The University of Edinburgh
School of Philosophy, Psychology and Language Sciences

Evidence in Support of a Cognitive Bias for Cross-Category Harmony between the Verb Phrase and the Adpositional Phrase in the Absence of Surface-Level Patterns

Cliodhna Hughes

Master of Arts with Honours Linguistics

Supervisors: Simon Kirby and Jenny Culbertson

Date of Submission: 30th March 2022

Word count: 9,844

Abstract

The vast majority of languages for which we have the data display syntactic harmony between the verb phrase and the adpositional phrase. Previous work has suggested that this may be due to the cross-linguistic tendency for adpositions to derive from the verbs of verb phrases. Recent experimental work has indicated that there may be a cognitive bias for harmony which also contributes to its cross-linguistic prevalence. These studies, however, all use the same objects for both the priming and testing stimuli. This leaves open the possibility of participants noticing surface-level patterns such as *the saucepan is always gestured first*, and using these in their responses, thus giving the appearance of a preference for syntactically harmonic patterns. This paper describes three experiments that were designed to establish whether there is a cognitive bias for cross-category harmony between the adpositional phrase and the verb phrase, when the possibility of using surface-level patterns is removed. Experiment 1 investigates the possibility of a baseline preference for adpositional phrase order in English-speaking participants, because it is the extent of their preference for particular adpositional orders which is manipulated in experiments 2 and 3. Results of experiment 1 show no evidence for a baseline preference for either prepositions or postpositions. Experiment 2 finds that the experimental methods employed, namely silent gesture perception in combination with artificial language learning, are sufficient to demonstrate a harmony effect when the elements in the priming and testing stimuli are the same. This replicates the results of previous experiments which utilised different methods. Experiment 3 reveals that a preference for harmonic patterns is also present when there are no repeated elements in the priming and testing stimuli. However, this is to a somewhat lesser extent than in experiment 2. This indicates that there is a bias for cross-category harmony between the verb phrase and the adpositional phrase which at least partially contributes to the typological patterns we see. It also highlights the importance of ensuring that stimuli for such studies are designed to remove the possibility of surface-level patterns being used.

Contents

Chapter 1

Introduction

Languages differ widely from each other, yet there are a number of typological regularities which hold across languages (e.g. Greenberg, 1963). One such regularity is the tendency for dependents to consistently appear on one side of the head of the phrase. Languages which have this property are said to be *harmonic*. In this paper I investigate a particular type of harmony, specifically that between the verb phrase and the adpositional phrase. This is an instance of cross-category harmony, as it acts across phrase types. Previous work has suggested that harmony between the verb phrase and the adpositional phrase could be a product of the historical tendency for adpositions to derive from verb phrases or from genitives, in a process which preserves linear order (Moravcsik, 2010). Research using experimental methods, on the other hand, has found some evidence suggestive of a cognitive bias for cross-category harmony using the artificial language learning paradigm (Cook, 1988; Zhao and Fedzechkina, 2020; Wang et al., 2021). Zhao and Fedzechkina (2020) only find such evidence in the head-final direction, but this is likely due to the fact that there is a baseline preference for SOV for the type of events stimuli used in their experiments (Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014), which competes with the harmony bias, preventing the effect from being demonstrated in the prepositional condition. The studies in Cook (1988), Zhao and Fedzechkina (2020), and Wang et al. (2021) also fail to consider the possibility of a similar baseline adpositional order preference, either for prepositions or for postpositions, which may similarly be affecting their results. Moreover, in all three studies the items used in the test phase shared elements with those used in the training phase. It is therefore possible that participants used surface level rules in their responses, (eg. *the word for table always comes last*) rather than a more general syntactic rule (eg. *the head comes before the dependent*).

This paper describes three experiments which together aimed to establish whether there is evidence for a bias for cross-category harmony between the verb phrase and the adpositional phrase, on a syntactic rather than surface level. Experiment 1 of this paper uses silent gesture perception methods to demonstrate that there is no baseline preference for either prepositions or postpositions when testing native English speakers using this method. This is crucial for comparison with experiments 2 and 3, in which this preference is manipulated. Experiment 2 combines silent gesture perception with artificial language learning methods,

in an extrapolation paradigm, finding that evidence of a bias for cross-category harmony can be found using these methods on native English speakers, when the objects in the training stimuli are the same as those in the testing stimuli. This replicates the findings of Wang et al. (2021), which used different methods. Finally, experiment 3 tests whether such evidence is still found using these methods when there is no overlap in the objects used for the training stimuli and testing stimuli. There is evidence for a stronger harmony preference in experiment 2 than experiment 3, demonstrating the importance of ensuring surface-level patterns are avoided when designing stimuli for such experiments. Nevertheless, a preference for harmonic patterns is still found in experiment 3, supporting the hypothesis that a cognitive bias for harmony between the verb phrase and the adpositional phrase exists.

Chapter 2

Background

2.1 Word-Order Harmony

Word-order harmony is the phenomenon whereby dependents consistently appear on one side of the head within a language, across phrase types. Harmony is prevalent among the world's languages and is one of a number of statistical language universals first noted by Greenberg (1963). This paper deals specifically with harmony in relation to the verb phrase and the adpositional phrase. For our purposes, therefore, harmonic languages are those which have Verb-Object (VO) order and prepositions (head-initial), or which have Object-Verb (OV) order and postpositions (head-final). According to the World Atlas of Linguistic Structures (WALS), this instance of harmony is quite cross-linguistically robust, with only 56 languages showing an explicitly non-harmonic pattern, in a sample of 984 languages, as table 2.1 indicates (Dryer, 2013c).

Table 2.1: Table showing correlation between adposition type and relative orders of object and verb cross-linguistically. Languages with no dominant order for one or both of the phrase types in question, as well as languages with inpositions or no adpositions at all, have been excluded. Data from Dryer (2013c).

	Prepositions	Postpositions
Object - Verb	14	472
Verb - Object	456	42

2.2 Explanations for Word-Order Harmony

A number of different explanations for the cross-linguistic prevalence of word-order harmony have been proposed. Broadly, these can be classified into two main categories:

1. Cognitive explanations, which appeal to some aspect of human cognition contributing to the cross-linguistic tendency for harmony that we see;
2. Historical explanations, which attribute harmony to the nature of historical processes of language change.

2.2.1 Cognitive Explanations

Cognitive explanations for cross-category harmony include that of the Principles and Parameters framework, which attributes harmony to a high-level parameter for head-directionality, which is set in the speaker's internal grammar only once per language acquired and therefore applies across all phrase types in the language (e.g. Chomsky, 2014). Hawkins (1990), on the other hand, proposes that harmony is due to the fact that harmonic syntactic structures are easier to process, because they minimise the number of words that need to be heard in order for the hearer to identify the immediate constituents of the sentence (Moravcsik, 2010).

Alternatively, Culbertson and Kirby (2016) argue that there is a domain-general cognitive bias for harmonic structures: specifically, a bias for simplicity, which has been suggested to drive a wide range of cognitive processes (Chater & Vitányi, 2003). Harmonic languages can be argued to be simpler than non-harmonic ones: for example, in harmonic languages, one rule describing head-directionality is sufficient to explain the word order in both the verb phrase and the adpositional phrase, whereas non-harmonic languages require two rules, and are thus more complex. This of course depends on the theoretical framework being used, however (Culbertson & Kirby, 2016).

Cognitive biases are believed to reveal themselves in the typological patterns that we see through a process of iterated learning, as each generation learns their language from the previous one. A number of computational simulations have indicated this to be the case, including that of Griffiths and Kalish (2007), who demonstrate that over several generations, the languages of people modelled as Bayesian learners directly reflect the learners' prior biases, when their learning strategy is modelled as sampling from the posterior distribution. Kirby et al. (2007) go further to demonstrate that, when the learners' strategy is modelled as one that involves choosing the language with the maximum a-posteriori probability, learners' prior biases are amplified in the resulting distribution of languages over the process of iterated learning. Moreover, Smith and Kirby (2008) argue that a maximum a-posteriori learning strategy has evolutionary advantages, and so is likely to be the strategy used by humans. Therefore, evidence for even a weak cognitive bias for harmony, would be sufficient to argue that this contributes to the cross-linguistic prominence of cross-category harmony.

Much experimental work has been undertaken investigating harmony within the noun phrase, with results supporting the influence of a cognitive bias on word order, acting upon the positions of the numeral, demonstrative and adjective in relation to the noun (e.g. Culbertson et al., 2012; Culbertson and Newport, 2015; Culbertson et al., 2020). Relatively little experimental work has investigated cross-category harmony, on the other hand. The few studies that have been carried out will be discussed in detail below.

Artificial Language Learning as a Method for Investigating Cognitive Biases

Artificial Language Learning (ALL) is a method commonly used for investigating the impact of cognitive biases on language. It involves giving participants

a miniature constructed language to learn, and then testing them on it. This method has been used widely to evaluate regularisation and extrapolation tendencies, and the learnability of certain language features (eg. Culbertson et al., 2012; Yin and White, 2018; Culbertson and Adger, 2014). Three experimental studies have been carried out using the ALL method to investigate harmony between the verb phrase and adpositional phrase, specifically those of Cook (1988), Zhao and Fedzechkina (2020) and Wang et al. (2021).

Cook (1988) investigated harmony between the verb phrase, adpositional phrase and noun phrase, through a series of ALL experiments on school-aged children using an extrapolation paradigm. In one of these experiments, participants were taught the vocabulary of an artificial language (AL), and shown simple transitive sentences, either of the order Subject-Object-Verb (SOV) or Verb-Subject-Object (VSO). They were then asked to translate adpositional phrases from English into the AL, and thus were being tested on whether they extrapolated the head-directionality of the adpositional phrase from that of the verb phrase. Results showed that participants had an overall preference for postpositions, regardless of whether they were trained on SOV or VSO orders, but that this preference was significantly stronger for participants trained on SOV. This could indicate that there is a baseline preference for postpositions, which will be discussed further in section 2.3.1. The difference in the extent of the preference could be due to a harmony bias increasing the preference for postpositions in the SOV condition, and acting against the baseline preference in the VSO condition. However, because the lexical items used in the test phase were the same as those in the training phase, this pattern of results may instead be due to participants noticing surface-level patterns, such as *the word for tiger always comes first*, rather than syntactic ones. Moreover, Cook (1988) points out that the participants may have approached the experiment as a problem solving task, rather than a language learning one, and Zhao and Fedzechkina (2020) note that Cook's method of asking participants to translate sentences from English into the AL may have caused them to consciously consider English word order when approaching the task, and potentially to adopt a strategy of making the word order maximally different to English by choosing postpositions.

Zhao and Fedzechkina (2020) conducted an ALL regularisation experiment in which participants were first exposed to either prepositional phrases or postpositional phrases in the AL. They were then exposed to a set of verb phrases, 50% of which displayed VO order, and 50% OV order. At test, participants were shown images depicting events and asked to describe them in the AL. Participants in the easy lexical retrieval condition were given the vocabulary items to choose from when doing this, and those in the hard lexical retrieval condition were not. Results showed that participants exposed to postpositions in the hard lexical retrieval condition were more likely to produce OV structures at test, thus favouring harmonic patterns. Participants exposed to prepositions, however, did not have such a preference for harmonic structures, and participants in the easy lexical retrieval did not have a harmony preference when exposed to either adposition type. Looking only at the hard lexical retrieval condition, these results indicate that a cognitive bias cannot fully account for the typological patterns we see crosslinguistically, as there is only experimental evidence for it favouring

harmony for one head direction. Zhao and Fedzechkina suggest that this may be because of a baseline preference for SOV orders, as suggested by the fact that it is the most common basic word order in WALS (Dryer, 2013b), and that experimental studies have shown it to be the most learnable (Tily et al., 2011). It is therefore possible that Zhao and Fedzechkina did not find evidence for a harmony bias in participants exposed to prepositions because the bias was interacting with this baseline preference for SOV order.

Wang et al. (2021), like Cook (1988), used an extrapolation paradigm to target harmony between the verb phrase and the adpositional phrase. Native speakers of English and of Chinese were trained on verb phrases in an AL either displaying OV or VO order, and then on the adpositional lexicon. At test they were asked to describe images depicting spatial relations in the AL. Results showed that both sets of participants had a strong preference for harmony in both the OV and the VO condition. This appears indicative of a cognitive bias for harmony, but, like the Cook (1988) study, the same lexical items were used in both the training and testing phases, meaning participants may have been utilising surface-level patterns.

Silent Gesture as a Method for Investigating Cognitive Biases

Another method which is often employed in the investigation of language universals is the silent gesture paradigm. This paradigm was first used by Goldin-Meadow et al. (2008) and has since been employed in many studies investigating word-order universals (e.g. Schouwstra and de Swart, 2014; Meir et al., 2010; Langus and Nespor, 2010). Typically, the method involves showing visual stimuli to hearing participants with no knowledge of any signed languages, and asking them to convey what they see using only gesture. The order in which participants gesture the relevant elements of the picture (e.g. those representing the subject, object and verb of the corresponding descriptive sentence in English) is subsequently analysed. Silent gesture studies produce comparable results when employed on participants with a range of different native languages, including English, Spanish, Mandarin, Turkish (Goldin-Meadow et al., 2008), Irish, Russian, Tagalog (Futrell et al., 2015), and Italian (Langus & Nespor, 2010). This indicates that this method is relatively successful in inhibiting the effects of the word order of the speakers' native language, and therefore the similarities in the orders that they produce may be caused deeper shared preferences, such as those caused by cognitive biases.

Basic word order and, by extension, the verb phrase, have been the focus of recent silent gesture studies on language universals. Goldin-Meadow et al. (2008), conducted a series of experiments, one of which was a silent gesture production study, which aimed to investigate how people communicate when they are prevented from using language. They found that for basic transitive events, participants overwhelmingly communicated the elements of the stimuli in an order that corresponded to SOV. Schouwstra and de Swart (2014) noted that all stimuli in Goldin-Meadow et al. (2008) depicted extensional events, that is, those in which the direct object is manipulated, and exists independently of the event (e.g. verbs like *throw, poke, drop*). They then ran a silent gesture

experiment which also tested participants on intensional events: events for which the direct object does not exist independently of the action, as for verbs like *paint, think about, dream of*. They found that participants were indeed more likely to produce SOV gestured orders for extensional events, but for intensional events they were more likely to produce SVO. This demonstrates that word order preferences can be affected by semantics (Schouwstra & de Swart, 2014). As all of Zhao and Fedzechkina's (2020) verb phrase stimuli represented extensional events, this strengthens the argument that a baseline preference for SOV contributed to the lack of evidence for a harmony bias in the prepositional condition of their study.

2.2.2 Historical Explanations

For the type of harmony of interest in this paper, between the verb phrase and the adpositional phrase, a historical explanation has also been put forward. Specifically, there is a diachronic tendency for adpositions to derive from the verbs of verb phrases, or from possessum constituents of possessive phrases (which also participate in cross-category harmony), via a process which preserves linear order (Moravcsik, 2010). Since verbs and adpositions are the heads of verb phrases and adpositional phrases respectively, this would help to explain the prominence of syntactic harmony between these phrase-types cross-linguistically. This historical tendency has been observed in several unrelated languages, including Abkhaz, Basque, Bihari, Buriat, Kui (Bybee, 1988, p. 354), Mandarin (Li & Thompson, 1974), and English (Moravcsik, 2010), and thus certainly contributes to the cross-linguistic patterns of syntactic harmony we see. The question is whether this historical tendency is the sole reason for the harmony universal, or if cognitive biases also play a role.

2.3 The Present Experiments

This paper describes three experiments, which were designed to together establish whether evidence for a cognitive bias for harmony between the verb phrase and the adpositional phrase can be found. If such evidence is found, this would indicate that a cognitive bias for cross-category harmony contributes to its cross-linguistic prevalence, alongside the historical explanation presented above.

2.3.1 Experiment 1 Aims

Experiment 1 first aims to identify whether there is an overall baseline preference for either prepositions or postpositions, like the baseline preference for SOV basic word order for extensional events found by Goldin-Meadow et al. (2008), and to quantify the extent of such a preference. This is crucial for comparison with experiments 2 and 3 of this paper, which aim to manipulate this preference through priming participants with verb phrases of either OV or VO order. If an overall baseline preference for prepositions or postpositions exists, the preferences found in the results of experiments 2 and 3 should be interpreted in

relation to this baseline preference, rather than just in relation to chance. The silent gesture paradigm is utilised here, as is typical for studies investigating such a baseline preference (e.g. Goldin-Meadow et al., 2008), because it minimises native language effects. Though such studies are typically done using silent gesture improvisation paradigms (e.g. Schouwstra and de Swart, 2014; Goldin-Meadow et al., 2008), here I opt for a silent gesture perception paradigm, as these are easier to implement online than improvisation studies, allowing more participants to be tested over a short period of time. Although it is a relatively new method, silent gesture production results have been shown to replicate using the silent gesture perception paradigm. Specifically the perception results of Motamedi et al. (2021) replicate the production results of Schouwstra and de Swart (2014) for a difference in basic word order preferences based on semantics, and the perception results of Verhoef et al. (2016) replicate the production results of Padden et al. (2015) investigating the preferences for handling gestures in comparison with action gestures.

2.3.2 Experiment 2 Aims

Experiment 2 then aims to replicate the results of Wang et al. (2021), but using silent gesture perception methods combined with ALL in the extrapolation paradigm, to establish whether these methods are sufficient to demonstrate a harmony bias when the objects used in the training stimuli are the same as those in the testing stimuli. This is required for comparison with experiment 3, in which the objects in the test stimuli are different to those in the priming stimuli. This is because, if a significant effect of condition is found in experiment 2 but not experiment 3, this would indicate that there is no harmony bias, and instead participants were just noticing surface-level patterns in experiment 2, and in the previous studies of Cook (1988), Zhao and Fedzechkina (2020) and Wang et al. (2021).

Participants in experiment 2 are first primed on either VO or OV verb phrase order, where in the VO condition, the stimuli exclusively represent intensional events, and in the OV condition the stimuli exclusively represent extensional events. This is to avoid the possibility of the differing baseline verb phrase order preferences found by Schouwstra and de Swart (2014) interfering with the strength of the prime in the two different conditions. Participants are then asked exactly the same questions as in experiment 1 to allow for direct comparison when evaluating whether the priming had an impact on their preferences for adpositional orders.

2.3.3 Experiment 3 Aims

Experiment 3 investigates whether the effect demonstrated in experiment 2 is the result of participants noticing surface-level patterns due to the repetition of objects in the training and testing stimuli, or whether there is evidence of a syntactic harmony bias at work. This is achieved by repeating experiment 2 but using different objects in the training stimuli to those in the testing stimuli.

Chapter 3

Methodology

3.1 Experiment 1

3.1.1 Participants

140 monolingual adult native English speakers were tested for experiment 1. All were recruited and tested on Prolific, and were paid the equivalent of £14.60/hr for their time. The experiment lasted roughly 3 minutes.

3.1.2 Materials

Nine drawings were made depicting adpositional relations between a human character and an inanimate object. Standard practice in silent gesture improvisation studies is to use a semantically-rich human character, such as a pirate or a chef (e.g. Schouwstra & de Swart, 2014), in order to encourage participants to express the subject in their gestures. In the present experiments a more generic representation of a person was used, as the relative position of the subject is not relevant to the research questions, and so was not overtly gestured. The objects in the stimuli were chosen to work semantically with the adpositions *in, under* and *in front of*, and to be easily gestured using a one-handed gesture. This is because using some one-handed gestures and some two-handed gestures for different elements of the sentence could affect the saliency of the different elements and thus impact the results. The objects chosen were a *teacup*, a *kettle* and a *saucepan*. The full list of stimuli is the following:

- [A person] in a teacup
- [A person] in a kettle
- [A person] in a saucepan
- [A person] in front of a teacup
- [A person] in front of a kettle
- [A person] in front of a saucepan

- [A person] under a teacup
- [A person] under a kettle
- [A person] under a saucepan

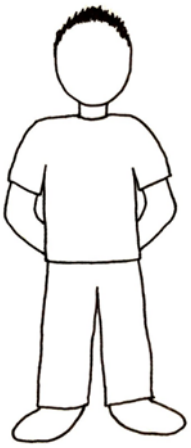

Figure 3.1: Drawing of generic human used in image stimuli.

Video stimuli were recorded of the experimenter gesturing interpretations of the image stimuli. Each video showed two distinct gestures, one for the object and one for the adposition, relying on the body being interpreted as the subject. For each stimulus, one video was made depicting the order adposition-object (i.e. a prepositional order), and one video depicting the order object-adposition (i.e. a postpositional order). All gestures were one-handed and involved movement. All gestures lasted 1 second each and all videos were exactly 4 seconds long in total.

3.1.3 Procedure

This and the following two experiments were coded in JsPsych, and ran on the participant's web browser. After consenting to participate in the study, participants were instructed that they would see an image, alongside two videos of gesture sequences representing the image, and that they should watch both videos, before clicking on the button corresponding to the video which best represents the image. They then proceeded to the task, where the image appeared on the screen, and both videos played beneath it simultaneously side-by-side, as shown in figure 3.2. The videos looped until the participant had selected one, and the order of the videos (left/right) was randomised for each participant.

Each participant received only one trial of this type, in order to capture their immediate reaction to the stimulus. The stimulus they received was assigned randomly from the nine available.

Participants were then given a slider-response question, for which they were shown the same image and videos as in the previous question, and asked to use the slider to indicate the strength of their preference for the video they chose,

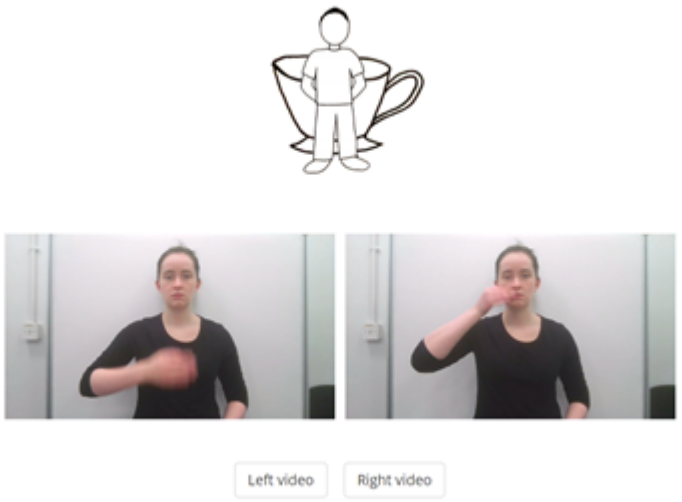

Figure 3.2: Screenshot of binary response question of experiment 1.

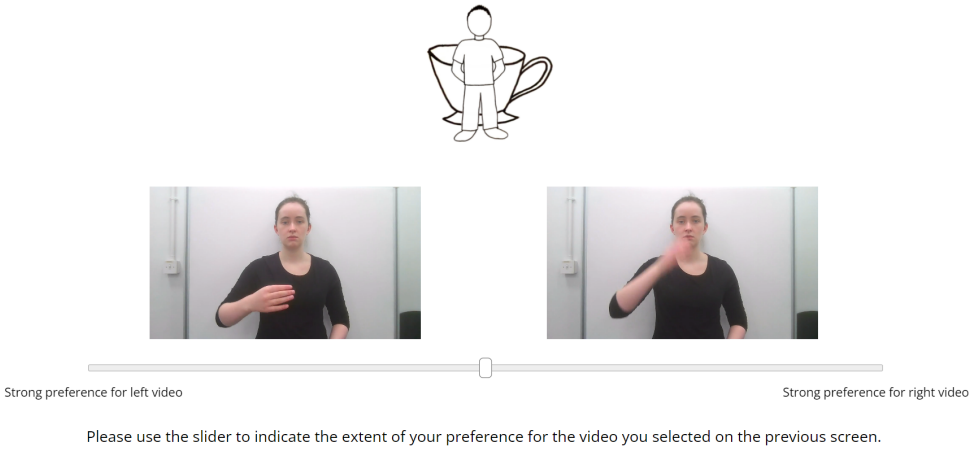

Figure 3.3: Screenshot of slider question of experiment 1.

with the leftmost limit of the slider representing a strong preference for the video on the left, and the rightmost limit a strong preference for the video on the right. Figure 3.3 shows a screenshot of this trial type.

Participants were then once again shown the image stimulus from the previous two questions, and asked to type a one-sentence description of the image in English.

3.2 Experiment 2

3.2.1 Participants

140 participants who were again monolingual adult native English speakers were recruited and tested on Prolific. Participants who took part in experiment 1 were prevented from participating in experiment 2. Experiment 2 lasted roughly 6

minutes and participants were paid the equivalent of £10.39/hr.

3.2.2 Materials

Nine drawings were made depicting intensional events, and nine depicting extensional events. The objects in the images were the same as in experiment 1, and again the subject was a generic human character. The intensional verbs used were *paints, thinks of,* and *dreams of*, and the extensional verbs were *throws, drops,* and *pokes*. The full list of events stimuli is given in table 3.1.

Table 3.1: Table showing events stimuli for experiment 2.

Extensional Events Stimuli	**Intensional Events Stimuli**
[A person] throws a teacup	[A person] paints a teacup
[A person] throws a saucepan	[A person] paints a saucepan
[A person] throws a kettle	[A person] paints a kettle
[A person] drops a teacup	[A person] thinks of a teacup
[A person] drops a saucepan	[A person] thinks of a saucepan
[A person] drops a kettle	[A person] thinks of a kettle
[A person] pokes a teacup	[A person] dreams of a teacup
[A person] pokes a saucepan	[A person] dreams of a saucepan
[A person] pokes a kettle	[A person] dreams of a kettle

Video stimuli were made in the same way as experiment 1. In this case, for each stimulus, one video depicted an OV gestured order, and the other a VO gestured order. The same adpositional stimuli from experiment 1 were used in experiment 2.

3.2.3 Procedure

Each participant was randomly assigned to either the VO condition or the OV condition. After consenting to participate in the study, they were instructed that they would first see some videos of gesture sequences representing images, and that they would be tested on these later so should pay attention. They proceeded to the training phase, where participants in the extensional condition saw all 9 extensional stimuli one by one, with the image presented at the top of the screen and the corresponding video for OV order underneath. The video played on a loop, but after it had played through once, a continue button was made clickable, allowing the participant to proceed to the next stimulus. Those in the intensional condition saw the 9 intensional stimuli with the corresponding videos for VO order, presented in the same manner as in the extensional condition.

In the reinforcement phase, participants were tested on how well they had learned the gesture orders in the training phase. They were shown each of the 9 events stimuli again one by one, in a random order, along with both the OV and VO gesture videos for that stimulus. Participants were asked to click the button corresponding to the video which best fit with the AL they had just learned, and were given feedback on their responses. This was to both reinforce their learning,

and to allow me to exclude participants from analyses if they did not learn the word order presented to them in the training phase.

Participants then proceeded to the critical test phase, where they were tested on the adpositional stimuli, with exactly the same questions and materials as in experiment 1. Like in experiment 1, each participant only saw one of the 9 adpositional stimuli.

3.3 Experiment 3

3.3.1 Participants

140 monolingual adult native English speakers were recruited and tested on Prolific for experiment 3. Participants who took part in experiments 1 or 2 were prevented from participating. Participants were paid the equivalent of £9.48/hr for their time. The experiment lasted around 6 minutes.

3.3.2 Materials

The same adpositional stimuli were used in experiment 3 as in experiments 1 and 2. Nine new extensional events stimuli and nine new intensional events stimuli were created for experiment 3. Here the objects were *hammer*, *spraycan* and *toothbrush*. The verbs were the same as in experiment 2. The full list of stimuli is in table 3.2:

Table 3.2: Table showing events stimuli for experiment 3.

Extensional Events Stimuli	**Intensional Events Stimuli**
[A person] throws a hammer	[A person] paints a hammer
[A person] throws a spraycan	[A person] paints a spraycan
[A person] throws a toothbrush	[A person] paints a toothbrush
[A person] drops a hammer	[A person] thinks of a hammer
[A person] drops a spraycan	[A person] thinks of a spraycan
[A person] drops a toothbrush	[A person] thinks of a toothbrush
[A person] pokes a hammer	[A person] dreams of a hammer
[A person] pokes a spraycan	[A person] dreams of a spraycan
[A person] pokes a toothbrush	[A person] dreams of a toothbrush

Image and video stimuli depicting these events were made in the same way as in experiment 1.

3.3.3 Procedure

The procedure for experiment 3 was identical to that of experiment 2, except that the stimuli for the training and reinforcement phases came from the set described in section 3.3.2.

Chapter 4

Results

4.1 Exclusion Criteria

In all experiments, participants were excluded from the analysis if they indicated in the slider response question that they preferred a different adpositional gesture video to the one they selected as their preferred video in the binary response question. For experiments 2 and 3, participants were also excluded if they gave more than two incorrect responses in the reinforcement phase. The number of participants excluded from the analysis for each reason in each experiment is listed in table 4.1.

Table 4.1: Table showing reasons for participant exclusions in each experiment.

Exclusion Criteria	**Number excluded from experiment 1**	**Number excluded from experiment 2**	**Number excluded from experiment 3**
Opposite preferences in binary response and slider questions	10	3	6
More than two incorrect answers in reinforcement phase	NA	5	7
Both of the above	NA	0	2
Total	10	8	15

4.2 Results of Experiment 1

Binary Response Data

I next analysed the responses to the critical questions in experiment 1, beginning with the binary response question, which asked participants to choose which gesture video they preferred to represent the image shown. Figure 4.1 shows the

proportions of participants who preferred the prepositional gesture video in each condition of each experiment.

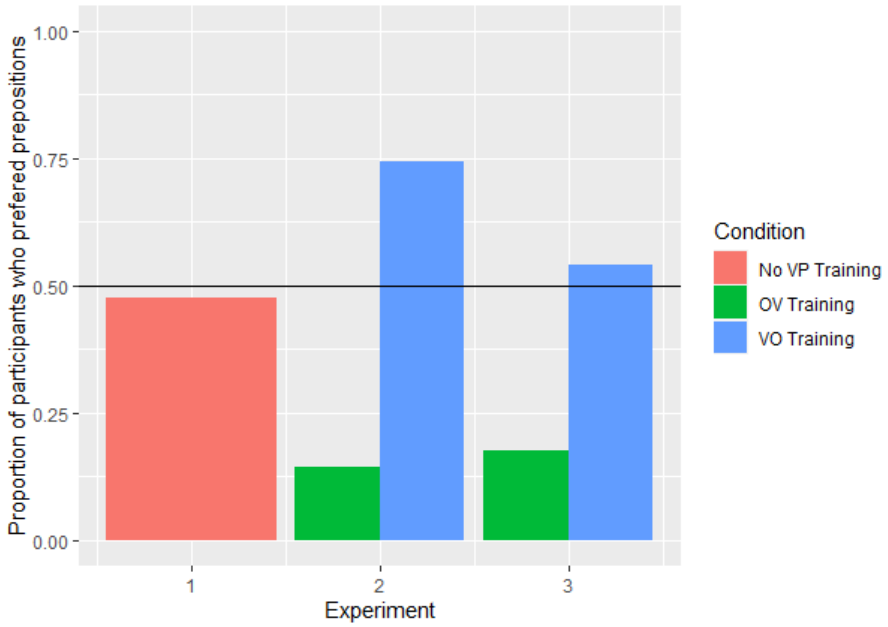

Figure 4.1: Bar chart showing the proportions of participants who preferred the prepositional gesture video in the binary response question in each condition of each experiment, with the horizontal line representing chance performance.

The proportion of participants who preferred the prepositional gesture video they were presented with in experiment 1 is 0.477, indicating a slight preference for postpositions. A logistic regression model was run on the experiment 1 data, using lme4 (Bates et al., 2015) in R (R Core Team, 2019) [1]. The outcome variable was a binary variable indicating whether the participant preferred the prepositional gesture video or not. For all statistical tests in this paper the standard alpha level of 0.05 is adopted. The model returned a non-significant intercept estimate (β = -0.09, se = 0.18, z = -0.53, p = 0.60). There is therefore insufficient evidence to reject the null hypothesis that participants have no preference for either prepositional or postpositional gesture videos.

Slider Data

I next analysed participants' responses to the slider question in experiment 1, which asked them to give a more graded indication of their preference for the gesture video they chose using a slider. Figure 4.2 is a violin plot showing the results of this question for all 3 experiments, where 100 represents a strong preference for the prepositional gesture video and 0 a strong preference for the postpositional one.

For experiment 1, figure 4.2 indicates a large number of participants had a very small preference for the video they chose, with many responses being

[1] As each participant only saw one stimulus and thus gave a single response for each question type, all statistical models were run without random effects.

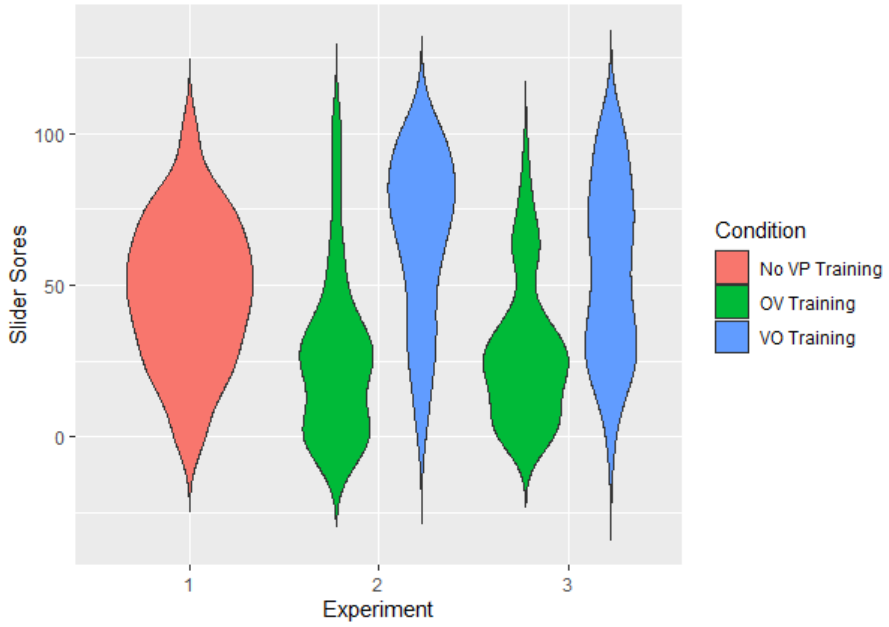

Figure 4.2: Violin plot showing participants' slider rating for the gesture video they chose, for each condition in each experiment. 100 represents a strong preference for the prepositional gesture video and 0 a strong preference for the postpositional gesture video.

centered around 50. After transforming the slider data to be in the range (0,1), a beta regression model was run on the experiment 1 data using mgcv (Wood, 2011) in R (R Core Team, 2019). The outcome variable was the participants' transformed slider scores for the adpositional videos. The model returned a non-significant intercept (β = -0.16, se = 0.12, z = -1.38, p = 0.17), indicating there is insufficient evidence to reject the null hypothesis that there is no preference for either prepositional or postpositional gesture videos, when preference is indicated using a slider.

Because for both the binary and slider data the prepositional preference was not significantly different from chance, there was no need for the following statistical tests to directly compare the results of experiment 1 with those of experiments 2 and 3.

4.3 Results of Experiments 2 and 3

4.3.1 Learning in Experiments 2 and 3

Before analysing the responses to the critical questions in experiments 2 and 3, I investigated how well participants learned the verb phrase order in these experiments (post exclusion), because if there is a difference between conditions this will affect how subsequent results are interpreted. Figure 4.3 shows the mean number of mistakes made by participants in each condition in the reinforcement phase of each of experiments 2 and 3, and indicates that more mistakes were

made by participants in the OV condition in both experiments.

A linear model was run on these data to investigate whether this difference was significant. The outcome variable was the number of mistakes made by participants in the reinforcement phase of the experiment, and the fixed effects entered were experiment, condition and their interaction. For this and all following models, condition was sum-coded and experiment treatment-coded, with the reference levels being experiment 2 and the VO condition. This means the intercept estimate represents the outcome variable averaged over both conditions in experiment 2. Table 4.2 shows the results of the model.

Table 4.2: Table showing the results of the linear model run on experiments 2 and 3 data with experiment, condition and their interaction as fixed effects, and the number of mistakes made by participants as the outcome variable. Condition was sum-coded, and experiment treatment-coded, and the reference levels were experiment 2 and the VO condition. * indicates significant results.

Effect	**Estimate**	**Standard Error**	**t-value**	**p-value**
Intercept	0.153968	0.038229	4.028	7.46e-05*
Condition	-0.068254	0.038229	-1.785	0.0754
Experiment	0.008248	0.055113	0.150	0.8812
Condition: Experiment	0.020792	0.055113	0.377	0.7063

As table 4.2 shows, none of the fixed effects were significant predictors of the number of mistakes made by participants. This indicates that the difference between conditions is not meaningful. Therefore subsequent analyses are carried out without taking this into account.

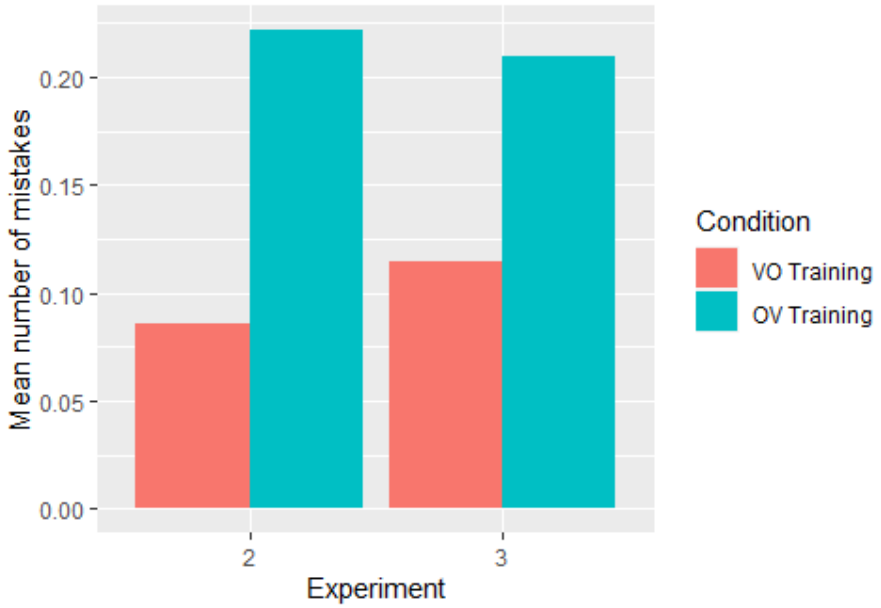

Figure 4.3: Bar chart showing the average number of mistakes made by participants in verb phrase testing in each condition in experiments 2 and 3.

4.3.2 Results of Experiment 2

Binary Response Data

For experiment 2, figure 4.1 clearly shows that harmonic patterns were preferred: only 14.3% of participants in the OV condition preferred the prepositional gesture video, compared with 74.3% in the VO condition. A logistic regression model was run on the experiment 2 data, with the outcome variable being whether the participant preferred the prepositional gesture video or not. Condition was entered as a sum-coded fixed effect. The model results are given in table 4.3.

Table 4.3: Table showing results of logistic regression model run on experiment 2 data, with the binary response indicating the preference for adpositional gesture video as the outcome variable, and condition as a sum-coded fixed effect. * indicates significant results.

Effect	**Estimate (log odds)**	**Standard Error**	**z-value**	**p-value**
Intercept	-0.3654	0.2261	-1.617	0.106
Condition	1.4263	0.2261	6.309	2.8e-10*

As table 4.3 shows, again the intercept was non-significant, indicating we do not have sufficient evidence to reject the null hypothesis that, averaged across the two conditions, participants' preference for prepositional gesture videos was not underlyingly different from chance. Table 4.3 also reveals condition to be a significant predictor of preference for prepositional gesture videos: VO training leads to more prepositional choices than OV training. In other words, the majority of participants chose the adpositional phrase order that was harmonic with the verb phrase order they were trained on.

Slider Data

From figure 4.2 it seems that in experiment 2 the preference for prepositions in the VO condition is weaker than the preference for postpositions in the OV condition, as the participants' slider responses were slightly more dispersed in the VO than in the OV condition. A beta regression model was run on the experiment 2 data, with condition entered as a sum-coded fixed effect, and the outcome variable being the participants' transformed slider scores for the adpositional videos. The results of this model are summarised in table 4.4.

Table 4.4: Table showing results of beta regression model run on experiment 2 data, with slider responses as the outcome variable and condition as a sum-coded fixed effect. * indicates significant results.

Effect	**Estimate (log odds)**	**Standard Error**	**z-value**	**p-value**
Intercept	-0.3465	0.1087	-3.189	0.00143*
Condition	1.0665	0.1087	9.815	<2e-16*

Here the intercept is significant so we have sufficient evidence to reject the null hypothesis that the average probability of a participant having a preference for prepositional gesture videos across both conditions in experiment 2 is not underlyingly different from chance, when preference is indicated by a slider response. The intercept estimate is negative, indicating that there is an overall postpositional preference in experiment 2. Condition is again a significant predictor of preference for prepositional gesture videos: participants in the VO condition preferred prepositions and those in the OV condition preferred postpositions. In other words participants showed a preference for harmonic patterns.

4.3.3 Results of Experiment 3

Binary Response Data

Figure 4.1 suggests that harmonic patterns were also preferred in experiment 3, though to a lesser extent than for experiment 2, with 17.7% of participants in the OV condition and 54.1% of participants in the VO condition preferring prepositions. A logistic regression model was run on the binary response data from experiment 3, with the same structure as the model for the binary response data from experiment 2. The model results are given in table 4.5.

Table 4.5: Table showing results of logistic regression model run on experiment 3 data, with binary response for adpositional preference as the outcome variable and condition as a sum-coded fixed effect. * indicates significant results.

Effect	**Estimate (log odds)**	**Standard Error**	**z-value**	**p-value**
Intercept	-0.6848	0.2101	-3.260	0.00111*
Condition	0.8491	0.2101	4.042	5.3e-05*

Unlike experiments 1 and 2, for experiment 3, the model intercept for the binary response data was significant. As the intercept term is negative, this indicates there is a reliable preference for postpositions averaged across the two conditions. In this model, condition is again revealed as a significant predictor of preference for prepositional gesture videos: participants in the VO condition were significantly more likely to choose the prepositional gesture video than those in the OV condition.

Slider Data

Figure 4.2 suggests that, when indicating their preference using the slider, participants in the OV condition of experiment 3 again had a preference for postpositions on average, but the overall preference of participants in the VO condition is less clear. A beta regression model was run on the experiment 3 data with the same structure as that for the model run on the experiment 2 slider data. The results of the model are summarised in table 4.6.

Here, the intercept estimate is significant and negative, providing sufficient evidence to reject the null hypothesis that the average probability of a partici-

Table 4.6: Table showing results of beta regression model run on experiment 3 data, with slider response as the outcome variable and condition as a sum-coded fixed effect. * indicates significant results.

Effect	Estimate (log odds)	Standard Error	z-value	p-value
Intercept	-0.3171	0.1163	-2.727	0.00639*
Condition	0.7443	0.1163	6.401	1.54e-10*

pant having a preference for prepositional gesture videos across both conditions in experiment 3 is not underlyingly different from chance, when prepositional preference is indicated using a slider. This indicates that there is an overall post-positional preference in experiment 3. Like in all previous models, condition was a significant predictor, so participants in the VO condition were more likely to prefer the prepositional gesture video than participants in the OV condition.

4.3.4 Comparison Between Experiments 2 and 3

Binary Response Data

To evaluate whether there was a difference in the effect of condition on prepositional preference between experiments 2 and 3, or an effect of experiment, a logistic regression model was run on the binary response data from both experiments, with experiment entered as a treatment-coded fixed effect, and condition as a sum-coded fixed effect. The interaction between experiment and condition was also entered into the model as a fixed effect. The outcome variable was again whether the participant preferred the prepositional gesture video or not. The results are reported in table 4.7.

Table 4.7: Table showing results of logistic regression model run on the data from experiments 2 and 3, with the binary response as the outcome variable and condition, experiment and their interaction as fixed effects. Condition was sum-coded and experiment treatment-coded. * indicates significant results.

Effect	Estimate (log odds)	Standard Error	z-value	p-value
Intercept	-0.3654	0.2261	-1.617	0.1060
Condition	1.4263	0.2261	6.309	2.8e-10*
Experiment	-0.3194	0.3086	-1.035	0.3007
Condition: Experiment	-0.5772	0.3086	-1.870	0.0614

Neither the main effect of experiment nor the interaction between experiment and condition were significant predictors of prepositional preference. There is thus insufficient evidence to reject the null hypothesis that there is a difference between experiments 2 and 3 in the preference for prepositions. There is also insufficient evidence to reject the null hypothesis that there is no difference in

the extent of the effect of condition on prepositional preference between the two experiments.

Slider Data

A final beta regression model was run to evaluate the main effect of experiment and the difference between experiments 2 and 3 in the effect of condition on prepositional preference when prepositional preference was indicated using a slider. This model was run on the data from experiments 2 and 3, and experiment, condition, and their interaction were entered into the model as fixed effects. Experiment was treatment-coded and condition was sum-coded. The outcome variable was the participants' transformed slider scores for the adpositional videos. The model results are summarised in table 4.8.

Table 4.8: Table showing results of beta regression model run on experiments 2 and 3 data, with slider response as the outcome variable and condition, experiment and their interaction as fixed effects. * indicates significant results.

Effect	**Estimate (log odds)**	**Standard Error**	**z-value**	**p-value**
Intercept	-0.36194	0.10727	-3.374	3.68e-07*
Condition	1.14792	0.10727	10.702	<2e-16*
Experiment	0.06511	0.15918	-0.409	0.68252
Condition: Experiment	-0.46162	0.15918	-2.900	0.00373*

Condition was a significant predictor of prepositional preference, with participants in both experiments being more likely to choose the prepositional gesture video in the VO condition than the in the OV condition. The interaction between condition and experiment was also significant, but not the main effect of experiment. This indicates the prepositional preference in the VO condition in experiment 2 did not differ significantly from that of experiment 3, and similarly for the OV condition, however the difference between the prepositional preference in the VO condition and OV condition differed significantly across experiments. Specifically, the effect of condition is stronger in experiment 2 than in experiment 3. This interaction effect is illustrated by figure 4.4, which shows the model's predictions for transformed slider scores.

4.4 Image Description

When asked to describe the adpositional image stimulus that they saw in English, the majority of participants in all three experiments correctly identified the object in the picture. For stimuli using the adpositions *in* and *in front of*, most participants also correctly identified the adpositional relation being targeted. This indicates that the images were clear enough for the purposes of the experiments. For the *under* stimuli, however, responses were more varied, with many participants giving descriptions like *saucepan above head*. This indicates that this set

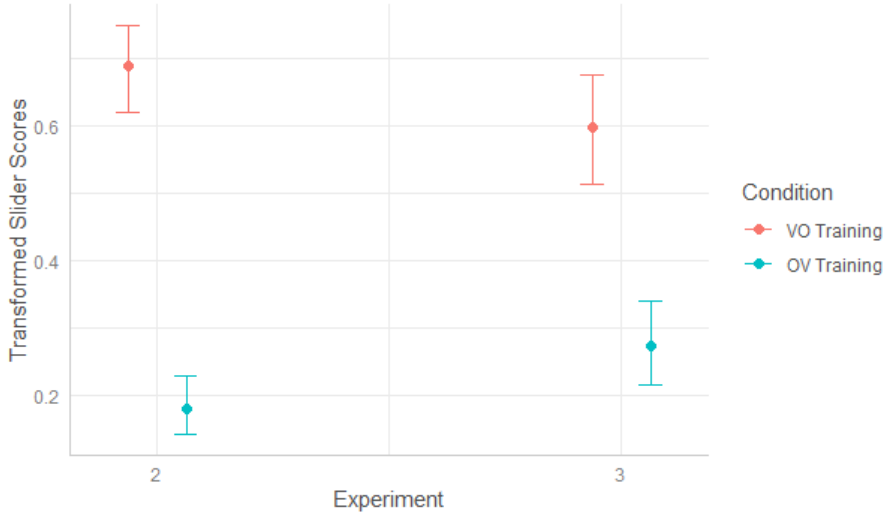

Figure 4.4: Graph plotting the model's predicted transformed slider scores for each experiment and condition, to illustrate the significant interaction effect between experiment and condition. Here 1 indicates a strong preference for prepositions and 0 a strong preference for postpositions.

of stimuli was not as well understood by the participants, and so maybe their responses were random. Just to be conservative, the above models were run again with all participants who received *under* excluded and the results followed the same key patterns, indicating that this potential misunderstanding of stimuli did not have an effect on the results. It is worth noting that although the participants did not give the intended response to the image description question, this does not necessarily mean that they did not correctly interpret the gestures in the videos as representing the intended adpositions and nouns. Participants may have correctly interpreted the gestures as *under* and *saucepan*, and responded to the critical questions accordingly, but when asked to describe the image in English, they found that describing the relation as *saucepan above head* felt more natural.

Chapter 5

Discussion

The present experiments aimed to evaluate whether there is evidence for a cognitive bias for syntactic harmony between the verb phrase and the adpositional phrase, which may contribute to the patterns we see in language typology. I carried out three experiments for this purpose, combining ALL with silent gesture perception methods. In experiment 1, participants were asked whether they preferred prepositional or postpositional gesture videos, in order to gauge what the baseline preference for the order of elements in an adpositional phrase is, without any priming. In experiment 2, participants were primed with gesture videos depicting verb phrases, either of the order VO or OV, before being asked the same questions as in experiment 1. In experiment 2 the objects used in the priming stimuli were the same as those used in the testing stimuli, and so experiment 3 was devised to rule out the possibility that participants were using surface-level rules (such as *the saucepan is always gestured first*) to answer the critical questions. This was achieved by repeating experiment 2, but using different objects in the verb phrase stimuli than in the adpositional stimuli, to test whether participants generalise beyond the individual items to the entire class of nouns.

Taken together, the results indicate that there is no baseline preference for either prepositions or postpositions, and that there is evidence for a harmony bias, both when the training and testing stimuli contain the same elements (experiment 2), and when they do not (experiment 3). The evidence for harmony in the VO condition of experiment 3 did however seem less strong than in the other condition and experiment 2, as shown in figure 4.1, and for the slider data, the effect of condition was significantly stronger in experiment 2 than in experiment 3. This suggests that the surface-level patterns visible in experiment 2 partially contributed to the harmonic preferences observed there.

5.1 Baseline Adpositional Preference

Experiment 1 served its purpose in providing a reference to compare the results of experiments 2 and 3 to, even if the baseline preference it revealed was not significantly different from chance and was thus not needed for direct statistical comparison. However, the non-significant result in experiment 1 is surprising, since Cook's (1988) experiments point to an underlying preference for postpo-

sitions. Moreover, Gentner and Boroditsky (2009, p. 5) argue that relational terms require the entities that they link to be introduced first, thus explaining why SOV is preferred for extensional verbs. All adpositional stimuli in the present experiments involve adpositions used to describe spatial relations between nouns. Following Gentner and Boroditsky's (2009) logic, there should therefore be a baseline preference for postpositions, since they allow both nouns in the spatial relation to be introduced, before the relation between them is expressed. One possible reason such a preference was not observed here is that some participants may have been experiencing native language influence, as perhaps the silent gesture perception method is not as successful as the production method at inhibiting this. If this were the case, then participants' native language influence which causes them to prefer prepositions, would have been interacting with a baseline preference for postpositions, resulting in the overall proportion of participants choosing prepositions not differing significantly from 0.5. However, the results of Motamedi et al. (2021) do not support this hypothesis, as in their silent gesture perception study which tested native English-speaking participants' preferences for basic word order, no overall preference for SVO was found. It seems likely, therefore, that no baseline preference for one adpositional order or another exists. The typological evidence supports this: the number of languages with postpositions (577) in WALS is roughly the same as the number of languages with prepositions (511) (Dryer, 2013a). Additionally, the results of Cook (1988), which suggest an overall postpositional preference, could be due to methodological issues with their experiment as discussed in section 2.2.1. More specifically, translating from English to the AL may have caused the participants in Cook's (1988) study to adopt a strategy of making the word order maximally different from English and therefore choosing postpositions.

5.2 Harmony

The results of experiment 2 provide evidence that the methods utilised in the present experiments, namely ALL combined with silent gesture perception, were sufficient to demonstrate a harmony bias when the objects in the priming and testing stimuli are the same, thus replicating the main results of Wang et al. (2021).

The results of experiment 3 are clear in demonstrating a harmonic preference in participants in the OV condition, who preferred postpositions. However, the preference for prepositions of those in the VO condition was only slightly higher than chance. One possible reason for this is that the priming stimuli followed the pattern of English, their native language, and also did not contradict their natural preferences (i.e. VO order was used to represent intensional events). This means that there is a chance many participants did not register it as a prime and thus responded to the critical question as they would if they had not been primed, like in experiment 1. This leaves a question of why this would happen in experiment 3 but not experiment 2. One possibility is that in experiment 2, the surface-level patterns were so obvious that they were easily recognised and used by the participants in choosing their responses to the critical questions. In experiment 3,

on the other hand, the analogies between the verb phrase stimuli and adpositional stimuli were less clear, so the participants could not rely upon such surface-level patterns and therefore responded as though they had not been primed at all, like in experiment 1.

Zhao and Fedzechkina (2020) similarly found evidence for harmony in the postpositional but not the prepositional condition, which may on the surface appear to be in conflict with the above hypothesis, since their test stimuli contained the same objects as their priming stimuli, like in experiment 2 of this paper. However, as discussed in section 2.2.1, their results were likely instead caused by the fact that their priming stimuli were adpositional phrases and their test stimuli were extensional verb phrases, which have a baseline preference for SOV order (Schouwstra and de Swart, 2014; Goldin-Meadow et al., 2008; Tily et al., 2011).

It can still be argued therefore that experiment 3 provides evidence for a cognitive bias for syntactic harmony between the verb phrase and the adpositional phrase. Although the preference for prepositions in the VO condition is only slightly higher than chance, it still differs significantly from the preference for prepositions in the OV condition. This difference in preference according to condition is sufficient evidence for a bias which at least partially contributes to the typological patterns we see, in combination with the historical explanation described in section 2.2.2. Moreover, as explained in section 2.2.1, it is possible that even a weak bias for harmonic patterns is sufficient to explain the prevalence of cross-category harmony between the adpositional phrase and the verb phrase in the world's languages. This is because the process of iterated learning over generations amplifies weak biases in the population over generations, as indicated by the results of the computational simulations of Kirby et al. (2007) and Smith and Kirby (2008).

Finally, the fact that the effect of condition was significantly stronger in experiment 2 than experiment 3 for the slider data indicates that that some participants in experiment 2 were relying on surface-level similarities between the training and test stimuli when choosing their preferred adpositional gesture video. This highlights the importance of ensuring that stimuli for experiments investigating harmony are carefully designed, to avoid the possibility of surface-level patterns being utilised by participants.

5.3 Directions for Further Research

The present study provides the foundation for a number of possible future studies. For example, all adpositional stimuli used in these experiments denote locations. However, different adpositional types may have different baseline preferences. Some evidence for this possibility comes from Mandarin, where adpositions which are path designators (e.g. *cóng*, 'from'; *dào*, 'to') are prepositions, and those which denote locations (e.g. *shàng*, 'on'; *xià*, 'under') are postpositions (Whitman et al., 2013). This suggests that different adpositional phrase order preferences are possible for different semantic categories of adpositions, similar to what we see for extensional and intensional verb phrases (Schouwstra & de Swart, 2014). The results of experiment 1 indicating no baseline preference for adpositional

phrase order are therefore not generalisable to all adposition types, and so future work could run similar experiments investigating baseline preferences for other semantic categories of adpositions.

English also displays cross-category harmony of the type investigated in this paper, and although it seems unlikely that there was native language influence at play in experiments 2 and 3, running the experiments on a population of speakers of a non-harmonic language would establish whether the key findings of this paper are robust.

I have proposed that the reduced prepositional preference in the VO condition of experiment 3 may be due to participants not being influenced by the prime in the VO condition. This is because the basic word order of their native language is VO, and there were no immediately obvious surface-level patterns between the priming stimuli and the test stimuli, like in experiment 2, so some participants responded as though they had not received any priming. This hypothesis could be directly tested by repeating experiment 3 on a population of speakers of an OV language. If the hypothesis is correct, then the participants should display a clear preference for harmony in the VO condition but not as strong a preference in the OV condition.

Finally, future research should also utilise the methods used here to investigate other types of cross-category harmony. Cross-category harmony describes a relationship between heads and dependents, but in harmony between the verb phrase and the adpositional phrase, the dependents of both phrase types are noun phrases. Thus even if further experimentation proved that participants were using a harmony bias in experiment 3, and that the less obvious preference for prepositions in the VO condition is simply due to the priming stimuli following the pattern of their native language and therefore having no effect, this would not provide conclusive evidence for a harmony bias targeting the abstract categories of head and dependent. This is because the harmony preference could still be due to them noticing that the syntactic category of noun phrase comes in the same relative position in both the verb phrase and the adpositional phrase. Therefore, future research should work to evaluate whether the results described in this paper are generalisable to other types of cross-category harmony.

Chapter 6

Conclusion

I have conducted three experiments utilising ALL and silent gesture perception methods which together aimed to establish whether there is evidence for a cognitive bias which contributes to the prevalence of harmony between the verb phrase and the adpositional phrase cross-linguistically. Experiment 1 demonstrated that, for native English speakers tested using silent gesture perception methods, there is no baseline preference for either prepositions or postpositions. Experiment 2 replicated results in Wang et al. (2021), thus demonstrating that ALL can be used in combination with silent gesture perception methods to show that when the priming and testing stimuli share elements, participants show a preference for harmonic patterns. In experiment 3, results showed that when there were no shared elements between the priming and testing stimuli, participants in the OV condition still showed a preference for harmony, but the preference for prepositions in the VO condition was only slightly above chance. I have argued that the reduced harmonic preference in the VO condition of experiment 3 may be due to the absence of immediately obvious surface-level patterns in combination with the fact that VO is the order in the participants' native language and so did not have the intended effects as a prime. Further experiments should be carried out to test this hypothesis, as well as to evaluate whether these findings are generalisable to other types of cross-category harmony. The fact that the effect of condition was stronger in experiment 2 than experiment 3 for the slider data indicated that the surface-level patterns visible in experiment 2 partially contributed to the significant preference for harmony observed. It is therefore important that stimuli of future experiments investigating biases for harmony using ALL methods are carefully designed to avoid this effect.

Bibliography

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bybee, J. (1988). The diachronic dimension in explanation. In J. Hawkins (Ed.), *Explaining language universals* (pp. 350–379).

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*, 19–22. https://doi.org/10.1016/S1364-6613(02)00005-0

Chomsky, N. (2014). *The minimalist program* (20th ed.). The MIT Press. https://doi.org/10.7551/mitpress/9780262527347.001.0001

Cook, V. J. (1988). Language learners' extrapolation of word order in micro-artificial languages. *Language learning*, *38*, 497–529. https://doi.org/10.1111/j.1467-1770.1988.tb00165.x

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

Culbertson, J., Franck, J., Braquet, G., Navarro, M. B., & Arnon, I. (2020). A learning bias for word order harmony: Evidence from speakers of nonharmonic languages. *Cognition*, *204*. https://doi.org/10.1016/j.cognition.2020.104392

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in psychology*, *6*, 1964–1964. https://doi.org/10.3389/fpsyg.2015.01964

Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, *139*, 71–82. https://doi.org/10.1016/j.cognition.2015.02.007

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

Dryer, M. S. (2013a). Order of adposition and noun phrase. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/85

Dryer, M. S. (2013b). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/81

Dryer, M. S. (2013c). Relationship between the order of object and verb and the order of adposition and noun phrase. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/95

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, *136*, 215–221. https://doi.org/10.1016/j.cognition.2014.11.022

Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: Evidence from navajo. *Routes to language: Studies in honor of Melissa Bowerman*, 5–32.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences - PNAS*, *105*, 9163–9168. https://doi.org/10.1073/pnas.0710060105

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, *2*, 73–113.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive science*, *31*, 441–480. https://doi.org/10.1080/15326900701326576

Hawkins, J. (1990). A parsing theory of word order universals. *Linguistic inquiry*, *21*, 223–261.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences - PNAS*, *104*, 5241–5245. https://doi.org/10.1073/pnas.0608222104

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive psychology*, *60*, 291–318. https://doi.org/10.1016/j.cogpsych.2010.01.004

Li, C. N., & Thompson, S. A. (1974). Coverbs in mandarin chinese: Verbs or prepositions? *Journal of Chinese linguistics*, *2*, 257.

Meir, I., Lifshitz, A., Ilkbasaran, D., & Padden, C. (2010). The interaction of animacy and word order in human languages: A study of strategies in a novel communication task. In A. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The evolotion of language* (pp. 455–456). World Scientific.

Moravcsik, E. A. (2010). Explaining language universals. In J. J. Song (Ed.), *The oxford handbook of linguistic typology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199281251.013.0005

Motamedi, Y., Wolters, L., Naegeli, D., Kirby, S., & Schouwstra, M. (2021). From improvisation to learning: How naturalness and systematicity shape language evolution. https://doi.org/10.31234/osf.io/7ed89

Padden, C., Hwang, S.-O., Lepic, R., & Seegers, S. (2015). Tools for language: Patterned iconicity in sign language nouns and verbs. *Topics in cognitive science*, *7*(1), 81–94.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*, 431–436. https://doi.org/10.1016/j.cognition.2014.03.004

Smith, K., & Kirby, S. (2008). Cultural evolution: Implications for understanding the human language faculty and its evolution. *Philosophical transactions. Biological sciences*, *363*(1509), 3591–3603.

Tily, H., Frank, M., & Jaeger, F. (2011). The learnability of constructed languages reflects typological patterns. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*.

Verhoef, T., Padden, C., & Kirby, S. (2016). Iconicity, naturalness and systematicity in the emergence of sign language structure. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference (evolangx11)*. Online at http://evolang.org/neworleans/papers/47.html.

Wang, F., Kirby, S., & Culbertson, J. (2021). A cognitive bias for cross-category word order harmony. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*.

Whitman, J., Paul, W., & Djamouri, R. (2013). Postpositions vs prepositions in mandarin chinese: The articulation of disharmony. *Theoretical approaches to disharmonic word order*. Oxford University Press.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36.

Yin, S. H., & White, J. (2018). Neutralization and homophony avoidance in phonological learning. *Cognition*, *179*, 89–101. https://doi.org/10.1016/j.cognition.2018.05.023

Zhao, Y., & Fedzechkina, M. (2020). Learners' harmonic preferences are modulated by lexical retrieval difficulty. *Proceedings of the Linguistic Society of America*, *5*, 708–721.