Cross-linguistic patterns of morpheme order reflect cognitive biases: An experimental study of case and number morphology

Carmen Saldana [a,*], Yohei Oseki [b,1], Jennifer Culbertson [a]

[a] Centre for Language Evolution, The University of Edinburgh, 3 Charles Street, Edinburgh EH8 9AD, UK
[b] Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan

ABSTRACT

A foundational goal of linguistics is to investigate whether shared features of the human cognitive system can explain how linguistic patterns are distributed across languages. In this paper we report a series of artificial language learning experiments which aim to test a hypothesised link between cognition and a persistent regularity of morpheme order: number morphemes (e.g., plural markers) tend to be ordered closer to noun stems than case morphemes (e.g., accusative markers) (Universal 39; Greenberg, 1963). We argue that this typological tendency may be driven by learners' bias towards orders that reflect scopal relationships in morphosyntactic and semantic composition (Bybee, 1985; Rice, 2000; Culbertson & Adger, 2014). This bias is borne out by our experimental results: learners—in the absence of any evidence on how to order number and case morphology—consistently produce number closer to the noun stem. We replicate this effect across two populations (English and Japanese speakers). We also find that it holds independent of morpheme position (prefixal or suffixal), degree of boundedness (free or bound morphology), frequency, and which particular case/number feature values are instantiated in the overt markers (accusative or nominative, plural or singulative). However, we show that this tendency can be reversed when the form of the case marker is made highly dependent on the noun stem, suggesting an influence of an additional bias for local dependencies. Our results provide evidence that universal features of cognition may play a causal role in shaping the relative order of morphemes.

Introduction

Human languages are incredibly diverse in the way they combine meaningful units (i.e., morphemes) with each other. For example, many languages concatenate morphemes to lexical bases to create morphologically complex words, like the English word "neighbourhood-s". But languages differ in how many morphemes typically attach to a base, in whether they attach before or after that base, in what precise meanings are encoded in morphemes, in whether there are one-to-one mappings between morphemes and meanings, and so on. Nevertheless, certain regularities are apparent. For example, some patterns of *morpheme order* occur more frequently across the languages of the world, while others are rare or even unattested. There is a rich literature which aims to explain patterns of morpheme order within and across languages. Here we are interested in whether universal cognitive or psycholinguistic mechanisms might play a causal role in shaping morpheme order cross-linguistically.

The typological regularity in morpheme order we target here concerns number and case morphology, specifically, how these two classes of morphemes are ordered when there is a clear morphological boundary between them. For example, in agglutinative languages such as Hungarian, Turkish or Chintang, there are distinct sets of number morphemes (marking plurality) and case morphemes (marking grammatical functions like subjects and objects). In such languages, as shown in the examples in (1a-c) below, when overt number and case morphemes are both attached to a noun base and both follow or both precede it, the expression of number is almost always realised closer to the noun base than the expression of case (Universal 39; Greenberg, 1963).

(1)
    a. Hungarian
       könyv-ek-et
       book-PL-ACC

---

* Corresponding author at: Department of Comparative Language Science at the University of Zurich, Plattenstrasse 54, 8032 Zurich, Switzerland.
  E-mail address: carmen.saldanagascon@uzh.ch (C. Saldana).
1 Current address: Department of Language and Information Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo 153-8902, Japan.

   b. Turkish
      kitap-lar-i
      book-PL-ACC
   c. Chintang
      kitap-ce-ηa
      book-PL-ERG

There are a number of candidate explanations for this phenomenon, which intersect with high-level hypotheses about how morpheme (and word) order is determined in language more generally. For example, it has been proposed that semantic or compositional relationships among morphemes, sometimes called *scope*, determine linear order (Bybee, 1985; Rice, 2000; Wunderlich, 1993). Related theories argue that universal syntactic hierarchies, potentially reflecting semantics, determine order (Baker, 1985; Cinque, 2005). On one formulation, morphemes which more directly affect or modify the semantic content of the base have narrower scope (Baker, 1985; Bybee, 1985; Rice, 2000). Wider-scope morphemes modify the larger semantic constituent which includes any lower scoping morphemes. Take the morphologically complex word given above, "neighbour-hood-s", which contains a lexical root ("neighbour"), a derivational morpheme ("-hood"), and an inflectional morpheme ("-s"). Here, as in many languages, derivational morphemes are ordered closer to the root than inflectional ones. On a scope-based account, this is because derivational morphemes change the lexical meaning of the root thus creating new lexemes. Inflectional morphemes scope higher, modifying only grammatical properties of the unit including the root and any derivational morphemes (i.e., the *stem*). In some languages, morpheme order alternations directly reflect differences in scope. Rice (2011) gives two examples from Yup'ik (originally from Mithun, 1999), "yug-pag-cuar" (lit. 'person-big-little') and "yug-cuar-pag" (lit. 'person-little-big'), the first meaning someone who is small for a big person (e.g., a little giant), the other meaning someone who is big for a small person (e.g., a large dwarf). Similarly, it has been claimed that the linear order of nominal modifiers (e.g., adjectives, numerals, demonstratives) reflects morphosyntactic or semantic scope relations (Bouchard, 2002; Culbertson and Adger, 2014). In the case of Universal 39, the idea would be that case scopes higher than number because number directly modifies the entity referred to by the noun, while the case morpheme signals an external relationship between the entity and some event (Bybee, 1985). Following Culbertson and Adger (2014), we call orders which reflect scope relations *scope-isomorphic*.

Notably, however, there are clear cases where linear order of morphemes in a language is not scope-isomorphic. For example, some languages exhibit what looks like free variation in the order of morphemes. Ryan (2010) gives an example from Chintang (originally from Bickel et al., 2007) where three prefixes, "ma-" (negative), "u-" (3rd person non-singular agent), and "kha-" (1st person non-singular object) can occur prefixed to a stem in all six possible orders with no differences in meaning. This suggests the possibility that morpheme order may not be (primarily) driven by scope in such cases. Indeed, alternative explanations of morpheme order focus not on meaning, but rather on frequency, and its effects on language processing. Ryan (2010) shows that variation in morpheme order in Tagalog, which might at first glance appear random, reflects the frequency of particular stem+morpheme bigrams (see also Rice, 2011). Specifically, Ryan (2010) demonstrates that when the relative order of two or more morphemes is variable, the most frequent order patterns are those containing the bigrams with the highest frequencies.

Along similar lines, Hay (2001) argues that when a stem is more frequent alone than with a particular affix, then that affix is easier to parse (decompose) from the stem. This in turn determines linear order according to the *parsability principle*: an affix which can be easily parsed out during processing should not occur inside an affix which can not (see also Hay and Plag, 2004; Manova and Aronoff, 2010). How might this explain Universal 39? It could be that number tends to be expressed more often than case, or similarly, that stems occur without case

morphemes more than they occur without number morphemes (for example due to patterns of zero exponence or differences in the size of their respective paradigms). This would make case morphemes more easily parsable than number morphemes. On this account, there is nothing about the semantics of these morphemes that determines their relative order, only their distributional properties which may in turn affect and/or reflect processing. Recent work further shows that the degree of dependency between affixes and stem is a good predictor of the relative order of verbal affixes (Hahn et al., 2020a), just as the degree of dependency between a head and a dependent at the phrase and sentence level is a good predictor of word order (Futrell et al., 2015; Futrell et al., 2020; Hahn et al., 2020b). In particular, Hahn et al. (2020a) show that (at least in Japanese and Sesotho) affixes with higher co-occurrence with the verb stem are more likely to appear closer to it. It is nevertheless worth mentioning that, as noted in Hahn et al. (2020a), this does not strictly provide an alternative explanation to accounts based on semantic scope. Rather it is a potential operationalisation of such explanations: if an affix has a more direct impact on the meaning of the stem, its application might be more restricted, which in turn will determine a higher co-occurrence between such stems and the affix and therefore closer proximity. As such, these findings do not tell us whether or not semantic constraints are the ultimate explanation for morpheme order.

A third possibility is that the relative order of case and number morphemes reflects patterns of diachronic change. For example, it could be that languages tend to grammaticalise number before case (Givón, 1979), potentially due to asymmetries in usage frequency of the sort just outlined.[2]

To date, there is no direct behavioral evidence adjudicating among these potential explanations for Universal 39. In fact, there is no independent evidence beyond the typology to show that placing number closer to the noun stem than case is in fact preferred over the reverse.[3] In a series of four experiments, we test whether participants learning a miniature artificial language are indeed biased in favor of placing number morphemes closer to noun stems than case morphemes. These experiments are also designed to investigate the mechanism underlying any such bias—in particular whether the bias is driven by (absolute and relative) frequency, or by a cognitive preference, for example, for scope-isomorphic ordering. To preview, we uncover clear evidence for biased ordering across two populations (English and Japanese speakers). We also find that it holds independent of morpheme position (prefixal or suffixal), degree of boundedness (free or bound morphology), frequency, and which particular case/number feature values are instantiated in the overt markers (accusative or nominative, plural or singulative). All things equal, this suggests that the typology may reflect frequency-independent cognitive biases of learners. However, we also find that the presence of case allomorphy conditioned on the stem (which strengthens the dependency between case markers and the stem) can reverse participants' preferences. We interpret this as a competing bias for local dependencies (as independently shown by, e.g., White et al., 2018). This result adds to the growing body of work using experimental methods to investigate how learning and use shape typological patterns in morphology and word order (e.g., Culbertson et al., 2012; Culbertson

---

[2] In fact, a comment to this effect included in the Konstanz Universals Archive (Plank and Filimonova, 2000) for Universal 39 states "Number will always be grammaticalised before Case, hence bound Number exponents will always end up closer to the stem and bound Case exponents will always be more marginal. (Frans Plank)".

[3] On top of that, sample numbers in this case are relatively small. From the sample of 30 languages used in Greenberg (1963), only five contain (at least some instances of) bound number and case morphology with clear morpheme boundaries: Basque, Finnish, Turkish, Burushaski and Kannada. We could add two further languages to the count if we consider Japanese's extraordinary cases of associative plurals (Nakanishi and Tomioka, 2004; Vassilieva, 2005) as instances of number marking and the English genitive marker as an instance of case marking.

and Adger, 2014; Fedzechkina et al., 2018, 2012; Hupp et al., 2009; Martin and Culbertson, 2020; Tabullo et al., 2012).

Experiment 1

Methods

The artificial language learning experiments described here use an extrapolation paradigm (called poverty-of-the-stimulus paradigm in earlier work, e.g., Culbertson and Adger, 2014; Wilson, 2003). This means learners are trained on input that is designed to be ambiguous between (at least) two hypotheses (or patterns) of interest: here, two potential ways of ordering case and number morphemes. Learners are exposed to a miniature artificial language with nouns, and case (accusative) and number (plural) markers. Crucially, their input indicates whether these morphemes precede or follow the noun, but does not include any examples in which the two morphemes co-occur within the same noun phrase. At test, they are asked to produce utterances, including these held-out examples. The order they infer will indicate whether they have a preference for placing number closest to the noun stem (e.g., Noun-Number-Case rather than Noun-Case-Number). All experimental materials and data reported here are available at https://github.com/CarmenSaldana/NounNumberCaseOrder, and the preregistered hypothesis and analysis plan for Experiment 1 are accessible at http://osf.io/8xuc9.

Input language

The lexicon included three semi-nonce verbs, four nonce nouns, and two nonce markers (one number marker indicating plural; one case marker indicating accusative). All words were produced with initial stress. The three semi-nonce verbs are taken from the English-based creole Tok Pisin (spoken in Papua New Guinea): "kikim"(ˈkʰɪkɪm), "poinim"(ˈpʰɔmɪm) and "straikim"(ˈstraɪkɪm), which refer to 'kicking', 'pointing' and 'punching' respectively. The nouns are "negid"(neʒɪd), "nork"(ˈnɔrk), "tumbat" (ˈtʰʌmbət), "vaem" (ˈvæm) (based on Fedzechkina et al., 2012), naming four characters: a burglar, a chef, a cowboy, and a waitress. The noun-character mappings were randomised for each participant. The two markers were also randomly mapped to number and case morphemes from the following set: "gu" (ˈguː), "sa"(ˈsɑː), and "ti"(ˈtʰiː). Word order in sentences was always Verb-Agent-Patient. We used a verb-initial order because it allowed us to prompt participants' responses by providing the verbal forms in test trials; doing so allowed us to reduce the number of lexical items participants had to retrieve from memory and allowed participants to focus on the nominal morphology which constitutes our response variable. Half of participants were trained on a language with post-nominal morphemes (case and number morphemes appeared after the noun stem), half with pre-nominal morphemes (case and number morphemes appeared before the noun stem).[4]

Participants were trained on three different Noun Phrase (NP) types: NPs consisting of a bare noun, NPs consisting of a noun with overt number morphology, and NPs consisting of a noun with overt case morphology. Note that singular number, and agent (nominative) case were unmarked in the language. During training, participants got descriptions of characters in isolation (singular or plural), or events with a singular patient; plural patients (requiring both number and case morphology) were *held-out until testing*. See Fig. 1 for examples. Crucially, number and case markers appeared with the exact same frequency—absolute, and relative to each given noun—both during

training and testing phases. Therefore, any potential preferences for morpheme order revealed in this task cannot be explained by frequency.

The input language was presented both orthographically and auditorily during training. Auditory stimuli were recorded in a sound-attenuated room by a 26yo male speaker of American English. Noun phrases were recorded without a pause between nouns and markers (i.e., they were produced as bound morphemes) but each marker was presented orthographically surrounded by spaces and thus not bound to the noun. We start with orthographically unbound morphology in order to avoid any issues with segmentation. However, note that versions of the experiment with orthographically bound morphology will be also presented here (Experiments 3 and 4).

Experimental procedure

The experiment was conducted on computer terminals in sound-attenuated individual booths, with all instructions provided in English, and an English-speaking experimenter. Participants were told that they would be learning part of a foreign language. The session proceeded as follows:

Phase 1, noun training and testing. Participants were first trained on the four nouns in isolation (Fig. 1, top row) during a block of 24 trials (6 per noun). In each trial, a single character appeared, and after a second its description (a bare noun) was displayed (orthographically and auditorily). After the audio of the description finished playing, the character and its text description remained on the screen for two seconds before moving to the next trial. Participants were instructed to repeat each description aloud. Participants were then tested on the noun vocabulary using a noun-selection task and an oral production task (12 trials per block, 3 per noun). In noun-selection trials, a character appeared, and participants had to select the correct noun from two choices. The foil noun was randomly selected at each trial. Feedback was provided (an in/correct-answer sound effect along with the image and correct noun; if incorrect, the audio of the noun was also played). In oral production trials, a character appeared, and participants had to say the corresponding noun aloud and press enter after they were finished. Feedback was also provided: the correct noun was displayed visually and auditorily after participants pressed enter.

Phase 2, one-marker NP training. Participants were next trained on noun phrases with a single marker, either number or case. There were three trial types (Fig. 1, middle row): (1) a group of 2–4 of the same characters in isolation (Number only), (2) an event with a (different) singular agent and a singular patient (Case only), or (3) an event with a plural agent, and a singular patient (Number & Case, where crucially each marker occurs with a different noun phrase). On each training trial, participants saw an image, and after a second its description was presented (orthographically and auditorily). After the audio description finished playing, the image and its text description remained on screen for three seconds (or two, for Number Only trials) until the next trial. Participants were instructed to repeat each description aloud. There were 62 trials total (randomised): 8 bare noun, 18 Number Only (each character appeared at least four times, and two of them appeared five times), 18 Case Only (randomly chosen from the 36 possibilities), and 18 Number & Case images (again randomly chosen).

Phase 3, one-marker NP comprehension test. Participants were then tested on their comprehension of one-marker NPs in an image-selection task. On each trial, they got a description and had to select the corresponding image out of an array of two. Feedback was provided (an in/correct-answer sound effect along with the image and correct orthographic description; if incorrect, the audio description was also played). The foil image was selected according to the trial type. For bare nouns and Number Only trials, the foil image was the same character with the wrong numerosity (e.g., singular instead of plural). For Case Only and Number & Case trials, the foil was the same event type with agent and patient reversed. There were 34 trials total (randomised): 4 bare nouns, and 10 of each of the three one-marker NP trial types. Note that the results of comprehension tests will not be reported in the main text as

---

[4] We use the terms *pre-* and *post-nominal* instead of *prefixal* and *suffixal* morphology to account for both bound and unbound orthographic representations of case and number morphology.
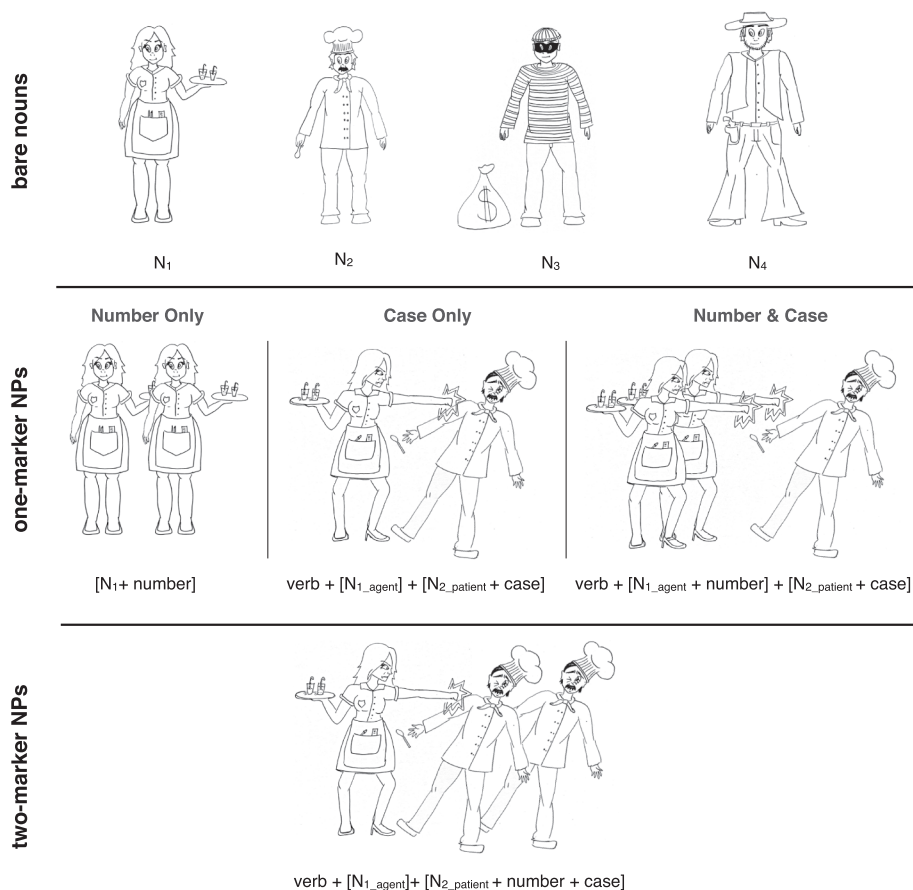
Fig. 1. Example visual stimuli and corresponding descriptions. Top to bottom: the four characters (nouns) in the miniature language in isolation; example events with one marker within a given noun phrase (either number or case); example events requiring two markers within a noun phrase (number and case, testing only).

they were designed to train participants on the language; results are nevertheless summarised in Fig. S4 in the supplementary material. With very few exceptions, participants performed at ceiling: mean accuracy was consistently above 95% across experiments.

Phase 4, one-marker NP written production test. Participants were then tested on their ability to produce one-marker NP descriptions. On each trial, participants saw an image and had to type in the corresponding NP(s). Verb forms were provided for Case Only and Number & Case trials, thus trials were essentially fill-in-the-blank. Participants were asked to press enter to submit their answer and feedback was then provided (an in/correct-answer sound was played, along with the image and correct description). There were 16 trials total (randomised): 4 trials for each of the trial types participants had been trained on so far.

Phase 5, two-marker NP production tests. In the two testing blocks, participants had to provide first written, then oral descriptions of events which required them to extrapolate to the held-out phrase type: two-marker NPs, with plural patients (Fig. 1, bottom row). The written production task was identical to Phase 4, except it only included the held-out trial types (12 trials, randomly chosen) and no feedback was given. This written task was added in order to familiarise participants with the held-out trial types prior to the final oral production test phase. While we show results of this phase in all figures, we therefore do not include these data in our statistical models.

In the final critical testing block, participants were asked to produce oral descriptions for *all* trial types in the language, including the critical held-out type. On each trial, participants saw an image and were asked to provide a description aloud. Verb forms were provided for the critical held-out sentence type as well as for Case Only and Number & Case trials, thus these trials were essentially fill-in-the-blank. As in previous

trials, production was self-paced: participants were asked to press enter after describing the image and only then did they move onto the next trial. Feedback was provided (as described for previous trials), but *only* when the target description did not contain a two-marker NP (i.e., for one-marker NP and bare noun trials). There were 58 trials total (randomised): 36 two-marker NP trials, 6 trials of each of the three one-marker NP trial types, and 4 bare noun trials. The oral production test lasted approximately 5-6 min; the written test lasted 1-2 min. For all experiments, information about individual duration times of critical production tests (oral and written) is given in Fig. S5 in the supplementary material.

Post-experimental questionnaire. Participants were asked to answer a series of questions about the language they had learnt right after the completion of the experiment. The questions were presented in text form on the computer screen (one at a time) and participants provided written responses. Firstly, participants were asked about the meaning of each marker individually (order randomised). Secondly, they were asked about the position of each of the markers in relation to the noun, that is, whether the marker appeared before or after the noun that referred to the character. Lastly, participants were asked to type in the languages they speak fluently.

Participants

Forty-one native English speakers were recruited from the University of Edinburgh's Careers Services database. Participants were paid £6 for a 35-min-long experimental session. In accordance with our preregistered analysis plan, participants (N = 1) whose accuracy was lower than 2/3 in non-critical production tests (i.e., on average across Phase 1 and Phase 4) were excluded. We further excluded from the analysis any critical testing trials (in Phase 5) with incomplete sentences (i.e., missing
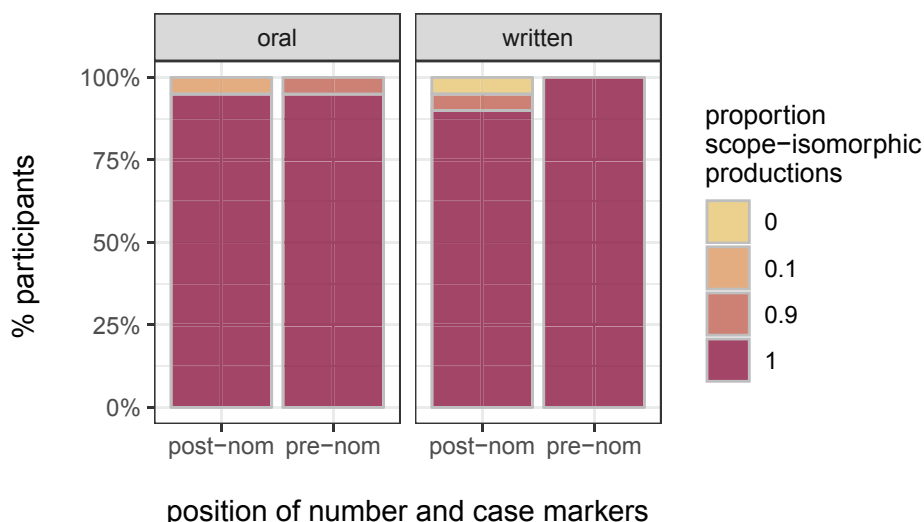
Fig. 2. Results from Experiment 1. Percentage of participants in Experiment 1 (English speakers) who produced scope-isomorphic responses a given proportion of the time, ranging from 100% of the time (darkest) to 0% of the time (lightest) in intervals of 10% (i.e., proportions are rounded to one decimal). Results are split by Marker Position (pre- vs. post-nominal) and faceted by Oral and Written production trials.

Table 1
Model output for Experiment 1 (oral production only).

	β	SE	z	$Pr(> |z|)$
(Intercept)	13.398	3.322	4.033	< 0.001
Marker Position	−0.219	2.437	−0.090	0.928

either number and/or case morphology in the patient NP). Summary statistics on (included) participants' accuracy scores in non-critical production trials are shown in Fig. S2 in the supplementary material; the proportions of excluded critical testing trials (mean < 3% across experiments) are shown in Fig. S3.

Results

Recall that, based on Universal 39 (Greenberg, 1963), participants were predicted to produce number markers closer to the noun stem than case markers. This should hold for both the pre- and post-nominal conditions. Our working hypothesis is that these orders are preferred because they reflect the scopal relations among morphemes.

Fig. 2 is a stacked histogram, showing the percentage of participants whose oral productions follow scope in 0–100% of trials across both conditions. For critical trials during the oral production task (in Phase 5), 95% of participants were (almost) perfectly consistent, producing two-marker NPs in the predicted order 95–100% of the time. Similar results were found during the written production task (also in Phase 5). In accord with our preregistered analysis plan, we ran a logistic mixed-effects regression model predicting use of scope-isomorphic morpheme order by Marker Position (pre-nominal vs. post-nominal) in two-marker NPs during the oral production phase. In all models, fixed effects were sum coded (unless stated otherwise), and random intercepts for participants were included; further random intercepts for item (noun) were included where possible (i.e., where variance was not 0).[5] Morpheme order was a binary variable (coded as 1 for a scope-isomorphic pattern, 0 otherwise). As shown in Table 1, the intercept (grand mean of scope-isomorphic productions across participants in both conditions) is positive and significant, confirming that the log-odds of producing scope-

isomorphic order is above chance ($P \approx 1$). The effect of Marker Position is not significant, confirming that this preference holds regardless of the pre- or post-nominal positioning of the markers.[6]

Discussion

The results of Experiment 1 are consistent with the hypothesis that participants' preferences reflect scope relations—here between number and case morphemes—which in turn determine proximity to the stem (i.e., number is placed closer to the stem than case). Under this hypothesis, participants infer scope-isomorphic orders without explicit exposure to them on the basis of the distinct meaning and function of the two markers: number morphemes directly modify the entity referred to by the noun, while case morphemes signal an external relationship between the entity and some event. This hypothesis is strengthened by the fact that we can rule out the effect of both raw and bigram frequency in driving our results, since these were held constant in our stimuli.

However, there is at least one alternative explanation for our results. In particular, they may reflect the fact that English overtly marks (plural) number with a bound morpheme but it does not have morphological case marking on nouns (case marking is restricted to pronouns and perhaps the genitive). Exactly how this would lead to a preference for placing number closer to case is not totally clear. One possibility is that familiarity with, or accessibility of the number marker leads English speakers to place it closer to the noun. Note that in the post-experimental questionnaire, 100% of participants assigned a meaning of plurality to number markers but only 60% (24/40) of participants assigned a meaning of object or direction-of-action to case markers (see Table S2 in the supplementary material for further details on post-experiment questionnaire reports for all experiments). Interestingly their interpretation of the case marker depended on the position of the markers: 85% (17/20) of participants assigned a correct meaning when it was pre-nominal, compared to only 35% (7/20) when it was post-nominal. This is likely because English speakers are familiar with

---

[5] Please note that the analyses outlined in the preregistration erroneously included by-participants random slopes for the effect of Condition, which we have not included because we use a between-subjects design.

[6] In principle, it would be possible to explore, using these data, whether participants' behavior differs in written and oral production. We do not report further such exploratory analyses here because results across tasks (written and oral production) would be perfectly (or almost perfectly) correlated, and models would be consequently deprecated.

pre-nominal words that can indicate this type of relationship, namely prepositions—of the 85% who interpreted the marker correctly in pre-nominal position, around 60% explicitly described case marker meanings with prepositions. To summarise, participants may consistently assume that plural is marked as a bound morpheme (regardless of its orthographic representation), and place case outside it for different reasons (other than scope) depending on marker order: in the pre-nominal condition they may place case outside number because they are treating the case marker as a preposition; and in the post-nominal condition they may place it outside number simply because they are uncertain about its meaning.

Importantly, there is no clear relation between differences in participants' interpretation of case morphemes and their use of isomorphic order. Participants who perfectly understood both morphemes do not behave any differently than those who do not. Nevertheless, in Experiments 2 and 3, we address the issue of marker familiarity/accessibility in two different ways. First, in Experiment 2 we test speakers of a language with overt case marking. Then, in Experiment 3, we return to English speakers, but alter the morphological system, training stimuli, and procedure in a number of ways to provide a stronger test of our hypothesis.

Experiment 2

To rule out the learners' unfamiliarity with case marking as a driver of ordering preferences in our task, we replicated Experiment 1 with native speakers of Japanese. In contrast to English, Japanese overtly marks cases (including accusative) via suffixation; however, the marking of plurality is exceptional (Nakanishi and Tomioka, 2004). The closest thing to number marking *on nouns* are associative plural classifiers or collectivising suffixes ("-kata", "-tachi", "-ra", "-domo") which are placed between the noun stem and the case markers; associative plurals indicate that a word refers to a group associated with an individual (e.g., "Tanaka-tachi" meaning 'Tanaka and his associates'). Cross-linguistically, these are restricted to pronouns, proper names and human nouns, with the focal referent interpreted as definite (Vassilieva, 2005). However, standard numerosity in Japanese is typically expressed instead via quantifiers (e.g., "hon-ga takusan" meaning 'many books'), classifiers (i.e., "hon-ga ni-satsu" meaning 'two books') or numerals (which usually combine with classifiers within the noun phrase, e.g., "go-rin-no hana" meaning 'five flowers'). All of these are outside the case-inflected noun. Japanese speakers should therefore have no trouble

acquiring a novel accusative case marker, and if anything should find the case marker more familiar/accessible than the number marker.

Methods

Experiment 2 was identical to Experiment 1, with one difference: the input lexicon. Rather than using a language with English-like phonotactics, the lexicon for Experiment 2 matched Japanese phonotactics. The preregistered hypothesis and analysis plan for Experiment 2 is accessible at http://osf.io/akcyp.

Input language

Lexical items in the language were displayed in Katakana (instead of Latin) script. The three semi-nonce verbs (which contain the stem of the existing verbs in Japanese) are: ケルラ ([ke↓rura]), ナグラ ([na↓gura]) and サ;スラ ([sa↓sura]), which refer to 'kicking', 'punching' and 'pointing' respectively. The (trisyllabic) nonce nouns are: ソ;ギナ ([sogi↓na]), ダクメ ([daku↓me]), ネチビ ([ne↓tɕibi]), and タソヌ ([taso↓nu]), naming four characters (a burglar, a chef, a cowboy, and a waitress). The two nonce markers (one for number, one for case) are randomly chosen from the following set: セヒ ([se↓hi]), ギト ([gi↓to]), ヨザ ([yo↓za]). Word order in sentences was always Verb-Agent-Patient. Half of the participants were assigned to each of two conditions as per Experiment 1 (i.e., pre-nominal or post-nominal morphology). Auditory stimuli were recorded in a sound-attenuated room by a 28yo female speaker of Japanese.

Procedure

The experiment was conducted in a quiet room, with all instructions provided in Japanese, and a Japanese-speaking experimenter. Participants were told that they would be learning part of a foreign language. The session proceeded exactly as outlined for Experiment 1.

Participants

Forty native Japanese speakers were recruited from Waseda University (Tokyo, Japan). Participants were paid ¥1000 for a 35-min-long experimental session. Note that all participants spoke English as a second language, which means that if these participants access knowledge from both their first and second languages when learning a new language, they will be familiar with both number and case. Though notice that the materials were designed to resemble Japanese, as described
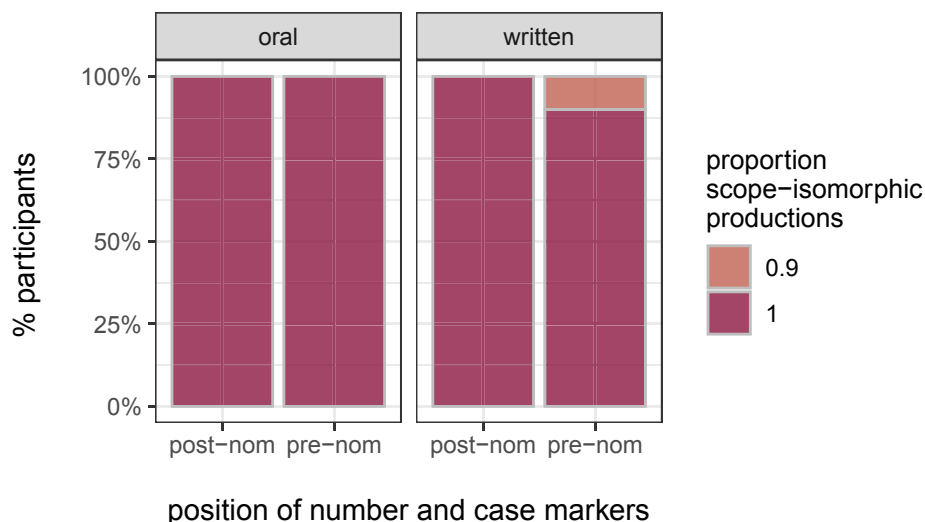

Fig. 3. Results from Experiment 2. Percentage of participants in Experiment 2 (Japanese speakers) who produced scope-isomorphic responses a given proportion of the time, ranging from 100% of the time (darkest) to 0% of the time (lightest) in intervals of 10% (i.e., proportions are rounded to one decimal). Results are split by Marker Position (pre- vs. post-nominal) and faceted by Oral and Written production trials.

above.

Results

The proportion of participants whose productions followed scope in 0–100% of trials is shown in Fig. 3.

All participants produced number consistently (95–100%) closer to the noun than case during oral production trials. This was true in both the pre-nominal or post-nominal marker conditions. Similar results were found for written productions. We ran a logistic mixed-effects regression model predicting scope-isomorphic productions by Marker Position (pre- vs. post-) and Experiment (English vs. Japanese). As shown in Table 2, the intercept is positive and significant, confirming above-chance production of scope-isomorphic order. The non-significant effects of Marker Position and Experiment confirm that this preference holds regardless of pre- or post-nominal positioning of the markers, and regardless of the native and test languages of participants.

Discussion

Experiments 1 and 2 are consistent with the hypothesis that learners have a natural preference to produce number morphology closer to the noun stem than case. These results hold for pre- and post-nominal orders, suggesting that the preference is not driven by linear order: number appears before case in post-nominal orders, but after case in pre-nominal orders. Our results hold for speakers of both English and Japanese, and thus do not appear to be driven by familiarity with a particular morpheme (number or case respectively); the fact that case affixes are more prominent in Japanese than number affixes (which recall are used less often only to indicate associates) did not lead participants to place them closer to the stem. As in Experiment 1, frequency of the markers cannot explain the preference for isomorphic order either: markers for case and number occurred with equal frequency in the input language, as did each stem + morpheme bigram. Consequently, the ratio of stem + number to stem alone was equal to the ratio of stem + case to stem alone, prior to starting the critical testing phases ($ratio \approx 1.00$) and overall ($ratio \approx 1.12$). The parsability of the morphemes is therefore also ruled out as an explanation, since the frequencies of stem + morpheme forms relative to stems alone are the same for each.

However, it is worth again discussing participants' interpretation of these elements. In Experiment 2, as in Experiment 1, there was a difference in the degree to which participants provided the correct interpretation of number and case morphemes. All participants assigned a meaning of plurality to number markers in the post-experimental questionnaire, while 68% (27/40) of participants assigned a meaning of object or direction-of-action to case markers. This was again mediated by position, with participants in the pre-nominal condition interpreting the case marker largely correctly: 75% (15/20) explicitly identified it as marking the object/patient or as an accusative marker, 15% (3/20) identified it as direction-of-action. In the post-nominal condition, 35% (7/20) participants interpreted it as marking the object/patient or as an accusative marker, and 10% (2/20) identified it as direction-of-action, but 35% (7/20) interpreted it as a politeness marker (see Table S2 in the supplementary material for the full distribution). This suggests that what drives Japanese speakers to differ in their interpretation of the case

Table 2
Model output comparing Experiments 1 and 2 (oral production only).

	β	SE	z	$Pr(> |z|)$
(Intercept)	12.112	1.958	6.187	< 0.001
Marker Position	−0.295	1.302	−0.227	0.821
Experiment	−0.012	1.302	−0.009	0.993
Marker Position × Experiment	0.05	1.302	0.038	0.970

marker pre- and post-nominally is the availability of a politeness marker interpretation; Japanese has phrase- and sentence-final politeness markers (Tsujimura, 1996).

Importantly though, as for Experiment 1, participants who correctly interpreted both markers nevertheless always chose the scope-isomorphic order. The results obtained so far thus continue to suggest a bias favoring scope-isomorphism. Participants assume that number is ordered closer to the noun stem than case because number directly modifies the entity referred to by the noun, while case signals the role that (modified) entity plays in an event. In Experiment 3, we set up a more stringent test of the scope hypothesis and the alternative hypothesis based on our post-experiment questionnaires—that participants' ordering preferences are at least partly motivated by differences in how familiar or accessible marker meanings are.

Experiment 3

In Experiment 3, we alter several aspects of the language training and procedure in order to ensure that case and number marking do not differ in terms of how easily participants can access or understand them. Specifically, we make four changes: (i) we use overt markers that are both unfamiliar to English speakers (described below), (ii) we present the markers as affixes bound to the noun stem in text, (iii) we explicitly introduce the meanings of both markers (and only include in our analysis participants who correctly interpret the markers in our post-experiment questionnaire), and (iv) we eliminate all trials in which stem + number NPs occur in isolation (Number Only) both in training and testing. The latter gets rid of an additional concern: in Experiments 1 and 2, NPs with number marking were presented in isolation, while NPs with case were not (they were always in the context of an event). This might have had some effect on morpheme order, potentially biasing learners towards producing stem + number forms as a unit before inflecting them for case. Eliminating Number Only trials will also have the effect of skewing frequency toward the case marker: stem + number bigrams will be half as frequent as stem + case bigrams, and the ratio of stem + number to stem alone (average mean ratio across stems = 0.60, maximum SD within a participant's input = 0.17) will also be much lower than the ratio of stem + case to stem alone (average mean ratio across stems = 1.19, maximum SD = 0.21). This makes number more parsable than case, therefore according to Hay (2001)'s parsability principle, number might be more likely to appear further away form the stem than case.

If we still find a preference for scope-isomorphic order in Experiment 3, we can conclude that a cognitive bias for ordering number closer to the stem than case is both stronger than (absolute and relative) frequency effects, and not dependent on any special status of the number marker in the input. The hypotheses and analysis plan for Experiment 3 were not preregistered, however they follow those outlined for the previous experiments.

Methods

Input language

The input language was as per Experiment 1 but with a distinct morphological system in featuring nominative and singulative markers instead of accusative and plural markers. Singulative marking—wherein nouns are overtly marked when they are singular rather than when they are plural—does not occur in English, and in fact it is cross-linguistically rare to find it along with plurality that is exclusively unmarked (Universal 35; Greenberg, 1963). Nominative, like accusative marking on nouns, is also not present in English; however, it cannot be as easily interpreted or translated as a preposition. The visual stimuli were adapted accordingly: Case Only trials (i.e., one-marker NP trials with case marking but no number marking) featured both plural agents and patients—plural number was unmarked and agents were overtly marked. Unlike in Experiment 1, both case and number markers

appeared as affixes bound to the noun stem when presented in text form (i.e., no spaces). The audio input stimuli remained the same as in Experiment 1 nonetheless, without pauses between nominal morphology. With the inclusion of orthographically bound morphology, we expected participants to be less primed to reinterpret case markers as adpositions.

Experimental procedure

The experimental procedure was identical to Experiment 1 with two exceptions. First, participants were explicitly told the grammatical features they would be learning prior to the start of the experiment (and again after noun training and testing, Phase 1). Fig. 4 shows the instructions presented to participants. Second, participants were never presented with Number Only trials. This meant that bare noun trials in Phase 3 (one-marker NP comprehension test) were also excluded. Consequently, the overall number of trials in Experiment 3 was lower than in Experiment 1 (188 vs 230 trials). There was an additional minor difference in procedure: for two-marker written trials, participants could not advance to the next trial until they typed the correct number of characters. This encouraged participants to produce both bound markers together.

Participants

Thirty-four English speakers were recruited and compensated as per Experiment 1. They were again randomly allocated to one of two conditions, pre- and post-nominal inflection. We added an additional exclusion criterion to those described in Experiments 1 and 2 (which also apply here): we excluded the data from participants who in the post-experimental questionnaire did not correctly report the meaning of the markers as explicitly taught (i.e., we only included participants that described case markers as subject/agent/nominative markers and singulative morphology as a singular marker). Following these criteria, the data of 13 participants were excluded, leaving the data of 21 participants (N = 11 pre-nominal, N = 10 post-nominal) for analysis.

Results

Fig. 5 shows the percentage of participants whose productions follow scope in 0–100% of trials across conditions. As in Experiments 1 and 2, participants show a very strong preference for scope-isomorphic orders, with the number marker closer to the noun stem than case. A logistic mixed-effects regression model predicting use of scope-isomorphic order by Marker Position confirms that scope-isomorphic orders are produced significantly above chance regardless of the marker position (see Table 3).

Discussion

In Experiment 3 we have again replicated the results from Experiments 1 and 2: participants assume that number marking is placed closer to the stem than case. These results allow us to more confidently rule out the possibility that prior linguistic knowledge is driving the results we observed in Experiments 1 and 2. On the one hand, the number and case markers participants were trained on (i.e., singulative and nominative respectively) were equally unfamiliar to them prior to the experiment; it is therefore unlikely that a priori familiarity or accessibility of the number marker led participants to place number closer to the noun stem. On the other hand, we presented both markers as bound morphemes, and gave participants explicit descriptions of the marker meanings (and recall that only participants who provided correct descriptions in our post-experiment questionnaire were considered); this makes it unlikely that any alternative conceptualisation of case markers (e.g., as adpositions) led participants to place case further from the noun stem than number.

On top of this, because we removed trials in which the stem + number occurred in isolation, we can also rule out the possibility that these trials drove the preference to place number closer to the stem in
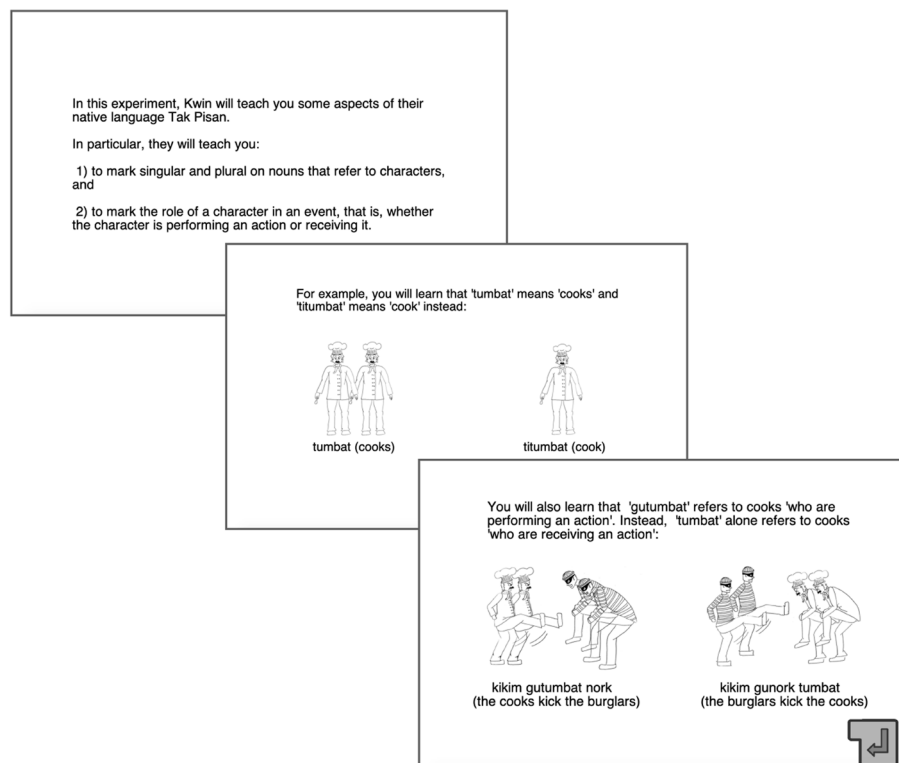

Fig. 4. Instruction trials explaining the function of number and case markers in Experiment 3. We provided participants with explicit descriptions of the grammatical features to be learned prior to the start of the experiment. The position of the markers in relation to the nouns varied according to the assigned condition; this example is taken from the pre-nominal condition.

Experiments 1 and 2. Further, we can conclude that participants' bias for scope-isomorphic patterns is stronger than any distributional effects of absolute and relative frequency in this experiment (cf. Hay, 2001). To strengthen this conclusion we also ran a version of Experiment 3 (which we will refer to as Experiment 3') using the same morphological marking system as in Experiments 1 and 2, namely plural and accusative markers, presented as independent non-bound morphemes, and no explicit training on marker meanings. In other words, the input language and procedure were identical to Experiment 1, the only difference being the exclusion of Number Only trials, in which stem + number occurred in isolation. Twenty-one English speakers were recruited and compensated as for Experiment 1. They were divided between two conditions, pre-nominal and post-nominal inflection (N = 10 and N = 11 respectively). Following the exclusion criteria set out for Experiments 1 and 2, the data of further two participants (pre-nominal condition) were excluded from analysis. Results again revealed a strong bias for scope-isomorphic order of the case and number markers. In fact *all* participants placed number closer to the noun stem than case. Further details on Experiment 3' can be found in the supplementary material A.

Of course, these results should not be taken to suggest that frequency cannot affect morpheme order regularities in the absence of competing cognitive biases; as outlined above, a number of corpus studies show that absolute and relative bigram frequencies are a good predictor of language-specific morpheme order distributions in the presence of variability (e.g., Ryan, 2010) and/or in the absence of clear competing cognitive biases or other grammatical principles (e.g., Baayen, 1993; Hay, 2001; Ryan, 2010).[7]

---

[7] However, to this date, there is no evidence from linguistic corpora supporting the role of frequency effects on number and case linearisation specifically, where a bias favouring scope-isomorphic orders can be postulated (however, cf Hahn et al., 2020a). We analysed the distributional properties of number and case morphology in the Universal Dependencies (UD) corpus data (version 2.1., Nivre et al., 2017) for Turkish and Hungarian—two available agglutinative languages which have number and case morphology and for which the corpora are morphologically annotated. We did not find evidence in the Turkish data supporting the parsability principle proposed in Hay (2001): $\frac{bigram_{base+case}}{base}$ and $\frac{bigram_{base+number}}{base}$ ratios are not statistically different ($median_{num} = 0.333$; $median_{case} = 0.4$; $U = 292783.5$, $p = 0.133$). However, we did find a significant difference between the ratios in Hungarian ($median_{num} = 1.0$; $median_{case} = 0.5$; $U = 135953.0$, $p < 0.001$). This difference could nevertheless be driven by the long tail of the frequency distribution of the different cases (i.e., there are more case markers with fewer instances). If we look at specific case markers such as accusative, we find similar ratios as for plural number both in Turkish ($median_{pl} = 0.333$; $median_{acc} = 0.333$; $U = 62123.0$, $p = 0.385$) and Hungarian ($median_{pl} = 1.0$; $median_{acc} = 1.0$; $U = 45569.0$. $p = 0.093$). Moreover, we did not find evidence either for differences in the degree of statistical dependence between number and case markers. Following Dyer (2018), we further calculated the entropy of the distribution of noun lexemes or lemmas that contained either number or case morphology. The hypothesis in Dyer (2018) is that "a dependent whose heads form a peaked probability distribution is easier to integrate—and therefore has a lower entropy—than a dependent whose heads form a flatter distribution". If we could extrapolate this to case and number morpheme order we should expect lower entropy for number values than for case values. The scores obtained for the overall features are identical as is to be expected given that all nouns need to be marked for both features (Turkish: $num = 0.736$ bits, $case = 0.736$ bits; Hungarian: $num = 0.812$ bits, $case = 0.812$ bits). The scores by feature value further show very similar scores for plural number and for accusative case for instance (Turkish: $pl = 0.809$ bits, $acc = 0.847$ bits; Hungarian: $pl = 0.860$ bits, $acc = 0.903$ bits): the latter is only greater by approximately 0.04 bits (further details are provided in Table S1 in the supplementary material). The results from this basic corpus analysis do not reveal a sharp enough contrast between the (synchronic) distributional properties of number and case morphology to provide an explanation (or operationalisation) of a constraint on case-number linearisation; further (cross-linguistic) work is nevertheless required to systematically assess this.

Altogether, results from Experiments 1–3 reveal a very strong bias in favor of scope-isomorphic orders of case and number morphology. This bias is not easily overridden by frequency effects and cannot be straightforwardly explained away by participants' prior linguistic knowledge. However, in natural language, aside from frequency effects and marker availability (mediated by prior linguistic knowledge), alternative competing formal cognitive biases may be present. One such bias, prominent in models of morphological learning, comes from the notion of locality. Dependencies between morphemes (e.g., between an allomorph and the stem that triggers it) tend to be local, or linearly adjacent (Bobaljik, 2012; Embick, 2010; Moskal, 2015). Locality constraints might in turn potentially reflect an additional general bias for local dependencies (see e.g., Gomez, 2002; White et al., 2018). In Experiment 4 we test the strength of the uncovered scope-isomorphism bias in the presence of an alternative competing locality bias.

Experiment 4

Previous work within Distributed Morphology has investigated contextual allomorphy in the presence of number and case morphemes. In this research, it is typically assumed that case hierarchically outscopes number (Bonet and Harbour, 2012; Halle, 1990; Halle and Marantz, 1993; Halle and Marantz, 1994). In line with locality restriction outlined above (i.e., dependencies between morphemes tend to be local and linearly adjacent; Bobaljik, 2012; Embick, 2010), in the presence of number morphology, case does not tend to trigger root suppletion in nouns for instance (Moskal, 2015, but see Radkevich, 2010). However, in the presence of stem-triggered case allomorphy, would learners rather violate isomorphism to satisfy alternative locality constraints on linearisation? In Experiment 4 we teach participants a language in which the form of the case marker is in fact dependent on (the lexical and phonological identity of) the noun. This linguistic system thus contains both stem-dependent case allomorphy, and number morphology (a typologically rare nominal system). This input language allows us to again test the strength of the scope-isomorphism bias, here in the face of a competing locality bias. Because such a system creates a dependency between the noun stem and the case marker, a locality bias would predict that learners will prefer to have these two elements linearly adjacent to one another. The effect of the scope-isomorphism bias uncovered in Experiments 1–3 may override this pressure from locality, or alternatively, the locality bias may interfere with the placement of number in closer proximity to the noun stem. In the latter case, we should observe a higher proportion of anti-scopal order productions of case and number morphology (typologically rare) in the presence of stem-dependent case allomorphy. The hypotheses and analysis plan for Experiment 4 were not preregistered, however they generally follow those described for the previous experiments.

Methods

Input languages

This was a 2x2 design, with Marker position (pre- and post-) and Allomorphy (no allomorphy vs. case allomorphy) varying between-subjects. The input language in no-allomorphy conditions was as in Experiment 1, but case and number markers were orthographically bound as in Experiment 3. The input language in the case allomorphy conditions differed additionally in having *two* accusative case markers, which alternated based on the length of the noun: one case marker appeared with disyllabic nouns ("negid", "tumbat"), and the other case
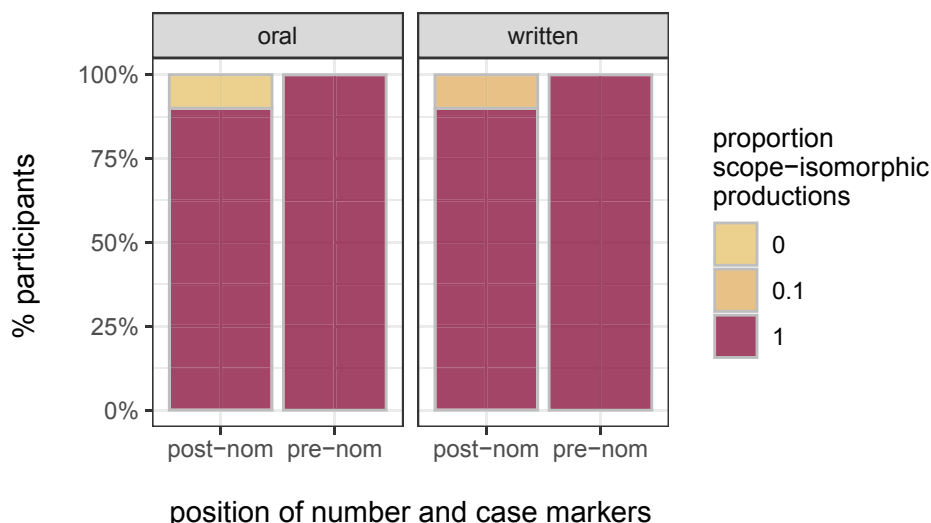
Fig. 5. Results from Experiment 3. Percentage of participants in Experiment 3 who produced scope-isomorphic responses a given proportion of the time, ranging from 100% of the time (darkest) to 0% of the time (lightest) in intervals of 10% (i.e., proportions are rounded to one decimal). Results are split by Marker Position (pre- vs. post-nominal) and faceted by testing Phase (Oral and Written production trials).

Table 3
Model output of the results of Experiment 3 (oral production only).

	β	SE	z	$Pr(> |z|)$
(Intercept)	24.311	7.259	3.349	< 0.001
Marker Position	−6.021	7.259	−0.829	0.407

marker appeared with monosyllabic nouns ("vaem", "nork").[8]

Experimental procedure

The procedure was identical to Experiment 1 for the no allomorphy conditions and identical to Experiment 3 (but without explicit training) for the case allomorphy conditions so that all (number and case) markers could appear with the same exact absolute frequency across conditions. Case marking needs to appear double the times than number marking in the case allomorphy conditions for each marker to appear with the same absolute frequency since it has three markers in total (two case, one number). We achieved this by removing the Num Only trials as we did in Experiment 3. In the no allomorphy conditions, however, case and number need to appear with the same absolute frequency, just as in Experiment 1. This difference between conditions is unlikely to be problematic because Experiments 1 and 3 (as well as Experiment 3') indicate that differences in the absolute frequency of case marking do not lead to differences in participants' behaviour. For all conditions, and as per Experiment 3, participants were required to type in the correct number of characters for two-marker written trials to advance to the next trial. Lastly, to mitigate the added complexity of learning case allomorphy (and the reduced number of exposure trials due to removal of Num Only trials), we added four additional trials for participants in the case allomorphy conditions in Phase 4 (one-marker NP written production test) prior to the critical testing phases (i.e., 20 instead of 16 trials).

Participants

Forty-four English speakers were recruited and compensated as for

Experiments 1 and 3. They were evenly divided between four conditions, as described above. The data of four participants were excluded from analysis based on exclusion criteria which are slightly more stringent than in Experiments 1–3. Here we excluded participants whose accuracy was lower than 2/3 during non-critical production tests in Phase 1 (bare nouns, oral) *and* Phase 4 (bare nouns and one-marker NPs, typed) separately (cf. Experiments 1–3 where accuracy had to be higher than 2/3 overall). We used this stricter criterion to ensure the exclusion of participants who did not learn the correct conditioning for case markers. As in previous experiments, testing trials with incomplete sentences were also excluded.

Results

Fig. 6 shows the percentage of participants whose oral productions follow scope in 0–100% of trials across all four conditions. For the no allomorphy conditions, participants strongly prefer the scope order, with the number marker closer to the noun stem than case. This replicates our previous results. By contrast, in the case allomorphy conditions, we find that more than half of the participants produce case closer to the noun stem. A logistic mixed-effects regression model predicting use of scope-isomorphic order by Marker Position (sum coded) and Allomorphy (treatment coded)[9] confirms that while participants in the no allomorphy conditions produce scope-isomorphic orders significantly above chance, there is a significant drop in the use of these orders in the condition with case allomorphy (see Table 4).

Discussion

These results indicate that the scope-isomorphic bias can be overridden by a locality bias; however, almost half of the participants (40%) still produced scope-isomorphic orders with non-adjacent case allomorphy. Our results support the idea that locality constraints on case allomorphy are one case of a more general bias favoring local dependencies (Futrell et al., 2015; Gibson, 1998; Gildea and Temperley,

---

[8] An example of this type of conditioned allomorphy based on syllable counting would be the Estonian genitive plural affixes "-te" and "-tte" and partitive plural affixes "-sit" and "-it". The plural suffixes in both cases vary according to whether the stem is even-numbered or odd-numbered (see Murk, 1992, pp. 295–296).

[9] With treatment coding for the fixed effect of Allomorphy, the intercept is the mean across pre- and post- nominal conditions only for the no-allomorphy conditions. This allows us to interpret the model's intercept as in the previous experiments as well as to compare case allomorphy to no allomorphy directly with simple effects.

2010; White et al., 2018). In sum, results form Experiment 4 suggest the existence of two biases, one favoring scope-isomorphic order (based on morphosyntactic and semantic composition), and one favoring local contextual allomorphy. It is worth noting that the results from the case allomorphy condition serve as a proof-of-concept—they indicate that both of these mechanisms are in principle at play in morphological learning—rather than as a reflection of typology. The patterns of allomorphy participants were trained on in this case appear to be very rare cross-linguistically. In other words, it appears that locality and scope biases tend to align in natural languages with distinct case and number morphemes.

It is also worth noting that here, participants were highly accurate in interpreting both number *and* case morphemes across conditions. Importantly, 77% (31/40) of participants assigned a meaning of object or direction-of-action to case markers, with the majority interpreting them as marking the object/patient rather than direction-of-action (see Table S2 in the supplementary material for details). On the one hand, these results (as those in Experiment 3) suggest that the misinterpretations of the case marker in previous experiments cannot explain participants' ordering preferences; results from the no allomorphy condition are comparable to those from Experiment 1. On the other hand, they further suggest that when the markers are orthographically represented as bound morphemes, English speakers are less likely to generate spurious interpretations of case markers.

General discussion

In the experiments reported here, participants were trained on languages with distinct number and case morphemes. Descriptions involving events that required a single overt morpheme (either number or case) allowed participants to learn the morphemes themselves, and how they were ordered relative to noun stems in the language (i.e., pre- or post-nominally). However, descriptions that required using both overt morphemes were held out, allowing us to investigate learners' implicit assumptions about the relative order of case and number

Table 4
Model output for Experiment 4 (oral production only).

	β	SE	z	$Pr(>|z|)$
(Intercept)	15.148	4.329	3.500	< 0.001
Marker Position	0.381	4.180	0.091	0.927
Allomorphy	−27.506	5.064	−5.432	< 0.001
Marker Position × Allomorphy	−0.529	4.527	−0.117	0.907

markers. We found that participants' inferences were consistent and strong: they placed number closer to the noun stem than case (regardless of whether the markers were pre- or post-nominal). This bias mirrors a typological generalisation known as Universal 39 (Greenberg, 1963), providing a potential causal link between human cognition and this recurrent cross-linguistic pattern. We have suggested that scope relations among morphemes can explain why this order is preferred. In particular, case (which marks the grammatical role of the noun in the event) scopes higher than number (which modifies the set properties of the entity). In other words, participants' ordering preferences match a structure-preserving linearisation of the semantic and morphosyntactic scope relations: case out-scopes number potentially because number morphemes directly modify the entity referred to by the noun, while case morphemes signal an external relationship between the entity and some event. The idea that, all things equal, linear proximity should reflect scope (either directly, or as mediated through syntax) has been a central hypothesised explanation in morpheme order for many years (Baker, 1985; Bybee, 1985; Grimshaw, 1986; Rice, 2000). More recent work has made the same argument for word order (Bouchard, 2002; Cinque, 2005), and this is indeed also supported by experimental evidence (Culbertson and Adger, 2014; Culbertson et al., 2020; Martin et al., 2020).

Importantly, we found strong evidence of a bias for scope-isomorphic order across two populations which differ in terms of their prior experience with case and number markers; English marks number but not
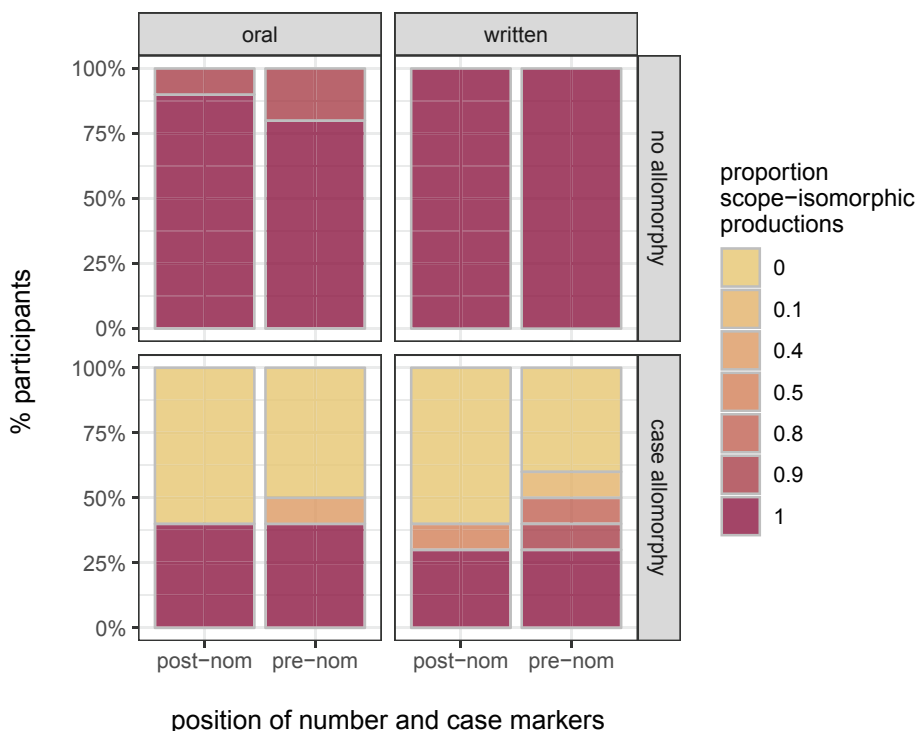

Fig. 6. Results from Experiment 4. Percentage of participants in Experiment 4 who produced scope-isomorphic responses a given proportion of the time, ranging from 100% of the time (darkest) to 0% of the time (lightest) in intervals of 10% (i.e., proportions are rounded to one decimal). Results are split by Marker Position (pre- vs. post-nominal) and faceted by Allomorphy (no allomorphy vs. case allomorphy) and testing Phase (Oral and Written production trials).

case, while Japanese marks case but not number. This suggests our results cannot be explained by learners' a priori familiarity with these markers. We also showed that these preferences are not likely to be explained (at least solely) by flexibility in how participants interpreted the markers—mediated by their prior linguistic knowledge. In Experiments 1 and 2 (and Experiment 3') our post-experiment questionnaire indicated that some participants were unsure about the interpretation of accusative case markers, or interpreted them as prepositions (English speakers), or politeness markers (Japanese speakers). This introduced the possibility that placing case outside number was driven by either the relative ease with which participants could identify the marker meanings, or by unintended interpretations of the case marker. However in Experiment 3, where we explicitly trained English participants on the meanings of the two markers, both of which were a priori unfamiliar to them (singulative and nominative), participants still showed a strong bias for scope-isomorphic order. Moreover, in Experiment 4, where learners were trained on *bound* plural and accusative morphemes (and no contextual allomorphy), almost all participants correctly described case-markers, and again showed a strong bias for isomorphic order. Finally, and perhaps surprisingly, the observed preference was not dependent on distributional information in the input: case and number markers never appeared together, and regardless of whether they had the same frequency during training, or the case marker was in fact more frequent (and more parsable; Hay, 2001), marker ordering preferences remained constant.

While the preference for scope-isomorphic order of case and number morphemes was very strong (regardless of morpheme frequency) in the absence of competing biases, results from Experiment 4 revealed that at least one such bias can override this default behavior. In particular, we showed that introducing stem-dependent contextual allomorphy for case led many participants to place the case morpheme closer to the conditioning noun. This suggests that scope-isomorphism in principle interacts with other constraints—i.e., imposed by morphophonological rather than semantic dependency relationships—as predicted by theories of linear locality (e.g., Embick, 2010), again supported by experimental research (White et al., 2018). Whether such allomorphy patterns are sensitive to locality in natural language points to the need for additional typological research (although see Božič, 2018; Moskal, 2015).

Conclusion

In this paper, we investigated a hypothesised link between a well-known typological generalisation of morpheme order, and cognitive biases stemming from semantic and morphosyntactic relations. Among languages with independent number and case morphemes, number shows a strong tendency to be more proximal to the noun stem than case (Greenberg's Universal 39). Across a series of artificial language learning tasks, we showed that adult speakers of English and Japanese consistently infer this relative order of morphemes after exposure only to the order of single overt morphemes in the language (i.e., case or number alone, preceding or following the noun stem). In other words, our results show that in the absence of explicit evidence, language learners default to a typologically common order of morphemes: with number closer to the noun stem than case. These results held regardless of the frequencies of individual morphemes or stem + morpheme bigrams (both argued to drive morpheme order effects in various languages). Our findings therefore support the hypothesised link between human cognition and Greenberg's Universal 39. However, this observed bias towards scope-isomorphism likely interacts with other complex factors. Here we identify one such factor, a bias for local relationships between morphemes and stems which condition allomorphy. We find that learners in some cases reverse their preference and place case closer to the noun stem in the presence of stem-dependent contextual allomorphy for case.

References

Baayen, H. (1993). On frequency, transparency and productivity. In *Yearbook of morphology 1992* (pp. 181–208). Springer.

Baker, M. (1985). The mirror principle and morphosyntactic explanation. *Linguistic Inquiry, 16*, 373–415. http://www.jstor.org/stable/4178442.

Bickel, B., Banjade, G., Gaenszle, M., Lieven, E., Paudyal, N. P., Rai, I. P., Rai, M., Rai, N. K., & Stoll, S. (2007). Free prefix ordering in chintang. *Language*, 43–73. https://doi.org/10.1353/lan.2007.0002.

Bobaljik, J. D. (2012). *Universals in comparative morphology: Suppletion, superlatives, and the structure of words* (Vol. 50). MIT Press.

Bonet, E., & Harbour, D. (2012). Contextual allomorphy. In J. Trommer (Ed.), *The morphology and phonology of exponence* (pp. 195–235). Oxford University Press.

Bouchard, D. (2002). *Adjectives, number and interfaces: Why languages vary*. Elsevier. https://doi.org/10.1353/lan.2006.0187.

Božič, J. (2018). Generalizations on root suppletion: Motivating a theory of contextual allomorphy. In W. G. Bennett, L. Hracs, & D. R. Storoshenko (Eds.), *35th West Coast Conference on Formal Linguistics* (pp. 114–123). Cascadilla Proceedings Project. http://www.lingref.com/cpp/wccfl/35/paper3381.pdf.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). John Benjamins Publishing.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry, 36*, 315–332. https://doi.org/10.1162/0024389054396917.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences, 111*, 5842–5847. https://doi.org/10.1073/pnas.1320525111.

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: deriving biases in noun phrase order from statistical properties of the world. *Language, 96*, 696–717. https://doi.org/10.1353/lan.2020.0045.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition, 122*, 306–329. https://doi.org/10.1016/j.cognition.2011.10.017.

Dyer, W. (2018). Integration complexity and the order of cosisters. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)* (pp. 55–65). https://www.aclweb.org/anthology/W18-6007/.

Embick, D. (2010). *Localism versus globalism in morphology and phonology*. MIT Press.

Fedzechkina, M., Chu, B., & Jaeger, T. F. (2018). Human information processing shapes language change. *Psychological Science, 29*, 72–82. https://doi.org/10.1177/0956797617728726.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences, 109*, 17897–17902. https://doi.org/10.1073/pnas.1215776109.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science, 44*, e12814. https://doi.org/10.1111/cogs.12814.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences, 112*, 10336–10341. https://doi.org/10.1073/pnas.1502134112.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1.

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science, 34*, 286–310. https://doi.org/10.1111/j.1551-6709.2009.01073.x.

Givón, T. (1979). *On Understanding Grammar*. John Benjamins.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436. https://doi.org/10.1111/1467-9280.00476.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of Human language* (pp. 73–113). MIT Press.

Grimshaw, J. (1986). A morphosyntactic explanation for the mirror principle. *Linguistic Inquiry, 17*, 745–749.

Hahn, M., Degen, J., & Futrell, R. (2020a). Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. *Preprint.* https://doi.org/10.31234/osf.io/nu4qz.

Hahn, M., Jurafsky, D., & Futrell, R. (2020b). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences, 117*, 2347–2353. https://doi.org/10.1073/pnas.1910923117.

Halle, M. (1990). An approach to morphology. In J. Carter, R. Dechaine, B. Philip, & T. Sherer (Eds.), *Proceedings of the Annual Meeting of the North East Linguistics Society* (Vol. 20, pp. 150–184). GLSA.

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale, & S. J. Keyser (Eds.), *The view from Building 20: Essays in Linguistics in honour of Sylvain Bromberger* (pp. 111–176). MIT Press.

Halle, M., & Marantz, A. (1994). Some key features of distributed morphology. In A. Carnie, H. Harley, & T. Bures (Eds.), *Papers on phonology and morphology* (pp. 275–288). MIT Press.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics, 39*, 1041–1070. https://doi.org/10.1515/ling.2001.041.

Hay, J., & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory, 22*, 565–596. https://doi.org/10.1023/B:NALA.0000027679.63308.89.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes, 24*, 876–909. https://doi.org/10.1080/01690960902719267.

Manova, S., & Aronoff, M. (2010). Modeling affix order. *Morphology, 20*, 109–131. https://doi.org/10.1007/s11525-010-9153-6.

Martin, A., & Culbertson, J. (2020). Revisiting the suffixing preference: Native-language affixation patterns influence perception of sequences. *Psychological Science, 31*, 1107–1116. https://doi.org/10.1177/0956797620931108.

Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (2020). Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa: A Journal of General Linguistics, 5*. https://doi.org/10.5334/gjgl.1085.

Mithun, M. (1999). *The languages of Native North America*. Cambridge University Press.

Moskal, B. (2015). Limits on allomorphy: A case study in nominal suppletion. *Linguistic Inquiry, 46*, 363–376. https://doi.org/10.1162/LING_a_00185.

Murk, H. W. (1992). *The structure and development of Estonian morphology*. Ph.D. thesis Indiana University.

Nakanishi, K., & Tomioka, S. (2004). Japanese plurals are exceptional. *Journal of East Asian Linguistics, 13*, 113–140. https://doi.org/10.1023/B:JEAL.0000019058.46668.c1.

Nivre, J., Abrams, M., Željko Agić, Ahrenberg, L., Aleksandraviiute, G., Antonsen, L., Aplonova, K., Aranzabe, M.J., Arutie, G., Asahara, M., Ateyah, L., Attia, M., Atutxa, A., Augustinus, L., Badmaeva, E., Ballesteros, M., Banerjee, E., Bank, S., Mititelu, V. B., Basmov, V., …, & Zhu, H. (2017). Universal Dependencies 2.1. (LINDAT/CLARIN Digital Library, Institute of Formal and Applied Linguistics [UFAL], Charles University, 2017).

Plank, F., & Filimonova, E. (2000). The universals archive: A brief introduction for prospective users. *STUF-Language Typology and Universals, 53*, 109–123. https://typo.uni-konstanz.de/archive/intro/index.php?pt=1.

Radkevich, N. V. (2010). *On location: The structure of case and adpositions* (Ph.D thesis). University of Connecticut Storrs.

Rice, K. (2000). *Morpheme order and semantic scope: Word formation in the Athapaskan verb* (Vol. 90). Cambridge University Press.

Rice, K. (2011). Principles of affix ordering: An overview. *Word Structure, 4*, 169–200. https://doi.org/10.3366/word.2011.0009.

Ryan, K. (2010). Variable affix order: Grammar and learning. *Language, 86*, 758–791. https://doi.org/10.1353/lan.2010.0032.

Tabullo, Á., Arismendi, M., Wainselboim, A., Primero, G., Vernis, S., Segura, E., Zanutto, S., & Yorio, A. (2012). On the learnability of frequent and infrequent word orders: An artificial language learning study. *The Quarterly Journal of Experimental Psychology, 65*, 1848–1863. https://doi.org/10.1080/17470218.2012.677848.

Tsujimura, T. (1996). *Introduction to Japanese Linguistics*. Blackwell Publishing.

Vassilieva, M. B. (2005). *Associative and Pronominal Plurality* (Ph.D thesis). Stony Brook University.

White, J., Kager, R., Linzen, T., Markopoulos, G., Martin, A., Nevins, A., Peperkamp, S., Polgárdi, K., Topintzi, N., & van de Vijver, R. (2018). Preference for locality is affected by the prefix/suffix asymmetry: Evidence from artificial language learning. In *Proceedings of NELS 48* (pp. 207–220). Graduate Linguistics Student Association.

Wilson, C. (2003). Experimental investigation of phonological naturalness. In G. Garding, & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)* (Vol. 22, pp. 533–546). Cascadilla Press. https://colincwilson.github.io/papers/WilsonExperimentalNaturalness2003.pdf.

Wunderlich, D. (1993). Funktionale kategorien im lexikon. In F. Beckmann, & G. Heyer (Eds.), *Theorie und Praxis des Lexikons* (pp. 54–73). Walter de Gruyter.