *Journal of East Asian Linguistics*, 23:1–25.

- James McCawley. 1968. *The Phonological Component of a Grammar of Japanese*. Mouton, The Hague.
- Andrei Mikheev. 1997. Automatic rule induction for unknown-word guessing. *Computational Linguistics*, 23:405–423.
- Keiko Murasugi, Chisato Fuji, and Tomoko Hashimoto. 2010. What’s acquired later in an agglutinative language. *Nanzan Linguistics*, 6:47–78.
- Timothy O’Donnell. 2015. *Productivity and Reuse in Language*. MIT Press, Cambridge, MA.
- Steven Pinker and Alan Prince. 1988. On language and connectionism. *Cognition*, 28:73–193.
- Steven Pinker and Michael Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences*, 6:456–462.
- David Rumelhart and James McClelland. 1986. On learning the past tenses of English verbs. In David Rumelhart and James McClelland, editors, *Parallel distributed processing: Explorations in the microstructure of cognition*, pages 216–271. MIT Press, Cambridge, MA.
- Gregory Stump. 2001. *Inflectional Morphology: A eory of Paradigm Structure*. Cambridge University Press, Cambridge.
- Peter Suski. 1942. *Conjugation of Japanese Verbs in the Modern Spoken Language*. P.D. and Ione Perkins, South Pasadena.
- M. Taft. 1979. Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7:263–272.
- M. Taft. 2004. Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, 57:745–765.
- Yu Tanaka and Jun Yashima. 2013. Deliberate markedness in Japanese hypocoristics. *Proceedings of GLOW in Asia IX*, pages 283–297.
- Tomoko Tatsumi, Ben Ambridge, and Julian Pine. 2018. Testing an input-based account of children’s errors with inflectional morphology: an elicited production study of Japanese. *Journal of Child Language*, 45:1144–1173.
- Timothy Vance. 1987. *An introduction to Japanese phonology*. State University of New York Press, New York.
- Timothy Vance. 1991. A new experimental study of Japanese verbal morphology. *Journal of Japanese Linguistics*, 13:145–156.
- Ken Wexler. 1994. Optional Infinitives, Head Movement, and Economy of Derivation. In *Verb Movement*, pages 305–350. Cambridge University Press, Cambridge.
- Charles Yang. 2002. *Knowlwdge and Learning in Natural Language*. Oxford University Press, Oxford.
- Charles Yang. 2016. *The Price of Linguistic Productivity*. MIT Press, Cambridge, MA.
- Charles Yang. 2017. Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition*, 24:100–125.
- Sora Heng Yin and James White. 2018. Neutralization and homophony avoidance in phonological learning. *Cognitive Science*, 179:89–101.
- George Zipf. 1949. *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley, Cambridge, MA.