



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# Referent properties and word order in emerging communication systems

*Fiona Kirton*



Doctor of Philosophy  
School of Philosophy, Psychology and Language Sciences  
The University of Edinburgh  
2020



# Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. The work presented is entirely my own, except where stated otherwise in the text.

Fiona Kirton



# Abstract

Why do languages look the way they do? This question lies at the core of much of linguistics research, and answering it can shine a light on the relationship between individual cognitive preferences and linguistic structure. One area that has attracted particular attention is basic word order. Many mature languages exhibit a fixed or dominant ordering of subject (S), object (O), and the verb (V). However, evidence from restricted communication systems, such as emerging sign languages, shows that even before such conventions have been established, language producers show strong ordering preferences. It has been suggested that SOV is the natural ordering of entities in an event, and the default order used by all newly emerging languages. Over the past decade or so, a growing body of research has endeavoured to investigate this question using the silent gesture paradigm in which participants describe events using only their hands. This work has been instrumental in uncovering a range of factors that influence the way people convey information about events in the absence of linguistic conventions, challenging the view that there is a single natural order.

In this thesis, I present a series of experimental studies, implementing new techniques for data collection and analysis, showing how properties of individual referents influence the word orders people use to convey information about simple transitive events. I start, in Chapter 2, with a detailed review of the silent gesture literature, highlighting the numerous accounts that have been offered to explain word order preferences. One theme common to many of these accounts is that there is a direct relationship between word order and the structural and semantic properties of events. In Chapter 3, I report a silent gesture experiment in which I investigate an additional, complementary factor, namely, the salience of entities in an event. In the data analysis, I develop a novel computational method for inferring word order preferences based on incomplete gesture strings. The results of this study suggest that the relationship between salience and word order is not necessarily linear. Rather, manipulating the salience of referents influences the perspective from which a producer frames an event, which in turn influences structural choices. The results, however, are inconclusive about whether these structural choices reflect a direct mapping from conceptual structure to word order. In Chapter 4, I investigate the role of another operational factor: biases specific to the gestural modality. I report three experiments in which participants conveyed information about events by selecting pictorial representations of event components. Although the findings from this study are inconclusive, they nevertheless highlight important questions about the effects of other task-specific factors such as the way elicitation stimuli are presented.

In Chapter 5, I focus on the relationship between word order and one of the most fundamental determiners and drivers of linguistic structure - animacy. Using a series of artificial language learning experiments, I test two existing accounts that have been proposed to explain

animacy-based word order variation. One emphasizes communicative pressures arising from the potential ambiguity of events involving two human referents; the other focuses on the salience of humans relative to inanimate objects. The results of this study offer tentative support for the salience-based hypothesis. Nevertheless, I suggest that further work is required to understand how language producers negotiate the communicative challenge of accurately conveying information about events where the role played by each of the noun referents is ambiguous. Echoing the conclusions from the previous two studies, I also highlight the need for more research to better understand the role played by other factors, such as modality and native language.

Overall, the studies reported in this thesis demonstrate that word order in newly developing languages not only reflects structural and semantic properties of events, but is also influenced by properties of referents interacting in an event, for example, salience and humanness. While the overall findings are inconclusive about the precise nature of this relationship, they nevertheless add to a growing body of literature showing that structural choices in the absence of linguistic conventions do not conform to a single natural order, but are subject to the effects of a range of potentially interacting factors.

# Lay Summary

This thesis is about how people convey information about events when their communication system has no rules or conventions. In a fully developed language such as English, when someone says *The girl pushed the boy*, we know that the girl did the pushing and the boy got pushed. The sentence *The boy pushed the girl*, on the other hand, refers to a different event: the boy did the pushing and the girl got pushed. We know this, in part, because of word order. The one doing the action – the ‘agent’ – comes before the verb (*pushed*), and the one affected by the action – the ‘patient’ – comes after. Take another example: *The girl was pushed by the boy*. The order of the agent and patient has been swapped, but we still know who is doing what to whom because the verb is now sandwiched between the words *was* and *by*. Now imagine a newly emerging communication system that has words for referring to things and actions (e.g., *girl*, *boy*, and *push*), but no rules for how to put them together, and no function words like *was* and *by* for expressing relations between the things referred to, e.g., *boy* or *girl* – the ‘referents’. In what order would people express the words, and what factors would influence their choices? These questions form the basis of this thesis. Answering them can shine a light on the preferences and pressures that shape linguistic structure.

Evidence for how people describe events in the absence of linguistic rules comes from studies of emerging sign languages that have developed naturally within communities of deaf individuals. What these studies demonstrate is that word order regularities appear early in the development of a new language. In the last decade or so, this work has been augmented by laboratory-based studies that have investigated word order when people describe events using only gesture and no speech. This ‘silent gesture’ methodology, together with work on emerging sign languages, has uncovered a range of factors that influence word order in emerging communication systems. Of particular relevance to this thesis is a body of work that has found a relationship between word order and the animacy of referents, that is, whether a referent is animate or inanimate. A number of hypotheses have been proposed to explain this relationship, each of which focuses on a different mechanism: a desire to clearly indicate who is doing what to whom; a tendency to ‘embody’ the role of human referents when describing events using gestures; and the tendency to express more salient human referents before less salient inanimate objects. In this thesis, I investigate these hypotheses in more detail. I also explore the concept of salience, or ‘interestingness’, in a broader sense and how it relates to word order.

Overall, the experiments reported in this thesis support the claim that word order in emerging communication systems reflects the relative salience of referents. This work also highlights the need for further research to understand how this effect interacts with other factors – the pressure to communicate clearly and the mode of communication (manual or spoken) – in shaping linguistic structure.



# Acknowledgements

Gushing emotion is not my style, but here goes...

First, I want to thank the Carnegie Trust whose generosity made this thesis possible. A special thank you to Professor Dame Anne Glover for your warm and inspiring words.

A huge thank you to my supervisors, Simon Kirby, Kenny Smith, Jenny Culbertson, and Marieke Schouwstra (there's so many of you – I hope I haven't forgotten anyone!) for your support and advice throughout my PhD. To Simon and Kenny, a special thank you for getting me through the gruelling PhD funding application process all those years ago, and for encouraging me to try again when I got nowhere first time round. I dedicate my Carnegie Robertson Medal to you! To Marieke, yours was the first CLE talk I ever attended as a Master's student – it was fascinating and ultimately was the inspiration for this PhD thesis. Jenny, I don't know when you became my supervisor – the first I knew of it was when your name appeared on my first year review form. Irrespective, from the very beginning your astute and pragmatic advice has been invaluable. To all of you, thank you for your positivity, encouragement, frightening levels of expertise, and tireless support.

The CLE has been a great environment to do a PhD. It would have been a far lonelier experience without my fellow CLE PhD students, past and present. Thank you to Andres Karjus, Annie Holtz, Asha Sato, Carmen Saldana, Cathleen O'Grady, Fang Wang, Fausto Carcassi, Henry Conklin, Jasmeen Kanwal, Jon Carr, José Segovia Martín, Jonas Nölle, Marc Meisezahl, “wee” Marieke Woensdregt, Svenja Wagner, Tamar Johnson, and Yasamin Motamedi. I already miss the Friday evening jaunts to the Dagda and the PhD retreats (fingers crossed for 2021 – I'm looking at you, SARS-CoV-2). I also want to say a big thank you my non-CLE office mates and Dagda buddies: Daniel Lawrence, E Jamieson, George Starling, Michela Bonfieni, and so many others who contributed to making the last four years such a positive experience. A massive thank you also to all the current and former CLE folks whom I've had the great pleasure of knowing over the years: Alex Martin, Chris Cummings, Chrissy Cuskley, Danielle Naegeli, Hannah Järvinen, Isabelle Dautriche, James Winters, Jia Loy, Kevin Stadler, Mark Atkinson, Matt Spike, Molly Flaherty, Mónica Tamariz, Mora Maldonado, Rachel Kindellen, Richard Blythe, Rob Truswell, Olga Fehér, Stella Frank, and too many others to mention. The friendliest, most interesting bunch of people anyone could ever hope to meet!

Last but by no means least, a very special thank you to my Mum, my partner Peter, and my brother-in-law Michael for your support. You have always encouraged me to follow my own path and have supported me in whatever I have done. Thank you for never telling me to “get a proper job”. Your faith in my abilities is unwavering, if a bit over the top at times. For that I am immensely grateful. Finally-finally, a very special dedication: to my sister Nicola, who I know would have been hugely proud of me for completing this thesis.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Word order: a fundamental structuring device in emerging languages . . . . .	2
1.2.1	Ordering principles in restricted communication systems . . . . .	3
1.2.2	Animacy and language evolution . . . . .	4
1.3	Word order in fully developed languages: the influence of referent properties . . . . .	6
1.3.1	Animacy and word order in fully developed languages . . . . .	7
1.4	Word order in homesign and emerging sign languages . . . . .	8
1.4.1	Homesign: structural regularities and individual cognition . . . . .	9
1.4.2	Emerging sign languages . . . . .	10
1.4.3	Silent gesture: moving from the field to the laboratory . . . . .	12
1.5	Thesis outline . . . . .	12
<b>2</b>	<b>Word order and silent gesture</b>	<b>15</b>
2.1	Introduction . . . . .	15
2.2	Early investigations into word order in silent gesture . . . . .	15
2.3	SOV: The natural order of events? . . . . .	16
2.4	Beyond SOV . . . . .	19
2.4.1	Event type and word order . . . . .	19
2.4.2	Animacy, semantic reversibility and word order . . . . .	21
2.4.3	SVO and the availability of a gestural lexicon . . . . .	24
2.5	Conclusion . . . . .	26
<b>3</b>	<b>Constituent Order in Silent Gesture Reflects the Perspective of the Producer</b>	<b>29</b>
3.1	Introduction . . . . .	29
3.2	Experiment 1: Silent gesture task . . . . .	32
3.2.1	Participants . . . . .	32
3.2.2	Materials . . . . .	32
3.2.3	Procedure . . . . .	33
3.2.4	Coding . . . . .	34
3.2.5	Results . . . . .	34
3.2.6	APV responses: a comparison with previous studies . . . . .	42
3.2.7	Discussion . . . . .	44
3.3	Modelling the underlying word order distribution . . . . .	44
3.3.1	Generating the surface distribution . . . . .	45

3.3.2	Inferring the underlying distribution . . . . .	46
3.3.3	Results . . . . .	46
3.3.4	Discussion . . . . .	49
3.4	General Discussion . . . . .	49
3.4.1	Saliency and word order . . . . .	50
3.4.2	Silent gesture, native language, and word order . . . . .	52
3.5	Conclusion . . . . .	53
<b>4</b>	<b>Ordering Preferences in a Picture Selection Task: Investigating the Role of Modality</b>	<b>55</b>
4.1	Introduction . . . . .	55
4.1.1	Picture reconstruction tasks . . . . .	56
4.2	Experiment 2: Picture selection . . . . .	59
4.2.1	Participants . . . . .	59
4.2.2	Materials . . . . .	60
4.2.3	Procedure . . . . .	61
4.2.4	Results . . . . .	61
4.2.5	Discussion . . . . .	65
4.3	Experiment 3: Scene reconstruction . . . . .	66
4.3.1	Participants . . . . .	66
4.3.2	Materials . . . . .	67
4.3.3	Procedure . . . . .	67
4.3.4	Results . . . . .	69
4.3.5	Discussion . . . . .	71
4.4	Experiment 4: Reconstructing scenes from memory . . . . .	73
4.4.1	Participants . . . . .	73
4.4.2	Materials . . . . .	73
4.4.3	Procedure . . . . .	73
4.4.4	Results . . . . .	73
4.4.5	Discussion . . . . .	76
4.5	General discussion . . . . .	78
4.6	Conclusion . . . . .	80
<b>5</b>	<b>Cognitive Saliency vs Communicative Pressure: Investigating the Effects of Animacy on Word Order using Artificial Language Learning</b>	<b>83</b>
5.1	Introduction . . . . .	84
5.1.1	Animacy and linguistic structure in emerging sign languages . . . . .	84
5.1.2	Animacy and word order in silent gesture . . . . .	85
5.2	Testing the noisy channel and cognitive saliency hypotheses . . . . .	88
5.2.1	Predictions . . . . .	90
5.3	Experiment 5: Free production task . . . . .	91
5.3.1	Method . . . . .	92
5.3.2	Results . . . . .	94
5.3.3	Discussion . . . . .	96
5.4	Experiment 6: Restricted production task . . . . .	97

5.4.1	Method . . . . .	97
5.4.2	Results . . . . .	99
5.4.3	Discussion . . . . .	99
5.5	Image orientation and word order in Experiments 5-6 . . . . .	101
5.5.1	Discussion . . . . .	103
5.6	Experiment 7: Eliminating the influence of image orientation . . . . .	103
5.6.1	Method . . . . .	104
5.6.2	Results . . . . .	104
5.6.3	Discussion . . . . .	105
5.7	Combined analysis . . . . .	105
5.7.1	V-final . . . . .	106
5.7.2	P-first . . . . .	107
5.7.3	Discussion . . . . .	107
5.8	General Discussion . . . . .	108
5.8.1	Cognitive salience hypothesis . . . . .	108
5.8.2	Noisy channel hypothesis . . . . .	108
5.8.3	Deviations from predicted results: native language and other influences . . . . .	109
5.8.4	Outstanding questions . . . . .	110
5.9	Conclusion . . . . .	111
<b>6</b>	<b>Summary and Conclusion</b>	<b>113</b>
6.1	Aims and contributions . . . . .	113
6.2	Summary of experiments . . . . .	115
6.3	Outstanding questions and areas for further research . . . . .	118
6.4	General conclusions . . . . .	121
<b>A</b>		<b>133</b>
A.1	Chapter 3: List of items . . . . .	133
A.2	Chapter 3: Patient-before-action in block 1 . . . . .	134
A.3	Chapter 3: Model supplementary data . . . . .	135
A.3.1	Generic-agent events . . . . .	135
A.3.2	Character-agent events . . . . .	136
<b>B</b>		<b>137</b>
B.1	Chapter 4: Experiment 3 reconstruction steps . . . . .	137
B.2	Chapter 4: Experiment 4 layout and steps . . . . .	139
<b>C</b>		<b>141</b>
C.1	Chapter 5: Summary of silent gesture results . . . . .	141
C.2	Chapter 5: Word order coding procedure . . . . .	143
C.3	Chapter 5: Image orientation: Experiments 5-6 . . . . .	144
C.3.1	Word order . . . . .	144
C.3.2	Word order errors . . . . .	144
C.4	Chapter 5: Word order preferences in Experiments 6 and 7: candidate explanations	146



# List of Figures

3.1	Experiment 1: Example items showing a generic-agent and character-agent event	33
3.2	Experiment 1: Word orders used to describe character-agent events and generic-agent events across both blocks . . . . .	35
3.3	Experiment 1: The distribution of word orders by agent type in each block . . .	36
3.4	Experiment 1: The proportion of orders used by each participant grouped by condition and block . . . . .	37
3.5	Experiment 1: The proportion of block 1 trials in which the agent was expressed before the patient . . . . .	37
3.6	Experiment 1: The proportion of agent and patient omissions in block 1 . . . . .	38
3.7	Experiment 1: The proportion of block 2 trials in which the agent was expressed before the patient . . . . .	40
3.8	Experiment 1: The proportion of agent and patient omissions for each agent type in block 2 . . . . .	41
3.9	Experiment 1: The proportion of responses coded as consistent with APV, following Goldin-Meadow et al. (2008), and patient-before-action, following Gibson et al. (2013) . . . . .	43
3.10	Model: Bootstrap mean proportions of each word order in the observed distribution across surface orders, the best-fit surface distribution, and the inferred underlying distribution for generic-agent events and character-agent events . . .	47
4.1	Experiment 2: Three component images depicting an action, an object, and a character human . . . . .	61
4.2	Experiment 2: Example layout . . . . .	62
4.3	Experiment 2: Distribution of selection orders in each condition . . . . .	62
4.4	Experiment 2: Proportion of trials in each condition where the agent was selected before the patient . . . . .	63
4.5	Experiment 2: Proportion of trials in each condition where the patient was selected before the action . . . . .	64
4.6	Experiment 3: Example trial in which the event components were selected in the order APV . . . . .	68
4.7	Experiment 3: Distribution of selection orders in each condition . . . . .	69
4.8	Experiment 3: Proportion of trials in each condition where the agent was selected before the patient . . . . .	70
4.9	Experiment 3: Proportion of trials in each condition where the patient was selected before the action . . . . .	72
4.10	Experiment 4: Distribution of selection orders in each condition . . . . .	74

4.11	Experiment 4: Proportion of trials in each condition where the agent was selected before the patient . . . . .	74
4.12	Experiment 4: Proportion of trials in each condition where the patient was selected before the action . . . . .	76
5.1	Schematic illustrating the overall pattern of responses predicted by the noisy channel hypothesis and the cognitive salience hypothesis for a P-first language and a V-final language . . . . .	91
5.2	Experiment 5: Examples of images presented during vocabulary training . . . . .	92
5.3	Experiment 5: Example stimuli showing a non-reversible event in which a girl is kicking a potted plant, and a reversible event in which a woman is pushing a boy . . . . .	92
5.4	Experiment 5: The proportion of PAV responses by event type for the P-first group and the V-final group . . . . .	95
5.5	Experiment 6: The proportion of PAV responses by event type in the P-first group and the V-final group . . . . .	100
5.6	Experiments 5 and 6: The proportion of PAV responses by image orientation in the P-first group and the V-final group . . . . .	101
5.7	Experiments 5 and 6: The proportion of invalid responses in the P-first group by image orientation . . . . .	102
5.8	Experiment 7: The proportion of PAV responses by event type in the P-first group and the V-final group . . . . .	105
5.9	Experiments 5, 6, and 7: The proportion of PAV responses by event type across all three experiments in the P-first group and the V-final group . . . . .	106
B.1	Experiment 3: Reconstruction steps – AVP . . . . .	137
B.2	Experiment 3: Reconstruction steps – PAV . . . . .	137
B.3	Experiment 3: Reconstruction steps – PVA . . . . .	137
B.4	Experiment 3: Reconstruction steps – VAP . . . . .	138
B.5	Experiment 3: Reconstruction steps – VPA . . . . .	138
B.6	Experiment 4: Layout and steps . . . . .	139
C.1	Coding procedure used when a word supplied by the participant did not match any of the event constituents . . . . .	143

# List of Tables

3.1	Experiment 1: Number of participants who used APV or AVP as their most common order overall across blocks and by agent type . . . . .	35
3.2	Experiment 1: Word order proportions in each condition and block . . . . .	36
3.3	Experiment 1: Number of participants in each condition and block who produced predominantly APV, AVP, or incomplete orders . . . . .	36
3.4	Experiment 1: Mixed effects logistic regression analysis of agent-before-patient responses in block 1 . . . . .	38
3.5	Experiment 1: Mixed effects logistic regression analysis of agent omissions in block 1 . . . . .	39
3.6	Experiment 1: Mixed effects logistic regression analysis of patient omissions in block 1 . . . . .	39
3.7	Experiment 1: Mixed effects logistic regression analysis of APV responses across blocks . . . . .	40
3.8	Experiment 1: Mixed effects logistic regression analysis of agent-before-patient responses in block 2 . . . . .	41
3.9	Experiment 1: Mixed effects logistic regression analysis of agent omissions across blocks . . . . .	42
3.10	Experiment 1: Mixed effects logistic regression analysis of patient omissions across blocks . . . . .	42
3.11	Model: Output for each agent type using two model parametrizations . . . . .	47
3.12	Model: Proportion of each incomplete order that derived from each complete order	48
4.1	Experiment 2: Mixed effects logistic regression analysis of agent-before-patient responses . . . . .	63
4.2	Experiment 2: Mixed effects logistic regression analysis of AVP responses . . . . .	64
4.3	Experiment 2: Mixed effects logistic regression analysis of APV responses . . . . .	64
4.4	Experiment 2: Mixed effects logistic regression analysis of patient-before-action responses . . . . .	65
4.5	Experiment 3: Mixed effects logistic regression analysis of agent-before-patient responses . . . . .	69
4.6	Experiments 2 and 3: Mixed effects logistic regression analysis of agent-before-patient responses . . . . .	70
4.7	Experiment 3: Mixed effects logistic regression analysis of AVP responses . . . . .	71
4.8	Experiments 2 and 3: Mixed effects logistic regression analysis of AVP responses	71
4.9	Experiment 3: Mixed effects logistic regression analysis of APV responses . . . . .	71

4.10	Experiment 3: Mixed effects logistic regression analysis of patient-before-action responses . . . . .	72
4.11	Experiments 2, 3 and 4: Mixed effects logistic regression analysis of agent-before-patient responses . . . . .	75
4.12	Experiment 4: Mixed effects logistic regression analysis of AVP responses . . . . .	75
4.13	Experiments 2, 3 and 4: Mixed effects logistic regression analysis of AVP responses	76
4.14	Experiment 4: Mixed effects logistic regression analysis of APV responses . . . . .	76
4.15	Experiments 2, 3 and 4: Mixed effects logistic regression analysis of patient-before-action responses . . . . .	77
5.1	Summary of the three main hypotheses explaining animacy-based word order variation in silent gesture . . . . .	88
5.2	Experiment 5: Summary of predictions of the statistical analysis . . . . .	94
5.3	Experiment 6: List of item-label mappings . . . . .	98
A.1	Experiment 1: Set of humans (agents), inanimate objects (patients) and actions .	133
A.2	Experiment 1: Mixed effects logistic regression analysis of patient-before-action responses in block 1 . . . . .	134
A.3	Model: Bootstrap means and 95% CIs for each word order in the observed surface distribution, best-fit surface distribution, and inferred underlying distribution based on block 1 descriptions of generic-agent events . . . . .	135
A.4	Model: Bootstrap means and 95% CIs for each word order in the observed surface distribution, best-fit surface distribution, and inferred underlying distribution based on block 1 descriptions of character-agent events . . . . .	136
C.1	Summary of animacy-based word order variation in silent gesture and signing tasks. . . . .	142
C.2	Experiments 5 and 6: Analysis of the effect of image orientation on word order for each language type . . . . .	144
C.3	Experiments 5 and 6: Analysis of the effect of image orientation on invalid orders for each language type . . . . .	145
C.4	Experiments 6 and 7: Word order preferences – candidate explanations . . . . .	146

# Chapter 1

## Introduction

### 1.1 Background

Human languages exhibit a striking degree of diversity at every level – in their sound systems, morphology, word classes, semantic categories, and so on (Evans and Levinson, 2009). Despite this diversity, it has long been noted that certain structural features are more typologically frequent than others (Greenberg, 1963). Understanding both the diversity and the commonalities across languages lies at the core of much of linguistics research. Identifying the causes can shine a light on our shared cognition, unveiling the biases, constraints, and pressures that shape linguistic structure.

In this thesis, I focus on one of the most widely studied linguistic phenomena: word order in a transitive clause. On the one hand, languages are hugely diverse in terms of basic word order, with all six logically possible orderings of subject (S), object (O), and verb (V) being attested in spoken languages. On the other hand, there is considerable cross-linguistic commonality: the vast majority (around 88%) of spoken languages identified as having a dominant word order use either SOV or SVO, while other orders are apparently quite rare (Dryer, 2013). In sign languages too, SOV and SVO are disproportionately represented (Napoli and Sutton-Spence, 2014). A vast amount of research has been dedicated to understanding this asymmetry, often with an emphasis on investigating why SOV or SVO are functionally better than other orders. A number of functional pressures have been proposed to account for the predominance of one or both of these orders. These include: processing during interaction (Roberts and Levinson, 2017); a drive to convey information at a uniform rate (Maurits and Griffiths, 2014); and learnability (Lupyan and Christiansen, 2002; Tily et al., 2011; Tabullo et al., 2012).

These studies provide valuable insight into how and why languages, as shared communication systems transmitted across generations, converge on one dominant word order rather than another. Nevertheless, the headline statistics mask a considerable degree of flexibility in how people structure information about events. Even languages that adhere to a rigid word order from a syntactic viewpoint are more flexible when looked at through the lens of semantic or pragmatic categories. English, for example, is a strict SVO language, but speakers can choose between active and passive voice thereby varying the relative positioning of the semantic agent and patient. Further, many languages identified as having a dominant word order permit alternative orders (Dryer, 2013). For example, Japanese is predominantly SOV but also allows OSV. Still other languages lack a dominant order, and permit any ordering of the three basic

constituents. Examples include Odawa (Christianson and Ferreira, 2005), Cayuga, Ngandi, and Coos (Mithun, 1992).

Nevertheless, word order variation is not unrestricted or random, and cross-linguistic research has uncovered numerous factors that condition ordering preferences in sentence production. Many of these factors relate to properties of referents interacting in an event, for example, their animacy status or contextually derived salience (see Jaeger and Norcliffe, 2009, for an overview). As I will discuss in §1.4, some of these factors also play a key role in the early development of word order regularities in emerging communication systems. One such factor – animacy – is not only one of the most widely studied topics in linguistics, but, as I argue in §1.2.2, is also of particular interest to the study of language evolution: the distinction between animate and inanimate entities is both fundamental to human cognition and, in addition, has implications for the extent to which the meaning of an utterance can be unambiguously recovered in the absence of linguistic conventions for expressing who is doing what to whom.

In recent years, an increasing number of researchers have turned to the silent gesture paradigm to investigate how animacy and, relatedly, semantic reversibility influence word order in emerging communication systems (e.g., Gibson et al., 2013; Hall et al., 2013; Meir et al., 2017). In these experiments, participants convey information about events using only gestures and no speech. By removing the restrictions of their native language, this paradigm can uncover the biases and pressures that operate in the context of language emergence. This work in the silent gesture paradigm forms the basis of the research presented in this thesis. In the experimental work presented here, I explore the relationship between animacy and word order in more detail and, drawing on findings from language production studies, investigate how other properties of referents – specifically, contextually derived salience – influence word order in emerging communication systems.

In the remainder of this chapter, I first set out why word order is important to the study of language evolution. I then describe evidence from restricted communication systems (i.e., communication systems that are not fully developed languages) in the spoken modality that provides clues as to the overarching ordering principles that may characterize emerging communication systems. In §1.2.2 I discuss the concept of animacy and explain its fundamental importance to understanding word order in emerging communication systems. This discussion will highlight a common theme of this thesis, namely, the extent to which word order in emerging communication systems is driven by a pressure to express who is doing what to whom in an event, or individual cognitive biases for structuring information. I next present a discussion of structural variation in fully mature languages, focusing in particular on the effects of referent properties such as animacy. I argue that findings from these studies can provide important insights into the factors that influence word order in emerging communication systems. Following this, I discuss the literature on word order in restricted and emerging communication systems in the manual modality. Finally, in §1.5, I provide an outline of the thesis.

## 1.2 Word order: a fundamental structuring device in emerging languages

All fully formed modern languages, both spoken and signed, offer users some mechanism for indicating who is doing what to whom in an event. English, for example, relies primarily on a

combination of word order and grammatical voice, while other languages may use a system of morphological case marking to indicate participant roles (e.g., German). Drawing on grammaticalization theory, Heine and Kuteva (2002) argued that the earliest forms of spoken language lacked morphosyntax and other grammatical units such as pronouns, adpositions, and markers for negation and questions. They proposed that the earliest languages comprised only two linguistic categories: words for thing-like, time-stable entities, and words denoting non-time-stable entities such as actions and activities. Going further, they noted that the corollary of this conclusion is that word order would have been the only productive means of specifying relations between entities.

Word order is therefore one of the most fundamental and possibly earliest structuring devices in language. The question that concerns this thesis is: what factors determine the choice of word order in the early stages in the development of a new language?

### 1.2.1 Ordering principles in restricted communication systems

Although we have no direct evidence of how the earliest spoken languages were structured, Jackendoff (2002) has suggested that certain forms of modern-day restricted communication systems exhibit some of the same ordering principles. One such example is the Basic Variety, a simple form of communication developed by adults acquiring a second language outside of a classroom setting. In a longitudinal study involving 40 respondents, Klein and Perdue (1997) reported a number of features that were characteristic of the Basic Variety and that showed intriguing parallels with the posited state of the earliest languages (Heine and Kuteva, 2002). First, they found that the systems had no inflectional morphology. Thus, lexical items were not marked for case, number, gender, tense, aspect, or agreement. Second, the lexicon of the Basic Variety consisted predominantly of noun-like and verb-like words. Looking at word order, Klein and Perdue (1997) argued that utterances in the Basic Variety are organized according to two general principles. The first is a semantic constraint whereby the NP referent with the highest control is expressed first. For example, agents instigate and perform actions and are therefore mentioned first in an agent-patient event. The second constraint is that the focus constituent is expressed last. Another example of restricted language offered by Jackendoff (2002) is pidgin languages. These are rudimentary systems that develop among speakers with no shared language who are brought together and have to communicate.<sup>1</sup> As in the Basic Variety, Jackendoff (2002) noted that the organizing principles of Agent First and Focus Last are typical of these communication systems.

Jackendoff (2002) argued that for language users lacking recourse to a fully developed system of grammatical machinery, Agent First provides a simple yet powerful means of specifying who is doing what to whom in an event. To illustrate why this would be useful, Jackendoff (2002) noted that the meaning of an utterance such as *eat apple Fred* can be recovered independent of the order in which constituents are expressed since the action *eat* is semantically non-reversible – humans can eat apples, but not vice versa. However, the meaning of *hit tree Fred*, he argued, cannot be recovered so easily. In this example, the action *hit* is semantically reversible – both entities could plausibly perform the action, or be affected by it.

The argument presented by Jackendoff (2002) hinges on the notion that semantic reversibil-

---

<sup>1</sup>Pidgins are typically thought of as the precursors of creoles, more developed systems acquired by children as a first language. The distinction and the relationship between pidgins and creoles is controversial, however. See Hurford (2012) for a detailed discussion.

ity imposes a communicative pressure. Under this view, a community might adopt a particular word order early in the development of a language because it is communicatively adaptive. In §1.4 and in Chapter 2, however, I discuss evidence from the manual modality that demonstrates that word order in restricted and emerging communication systems is more variable than the broad ordering principles highlighted by Jackendoff (2002) might imply.

Studies of these systems have uncovered a range of factors that influence word order. Of particular interest is the relationship between animacy and word order. The effects of animacy on language are well-documented and widely studied (see §1.3.1). Further, understanding how and why animacy shapes linguistic structure is of particular interest to the study of language evolution. As I will discuss in the next section, the animacy distinction is not only fundamental to human cognition, but is also intimately related to the notion of semantic reversibility. In the example *hit tree Fred* discussed above, a receiver might infer that Fred is performing the action since humans, or animate entities, are more likely to be agents and possess more agent-like properties than do inanimate objects (Dowty, 1991; Dahl, 2008). If both entities are animate, however, the meaning is less easily recovered, since each could plausibly be the agent or patient. The interplay between animacy and semantic reversibility therefore presents a communicative challenge for individuals who do not share a conventionalized means of marking participant roles. Investigating how animacy influences word order therefore presents a valuable test for the relative contributions of general cognition and communication in shaping language.

## 1.2.2 Animacy and language evolution

The relationship between animacy and linguistic structure is one of the most widely studied topics in linguistics. The effects of animacy on language are of particular interest to the study of language evolution for a number of reasons. First, the distinction between animate and inanimate entities is fundamental to human cognition and develops early in infancy (Mandler, 1992; Opfer and Gelman, 2011; Becker, 2014). Studies have found that even within the first hour of life, infants preferentially attend to pictures of faces with a normal configuration than to faces with a mixed-up configuration or with no features (see Becker, 2014, and references therein). As I will discuss in the next section and in §1.4.2.1, the tendency for humans to preferentially attend to other humans – that is, the *salience* of humans – has been proposed to play a basic role in determining structural preferences in emerging communication systems (Meir et al., 2017).

Second, there is a fundamental correspondence between the characteristics of animate and inanimate entities and their potential to fulfil certain semantic roles. Opfer and Gelman (2011) cite evidence showing that from an early age, children hold certain beliefs and expectations about animate and inanimate motion. Specifically, children believe that animate motion, but not inanimate motion, is self-generated, self-sustained, and goal-directed. In contrast, studies involving young infants suggest that they have the same expectations about contingent motion in animates as in inanimates, that is, about motion or behaviour in response to an external stimulus. Thus, fundamental to the animate-inanimate distinction is the notion that only animates can act intentionally and in a goal directed fashion, whereas both animate and inanimates can be acted upon. In addition to these dynamical aspects, animates are distinct from inanimates in that only the former can have mental states: they can know, perceive, emote, learn, and think (Gelman and Spelke, 1981). These properties are closely aligned with the entailments of

the high-level generalized semantic roles proposed by Dowty (1991) – Proto-Agent and Proto-Patient. Proto-Agents can act volitionally, are sentient, can cause an event or change of state, and can undergo movement relative to another participant. Proto-Patients, on the other hand, can undergo a change of state, can be causally affected by another participant, and are more likely to be stationary with respect to another participant.

The relationship between animacy and semantic roles was highlighted by Dahl (2008), who argued that animacy is an ontological type, that is, a natural kind similar to numbers, times, locations, etc. Membership within the animacy type therefore determines what can be true about an entity. For example, Dahl (2008) noted that the semantic roles of agent and experiencer can be fulfilled by animate, but not inanimate, entities. A corollary of this, as I noted in the previous section, is that animacy may act as a cue to semantic roles in an otherwise ambiguous utterance (de Hoop and Lamers, 2006; Opfer and Gelman, 2011; Primus, 2012).

This point brings us to the third reason why animacy is significant for language evolution, namely, its effect on the interpretability of utterances. Recall from §1.2.1 that Jackendoff (2002) suggested that an Agent First organizing principle may serve a functional purpose in disambiguating participant roles in a semantically reversible event such as *hit tree Fred*. However, as we have seen, a hearer might infer that the human referent is the agent in this instance, since it is the more agent-like of the two referents. However, what if both referents are equally agent- or patient-like, for instance, if both are human as in *hit Fred John*? In this case, the combined effects of semantic reversibility of the verb and the properties of the referents – their animacy – would make such an utterance fully ambiguous (absent any other contextual information, for instance, that John is typically more prone to violence than Fred).

To summarize this discussion, in transitive clauses, the semantic properties of the verb combined with the properties of referents can serve to identify the agent and patient irrespective of the order in which constituents are expressed. For semantically reversible events, a hearer may still recover the meaning of an utterance if one entity possesses more agent- or patient-like properties than the other. However, where both entities are equally plausible candidates for those roles, for example, if both are animate, the meaning is less easily recovered. Under the assumption that language producers exploit word order to mark participant roles, we might expect that semantically reversible events involving two human participants would tend to elicit more consistent ordering strategies in emerging communication systems. In fact, as I will discuss in §1.4 and in Chapter 2, this is not the case.

In the next two sections, I discuss evidence from typology and language production studies (§1.3) and from restricted and emerging communication systems in the manual modality (§1.4) demonstrating that a pressure to transparently convey information about who is doing what to whom is only one of a number of factors that shape word order choices. These studies show that the order in which constituents are expressed is also driven by a preference for expressing certain entities before others, independent of communicative need. More specifically, they show that animacy-based word order variation is driven by a preference for expressing more conceptually accessible, or more salient, entities earlier (Branigan et al., 2008; Meir et al., 2017).

### 1.3 Word order in fully developed languages: the influence of referent properties

A vast body of research has been dedicated to investigating the factors that determine how language producers structure information about events. In the psycholinguistics literature, much of this work has focused on the effects of *conceptual accessibility* (see Jaeger and Norcliffe, 2009, for an overview). This notion was introduced by Bock and Warren (1985) and refers to the ease with which concepts are represented in thought and/or retrieved from memory. In their original conception, Bock and Warren (1985) assumed that more accessible concepts were those which were more ‘thinkable’, that is, ‘those whose mental representations are learned earliest and are most richly detailed in adult representations of knowledge.’ (p. 50) Under this definition, conceptual accessibility relates primarily to inherent properties of referents such as animacy or imageability. In its broader sense, however, the notion has also been applied to contextually derived factors such as thematic role. According to the notion of thematic role accessibility, the agent role is more accessible, or more prominent, than other roles such as the patient (e.g., Hwang, 2017).

Bock and Warren (1985) argued that there is a correlation between syntactic role assignment and conceptual accessibility such that more accessible entities tend to occupy more prominent syntactic roles. For example, animate entities are more likely to be assigned the subject role than are inanimate entities. Similarly, agents are more likely to appear as syntactic subjects compared with patients. Later work has uncovered evidence that conceptual accessibility also has a more direct effect on positional processing independent of functional role assignment, that is, more accessible entities tend to be mentioned earlier in an utterance (e.g., Prat-Sala and Branigan, 2000; Gleitman et al., 2007; Branigan et al., 2008; Hwang, 2017).

In much of the literature exploring the effects of conceptual accessibility, the notion is often used interchangeably with that of *salience*. As noted by Ferreira and Rehrig (2019), while in the scene literature the term refers to measurable, low-level visual properties such as luminance and size, elsewhere in the psycholinguistics literature it is applied more loosely to a range of phenomena that can broadly be summarized as referring to factors that make an entity more prominent, important, or interesting, and therefore more likely to attract the attention of the individual describing the scene (see also Chapter 3). Accordingly, the notion of conceptual accessibility has been extended to include a range of factors including discourse salience (e.g., Prat-Sala and Branigan, 2000), visual properties such as size, contrast, or colour (e.g., Coco et al., 2014; Clarke et al., 2015), or visual prominence associated with implicitly or explicitly cueing a speaker’s visual attention towards a particular referent in a scene (e.g., Antón-Méndez, 2017; Gleitman et al., 2007; Myachykov and Tomlin, 2008; Myachykov et al., 2012; Vogels et al., 2013).

Conceptual accessibility provides a processing account of language production. However, Bock and Warren (1985) noted that it is closely related to cognitive theories of language such as the perspective hypothesis proposed by MacWhinney (1977). According to this account, the starting point of an utterance can serve four functions: it can establish the attentional focus of the producer, which MacWhinney (1977) argued is invariant; the perspective from which an event is construed; the agent in an event; and/or given information in discourse. The first of these functions – establishing the attentional focus of the producer – echoes the notion of salience described above. That is, more salient entities are those which attract the attention of

the producer. Notice also that the last two functions are precisely the ordering principles that Jackendoff (2002) argued are characteristic of restricted communication systems (given, or topic, first is the mirror of focus last). This is an important observation, since it suggests that the biases and pressures that influence linguistic structure are operational from the earliest stages in the emergence of a communication system and remain active as it develops and matures into a full language. Consequently, evidence from typology and language production studies provides valuable insight into the factors that shape word order in emerging languages.

Of particular relevance to this thesis is the finding that more salient entities tend to be mentioned first. This can shed light on the mechanisms underlying the ordering principles found in emerging communication systems since, as I discussed in §1.2, word order may be the earliest structuring device available in a new language (Heine and Kuteva, 2002; Jackendoff, 2002). Further, this salience-based conception of word order preferences offers an alternative, or additional, explanation for the communication-based Agent-First principle proposed by Jackendoff (2002) (see §1.2.1). That is, rather than serving a communicative function of disambiguating participant roles, the principle may reflect a general cognitive preference for expressing more conceptually accessible, or salient, entities first. Accordingly, agents are more accessible than patients and therefore occupy an earlier position. This ordering preference may alternatively derive from the salience of human entities over inanimate entities. Recall that there is a close correspondence between the agent role and animacy such that agents are typically animate rather than inanimate (see §1.2.2). This animate-first or, more specifically, human-first bias has been proposed by Meir et al. (2017) as a fundamental determiner of word order preferences in emerging communication systems. I discuss this hypothesis in more detail in §1.4.2.1. In the next section, I discuss evidence that animacy-based salience, or accessibility, plays a key role in shaping word order preferences in fully developed languages.

### 1.3.1 Animacy and word order in fully developed languages

Evidence that conceptual accessibility has a direct influence on word order comes from studies investigating the effects of animacy. As previously mentioned, animacy is of particular relevance to the study of language evolution. Evidence from typology and language production demonstrates that these effects are numerous and widespread. Detailed treatments of the subject are provided in Yamamoto (1999), de Swart et al. (2008), and Becker (2014). Here, I provide a brief overview of the relationship between animacy and argument encoding, particularly word order.

de Swart et al. (2008) cite numerous examples of languages in which animacy distinctions have been grammaticalized in the way they encode arguments. One such example is Navajo, which is an SOV language but also permits the order of the subject and object to be reversed. Rules of reversal are governed by the relative animacy of the referents: if the subject and object occupy the same position on the animacy scale, reversal is optional; if the subject is higher, reversal is not permitted; if the object is higher, reversal is obligatory. de Swart et al. (2008) also note that in languages such as Jakeltek and Lakhota, inanimate subjects are not typically permitted with active, transitive verbs. In others, Mam-Maya for instance, the subject of an active, transitive verb must be at least as high on the animacy scale as the object. Another example of the grammaticalization of animacy distinctions is differential object marking where, in some languages, only objects high on the animacy scale are explicitly marked for case (Aissen,

2003; Primus, 2012).

The effects of animacy are also reflected in statistical tendencies in the ordering of constituents. Cross-linguistic experimental work provides strong evidence that speakers tend to use syntactic constructions that allow an animate entity to precede an inanimate entity (see Branigan et al., 2008, for an extensive overview). For example, in a picture-description study involving native speakers of English and Spanish, Prat-Sala and Branigan (2000) found that participants were more likely to produce passive constructions when the patient of an event was animate and the agent was inanimate compared with when both entities were inanimate. In addition, they found that Spanish-speaking participants were more likely to express animate patients than inanimate patients as direct objects in left-dislocated sentences (for example, *A la mujer la atropelló el tren*, literally, *The woman, her ran over the train*). Similar findings have been reported in studies involving speakers of German (van Nice and Dietrich, 2003; Esaulova et al., 2019), Dutch (van de Velde et al., 2014), and Korean (Dennison, 2008). Other studies have found that people are more likely to recall sentences in a form in which an animate entity precedes an inanimate entity (e.g., Branigan and Feleki, 1999; Tanaka et al., 2011).

In summary, what this brief discussion demonstrates is that even where languages have conventionalized devices for marking participant roles, animacy imposes grammatical restrictions and/or influences word order choice. Focusing on word order, the evidence discussed above shows that animate entities are more likely to be mentioned first in an utterance, independent of their syntactic or semantic role. This points to the possibility that a pressure to accurately convey information about who is doing what to whom is only one of a number of factors that determine word order in emerging communication systems. In the next section, I discuss word order in restricted and emerging communication systems in the manual modality. As we will see, word order regularities appear early in the development of these systems. In addition, at least some of these regularities are evident even in the absence of communicative interaction.

## 1.4 Word order in homesign and emerging sign languages

The evidence described in the previous section reveals much about the factors that shape language systems. In particular, as we saw, the effects of referent properties such as animacy have become grammaticalized in some languages, while in others they are felt through statistical tendencies and structural preferences. This fact points to a key role for such factors in shaping language. Nevertheless, understanding the precise nature of this role is complicated by the fact that such effects are filtered through the constraints and conventions of individual languages. In addition, every one of the world's approximately 6000 spoken languages is the product of thousands of years of change and evolution. Consequently, some authors contend that we can deduce very little about the structure of the earliest languages from looking at present day phenomena (e.g., Campbell and Poser, 2008).

Fortunately, this is not true of all languages. Unlike spoken languages, whose lineages stretch back thousands of years, in the visual-manual modality there are numerous examples of language systems that have emerged independently within the last century. These are *homesign* systems and emerging sign languages. Homesign systems are idiosyncratic gestural systems created by deaf children with no exposure to a conventional signed or spoken language. Emerging sign languages, on the other hand, develop either within geographically or socially isolated communities with a high incidence of hereditary deafness, or when groups of deaf individuals

who do not share a sign language form a community, for example, when a school for the deaf is established.

In the sections that follow, I discuss findings from studies of these systems showing that word order regularities appear from the very earliest stages in the development of a language. These studies also provide evidence that properties of referents, animacy in particular, are instrumental in shaping structural preferences.

#### 1.4.1 Homesign: structural regularities and individual cognition

Homesign systems are idiosyncratic systems of communication developed and used within a family setting by deaf individuals who have little or no access to a signed or spoken language (Coppola, 2002; Goldin-Meadow et al., 2009). Unlike contexts of typical language acquisition and development, homesign systems are created by individuals who lack a language model and, typically, a communicative partner who shares and uses the same system. Studies have found that homesign systems exhibit a number of features typical of language, such as segmentation and combination (Goldin-Meadow et al., 1996), morphology for encoding causal structure (Rissman and Goldin-Meadow, 2017), and a range of other linguistic devices (see Carrigan and Coppola, 2017, and references therein). The structures that have been found in homesign systems provide evidence that individual cognitive preferences play a fundamental role in shaping newly developing languages.

Of particular interest to the present discussion are studies of word order in homesign systems. These have uncovered striking similarities across individual systems. For example, homesigners living in different linguistic cultures show a common tendency to omit agents in transitive events and to express patients and intransitive agents before actions (Goldin-Meadow, 1985; Goldin-Meadow and Mylander, 1998; Goldin-Meadow et al., 2009). In addition, Coppola and Newport (2005) reported that when homesigners described two-argument events, they tended to express the primary argument (i.e., the agent in an agent-patient event, or non-agents such as experiencer in an experiencer-theme event) in clause-initial position.

This finding recalls the Agent First principle characteristic of restricted communication systems in the spoken modality (Jackendoff, 2002). However, a more detailed look at homesign systems reveals a more complex picture. In a study of the homesign systems created by three deaf children living in Nicaragua, Coppola (2002) found evidence that homesigners vary their structural choices based on the animacy of entities interacting in an event. Two individuals in the study typically produced SOV<sup>2</sup> for events involving an animate subject and an inanimate object. Of these, one preferred SVO when the object was animate. The other individual typically produced a construction that can be glossed as O-SV, where the object was fronted and prosodically set off from the clause containing the subject and verb. A third individual produced predominantly SVO for events involving an animate subject and inanimate object and SOV when the object was animate.

The gestural regularities produced by these homesigners cannot be attributed to influence from co-speech gestures produced by their primary caregivers (e.g., Goldin-Meadow et al., 2009). Further, a study by Carrigan and Coppola (2017) found that mothers and other family members of four adult homesigners living in Nicaragua showed poor comprehension of out-of-context

---

<sup>2</sup>Primary arguments such as agents in agent-patient events or experiencers in experiencer-theme events were analysed as syntactic subjects; secondary arguments were analysed as objects.

descriptions of one- and two-argument events. These results suggest that word order preferences in homesign systems do not develop in response to communicative pressures, that is, the structures do not appear to have a functional role in facilitating efficient or transparent communication. Interestingly, when Carrigan and Coppola (2017) looked at the pattern of errors, they found that for reversible events, participants most often selected a foil image depicting the ‘reverse’ event, for example, a woman kissing a man in response to gesture sequence describing a man kissing a woman. Based on this observation, the authors concluded that family members did not exploit the systematicity of homesign systems to determine who was doing what to whom.

#### 1.4.2 Emerging sign languages

Despite the language-like properties identified in homesign systems, the circumstances under which they are created and used are atypical of the context in which languages emerge, develop, and are learned (Botha, 2007). In addition, detailed studies of homesign systems are relatively small in number and typically involve only a handful of respondents. Consequently, there are limits to how much insight we can gain from these systems. In contrast, emerging sign languages are shared by communities of signers, and transmitted across generations of new learners. Moreover, there are numerous examples of languages that have emerged within the last century for which we have multi-generational data. Emerging sign languages therefore are a rich seam of information on the biases and pressures that shape newly developing communication systems.

As in homesign, studies of emerging sign languages demonstrate that ordering preferences appear early. In Al-Sayyid Bedouin Sign Language (ABSL), for example, SOV emerged as the dominant word order by the second generation (Sandler et al., 2005). Similarly, Ergin et al. (2018) reported that SOV was the most common order used by second- and third-generation signers of Central Taurus Sign Language (CTSL).

These findings are often cited as evidence that SOV enjoys a special status in the early stages of language development (e.g., Goldin-Meadow et al., 2008). However, an early convergence on SOV is by no means universal. In one of the earliest studies of Nicaraguan Sign Language (NSL), for example, Senghas et al. (1997) found that first- and second-generation signers typically described events involving an animate subject and an inanimate object using verb-final constructions. However, there was no preferred ordering of the subject and object. A later study of NSL found evidence for a shift towards more consistent use of SOV for events of this kind, however, there was considerable individual variation, with some participants showing no preferred ordering (Flaherty, 2014). In a study of Israeli Sign Language (ISL) Meir (2010) found that word order was variable, although third generation signers used SOV more frequently than other orders. More broadly, in a review of rural sign languages, de Vos and Pfau (2015) found that word order in young sign languages can be highly variable, both within and between language. In Kata Kolok, for example, SVO, SOV, and OVS appear in equal proportions. Based on these findings, de Vos and Pfau (2015) concluded that there was no evidence that young sign languages share a strong predisposition for SOV.

These findings demonstrate that word order in emerging languages is variable. However, as previously highlighted, focusing on the dominant order masks important phenomena: just as in fully formed languages, word order variability in emerging languages is not random and there are striking cross-linguistic similarities that reveal interesting facts about the biases and

pressures that shape the structure of newly developing languages. In many of the languages discussed above, for example, the preference for SOV disappears when people describe semantically reversible events.

#### 1.4.2.1 Animacy, semantic reversibility, and word order

Meir (2010) provided a detailed analysis of word order in ABSL, looking in particular at the relationship between structural choices and the properties of the referents involved in an event. Of particular interest here is the finding that when signers described events involving more than one argument, they tended to break the description down into a string of single-argument clauses. For example, an event in which a girl feeds a woman may have been described as WOMAN SIT; GIRL FEED. Findings from other emerging sign languages show that this strategy is common. Crucially, in all languages where this phenomena has been described, these so-called *paired verb constructions* (Flaherty, 2014) are especially common when people describe semantically reversible events, typically involving two human referents (ABSL: Meir, 2010; CTSL: Ergin et al., 2018; ISL: Meir, 2010; NSL: Senghas et al., 1997; Flaherty, 2014). Meir (2010) proposed that deconstructing an event in this way – separately representing the agent-centred and patient-centred perspectives – may enable people to accurately convey information about participant roles without relying on explicit role-marking devices such as conventionalized word order or case marking.

There is little data relating to the relative ordering of clauses in paired verb constructions in emerging sign languages. Data from NSL suggests that agent-centred clauses typically precede patient-centred clauses (Senghas et al., 1997). This order seems intuitive, since it reflects the temporal structure of the event – cause precedes effect. Findings from ABSL, reported in Padden et al. (2010), however, suggest an alternative, or additional, ordering bias. Although signers were highly variable in the order in which clauses were expressed, Padden et al. (2010) noted that one type of event elicited a more uniform ordering strategy than others: where one participant was stationary, or passive, and the other more active, signers typically expressed the stationary participant before the active participant. The authors attributed this phenomenon to Figure-Ground assignment, that is, to the way a producer perceives events.

Where signers of emerging sign languages describe events using more than one argument within a clause, here too there is variation based on the animacy of referents. In a study involving signers of three emerging sign languages (ABSL, ISL, and Kafr Qasem Sign Language (KQSL)), Meir et al. (2017) found that descriptions of non-reversible events involving a human agent and an inanimate patient were predominantly SOV. In contrast, this order was less common for reversible events. Instead, non-literate signers with limited or no exposure to a second language tended to produce SOV and OSV in roughly equal proportions.<sup>3</sup> Similar findings have been reported in NSL (Flaherty, 2014) and CTSL (Ergin et al., 2018).

To account for these findings, Meir et al. (2017) proposed a human-first principle whereby human entities are more salient than inanimate entities and are therefore expressed first. This results in S-before-O when the subject is human and the object is inanimate, and no preferred ordering of S and O when both are human. This hypothesis echoes the discussion in §1.3 in relation to fully developed languages: more conceptually accessible entities and/or those that attract the attention of the producer are more likely to be expressed first in an utterance. Much

---

<sup>3</sup>Literate signers, in contrast, produced descriptions that reflected the canonical orders of the other language(s) with which they were familiar.

of the work presented in this thesis takes this hypothesis as its starting point. Throughout the thesis, I refer to this as the salience, or cognitive salience, hypothesis.

### 1.4.3 Silent gesture: moving from the field to the laboratory

As the preceding discussion demonstrates, the study of homesign and emerging sign languages is an invaluable tool in understanding the mechanisms that shape the structure of newly developing communication systems. Despite the clear benefits it provides, however, research in this field faces significant challenges. Charting the development of a natural language requires longitudinal studies that can involve years, or even decades of field work. In addition, signing communities are often geographically isolated, which can present considerable practical difficulties for researchers. An additional limiting factor is that signing communities are often small, thereby restricting the amount of individual data points available for analysis.

Over the past ten or fifteen years, an experimental paradigm has been developed that circumvents these issues and seeks to emulate some of the conditions under which new communication systems emerge – the silent gesture paradigm.<sup>4</sup> In such experiments, hearing participants with no knowledge of any sign language are asked to provide descriptions of stimuli using only gesture and no speech. The task facing participants is to convey information about some concept typically without a conventional lexicon, and in the absence of conventionalized strategies for structuring information. We might expect under these conditions that participants would simply replicate the conventions of their native, spoken language in the manual modality, for example, using canonical word order when describing events. However, as we will see in Chapter 2, numerous studies have found that this is not the case (e.g., Goldin-Meadow et al., 2008; Langus and Nespors, 2010; Schouwstra, 2012; Futrell et al., 2015; Özçalışkan et al., 2016; but see Meir et al., 2017 for conflicting evidence).

The silent gesture paradigm therefore provides a window into the process of natural language emergence where individuals improvise communication from scratch (Kirby, 2017). In Chapter 2, I review the silent gesture literature and discuss the insights that these studies provide into word order preferences in emerging communication systems.

## 1.5 Thesis outline

The remainder of the thesis is structured as follows. In Chapter 2 I present a detailed review of the silent gesture literature, highlighting the numerous accounts that have been offered to explain word order preferences. Two common threads run through these accounts: the role of a communicative pressure to accurately convey information on the one hand; and the effect of cognitive preferences for structuring information on the other. These two themes are particularly relevant to the debate around the relationship between animacy and word order variation in silent gesture. Hypotheses that have emerged from this body of work highlight: the relationship between word order and the potential ambiguity inherent to semantically reversible events in which both the agent and patient are human; modality-specific production constraints arising from the relative animacy of interacting entities; and the salience of human entities relative to inanimate entities.

---

<sup>4</sup>The paradigm is also referred to as improvised gesture (e.g., Schouwstra, 2012), or elicited pantomime (e.g., Hall et al., 2013).

In Chapter 3 I report a silent gesture study in which I investigate this latter hypothesis in more detail. Meir et al. (2017) argue that animacy-based salience is a fundamental determiner of word order in emerging communication systems. However, as the discussion in §1.3 demonstrates, the salience of a referent may derive from a number of different sources. An important question concerns whether salience deriving from one property, such as animacy, can interact with salience based on another to influence word order preferences. Evidence from a small number of language production studies suggests that animacy-based salience can be modulated, or overridden by, for example, making an inherently non-salient referent more salient in discourse (Prat-Sala and Branigan, 2000), or by manipulating the visual prominence of a human agent (Rissman et al., 2018). Nevertheless, such effects have not been taken into account in studies of emerging sign languages or in the silent gesture literature. In particular, the extent to which the ‘interestingness’ of the event agent and/or the visual prominence of the patient relative to the agent influence word order choices has not been explored. Consequently, the contributions of factors other than humanness and/or agency are poorly understood. For example, presenting visually or conceptually salient human agents (e.g., a pirate or a ballerina) interacting with relatively small, less visually prominent inanimate patients (e.g., a ball or a guitar) may maximize the salience difference between the agent and patient resulting in a strong agent-before-patient bias.

In this study, I therefore ask if the bias for expressing animate entities before inanimate entities can be modulated by manipulating the salience of a human agent in an event. I also develop a novel computational method for inferring word order preferences based on incomplete gesture strings. The results of this study suggest that the relationship between salience and word order is not necessarily linear. Rather, manipulating the salience of referents influences the perspective from which a producer frames an event, which in turn influences structural choices. The results, however, are inconclusive about whether these structural choices reflect a direct mapping from conceptual structure to word order. In Chapter 4, I pursue this question further and investigate the role of another operational factor: biases specific to the gestural modality. I report three experiments in which I test if the results of Experiment 1 can be replicated in a non-gestural modality. In the first experiment (Experiment 2), participants select pictorial representations of the agent, patient, and action of a target event. In Experiments 3 and 4 I implement changes to the design of the first experiment such that participants create a reconstruction of a target scene by selecting each of the three components. The findings from this study are inconclusive about the relationship between salience, modality and word order. They nevertheless highlight important questions about the effects of other task-specific factors such as the way elicitation stimuli are presented.

In Chapter 5, I focus on the relationship between word order and animacy. I report three artificial language learning experiments in which I further investigate the salience hypothesis and test this against a communication based account – the noisy channel hypothesis – which posits that word order is exploited by producers to mitigate against the the potential ambiguity of events involving two human referents (Gibson et al., 2013). The results of this study offer tentative support for the salience hypothesis, indicating that the properties of the referents (i.e., animate or inanimate) rather than the semantic reversibility of events influenced word order choices in these experiments. Nevertheless, I suggest that further work is required to understand how language producers negotiate the communicative challenge of accurately conveying information about events where the role played by each of the noun referents is potentially

ambiguous.

The closing chapter provides a summary of the work presented in this thesis and offers possible directions for future research.

## Chapter 2

# Word order and silent gesture

### 2.1 Introduction

Over the past ten to fifteen years the silent gesture paradigm has become increasingly popular as a means of tapping into the cognitive biases that shape linguistic structure. In removing people from the constraints of their native language, silent gesture offers a window into the individual biases that influence the emergence and evolution of linguistic structure, and can provide valuable insight into how languages are structured in the earliest stages of their development (Schouwstra, 2012; Motamedi, 2017; Kirby, 2017).

Silent gesture has proved a fruitful experimental paradigm for uncovering cognitive preferences in a range of domains, including word order within the noun phrase (Schouwstra et al., 2017; Culbertson et al., in press), how people convey temporal information in the absence of linguistic conventions (Schouwstra, 2017), and the emergence of systematic structure within language (Motamedi et al., 2018; Nölle et al., 2018; Motamedi et al., 2019; Schouwstra et al., 2020). In this chapter, I focus on a body of work that has adopted the silent gesture paradigm to investigate the word orders people use to convey information about events.

### 2.2 Early investigations into word order in silent gesture

The first study to adopt the silent gesture paradigm was Goldin-Meadow et al. (1996). English-speaking, hearing participants were assigned to one of two conditions: in one, they were shown a series of animations depicting events in which one object moved relative to a second, stationary object (e.g., a doll jumping into a hoop), and were asked to describe those events verbally; in the other, participants described the same events using only gesture and no speech. Two findings are of particular interest to this discussion. First, participants in the gesture-only condition used highly consistent word order across items. Second, in trials where participants expressed all three elements, the most common order was Stationary-Moving-Action, for example, *hoop doll jump*. Crucially, this order is not consistent with canonical English order (Moving-Action-Stationary, for example, *a doll jumps into a hoop*). More generally, participants showed a strong preference for expressing the action in final position.

To explain their findings, Goldin-Meadow et al. (1996) argued that the order Stationary-Moving-Action ‘follows an intuitively natural progression in terms of laying out the scene for the listener’ (p. 49). The stationary object is introduced first to set the scene; the moving

object represents the topic, that is, what the event is about; and finally, the action serves as the comment on that topic. Goldin-Meadow et al. (1996) also suggested that using a consistent word order has an additional functional role, namely, to distinguish the thematic roles played by entities in an event (see also Jackendoff, 2002).

The findings of this study were later replicated by Gershkoff-Stowe and Goldin-Meadow (2002) who found that the order Stationary-Moving-Action was robust across different communicative contexts (the signaller was sharing new knowledge with their communicative partner; the partner was permitted to provide feedback to indicate comprehension or request clarification; the signaller and receiver swapped roles). Noting the absence of an effect of communicative context, Gershkoff-Stowe and Goldin-Meadow (2002) argued that word order may not reflect a pressure for communicative transparency. Instead they proposed that the way people structure information may directly reflect the way they think about events. Another way of putting this is that structural choices are driven by event conceptualization and construal. I return to this theme in Chapter 3 where I discuss the relationship between word order and producer perspective in silent gesture.

Support for this hypotheses came from a second, non-gestural experiment in which participants reconstructed events by stacking a set of transparencies depicting each of the event components. I describe this experiment in detail in Chapter 4. Of interest to the present discussion is the finding that across different communicative contexts, participants were again consistent in their ordering strategies. Furthermore, the order in which they stacked pictures did not reflect canonical English order<sup>1</sup>. This finding is particularly striking, since the final reconstructed picture looked the same irrespective of the order in which the transparencies were stacked.

The preference for describing crossing-space events using the order Stationary-Moving-Action has an interesting parallel with the observation discussed in §1.4.2.1 that ABSL signers typically place clauses describing stationary or passive human participants before those describing active participants (Padden et al., 2010). As previously noted, the authors attributed this phenomenon to Figure-Ground assignment, that is, to the way a producer perceives events. Relatedly, in a review of 42 sign languages, Napoli and Sutton-Spence (2014) found that in locational expressions, larger immobile objects tend to precede smaller mobile objects. This preference is independent of the thematic role or grammatical function performed by the noun phrases in the clause and has similarly been related to Figure-Ground considerations.

## 2.3 SOV: The natural order of events?

More recent silent gesture studies have focused on the relative ordering of the three basic constituents in a transitive event: agent, patient and action. The main focus of much of this later work has been to investigate how cognitive biases present in individual producers can explain the prevalence of SOV and SVO as the basic word orders in spoken languages (Dryer, 2013) and in sign languages (Napoli and Sutton-Spence, 2014).<sup>2</sup>

---

<sup>1</sup>Participants in the Talk condition – where they were asked to describe what they were doing to the experimenter while they stacked the transparencies – were less consistent and were more likely to produce Moving-Action-Stationary, which is analogous to the canonical English order (e.g, the doll jumps into the hoop). This finding suggests that stacking order mirrored the order in which participants verbally expressed constituents.

<sup>2</sup>Reflecting this focus, many of these studies use the syntactic categories S, O, and V to classify gestures. It should be kept in mind, however, that this is a notational convenience: there has been no analysis to establish the syntactic status of gestures in these studies.

In one of the most widely cited silent gesture studies, Goldin-Meadow et al. (2008) recruited speakers of different languages – Chinese (which uses both SOV and SVO), English (SVO), Spanish (SVO), and Turkish (SOV) – to investigate if native language word order influenced ordering preferences. Participants were asked to provide gestured descriptions of four different types of event: intransitive actions in place (e.g., a man bends over), intransitive actions crossing space (e.g., a dog moves to a van), transitive actions in place (e.g., a boy stirs a spoon), and transitive actions crossing space (e.g., a man carries a chicken to scaffolding). The authors found that the relative ordering of the actor, patient, and action was highly consistent across all four language groups. For events involving intransitive actions, participants predominantly expressed the actor before the action. Events involving transitive actions were typically described using orders the authors classed as consistent with Actor-Patient-Action. These included Actor-Patient-Action, Actor-Action, and Patient-Action. For crossing-space events, participants expressed end-points (i.e., goals and recipients) either at the beginning or the end of the gesture string.<sup>3</sup> As in Gershkoff-Stowe and Goldin-Meadow (2002), these ordering preferences were replicated in a second experiment in which participants reconstructed scenes by stacking transparencies depicting each of the event components (see Chapter 4 for more detail).

To explain these ordering preferences, Goldin-Meadow et al. (2008) proposed two cognitive biases: first, they posited a close cognitive link between patients and actions such that they tend to be mentioned contiguously; second, they argued that concrete entities tend to be mentioned first because they are cognitively more basic and less relational than actions. This proposal stems from a hypothesis by Gentner and Boroditsky (2001) that states that certain aspects of an individual's experience of the world are more readily conflated into unified concepts than are others. According to this hypothesis, concrete objects and individuals form cohesive units that can be readily mapped to individual concepts. In contrast, relations between entities vary with experience and do not form a cohesive collection of percepts. Gentner and Boroditsky (2001) proposed their theory to explain cognitive development in children and, more precisely, how and why they acquire object names before relational terms. Goldin-Meadow et al. (2008) do not make clear why such biases should be operational in adult participants with fully formed concepts for both concrete entities and relations. Nevertheless, it is conceivable that more readily individuated referents in an event may be highlighted before more abstract relations. It is also interesting to note the parallels with the notion of conceptual accessibility discussed in §1.3 as relating to mental representations that are learned early and are more richly represented in adult knowledge (Bock and Warren, 1985).

In contrast to earlier silent gesture studies that emphasized the relationship between word order and the way people conceptualize particular events, Goldin-Meadow et al. (2008) argued that their findings provide evidence for a 'natural' ordering of constituents. They suggested that this order – Actor-Patient-Action – may be semantically clear, but also argued that it arises independently of communicative pressures. Going further, they noted the parallel between Actor-Patient-Action and Subject-Object-Verb (SOV) and suggested that this may be the default order adopted by all developing languages. In support of this conclusion, Goldin-Meadow et al. (2008) cited evidence from ABSL, which adopted a basic SOV order from an early stage in its development (Sandler et al., 2005). The authors also noted that homesign systems typically exhibit object-verb ordering. In addition, they highlighted the predominance

---

<sup>3</sup>The authors do not provide a break-down of end-point position by action type (intransitive or transitive).

of SOV and SVO in the world's languages (Dryer, 2013; Napoli and Sutton-Spence, 2014) and noted that SOV is somewhat more common than SVO. Newmeyer (2000) has argued that SOV was the order used by so-called Proto-World, the putative common ancestor of all spoken languages. In a more recent study, Gell-Mann and Ruhlen (2011) similarly argued for an ancestral SOV language. I return to these arguments below.

The ordering preferences reported in Goldin-Meadow et al. (2008) were replicated in a later silent gesture study by Langus and Nespors (2010), which involved native speakers of Italian (SVO) and Turkish (SOV). In interpreting their results, the authors assume that the faculty of language is modular, comprising the conceptual system (semantics), the sensory-motor system (phonology and phonetics), and the computational system of grammar (syntax). They proposed that SOV derives from the interaction between the sensory-motor and the conceptual systems, which they characterized as more primitive than the computational system of grammar. This latter system, they argued, prefers verb-object orders 'and is possibly limited to the SVO order' (p. 310). To support this claim, Langus and Nespors (2010) presented results from a second experiment in which participants provided gestured descriptions of more complex events, for example, *the man tells the child that the girl catches a fish*. In this experiment, both groups of participants tended to express the subordinate clause after the main clause, a pattern which is typical of SVO, but not SOV, languages. The authors concluded that 'the SOV order in improvised gesturing does not generalize to more complex SOV language-like constructions.' (p. 300). Hence, gesture production does not engage the computational system of grammar. As Schouwstra (2012) has previously highlighted however, this is a puzzling line of logic, since a core aspect of the argument put forward by Langus and Nespors (2010) is that SVO is preferred by the computational system. As such, one might expect that the appearance of SVO-like constructions when participants described more complex events would be taken as evidence that the computational system *was* engaged.

In summary, although drawing on fundamentally different conceptions of language, both Goldin-Meadow et al. (2008) and Langus and Nespors (2010) characterize SOV as more cognitively basic than, or primary to, SVO. This conclusion is, in my view, unsatisfactory for a number of reasons. First, the claim that SOV was the order used by a single ancestral language (e.g., Newmeyer, 2000; Gell-Mann and Ruhlen, 2011) is controversial. Indeed, attempts to reconstruct Proto-World have met with severe criticism (e.g., Campbell and Poser, 2008). In addition, a recent study by Maurits and Griffiths (2014) that used Bayesian phylogenetic methods to infer the ancestral word orders of seven language families found only weak evidence in favour of SOV. On the basis of their findings, Maurits and Griffiths (2014) concluded that 'SOV may be the safest bet for a common ancestral word order, but it is not an especially safe bet to take' (p. 13579).

A second criticism concerns the claim that emerging sign languages converge early on SOV. As I discussed in §1.4.2, this is true of some languages, for example ABSL and CTSL, but does not hold in general. Languages such as ISL and NSL, for example, have not converged on a fixed order, although there is evidence that they may be shifting towards SOV. Moreover, de Vos and Pfau (2015) found no evidence that young sign languages share a strong predisposition for SOV. Rather, they show a considerable degree of variation in this regard, both within and between languages.

Third, as I discussed in §1.4 and above in §2.2, evidence from homesign and emerging sign languages shows that certain properties of referents – such as animacy, or perceptual properties

in relation to Figure-Ground assignment – elicit different word orders. Similarly, a growing body of evidence from the silent gesture literature shows that SOV is not always the preferred order. As I discuss in the sections that follow, a number of factors have been found to influence the ordering of the three basic constituents. These include the semantic relations between entities (Schouwstra and de Swart, 2014), the structural properties of events (Christensen et al., 2016), the semantic reversibility of events and/or the animacy of interacting entities (Gibson et al., 2013; Hall et al., 2013; Futrell et al., 2015; Meir et al., 2017; Kocab et al., 2018), and even the availability of a lexicon (Hall et al., 2014; Marno et al., 2015)

## 2.4 Beyond SOV

### 2.4.1 Event type and word order

Schouwstra (2012), and later Schouwstra and de Swart (2014), noted that the study by Goldin-Meadow et al. (2008) used a restricted set of event types to elicit gestured descriptions. Specifically, all of the events involved one concrete entity acting on another, where the ontological status of the actor and patient was equal and independent of the other. For example, in the event *Man plays guitar*, both the guitar and the man exist in the world and are co-present, independent of the playing action being performed. Schouwstra (2012) contrasted these so-called *extensional* events with *intensional* events, where the ontological status of the actor and patient are not equal. For example, consider the event *a princess wants an apple*. Schouwstra notes that:

in order for a sentence describing this situation to be true, we need the princess to exist, but the ‘ontological demands’ on the apple are different: a princess can want an apple without the actual apple being around, or she can want an apple but not a particular one. It is even possible for the princess to want something that does not exist at all, as in ‘The princess wants a unicorn’. (p. 131)

Thus, in contrast to extensional events, the patient in an intensional event is in some sense dependent on the agent and the action being performed. Schouwstra (2012) hypothesized that this semantic relationship would be reflected in word order; specifically, patients in intensional events should be expressed after the action, yielding Agent-Action-Patient, glossed as SVO. To test this prediction, and to rule out influences from the participants’ native language, Schouwstra and de Swart (2014) recruited Turkish (SOV) and Dutch (SVO) speakers and asked them to describe a series of extensional and intensional events using only gesture and no speech. As per their prediction, they found a strong effect of event type whereby extensional events were predominantly described using SOV, while SVO was the most common order for intensional events.

These findings were replicated in a silent gesture study by Christensen et al. (2016) involving native speakers of Danish (SVO). They contrasted the word orders used to describe object manipulation events, such as *the chef eats a banana*, with those used for object construction events, which are a sub-category of intensional events (Schouwstra, 2012), for example, *the doctor bakes a cake*. As in Schouwstra and de Swart (2014), participants predominantly produced SOV when describing manipulation events and SVO for construction events. The authors offered an alternative explanation to the semantics-based account proposed by Schouwstra and

de Swart (2014). They hypothesized that word order in improvised communication is driven by structural iconicity, that is, individual signs are arranged in a way that reflects the relations between their referents and the temporal structure of events. Accordingly, in object manipulation events, the agent and patient must be physically co-present before the action can be performed, hence, the patient is more naturally expressed before the action. In contrast, actions logically precede patients in construction events.

Interestingly, a study by Napoli et al. (2017) found the same systematic conditioning of word order on event type (extensional vs intensional) in Brazilian Sign Language (Libras). Although the language is generally considered to adhere to an SVO order, the authors found that extensional events typically elicited SOV structures, while intensional events were described with SVO.<sup>4</sup> Napoli et al. (2017) interpreted these results, like Christensen et al. (2016), in terms of iconicity and argued that the order in which constituents are visually presented in both gesture and sign corresponds to the ‘chronology of the unfolding event’ (p. 643). This interpretation echoes a proposal by Gershkoff-Stowe and Goldin-Meadow (2002) that word order may reflect a producer’s temporal perspective on an event (see Chapter 4).

Although Schouwstra and de Swart (2014) and Christensen et al. (2016) interpreted their findings from different theoretical perspectives, they have in common the notion that structural choices in improvised communication reflect the way entities relate to one another. In this way, the accounts offered by these authors are more closely aligned with the proposals that emerged from early work in the silent gesture paradigm (Goldin-Meadow et al., 1996; Gershkoff-Stowe and Goldin-Meadow, 2002), than with later claims for a single, default order (Goldin-Meadow et al., 2008).

Nevertheless, Schouwstra (2012) argued that word order differences reflect more than just a simple mapping from the structure of a mental representation to the structure of an utterance. Rather, she suggested that producers use certain orders because they are communicatively transparent. To support this claim, Schouwstra et al. (2019) investigated whether receivers exploited word order to infer the meaning of an ambiguous gesture string in which the action gesture could potentially represent an extensional verb (e.g., climb), or an intensional verb (e.g., build). The main finding of this study was that comprehension matched production behaviour: participants were more likely to interpret a gesture sequence as an extensional event when it was presented in SOV order, and more likely to interpret a sequence as referring to an intensional event when the order was SVO. However, conflicting evidence comes from an earlier study by Hall et al. (2015) who investigated word order preferences for reversible and non-reversible events. In contrast to Schouwstra et al. (2019), they found that production and comprehension diverged, which they concluded was evidence that word order is not exploited by producers for transparent communication.

This debate highlights a common theme in the silent gesture literature: the extent to which word order reflects general cognitive preferences for structuring information, or results from a pressure to accurately convey information. In the next section, I discuss a body of work where this question occupies a central role.

---

<sup>4</sup>Napoli et al. (2017) used the same elicitation stimuli as Schouwstra and de Swart (2014).

## 2.4.2 Animacy, semantic reversibility and word order

As discussed in §1.2.2, the relationship between animacy and word order has particular relevance to the study of language evolution for a number of reasons. First, the animacy distinction is fundamental to human cognition and emerges very early in infancy (Opfer and Gelman, 2011). Second, there is an intimate relationship between animacy and the potential to fulfil certain semantic roles (Dahl, 2008): the properties that characterize animate entities align closely with the entailments of the generalized Proto-Agent semantic role, whereas inanimate objects are more closely aligned with the Proto-Patient role (Dowty, 1991). Third, the animacy of entities interacting in an event, combined with the semantic reversibility of verbs, has implications for the ease with which a receiver is able to determine the intended meaning of an utterance in the absence of conventionalized mechanisms for marking participant roles.

The effects of animacy on language are numerous and have been widely studied in typology and psycholinguistics (Yamamoto, 1999; de Swart et al., 2008; Becker, 2014) (see §1.3.1). Animacy has also been found to influence word order choices in restricted and emerging languages in the manual modality (Senghas et al., 1997; Coppola, 2002; Meir, 2010; Flaherty, 2014; Meir et al., 2017; Ergin et al., 2018), demonstrating that these effects play a key role in shaping the structure of a new language (see §1.4). For these reasons, the relationship between animacy, semantic reversibility, and word order is also one of the most widely studied topics in the silent gesture literature (Gibson et al., 2013; Hall et al., 2013, 2014; Futrell et al., 2015; Meir et al., 2017; Kocab et al., 2018). In this section, I present an overview of these studies. I provide a more detailed discussion in Chapter 5 and a summary of the main findings can be found in Appendix C.1.

An observation worth noting at this point is that the silent gesture literature often treats the notions of animacy and semantic reversibility as though they were interchangeable. However, as I discussed in the previous chapter, these are two separate, albeit intimately related, concepts. Semantic reversibility concerns properties of the verb and the types of argument that appear with it. In the example discussed by Jackendoff (2002) – *hit tree Fred* – the verb *hit* is semantically reversible not because of the animacy of the interacting entities, but because both could plausibly perform and be affected by the action irrespective of their animacy. As I noted, there is an interplay between reversibility and animacy such that in the absence of conventions for marking participant roles, the relative animacy of entities can either act as a cue to the meaning of an utterance, or, if they are the same, can make it fully ambiguous. The literature discussed here has focused exclusively on this latter case, i.e., where the agent and patient interacting in a reversible event have the same animacy status.

As in natural language, silent gesture studies have found that people tend to vary their choice of word order based on the animacy of interacting entities and/or the semantic reversibility of events. A common finding from these studies is that when people describe extensional, non-reversible events involving an animate agent and an inanimate patient, they tend to use SOV, or orders that are classed as consistent with SOV. However, these orders are typically, though not always, less common for reversible events, which in most studies involve two animate entities. Although it is often claimed that SVO is the preferred order for reversible events (e.g., Gibson et al., 2013), the evidence in fact paints a more complex picture, as I will outline briefly below.

In one study, Gibson et al. (2013) found that participants tended to describe reversible events using an order consistent with their native language: English speakers used SVO, while

Japanese and Korean speakers used SOV. In a follow-up study, Futrell et al. (2015) reported that native speakers of various languages – Irish (VSO), Russian (SVO), and Tagalog (VSO) – all used predominantly SVO to describe reversible events. Hall et al. (2013) (English speakers), and later Hall et al. (2014) (English and Turkish), similarly found that SVO was more common for reversible events compared with non-reversible events, although Turkish speakers used predominantly SOV for both types of event, consistent with canonical Turkish word order. Hall et al. (2013) also noted an increase in other orders in which, like SVO, the object did not appear immediately before the verb, for example, OSV, OSVO, and SOSV. Similar results were reported by Kocab et al. (2018) (English) for events involving an animate agent and an animate patient. However, for reversible events involving two inanimate entities, for example, *a car hits a truck*, SVO occurred in a minority of trials and OSV was the most common order (excluding those coded as *Other*). Interestingly, this order corresponds to what Gershkoff-Stowe and Goldin-Meadow (2002) coded as Stationary-Moving-Action, which was the preferred order for events involving one object moving relative to a second stationary object (see §2.2). In another study, involving hearing speakers of Arabic, Hebrew, and Turkish, and signers of ABSL, ISL, and KQSL, Meir et al. (2017) found that the distribution of word orders for reversible events varied according to the language experience of the participants. A summary of the relevant findings from these studies is presented in Appendix C.1.

A number of hypotheses have emerged from this body of work to account for the various findings. Each of these proposals is consistent with some of the data but falls short of providing a full explanation. I describe the three main proposals in detail in Chapter 5. Here, I provide a brief summary.

**Noisy channel:** This information theoretic account proposed by Gibson et al. (2013) is predicated on the assumption that communication takes place in the presence of noise, which may result from production errors, external noise that may corrupt a signal, or errors on the side of the receiver. Gibson et al. (2013) argue that producers convey information in a way that will maximize the receiver’s ability to recover a message. Accordingly, verb-medial orders are more robust against information loss for reversible events, since if one noun phrase is lost in transmission, it will still be possible to recover the role (agent or patient) of the received noun phrase. For example, assuming agents always precede the verb and patients always follow the verb, the message *push girl* can be reliably interpreted as referring to an event in which the girl is being pushed. If both entities always appeared on the same side of the verb, it would not be possible to recover this information.

**Role conflict:** Hall et al. (2013) noted that participants typically embodied the role of human agents and patients by anchoring gestures on their own body, for example, flexing their muscles to indicate a man. In addition, they typically enacted actions, miming a pushing action with their own hands. Hall et al. (2013) argued that it may feel more natural to enact the action while still embodying the role of the agent, that is, to produce an SV sequence. In contrast, enacting the action while in the role of the patient – OV – might feel like a conflict of roles.

**Cognitive salience**<sup>5</sup>: Meir et al. (2017) found that SVO was more common for reversible events only among participants with substantial exposure to an SVO language. Other participants showed a different pattern, namely, a roughly equal preference for SOV and

OSV. To explain this, and the preference for describing non-reversible events using SOV, Meir et al. (2017) proposed that word order in emerging languages reflects the relative salience of entities. Human entities are more salient than non-human entities and so tend to be mentioned first. When both entities are human, there is no preferred ordering.

Each of these accounts emphasizes different pressures and biases arising from the animacy of entities interacting in an event. The noisy channel hypothesis focuses on the inherent ambiguity of reversible events. The role conflict hypothesis, on the other hand, is a modality-specific explanation that emphasizes production-based constraints. In contrast to both of these proposals, the cognitive salience hypothesis highlights the importance of conspecifics to human cognition and, more broadly, the relationship between salience and order of mention. This hypothesis also raises the possibility that word order in improvised communication and emerging languages may be sensitive to other salience-related factors, such as discourse prominence (e.g., Prat-Sala and Branigan, 2000), visual properties of entities in a scene such as size, contrast, or colour (e.g., Coco et al., 2014; Clarke et al., 2015), or other factors that may draw a producer’s attention to a particular referent in a scene (e.g. Gleitman et al., 2007). I explore this possibility in Chapter 3.

One interesting but under-explored observation from these studies is that word order for reversible events is often *more* variable compared with non-reversible events (e.g., Hall et al., 2013, 2014; Meir et al., 2017; Kocab et al., 2018). For example, Hall et al. (2013) reported that in Experiment 1, approximately 80% of responses for non-reversible events were classed as either SOV or SVO. For reversible events, fewer than 50% of responses fell into these two categories. Moreover, more descriptions of reversible events were excluded from the analysis, or were classed as ambiguous, simultaneous or rare. They found a similar pattern in Experiments 2 and 3. In sign languages too, as discussed in §1.4.2.1, word order is often less consistent for reversible compared with non-reversible events. For example, in languages such as ABSL (Meir, 2010; Meir et al., 2017) and CTSL (Ergin et al., 2018), where SOV is preferred for non-reversible events, reversible events are more likely to also elicit paired verb constructions or OSV.

This seems counterintuitive. As previously mentioned, semantically reversible events are potentially ambiguous, hence, under the assumption that word order serves a functional role in marking who is doing what to whom (Jackendoff, 2002), we might expect a more consistent choice of word order for such events. One possibility, of course, is that this assumption is not correct and that word order is not subject to communicative pressures. An alternative possibility, though speculative, is that the pressure to elucidate participant roles may itself disrupt an otherwise stable system. More precisely, in the absence of conventionalized devices for explicitly marking participant roles, the pressure to be clear may cause people to try out alternative ways of communicating information. Some support for this suggestion comes from a study of constituent order in three sign languages – Auslan, Flemish Sign Language, and Irish Sign Language (Johnston et al., 2007). In all three languages, reversible events elicited a greater variety of constructions compared with non-reversible events. Further, some of the responses to reversible events consisted of complex, multi-clause utterances. Interestingly, the authors also observed the use of lexical prepositions, typically inserted after the main verb and before the second argument, usually the undergoer (e.g., COWBOY STAB TO AMERICAN-INDIAN). They suggested that this device may serve to mark the undergoer role.

---

<sup>5</sup>This is my own term. Meir et al. (2017) formulate their proposal drawing on a number of mechanisms that can be grouped under broad notions of salience and other cognitive biases.

It is also important to recognize that in addition to word order, the gestural modality affords other mechanisms for indicating participant roles. In a gestural communication task in which participants had to successfully convey information about two sentences to a partner, Motamedi et al. (preprint) identified two strategies specific to the gestural modality: using body position to distinguish between referents, and indexing referents to a particular location in space. Post-hoc analysis by both Gibson et al. (2013) and Hall et al. (2013) showed that the use of spatial indexing was more common when people described reversible events using silent gesture, suggesting that people were exploring other mechanisms for indicating participants roles. This phenomenon has, however, not been systematically investigated in relation to reversible vs non-reversible events in silent gesture and presents an interesting area for further research.

Teasing apart the relative contributions made by the various factors identified by these studies – communicative pressures, modality-specific production factors, salience and, as Meir et al. (2017) put it, the effect of being human – is an active area of research. So far, efforts to test the noisy channel hypothesis against the role conflict hypothesis have produced inconclusive results (Kocab et al., 2018; Kline et al., preprint). In Chapter 5, I present a study that seeks to pit the noisy channel hypothesis against the cognitive salience hypothesis.

### 2.4.3 SVO and the availability of a gestural lexicon

One thing to notice about the studies discussed so far is that they often demonstrate a dichotomy between conditions in which SOV is the preferred order, and those in which SVO is more frequent. The accounts that have emerged from this work emphasize individual cognitive preferences for linearising information (e.g., Schouwstra, 2012; Christensen et al., 2016), and, additionally or alternatively, the presence of a communicative pressure to convey information accurately (e.g., Gibson et al., 2013; Futrell et al., 2015). In parallel, some studies investigating the effects of animacy on word order have also found an increase in OSV for reversible events relative to non-reversible events (e.g., Hall et al., 2013; Meir et al., 2017; Kocab et al., 2018). Hall et al. (2013) explained this finding in terms of modality-specific production constraints, while Meir et al. (2017) highlighted the role of animacy-derived salience.

In this section, I look at two studies – Hall et al. (2014) and a later study by Marno et al. (2015) – that have offered another explanation for these findings, namely, the extent to which the conditions of gesture production were language-like. More specifically, these studies investigated the effect on silent gesture word order when individuals were equipped with a gestural lexicon.

In the first of these studies, Hall et al. (2014) highlighted two features of the ordering preferences reported in Hall et al. (2013) that were atypical of natural language. First, participants often produced ‘inefficient’ orders, that is, gesture sequences that were either under-informative (i.e., a constituent was omitted), or repetitious. Hall et al. (2014) observed that, among the world’s languages, ‘none are known to have a basic constituent order that is inefficient’<sup>6</sup> (p. 6). Second, as noted above, in addition to SVO, reversible events tended to elicit orders in which the object was expressed before the subject, for example, OSV (see §2.4.2). Although O-before-S languages are attested, they are much rarer than languages in which the subject is expressed before the object (Dryer, 2013).

---

<sup>6</sup>This is a puzzling argument. The relative ordering of the three constituents – S, O, and V – is definitional of a ‘basic order’, of which there are exactly six. Hence, a language would only be classified as having a basic order if its canonical order conformed to one of these six orders.

Hall et al. (2014) argued that these findings provided evidence that gestured descriptions in Hall et al. (2013) were not constrained by the same pressures that operate in the context of natural language emergence. They further posited that under more language-like conditions, participants would produce word orders that were more reflective of patterns found in natural language. In particular, they predicted that O-before-S would become less common for reversible events, resulting in an increase in SVO.<sup>7</sup> Whereas previous studies focused on word order preferences when individuals *improvised* descriptions of events, Hall et al. (2014) noted that in contexts of natural language emergence in the visual-manual modality, language users have access to a stable lexicon from an early stage. To make the gesture task more language-like, they therefore asked some participants to use a consistent lexicon when describing events. They further sought to emulate a natural context by dividing participants who used a consistent lexicon into two groups. In the shared group, participants devised a lexicon which they then taught to the experimenter.<sup>8</sup> Participants in the private group were asked to think of a lexicon, but were not required to share it with the experimenter. A third group – the baseline group – were simply asked to describe events using gesture.

Replicating findings from previous studies, English-speaking participants across all three groups produced fewer SOV responses for reversible events compared with non-reversible events. Crucially, however, and contrary to their prediction, Hall et al. (2014) found that SVO was more prevalent in the shared group compared with the baseline group for *both* types of event.<sup>9</sup> Similar findings were reported in a second experiment involving native speakers of Turkish (SOV).

The account offered by Hall et al. (2014) not only fails to explain why SVO increased for both reversible and *non*-reversible events, but also leaves a number of other questions unanswered. Absent from their hypothesis is a causal link between language-like conditions of production and a preference for S-before-O. Put another way, the authors do not explain why conditions of improvisation, or the absence of a communicative partner, should lead to more O-before-S responses. As we saw in the previous section, studies of numerous emerging sign languages have reported an increased use of OSV for reversible events relative to non-reversible events (e.g., Flaherty, 2014; Meir et al., 2017; Ergin et al., 2018). Nevertheless, these languages are used by communities of signers under language-like conditions. Hence, one cannot argue that the constraints of natural language emergence are not operational within these settings. Further, Hall et al. (2014) do not make clear why these effects should be more apparent when people describe reversible events compared with non-reversible events, which in the study by Hall et al. (2013) typically elicited S-before-O responses.

In a later study, Marno et al. (2015) also reported an increase in SVO when participants (Italian (SVO) and Persian (SOV) speakers) described non-reversible events using a consistent lexicon.<sup>10</sup> To explain these findings, Marno et al. (2015) proposed the same modular view of language espoused by Langus and Nespor (2010) according to which, SVO is preferred by the computational system of grammar (see §2.3). Marno et al. (2015) argued that this system can only operate when sufficient cognitive resources are available to express grammatical relations.

---

<sup>7</sup>Descriptions of non-reversible events in Hall et al. (2013) were predominantly S-before-O.

<sup>8</sup>The experimenter confirmed understanding but did not use the lexicon at any point in the experiment.

<sup>9</sup>The private group was numerically intermediate between the baseline and shared group, but was not statistically different from the baseline. Although participants in the private group were asked to think of gestures for each of the entities, the experimental procedure did not include any validation to ensure they had actually done so, or, if so, that they consistently used these gestures. It is therefore difficult to draw any conclusions about the findings from this condition.

<sup>10</sup>The elicitation stimuli in this study comprised a subset of the vignettes used by Langus and Nespor (2010) and included non-reversible events only.

This can happen, they claimed, only when a stable lexicon exists. They further hypothesized that the switch from SOV- to SVO-like constructions when people describe complex events using improvised gesture (Langus and Nespors, 2010; Gibson et al., 2013) could also be explained by appeal to modularity. Under this view, when events are more complex, people may start to use grammar to describe them. Marno et al. (2015) speculated that, to facilitate this, people may use more consistent gestures, thereby freeing up sufficient cognitive resources for the computational system of grammar to operate. Marno et al. (2015) also characterized intensional and reversible events as more cognitively demanding, which, they claimed, could explain why SVO is preferred for such events (although, as we saw in §2.4.2, SVO is not always the preferred order for reversible events).

The hypothesis proposed by Marno et al. (2015) has a number of shortcomings. First, the authors provided no evidence that gestures are more consistent when people describe complex, or cognitively demanding events compared with simple events. Second, it is not clear from their discussion *why* the computational system of grammar requires the presence of a lexicon to free up sufficient cognitive resources. Indeed, it could be argued that simple events are *more* likely to engage the computational system, since, according to Marno et al. (2015), they are less cognitively demanding. Third, assuming that people do rely more heavily on grammar under certain circumstances, a reasonable first assumption might be that they would fall back on the grammar of their own language (see also Meir et al., 2017), rather than switching to something consistent with an SVO language.

In summary, evidence from the two studies described in this section suggests that when people have access to a consistent lexicon, there is a clear shift towards SVO order. However, as we have seen, the explanations offered by both Hall et al. (2014) and Marno et al. (2015) fall short of providing an adequate explanation for this effect. In particular, neither of these accounts, in my view, offer a satisfactory alternative to the various hypotheses, already discussed in this chapter, that have been proposed to explain word order variation in silent gesture. Notwithstanding, the shift to SVO when producers have access to a lexicon is an intriguing and apparently robust finding. It has nevertheless received limited attention in the silent gesture literature. Uncovering its cause will undoubtedly add considerably to our understanding of the mechanisms underlying word order preferences.

## 2.5 Conclusion

Early work in the silent gesture literature suggested that word order in the absence of linguistic conventions reflects the way people mentally represent events (Goldin-Meadow et al., 1996; Gershkoff-Stowe and Goldin-Meadow, 2002). Later, Goldin-Meadow et al. (2008) argued that there is a single, natural order for representing events, namely, SOV. The authors further hypothesized that this is the default order adopted by all newly developing communication systems. In this chapter, I have reviewed findings from numerous studies that challenge this claim and reveal a range of factors that influence word order in silent gesture. Two common threads run through the various hypotheses that have been proposed to account for the findings of these studies: the relationship between word order and individual cognitive preferences for structuring information (e.g., Schouwstra, 2012; Hall et al., 2013; Christensen et al., 2016; Meir et al., 2017); and the role played by a pressure to accurately convey information about events (e.g., Gibson et al., 2013; Futrell et al., 2015).

The debate around whether cognitive preferences or communicative pressures underlie word order variation in silent gesture has attracted particular attention from researchers investigating the effects of animacy and semantic reversibility. The role conflict hypothesis proposed by Hall et al. (2013) emphasizes the role of modality-specific production factors deriving from the animacy of the agent and patient in an event. Meir et al. (2017), on the other hand, suggested that word order derives from the salience of humans relative to inanimate objects. In contrast to these hypotheses, Gibson et al. (2013), and later Futrell et al. (2015), suggested that producers exploit word order to reliably convey information about semantically reversible events where the role of two human referents is potentially ambiguous.

In the next chapter, I investigate the salience hypothesis in more detail and ask if word order preferences deriving from animacy-based salience can be modulated by manipulating the salience, or ‘interestingness’, of the agent in an event. In Chapter 4 I explore the role of modality and test if salience-driven word order preferences found in Chapter 3 (Experiment 1) can be replicated in a non-gestural modality. In Chapter 5 I turn to the question of whether cognitive preferences or communicative pressures drive word order variation. Using an artificial language learning paradigm, I test the salience hypothesis against the noisy channel hypothesis proposed by Gibson et al. (2013).



## Chapter 3

# Constituent Order in Silent Gesture Reflects the Perspective of the Producer

The material presented in this chapter forms the text of a manuscript submitted to the *Journal of Language Evolution* on 10 August 2020 following peer review and an invitation to revise and resubmit, received from the journal on 30 July 2020 (original submission data: 1 June 2020). The text has been formatted to be consistent with the rest of the thesis. The experiment and computational model described in this chapter were conceived in cooperation with all four of my supervisors – Simon Kirby, Kenny Smith, Jenny Culbertson, and Marieke Schouwstra – who are also co-authors on the paper. I designed and developed the experiment, collected the data, performed the analysis, developed the computational model, and wrote the first draft of the paper myself. All co-authors provided advice and guidance on the work presented in this chapter and contributed to editing the paper.

### 3.1 Introduction

How do people convey information about events in the absence of linguistic conventions? This is one of the central questions in language evolution research, and answering it can shed light on the biases and pressures that shape emerging languages. Over the past decade, a number of studies have investigated this question using the silent gesture paradigm in which participants describe events using only gesture and no speech. In one of the earliest studies of this kind, Goldin-Meadow et al. (2008) found that speakers of different languages (Chinese, English, Spanish, and Turkish) expressed event constituents in an order the authors classed as consistent with Agent-Patient-Action, irrespective of their native language. Similar findings were reported in a later study involving Italian- and Turkish-speaking participants (Langus and Nespors, 2010). To explain their findings, Goldin-Meadow et al. (2008) argued that concrete entities are cognitively more basic than actions and so tend to be mentioned first. In addition, they hypothesized a close cognitive link between the patient and action such that they tend to be mentioned contiguously, yielding Agent-Patient-Action. Going further, the authors drew a parallel between Agent-Patient-Action and Subject-Object-Verb (SOV) and proposed that this is the default order

used by all newly developing languages.

Studies of young sign languages, however, challenge this view. de Vos and Pfau (2015), for example, conducted a review of young rural sign languages and found no evidence that they share a single, preferred constituent order. In addition, a growing body of silent gesture literature has uncovered numerous factors that influence improvised word order. These include the semantic relation between interacting entities (Schouwstra and de Swart, 2014), the temporal properties of events (Christensen et al., 2016; Gershkoff-Stowe and Goldin-Meadow, 2002), the animacy of interacting entities (Futrell et al., 2015; Gibson et al., 2013; Hall et al., 2013, 2014; Kocab et al., 2018; Meir et al., 2017), and even the availability of a lexicon (Hall et al., 2014; Marno et al., 2015).

In a recent study investigating the relationship between animacy and word order in silent gesture and emerging sign languages, Meir et al. (2017) argued that SOV is no more cognitively basic than any other order. Rather, they suggest that word order in emerging languages reflects the relative salience of interacting entities. Specifically, they propose that human referents are more salient than inanimate entities and therefore tend to be mentioned first.<sup>1</sup> This ‘human-first’ principle is supported by findings from both emerging sign languages (Nicaraguan Sign Language: Flaherty, 2014; Central Taurus Sign Language: Ergin et al., 2018) and from language production studies (e.g., Prat-Sala and Branigan, 2000; van Nice and Dietrich, 2003; Branigan et al., 2008; Dennison, 2008; van de Velde et al., 2014; Esaulova et al., 2019), which have found that people tend to use constructions in which animate entities are expressed before inanimate entities.

But what exactly is meant by the term ‘salience’? For Meir et al. (2017) the salience of human entities derives from the central importance of conspecifics to human cognition. Elsewhere, the term has been applied to a range of phenomena that can be broadly summarized as referring to factors that make an entity more prominent, important, or interesting, and therefore more likely to attract the attention of the viewer (Ferreira and Rehrig, 2019).<sup>2</sup> Accordingly, the salience of a referent may derive not only from conceptual properties such as animacy, but from numerous other factors including discourse prominence (e.g., Prat-Sala and Branigan, 2000), visual properties such as size, contrast, or colour (e.g., Coco et al., 2014; Clarke et al., 2015), or visually drawing a speaker’s attention towards a particular referent (e.g., Antón-Méndez, 2017; Gleitman et al., 2007; Myachykov and Tomlin, 2008; Myachykov et al., 2012; Vogels et al., 2013). The general conclusion from these studies is that more salient entities tend to be mentioned earlier (but see Myachykov et al., 2009; Hwang and Kaiser, 2015, for evidence of language-specific differences that modulate the effect of visual cueing).

In this study, we investigated in more detail the relationship between salience and word order in communication systems that lack linguistic conventions. An important question concerns whether salience deriving from one property, such as animacy, can interact with salience based on another in influencing word order. Here, we asked if the human-first bias can be modulated by manipulating the contextual salience of entities in an event. Evidence from a small number of language production studies suggests that animacy-based salience can indeed be

---

<sup>1</sup>The authors only consider the distinction between human and inanimate entities, since their elicitation material did not include referents in other categories, for example, animate non-humans.

<sup>2</sup>Ferreira and Rehrig (2019) note that in the scene literature, ‘salience’ refers to measurable, low-level visual properties such as luminance and size. In the psycholinguistics literature, the term is applied more loosely. In this study, we use the term in this less formal sense as a terminological convenience. Salience has also been equated with the notion of conceptual accessibility, which refers to the how ‘thinkable’ a concept is and how easily it is retrieved from memory (Bock and Warren, 1985).

modulated, or overridden. For example, in a verbal sentence production task involving English and Spanish speaking participants, Prat-Sala and Branigan (2000) found that making an inherently non-salient entity, such as an inanimate object, more salient in discourse could override the preference for expressing animate entities earlier. In a more recent study in which English speaking participants provided written responses, Rissman et al. (2018) found that reducing the visual prominence of a human agent by occluding the face resulted in significantly more passive descriptions (i.e., the an inanimate patient was mentioned first), compared with events in which the face was visible. In the present study, we extended the scope of this approach by investigating the influence of contextually derived salience in silent gesture.

Previous studies in the silent gesture literature have not taken into account how the salience, or interestingness, of a human agent might influence word order choices. While some have used elicitation stimuli involving generic humans such as a man or a woman (e.g., Hall et al., 2013), others have featured more salient characters, such as a pirate or ballerina (e.g., Schouwstra, 2012; Christensen et al., 2016), or a mix of generic and character agents (e.g., Goldin-Meadow et al., 2008; Langus and Nespors, 2010; Gibson et al., 2013). In addition, inanimate patients have typically been small with respect to the human agent (e.g., a guitar, ball, food item, plant, etc.). This may have maximized the salience difference between the agent and patient by combining a size/visual prominence contrast (Clarke et al., 2015) with the animacy distinction. In the present study, we sought to eliminate the effects of size by using inanimate patients of a similar size and scale to the agents, making it easier to focus our investigation on the influence of agent salience.

The study had two main objectives. First, we sought to replicate the Agent-Patient-Action (APV)<sup>3</sup> ordering preference in a silent gesture study in which participants described simple transitive events involving human agents and concrete, inanimate patients. Second, we investigated if manipulating the salience of a human agent influenced word order choices.

In the silent gesture experiment detailed in the next section, we asked participants to describe simple transitive events involving human agents and inanimate patients. We manipulated the salience of the agents across two conditions such that participants described events involving either a ‘generic’ human, such as a man or a woman, or a more interesting character, such as a king or a pirate (see §3.2.2 for more details). The hypothesis we sought to test was that the tendency to express human agents before inanimate patients would be modulated by manipulating the salience of the agent. Both spoken and sign languages offer devices for backgrounding non-salient agents. For example, the passive form in English allows speakers to mention the patient before the agent, or to omit the agent entirely. Agent omission is also frequently used in sign languages as a backgrounding device (e.g., American Sign Language: Kegl, 1990; Janzen et al., 2001; Catalan Sign Language: Barberà et al., 2018; Nicaraguan Sign Language: Rissman et al., 2020). The focus of the present study was to investigate the proposal by Meir et al. (2017) that salience influences word order. Accordingly, we predicted that reducing the salience of the agent, presented along with a large, visually prominent patient, would lead to fewer APV responses, and correspondingly more PAV.

To pre-empt our results, we found that, across the board, APV was less common than we anticipated and that PAV was rare. In addition, the relative ordering of the patient and action

---

<sup>3</sup>It is common practice in the silent gesture literature to equate agents with subjects and patients with objects, ostensibly as a notational convenience. We do not follow this convention, since we make no assumptions about the syntactic status of gestured descriptions. In this study, we coded gestures according to the semantic role of referents. We used ‘A’ for agent, ‘P’ for patient, and, to avoid confusion with agent, ‘V’ for action.

was strongly dependent on event type such that patients tended to precede actions for generic-agent events and vice versa for events involving character agents. A third key finding was that participants showed a strong tendency to omit generic agents, but not character agents, from their descriptions. We will argue that, taken together, these findings provide evidence that salience influences structural choices through its effect on the perspective from which the producer frames an event.

As noted above, omitting referents, particularly agents, is a common feature of sign languages. This phenomenon is also found frequently in silent gesture studies (e.g., Goldin-Meadow et al., 2008). While many studies have simply excluded incomplete orders from the analysis (e.g. Christensen et al., 2016), others have incorporated them by classifying them as consistent with a complete order according to some criterion, such as the relative positioning of the expressed constituents (e.g., Goldin-Meadow et al., 2008; Gibson et al., 2013; Hall et al., 2013). In §3.3 we present an alternative solution: a computational model that exploits incomplete descriptions of events to infer the distribution of word orders that participants would have produced had they expressed all three constituents on every trial. This analysis suggests that in the present study APV was the preferred order for describing generic-agent events, and AVP was preferred for character-agent events.

## 3.2 Experiment 1: Silent gesture task

### 3.2.1 Participants

We recruited 28 participants via the University of Edinburgh’s Career Hub website. All participants were self-reported native English speakers with no knowledge of any sign language. They were paid £5 for their participation in the experiment.

### 3.2.2 Materials

The stimuli consisted of a set of cartoon images depicting simple transitive events. Stimuli were controlled for animacy such that all events involved a human agent acting on an inanimate patient. To control for effects of concreteness, or the semantic relation between interacting entities, all events were *extensional*, that is, both the agent and patient were concrete entities existing independently of the action (Schouwstra and de Swart, 2014).

We produced two sets of stimuli depicting either generic-agent or character-agent events. Generic-agent events involved human agents that were identifiable by gender or could be described with reference to basic physical characteristics such as facial hair, glasses or other accessories. Character agents were strongly associated with a profession and/or a distinctive cultural identity, for example, a king or a pirate. We expected that character agents would have high salience due to their distinctive and prominent physical features (for example, a pirate with an eye patch and bandanna), and to their being less prototypical representations of humans. All events depicted the same set of inanimate patients. To control for ordering effects based on the relative size of referents (Clarke et al., 2015), all patients were designed to be similar in size and scale to the agents. The full set of stimuli consisted of five character humans, five generic humans, four objects and four actions (see Appendix A.1). This gave 80 character-agent events and 80 generic-agent events. Figure 3.1 shows an example of (a) a generic-agent event and (b) a character-agent event.



Figure 3.1: Example items showing (a) a generic-agent event, and (b) a character-agent event.

### 3.2.3 Procedure

Participants were randomly assigned to one of two conditions. In the generic-first condition, they were first presented with a block of 40 generic-agent events followed by a block of 40 character-agent events. In the character-first condition the order of presentation was reversed.<sup>4</sup> Items were presented in pseudo-random order such that consecutive trials differed in all three constituents. A different pseudo-randomly ordered set was generated for each participant. The left-to-right arrangement of the agent and patient was randomized across trials. Participants were presented with written instructions asking them to describe each scene using only gestures. They were further instructed to not speak and to provide as much information as they could. Participants were not cued to the kind of information they were expected to produce (i.e., there was no explicit mention of agent, patient and action).

Prior to each testing block, participants completed a passive exposure phase in which they were shown 10 randomly selected events simultaneously. The event type (generic or character) depicted in the scenes corresponded to the block event type. Participants were provided with written instructions requesting that they pay close attention to the details of each scene and think about what the scenes had in common and how they differed. The purpose of this exposure phase was to prompt participants to notice that scenes contained different interacting entities and actions, and thereby encourage them to express all three constituents without providing an explicit cue to this effect.

Previous studies where participants were explicitly instructed to provide three gestures were not successful in preventing omissions (e.g., Langus and Nespors (2010) found that approximately 40% of responses contained two gestures). Moreover, this approach is, in our view, too informative as to the nature of the task (Schouwstra, 2012). An alternative strategy would be to use a director-matcher design, where one participant communicates about an event to a partner whose task is to identify the target event from a set of options. While this approach has been used in some silent gesture studies (e.g., Christensen et al., 2016; Meir et al., 2017), we chose not to adopt it, since contrasting the target agent against alternatives within a trial might have the effect of increasing its contextual salience. In addition, the effect of introducing a communicative pressure is not clear. There is some suggestive evidence that it may influence the word orders people use (e.g., Hall et al., 2015) and drive them to be more consistent in their

<sup>4</sup>Given our definition of salience, i.e., how interesting or prominent an entity is, it is not clear how this would be affected by mixing event types. On the one hand, presenting generic agents with character agents might make them more salient because of the contrast. Alternatively, they may become even less salient when contrasted with the more interesting characters. Given this uncertainty, we felt it was more appropriate to present the two event types separately.

choices and less improvisational (Schouwstra et al., 2020, preprint).

Participants completed the experiment seated alone in a booth. Stimuli and instructions were presented on a computer screen and responses were video recorded using a webcam. The experiment was developed using PsychoPy (Peirce, 2009).

### 3.2.4 Coding

Individual gestures within a sequence were coded according to the intended referent – agent (A), patient (P) or action (V). On a number of trials, participants also encoded the result of the action, for example, by indicating an object falling. These result gestures (‘R’) were excluded from the analysis (see below). As is common practice in the gesture literature, multiple consecutive gestures with the same referent were coded as a single constituent (e.g., Hall et al., 2013; Meir et al., 2017; Kocab et al., 2018). All gestures in a given trial were coded as a single utterance except where the participant returned to a neutral position for more than 2s before resuming their response. In these cases, multiple responses were recorded.

Two data cleaning procedures were applied. First, contiguous repeated sequences were replaced with a single occurrence of the sequence. For example, the string APAPV was recoded as APV. This decision was motivated by the observation that some participants appeared to repeat sequences as a way of filling thinking time. There were 57 such trials (5.2% of 1104 trials). Second, result gestures were removed from the sequence, since we were primarily interested in the relative ordering of the agent, patient and action. This included ‘R’ gestures as well as ‘PR’ sequences where the patient was expressed earlier in the string. For example, APVPR was recoded as APV. This decision was based on the assumption that the participant reintroduced the patient to provide context for the result gesture. In total, 216 (19.6%) responses contained an ‘R’ gesture, of which 36 (3.3%) were part of a ‘PR’ sequence.

Following these data cleaning procedures, responses were recoded for analysis as follows: (1) Strings containing one or more simultaneously produced gestures were coded *simultaneous*. There were 50 (4.5%) such trials; (2) For trials in which multiple responses were recorded, we retained the first sequence that contained the action and at least one occurrence of a noun referent. There were two (< 1%) such trials – in both, the retained sequence was the first of two; (3) Following Hall et al. (2013), orders which individually accounted for less than 2% of trials were coded *rare*. A total of 88 (8.0%) trials fell into this category.<sup>5</sup>

### 3.2.5 Results

Nine trials were excluded from the analysis due to a technical error and a further seven were excluded because the participant did not provide a response. The resulting data set comprised 1104 trials (557 character-agent trials and 547 generic-agent trials).

In the sections that follow, we first investigate the overall proportion of responses coded as APV for each agent type. We then present a more detailed analysis of the results from each block.

---

<sup>5</sup>All data cleaning and analysis scripts are available at <https://osf.io/j46kq/>.

### 3.2.5.1 Proportion of responses coded as APV

Figure 3.2 shows the distribution of word orders for each agent type across both blocks. It represents 557 character-agent events (278 from block 1; 279 from block 2) and 547 generic-agent events (280 from block 1; 267 from block 2).

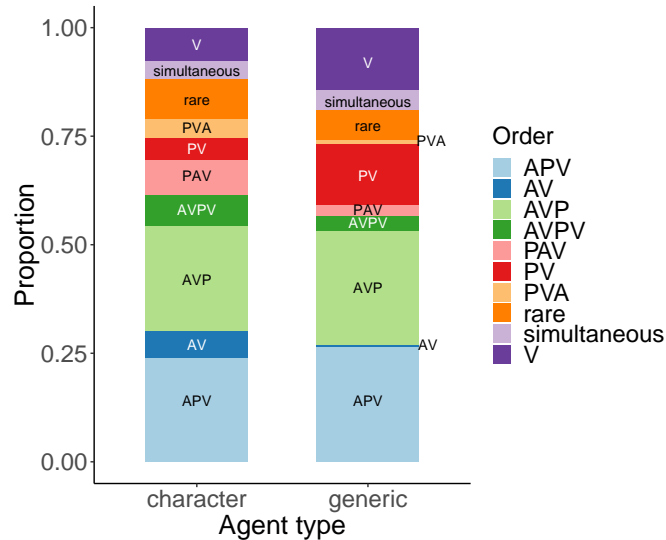


Figure 3.2: Word orders used to describe character-agent events and generic-agent events across both blocks. We found no evidence that APV was the preferred order. APV and AVP were the most frequent orders and occurred in roughly equal proportions.

Contrary to our expectations, we found no evidence that APV was the preferred order, although it was one of the most commonly used. This finding is at odds with previous studies; we return to this in §3.2.6. Participants produced APV on 25.3% of trials overall (24.1% of character-agent trials; 26.5% generic-agent). In addition, we found that AVP was used exactly as often as APV, accounting for 25.3% of trials overall (24.2% character-agent; 26.3% generic-agent).

The equal preference for APV and AVP was reflected at the participant level. Out of 28 participants, eight used AVP as their most common order overall across both blocks and nine used APV (seven as their most common order and two jointly with one other order). We found a similarly even distribution across agent types (see Table 3.1).

Table 3.1: Number of participants who used APV or AVP as their most common order overall across blocks, and by agent type.

	Overall	By agent type	
		Character	Generic
APV	9	7	9
AVP	8	9	8

### 3.2.5.2 Block 1

Figure 3.3 (left) shows the distribution of word orders in block 1 plotted by agent type. The results are summarized in Table 3.2. As we noted above, the overall proportion of APV trials

was unexpectedly low. In block 1, the proportion was lower still at 15.8% (13.3% of character-agent trials; 18.2% of generic-agent trials). Looking at individual responses, only five of the 28 participants used APV most often, or as often as another order in the first block (Figure 3.4, top row; counts for each condition and block are summarized in Table 3.3).

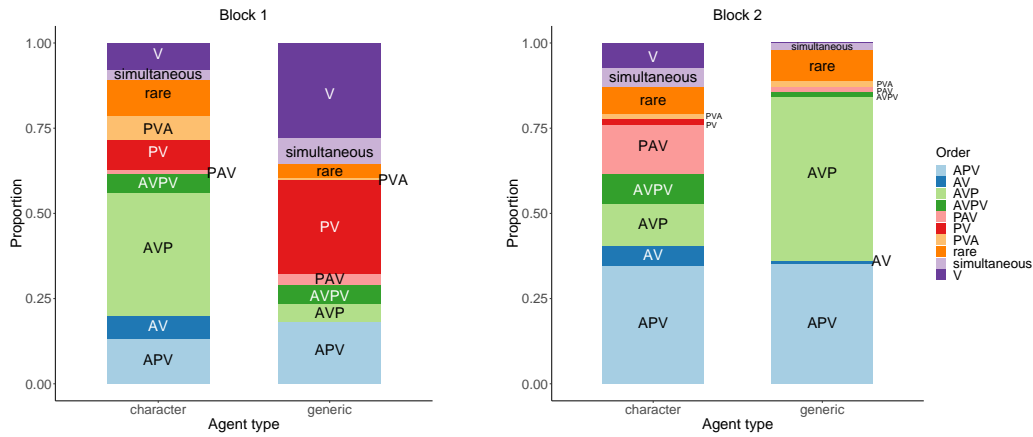


Figure 3.3: The distribution of word orders by agent type in block 1 (left) and block 2 (right). These plots demonstrate that structural choices were sensitive to agent type.

Table 3.2: Word order proportions in each condition and block. Cells in normal font indicate character-agent trials; cells in boldface show generic-agent trials.

	character-first		generic-first	
	Block 1	Block 2	Block 1	Block 2
APV	0.133	<b>0.352</b>	<b>0.182</b>	0.348
AV	0.065	<b>0.007</b>	<b>0.000</b>	0.057
AVP	0.363	<b>0.483</b>	<b>0.054</b>	0.122
AVPV	0.054	<b>0.015</b>	<b>0.054</b>	0.090
PAV	0.014	<b>0.015</b>	<b>0.032</b>	0.143
PV	0.086	<b>0.000</b>	<b>0.279</b>	0.018
PVA	0.072	<b>0.015</b>	<b>0.004</b>	0.014
V	0.079	<b>0.000</b>	<b>0.279</b>	0.072
simultaneous	0.029	<b>0.019</b>	<b>0.075</b>	0.057
rare	0.104	<b>0.094</b>	<b>0.043</b>	0.079

Table 3.3: Number of participants in each condition and block who produced predominantly APV, AVP, or incomplete orders. Cells in normal font indicate character-agent trials; cells in boldface show generic-agent trials.

	character-first		generic-first	
	Block 1	Block 2	Block 1	Block 2
APV	2	<b>6</b>	<b>3</b>	5
AVP	7	<b>7</b>	<b>1</b>	2
Incomplete	3	<b>0</b>	<b>8</b>	2

We also found that PAV was rare for both types of event (1.4% of character-agent events; 3.2% of generic-agent events). Recall, however, that our main prediction was that participants would express generic agents before the patient less often than character agents. To investigate this, independent of the positioning of the action, we analysed all trials for which it was possible to determine the relative positioning of the agent and patient. Trials in which one or both noun

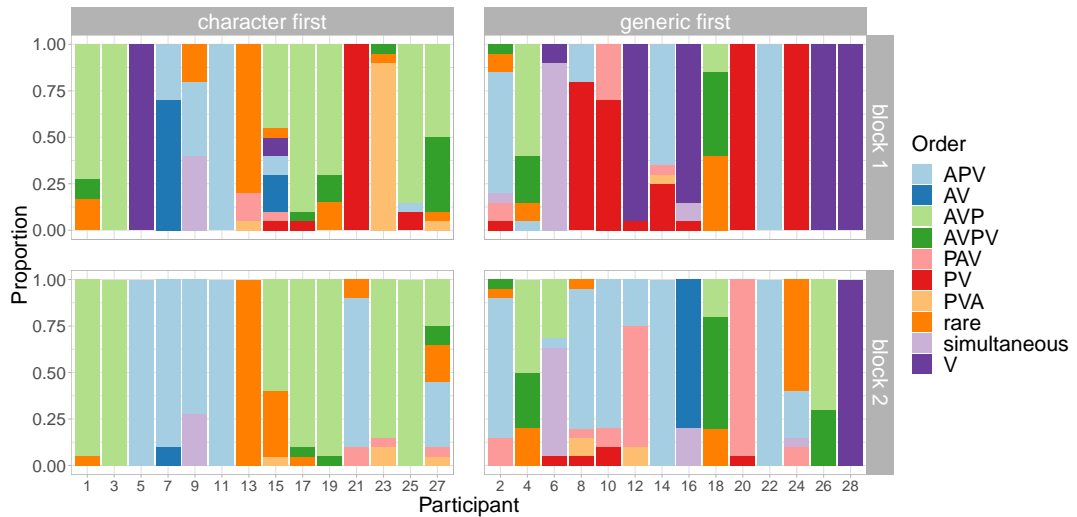


Figure 3.4: The proportion of orders used by each participant grouped by condition and block. A minority of participants used predominantly APV in block 1. Participants showed a strong tendency to omit noun referents, particularly agents, in block 1 of the generic-first condition. These results also demonstrate a tendency for individuals to persevere block 1 ordering preferences into block 2.

referents were omitted or where they were expressed simultaneously were therefore excluded. On some trials, participants expressed the agent or patient more than once. We categorized such responses according to the position of the first occurrence of each constituent. The resulting data set included 317 trials (213 character-agent and 104 generic-agent trials).

For both types of agent, participants expressed the agent before the patient in a majority of trials (80.8% of character-agent events; 87.5% of generic-agent events; see Figure 3.5). A mixed effects logistic regression analysis<sup>6</sup> found no significant difference between agent types (Table 3.4).

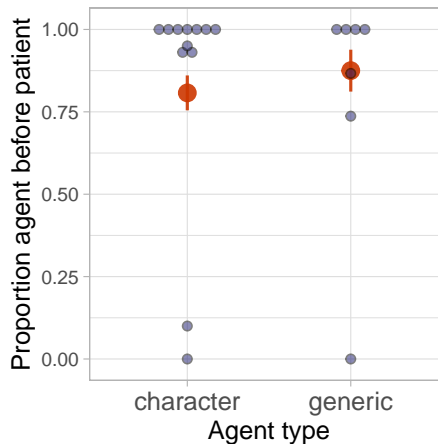


Figure 3.5: The proportion of block 1 trials in which the agent was expressed before the patient. Small circles represent the proportion of agent-before-patient responses for each participant. Large circles represent the overall means for each agent type (error bars show 95% CIs). For both agent types, participants expressed the agent before the patient in a majority of trials.

Nevertheless, Figure 3.3 clearly suggests that responses were indeed affected by the agent.

<sup>6</sup>All analyses in this study were performed using the R programming language (R Core Team, 2017) and the *lme4* package (Bates et al., 2015a).

Table 3.4: Mixed effects logistic regression analysis of agent-before-patient responses in block 1.

Predictor <sup>a</sup>	$\beta$	$SE$	$p$
Intercept	8.276	2.593	0.001**
agent_type	-0.503	2.993	0.866
orientation <sup>b</sup>	-0.707	2.785	0.800
agent_type:orientation	-0.736	1.991	0.712

Model:  $a\_before\_p \sim agent\_type*orientation + (1 + orientation | participant) + (1 | item)$

(a) Binary inputs were deviation coded in all models in this study. The reported random effects structures represent the maximal structures for which models converged without warnings.

(b) Orientation was a binary flag indicating if the agent appeared on the left (coded -0.5) or the right (coded 0.5) of the image.

Events depicting character agents were predominantly described using AVP (36.3% of responses). In contrast, this order was rare for generic-agent events (5.4%). Looking at individual responses, while seven people used AVP most often to describe character-agent events, only one used this order most frequently to describe generic-agent events (see Table 3.3).

Descriptions of generic-agent events were characterized by a high proportion of incomplete orders, that is, orders in which one or more constituents were omitted. Across all trials, 62.9% of generic-agent trials in block 1 elicited an incomplete order compared with 23.4% of character-agent trials. PV and V accounted for the majority of incomplete descriptions of generic-agent events, indicating that generic-agent omissions were more common than patient omissions.

Agent and patient omissions for each agent type are plotted in Figure 3.6. A mixed effects logistic regression analysis confirmed that participants were significantly more likely to omit generic agents (62.1% of trials) compared with character agents (16.5%) (see Table 3.5). We suggest that this reflects the lower salience of generic agents compared with character agents.

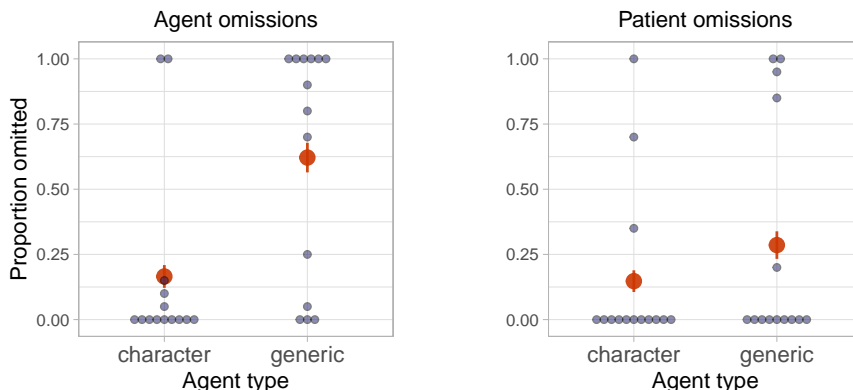


Figure 3.6: The proportion of agent and patient omissions in block 1. Blue circles indicate proportions for each participant. Large circles show the means for each agent type (error bars show 95% CIs). Participants were significantly more likely to omit generic agents compared with character agents.

Patient omissions were also more frequent for events depicting generic agents (28.6%) compared with character-agent events (14.8%). However, this difference was not significant (Table 3.6). The model additionally showed that there was a significant interaction between agent type and image orientation. A possible explanation for this finding is that left-positioned patients are more prominent than patients positioned to the right of the agent (e.g., Esaulova et al., 2019). This effect may have been greater when the agent had low salience (i.e., a generic agent) and was therefore less likely to compete with the patient for the viewer’s attention.

Table 3.5: Mixed effects logistic regression analysis of agent omissions in block 1.

Predictor	$\beta$	$SE$	$p$
Intercept	-0.727	1.400	0.604
agent_type	20.714	6.470	0.001**
orientation	1.025	0.969	0.290
agent_type:orientation	0.200	1.477	0.892

Model: a\_omitted  $\sim$  agent\_type\*orientation + (1 | participant) + (1 | item)

Table 3.6: Mixed effects logistic regression analysis of patient omissions in block 1.

Predictor	$\beta$	$SE$	$p$
Intercept	-11.953	2.128	<0.001***
agent_type	1.183	3.173	0.709
orientation	0.463	0.588	0.431
agent_type:orientation	-3.481	1.180	0.003**

Model: p\_omitted  $\sim$  agent\_type\*orientation + (1 | participant)

One possible explanation for why AVP responses were rare for generic-agent events is that this was a straightforward consequence of participants omitting the agent. However, a more detailed analysis of the data suggests that this is unlikely to be the case. Specifically, we found that participants were significantly more likely to express the patient before the action when describing generic-agent events (78.5% of trials) compared with character-agent events (46.7%). Details of the statistical analysis are provided in Table A.2 (Appendix A.2). This demonstrates that there was not only a significant difference in the rate of agent omissions, but also in the relative ordering of expressed constituents. In addition, as we will see in §3.3, the results of the computational model indicate that the majority of incomplete descriptions of generic-agent events derived from an underlying APV order rather than AVP.

### 3.2.5.2.1 Block 1 discussion

We can draw two general conclusions from the analysis of block 1 responses. First, as in the combined analysis, we found no evidence to support the claim that APV is the preferred order for describing events involving an animate agent and an inanimate patient (cf. Goldin-Meadow et al., 2008). Second, although the results did not support the hypothesis that generic-agent events would elicit fewer agent-before-patient responses compared with character-agent events, we found that structural choices in block 1 were clearly conditioned on agent type. AVP was the preferred order for character-agent events, but was rare for generic-agent events. For this type of event, participants showed a strong tendency to omit the agent, which we attribute to their lower salience. Among trials in which at least one noun referent was expressed, PV was the most common order for generic-agent events.

### 3.2.5.3 Block 2

In block 2, we saw an increase in the proportion of APV responses for both types of event. This order accounted for 35.0% of trials overall in block 2 (15.8% in block 1): 34.8% of character-agent trials and 35.2% of generic-agent trials (see Table 3.2). Although this increase did not reach

statistical significance (see Table 3.7), it is nevertheless notable that the number of participants using APV as their most common order also increased in both conditions (Figure 3.4, bottom row; see also Table 3.3). Interestingly, all of the participants who predominantly used APV in block 1 continued to use this as their most common order in block 2. This suggests that responses in the second block were influenced by the pattern established in the first.

Table 3.7: Mixed effects logistic regression analysis of APV responses across blocks.

Predictor	$\beta$	$SE$	$p$
Intercept	-5.215	2.415	0.031*
agent_type	-0.519	1.261	0.681
block	1.436	3.088	0.642
orientation	0.052	0.751	0.945
agent_type:block	-2.234	5.478	0.683
agent_type:orientation	0.211	0.588	0.720
block:orientation	-0.103	0.725	0.888
agent_type:block:orientation	-0.694	1.396	0.619

Model:  $is\_apv \sim agent\_type*block*orientation + (1 + block + orientation | participant)$

The use of AVP across blocks provides further evidence for this self-priming effect. As in block 1, participants in the character-first condition continued to use predominantly AVP in block 2. Consequently, this was the preferred order for describing character-agent events in block 1 (36.3%) and generic-agent events in block 2 (48.3%). Participants in the generic-first condition continued to use AVP in a minority of trials (12.2%). Of the eight participants who predominantly produced AVP in block 1, all but one (participant #27 in the character-first condition; see Figure 3.4) continued this pattern into block 2.

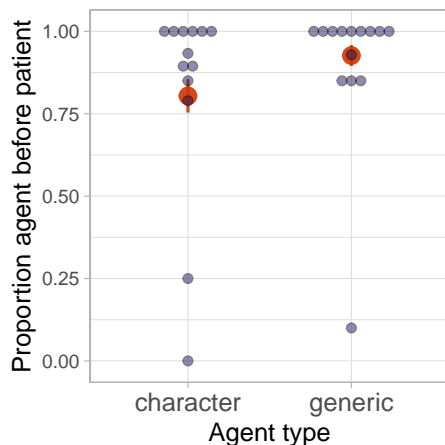


Figure 3.7: The proportion of block 2 trials in which the agent was expressed before the patient. Small circles represent the proportion of agent-before-patient responses for each participant. Large circles represent the overall means for each agent type (error bars show 95% CIs). For both agent types, participants expressed the agent before the patient in a majority of trials.

As in the first block, participants in block 2 typically expressed the agent before the patient (80.4% of 235 character-agent trials; 92.7% of 259 generic-agent trials; see Figure 3.7). Again, a mixed effects logistic regression analysis confirmed that there was no significant difference between agent types (Table 3.8).

Table 3.8: Mixed effects logistic regression analysis of agent-before-patient responses in block 2.

Predictor	$\beta$	$SE$	$p$
Intercept	4.719	1.416	<0.001***
agent_type	1.894	1.793	0.291
orientation	-0.707	0.447	0.113
agent_type:orientation	1.044	0.888	0.240

Model: a\_before\_p  $\sim$  agent\_type\*orientation + (1 | participant) + (1 | item)

The tendency to omit constituents was considerably lower in block 2 compared with the first block. Figure 3.8 shows the proportion of trials for each agent type in which the agent (left) or patient (right) was omitted. Interestingly, the two participants who produced predominantly incomplete orders in block 2 had also done so in block 1 (see Figure 3.4). This observation provides further evidence that participants tended to persevere the pattern of responses established in block 1 into block 2.

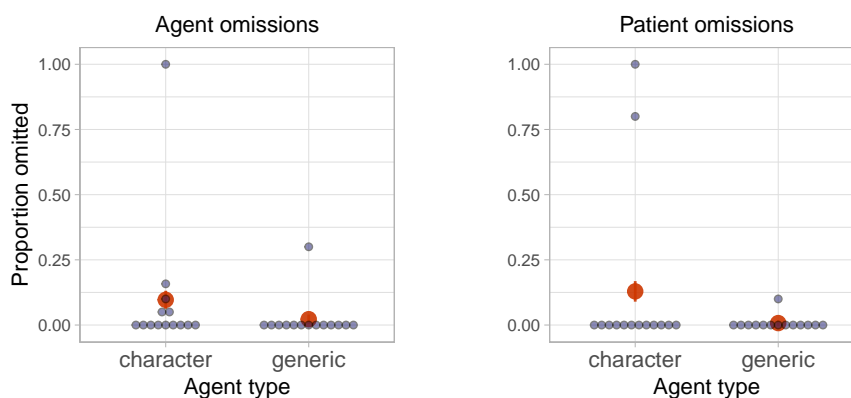


Figure 3.8: The proportion of agent and patient omissions for each agent type in block 2. Blue circles indicate the proportions for each participant. The large circles show the means for each agent type (error bars represent 95% CIs).

An analysis of omissions across blocks confirmed that participants were significantly less likely to omit agents in block 2 compared with block 1 (Table 3.9). This may reflect a novelty effect: participants who had not attended to agents in the first block, or who had not considered them sufficiently worthy of mention, may have been more likely to attend to and mention agents in block 2 because they differed from those in block 1. We also found a significant interaction between block and agent type, reflecting the fact that the rate of generic-agent omissions dropped markedly from block 1 to block 2, whereas character-agent omissions were already infrequent in both blocks.

As in block 1, a comparison of patient omissions across blocks revealed a significant interaction between agent type and image orientation (Table 3.10). There were no other significant effects or interactions.

Table 3.9: Mixed effects logistic regression analysis of agent omissions across blocks.

Predictor	$\beta$	$SE$	$p$
Intercept	-5.794	1.792	0.001**
agent_type	6.181	4.036	0.126
block	-8.250	3.223	0.011*
orientation	0.582	0.476	0.222
agent_type:orientation	1.139	0.953	0.232
agent_type:block	-20.684	7.305	0.005**
block:orientation	-0.416	0.928	0.654
agent_type:block:orientation	2.178	1.869	0.244

Model: a\_omitted  $\sim$  agent\_type\*block\*orientation + (1 + block | participant) + (1 | item)

Table 3.10: Mixed effects logistic regression analysis of patient omissions across blocks.

Predictor	$\beta$	$SE$	$p$
Intercept	-15.560	5.740	0.007**
agent_type	-0.254	2.849	0.929
block	-6.679	11.650	0.567
orientation	0.490	0.554	0.376
agent_type:orientation	-2.163	0.950	0.023*
agent_type:block	-2.837	7.578	0.708
block:orientation	0.000	1.05	1.000

Model: p\_omitted  $\sim$  agent\_type\*block\*orientation + (1 + block | participant)

### 3.2.5.3.1 Block 2 discussion

The analysis of block 2 word orders revealed evidence of a self-priming effect whereby participants tended to continue using the same overall pattern of responses that they had used in block 1. The second key result was that the tendency found in block 1 to omit generic agents was not seen in the second block.

## 3.2.6 APV responses: a comparison with previous studies

In this section, we turn our focus to the unexpectedly low occurrence of APV in our data and compare our findings with two previous silent gesture studies, namely, Goldin-Meadow et al. (2008) and Gibson et al. (2013).

Goldin-Meadow et al. (2008) found that APV (glossed as ArPA) was the preferred order for describing transitive events based on an analysis of gesture strings categorized as consistent with this order. Crucially, this category included the incomplete orders AV and PV.<sup>7</sup> Coding our own results according to this approach,<sup>8</sup> we found that orders consistent with APV in block 1 accounted for a majority of generic-agent trials (63.9%) and just under one third of

<sup>7</sup>Goldin-Meadow et al. (2008) reported that participants produced 501 gesture strings containing two elements, compared with only 113 complete strings.

<sup>8</sup>To be consistent with the results reported in Goldin-Meadow et al. (2008) we calculated all percentages based on the set of trials containing at least one occurrence of the agent or patient. Trials in which only the action was expressed were therefore not included.

character-agent trials (30.9%). In block 2, 36.0% of generic-agent trials were consistent with APV, and 45.6% of character-agent trials. These results are plotted in Figure 3.9(a).

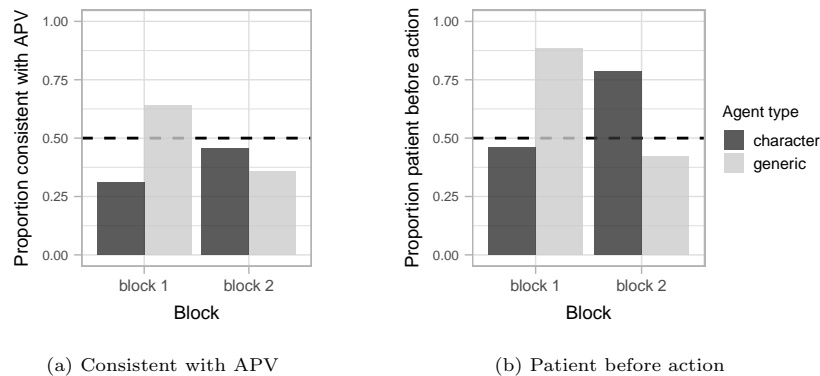


Figure 3.9: The proportion of responses coded as (a) consistent with APV, following Goldin-Meadow et al. (2008); and (b) patient-before-action, following Gibson et al. (2013). A majority of generic-agent trials in block 1 were categorized as APV-like, according to the coding strategies adopted in these two studies.

Our results coded according to an alternative strategy used by Gibson et al. (2013) are plotted in Figure 3.9(b). Under this coding scheme, all trials in which the patient was expressed before the action were coded as APV-like (glossed as SOV). The authors reported that this was the majority order produced by English-speaking participants. However, closer inspection of their data (reported in Futrell et al., 2015) shows that APV accounted for only 31.6% of analysed trials,<sup>9</sup> followed closely by AVP (28.7%). The proportion of responses coded as PV, and included in the APV-like category, was 23.6%. In our own data, block 1 responses in which the patient was expressed before the action comprised the majority of generic-agent trials (88.6%)<sup>10</sup> and a large proportion of character-agent trials (46.1%). In block 2, we found the reverse pattern: patient-before-action response accounted for a majority of character-agent trials (78.8%) and a substantial proportion of generic-agent trials (42.3%).

These analyses show that our results are more closely in line with findings from previous studies than our original analysis suggests. In particular, for generic-agent trials in block 1, orders classified as consistent with APV in other studies comprised the majority of trials. However, it's not clear that we should therefore interpret our findings as providing evidence for an APV bias for this type of event. The coding strategy adopted by Gibson et al. (2013), which equates all patient-before-action orders with APV, is not well motivated. In particular, it leads to the odd conclusion that both PAV and PVA, in addition to PV, are APV-like.

What about categorizing AV and PV as consistent with APV, as in Goldin-Meadow et al. (2008)? This approach is based on the assumption that incomplete orders represent surface manifestations of underlying predicate frames from which constituents have been dropped (Goldin-Meadow, 1985; Goldin-Meadow et al., 2009). For transitive events, these frames consist of an agent, patient, and action. Further, the strategy assumes that AV and PV derive from an underlying APV sequence where either the agent has been dropped from initial position (resulting in PV), or the patient from second position (resulting in AV). In other words, it assumes a priori that the agent would have preceded the patient had both constituents been expressed.

<sup>9</sup>These percentages are based on data collected from English-speaking participants and include trials involving an animate agent and inanimate patient. Futrell et al. (2015) do not report trials in which the patient was not expressed, or where the patient and/or action were expressed more than once.

<sup>10</sup>Consistent with Gibson et al. (2013), the figures reported in this section exclude trials in which the patient was not expressed, or the where the patient and/or action were expressed more than once.

Again, it is unclear whether this assumption is warranted. For this reason, in §3.3 we present details of an alternative, computational method for dealing with missing constituents that does not require these a priori assumptions to be made.

### 3.2.7 Discussion

Based on findings from previous silent gesture studies, we predicted that participants would predominantly describe events using APV. We further predicted that participants would express generic agents before patients less often than character agents, resulting in fewer APV responses and correspondingly more PAV. Our results were not consistent with either prediction. Overall, we found that APV constituted around one quarter of responses, roughly equal to the proportion of responses coded as AVP. In addition, we found no evidence that the tendency to express the agent before the patient was conditioned on agent type.

Nevertheless, our results did provide evidence that participants were sensitive to the properties of the event agent. In block 1, AVP was the most common order for describing character-agent events. For generic-agent events, on the other hand, AVP was rare and participants showed a strong tendency to omit constituents, particularly the agent. Another key finding was that results from block 2 pointed to the presence of a self-priming effect whereby the preferred word orders established in block 1 perseverated into block 2.

Finally in this section, we presented a reanalysis of our data where we found that our results were more aligned with previous findings, particularly for generic-agent events in block 1, if we recoded our data according to the strategies employed in two previous silent gesture studies. Nevertheless, these strategies are problematic: equating *all* patient-before-action responses with APV (Gibson et al., 2013) does not seem well motivated; while equating AV and PV with APV (Goldin-Meadow et al., 2008; Hall et al., 2013) relies on a priori assumptions about the relative positioning of dropped constituents.

We propose an alternative to this latter approach, making the weaker assumption that incomplete orders may in principle derive from *any* consistently ordered complete sequence, that is, sequences exhibiting the same relative ordering of expressed constituents. For example, PV may derive from dropping the agent from an underlying APV, PAV, or PVA sequence. In the next section, we describe details of a computational model that exploits this assumption to infer the distribution of complete orders that participants would have produced had they expressed all three constituents on every trial. The model also provides an estimate of the distribution of complete orders from which each incomplete order derived, for example, the proportions of PV responses that derived from APV, PAV, and PVA.

## 3.3 Modelling the underlying word order distribution

The results of the present study demonstrate that improvised, gestured descriptions of events can be messy. That is, rather than consistently producing three-element sequences from which the constituent order can be unambiguously determined, participants often repeat elements, or omit them altogether. Although some studies in the silent gesture literature excluded incomplete orders from their analysis (e.g., Schouwstra and de Swart, 2014), this approach was not appropriate in the present study, for two reasons. First, the proportion of incomplete orders was relatively high, which may tell us something important about how people responded to the

set of stimuli. Second, and more importantly, we saw in §3.2.5.2 that the tendency to omit event constituents, particularly the agent, was greater in block 1 when participants described generic-agent events compared with character-agent events. Excluding incomplete orders would obscure this conditioning on agent type.

As discussed in §3.2.6, Goldin-Meadow et al. (2008) attempted to deal with incomplete orders by binning them with one of the three-element orders. However, this approach is problematic since it makes a priori assumptions about the relative positioning of dropped constituents. In seeking to address the question of what factors influence the relative ordering of the three basic constituents, how else might we deal with situations where the rate of omissions is high? One obvious response to this question is that we should endeavour to encourage participants to express all three constituents. The passive exposure phase used in the present study was not successful in achieving this. As we noted in §3.2.3, alternative approaches taken in the silent gesture literature have been similarly unsuccessful and/or may not be appropriate under all circumstances.

In this section, we present details of a computational method for dealing with missing constituents that not only avoids these methodological problems, but, more importantly, does not require a priori assumptions about the positioning of omitted constituents. The model described here infers an *underlying distribution* across the six basic word orders (APV, AVP, PAV, PVA, VAP, VPA) based on the empirically derived distribution across the set of 11 complete and incomplete orders (APV, AVP, PAV, PVA, VAP, VPA, AV, PV, VA, VP, V)<sup>11</sup> based on the assumption that incomplete sequences may derive from any consistently ordered complete sequence.

Details of the model are provided below. In brief, it proceeds by sampling candidate underlying distributions from the space of possible distributions. Each candidate distribution is transformed into a *surface distribution* using the probabilities of omitting the agent, patient, or both. The inferred underlying distribution corresponds to the surface distribution that best fits the empirically derived data. For each incomplete order, the model also provides an estimate of the proportion that derives from each complete order.

### 3.3.1 Generating the surface distribution

For a given word order in a candidate underlying distribution, the model generates a sub-distribution consisting of the proportion of the original order and the proportions of each transformed order. For example, if the original order is APV, the resulting sub-distribution represents the proportions of APV, AV, PV and V sequences. The proportion of each transformed order is calculated using the set of omission probabilities, that is, the probabilities that the agent, patient, or both are omitted. More formally, the proportion of each transformed order  $t_{-c}$  is given by:

$$P(t_{-c}) = P(w) \times O(c),$$

where  $t_{-c}$  represents a transformed order that excludes constituent(s)  $c$ ,  $P(w)$  is the proportion of the original word order  $w$  within the candidate underlying distribution, and  $O(c)$  is the omission probability for constituent(s)  $c$ . Following the transformation procedure, the remaining proportion  $P_{remaining}(w)$  of the original order is calculated by subtracting each  $P(t_{-c})$  from

---

<sup>11</sup>Participants in our study expressed the action on every trial, hence, all incomplete orders include this constituent.

$P(w)$  such that  $P(w) = P_{remaining}(w) + \sum_{c \in C} P(t_{-c})$ , where  $C$  is the set of constituent(s) that can be omitted (i.e., agent, patient, and both).

As an example, suppose the proportion of APV sequences in the candidate distribution is  $P(\text{APV}) = 0.5$ . Suppose further that the probability of an agent omission  $O(A) = 0.2$ , the probability of a patient omission  $O(P) = 0.1$ , and the probability that both are omitted  $O(A+P) = 0.1$ . This means that 20% of APV responses are transformed to PV, 10% are transformed to AV, and 10% are transformed to V. The resulting sub-distribution across APV, AV, PV, and V then contains  $P_{remaining}(\text{APV}) = 0.3$ ,  $P(\text{AV}) = 0.05$ ,  $P(\text{PV}) = 0.1$ , and  $P(\text{V}) = 0.05$ , which sum to 0.5. This procedure is repeated for each of the six basic word orders and the resulting sub-distributions combined to give a transformed surface probability distribution across the 11 complete and incomplete orders. The model also records the proportion of each incomplete order within the surface distribution that derived from each complete order, for example, the proportions of PV that derived from APV, PAV, and PVA.

### 3.3.2 Inferring the underlying distribution

We used a least-squares method to determine the surface distribution that best fit the observed data. The model employed a basin-hopping global optimization algorithm in the Python SciPy optimize package.<sup>12</sup> The procedure repeatedly sampled candidate distributions across the six basic orders which were then transformed to a candidate surface distribution according to the procedure described above. The objective function (i.e., the function whose output was to be minimized) represented the squared residuals between the observed distribution and the candidate surface distribution. The inferred underlying distribution corresponded to the candidate surface distribution that minimized the objective function. Below, we present results from two model parametrizations. In one, the model inferred the underlying distribution and received the set of omission probabilities  $O(A)$ ,  $O(P)$ ,  $O(A + P)$  as fixed parameters. The probabilities were determined empirically, one set for each agent type. In the other model, the omission probabilities were free parameters estimated by the model.

### 3.3.3 Results

For this analysis, we focused on block 1 responses, since the rate of omissions in the second block was very low. The values specifying the initial estimate of the underlying distribution (used as a starting point by the model) were  $P(\text{APV}) = P(\text{AVP}) = P(\text{PAV}) = P(\text{PVA}) = 0.2$ , and  $P(\text{VAP}) = P(\text{VPA}) = 0.1$ . Where the omission probabilities were estimated by the model, initial values were  $O(A) = O(P) = O(A+P) = 0.2$ .

Table 3.11 shows the results of the model for generic-agent and character-agent events. We calculated the Akaike information criterion (AIC) to compare the two model parametrizations. AIC estimates the goodness of fit of a model and adjusts for the number of estimated parameters to reduce the risk of overfitting (Gelman and Hill, 2006, pp. 524–525). For generic-agent events, the free omission probabilities model provided a better fit to the data. For character-agent events, the fixed omission probabilities model resulted in a better fit.

Figure 3.10(a) shows bootstrap mean proportions for each word order in the observed and best-fit surface distributions for generic-agent events. These data were generated by drawing 10,000 samples of  $n = 232$  trials, where the probability of drawing a particular word order

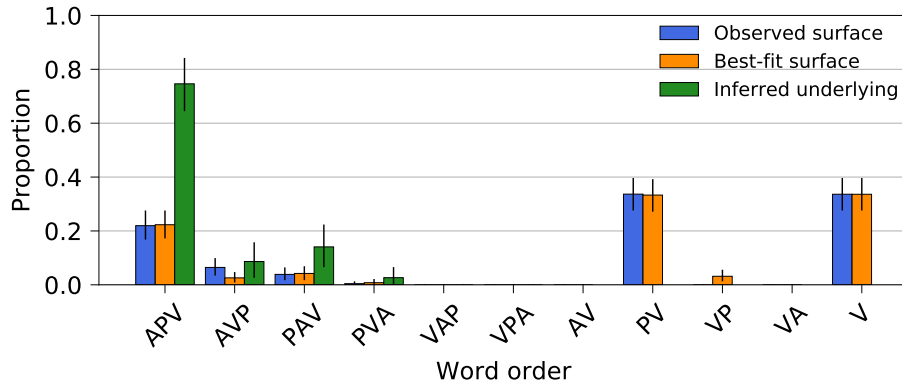
---

<sup>12</sup>SciPy package version 1.4.1.

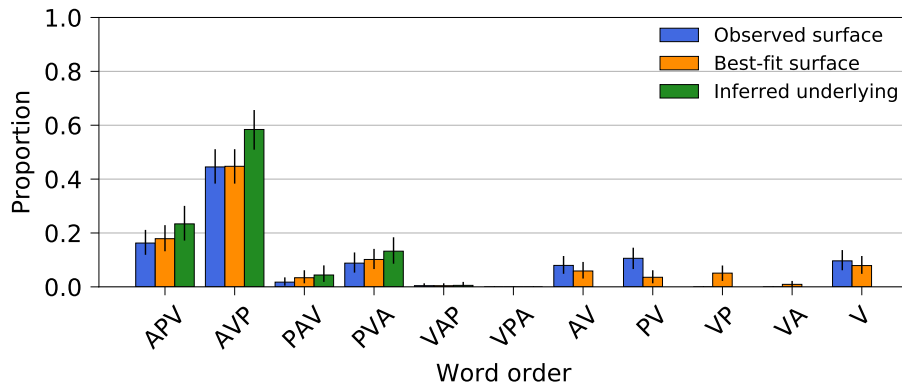
Table 3.11: Model output for each agent type using two model parametrizations.

Agent type	Omission probabilities	Objective function	AIC	$O(A)$	$O(P)$	$O(A+P)$
Generic	Fixed	0.0077	70.23	0.3429	0.0071	0.2786
	Free	0.0025	64.12	0.3645	$\sim 0$	0.3361
Character	Fixed	0.0091	95.59	0.0863	0.0683	0.0791
	Free	0.0081	155.43	0.0814	0.0893	0.0970

was given by the empirically derived proportion in the observed distribution and estimated proportion in the best-fit surface distribution, respectively. The sample size  $n$  corresponded to the number of trials in the observed data. The plot also shows data simulated from the inferred underlying distribution ( $n = 76$ ). Simulation results for character-agent events are plotted in Figure 3.10(b) (surface distribution:  $n = 227$ ; underlying distribution:  $n = 163$ ). The data are provided in Appendix A.3. For both types of event the proportion of each word order in the best-fit surface distribution closely matches the observed distribution, with some exceptions that we return to below.



(a) Generic-agent events



(b) Character-agent events

Figure 3.10: Bootstrap mean proportions of each word order in the observed distribution across surface orders, the best-fit surface distribution, and the inferred underlying distribution for (a) generic-agent events and (b) character-agent events. Error bars represent 95% confidence intervals.

Consistent with our analysis of block 1 responses, the model indicated that AVP was the preferred order for character-agent events (estimated mean=0.584, 95% CI=[0.509, 0.656]),

but was much less common for generic-agent events (estimated mean=0.087, 95% CI=[0.026, 0.158]). The model also suggested a strong APV preference for generic-agent events (estimated mean=0.746, 95% CI=[0.645, 0.842]), but not for character-agent events (estimated mean=0.234, 95% CI=[0.172, 0.301]).

Table 3.12 shows the proportion of each incomplete order that derived from each of the six complete orders. For generic-agent events, the model estimated that a majority of AV (0.766) and PV (0.817) derived from an underlying APV order. This finding may suggest that assuming a priori that AV and PV originate as APV, as in Goldin-Meadow et al. (2008), is a sound approach. However, model estimates based on data from character-agent events show that this assumption does not hold in general. Here, just over half of PV (0.570) derived from APV, while around one third originated as PVA (0.323); fewer than one third of AV (0.271) responses came from APV, while the majority (0.678) derived from AVP.

In fact, these findings are not surprising and are a direct consequence of the fact that the model described here transforms all orders according to the same set of omission probabilities. Thus, PV responses, for example, will be distributed across APV, PAV, and PVA according to the relative frequency of these orders in the underlying distribution. To make this more explicit, if the underlying distribution contains equal proportions of these three orders, then the transformed PV responses will be distributed evenly across them.

Table 3.12: Proportion of each incomplete order that derived from each complete order.

		APV	AVP	PAV	PVA	VAP	VPA
Generic <sup>a</sup>	AV	0.766	0.089	0.145	0	0	0
	PV	0.817	0	0.156	0.029	0	0
	V	0.746	0.087	0.141	0.026	0	0
	VA	0	0	0	1.0	0	0
	VP	0	1.0	0	0	0	0
Character <sup>b</sup>	AV	0.271	0.678	0.051	0	0	0
	PV	0.570	0	0.107	0.323	0	0
	V	0.234	0.584	0.044	0.133	0.006	0
	VA	0	0	0	0.960	0.040	0
	VP	0	0.991	0	0	0.009	0

<sup>a</sup>Omission probabilities free parameters of the model.

<sup>b</sup>Omission probabilities fixed parameters of the model.

We now return to the divergences between the best-fit surface distributions and the observed distributions. For generic-agent events, the model underestimated the proportion of AVP responses (observed: mean=0.065, 95% CI=[0.034, 0.099]; model estimate: mean=0.026, 95% CI=[0.009, 0.047]) and overestimated the proportion of VP (observed: zero occurrences; model estimate: mean=0.032, 95% CI=[0.013, 0.056]). For character-agent events, the model underestimated PV (observed: mean=0.106, 95% CI=[0.066, 0.145]; model estimate: mean=0.036, 95% CI=[0.013, 0.062]), and again overestimated VP (observed: zero occurrences; model estimate: mean=0.051, 95% CI=[0.022, 0.079]). These findings indicate that the model may not fully capture the mechanism by which surface orders are generated. It is notable, for example, that while AV and PV are frequently attested in studies of homesign systems (Goldin-Meadow et al., 2009) and emerging sign languages (e.g. Sandler et al., 2005; Padden et al., 2010), action-initial constructions are much rarer. One possibility is that the probability of dropping a constituent may vary between word orders. For example, there may be a smaller probability of dropping

the agent from an underlying AVP sequence compared with APV.<sup>13</sup>

### 3.3.4 Discussion

We developed a computational model that infers an underlying distribution across the six basic word orders by assuming that incomplete orders derive from complete orders where one or more noun referents have been dropped. The set of probabilities of dropping the agent, patient, or both was either provided to the model as fixed parameters, or was estimated by the model. The model also provided an estimate of the proportions of each incomplete order that derived from each complete order.

We used this model to infer the word orders that participants in block 1 would have produced had they expressed all three constituents on every trial. Consistent with the analysis of the empirically derived word order distributions, the model showed that AVP was the preferred order for describing character-agent events. Our model also indicated that for generic-agent events, the majority of incomplete orders derived from APV, resulting in an overall preference for this order in the inferred underlying distribution. However, we also saw that one cannot make a priori assumptions about the source of incomplete orders. In particular, under the assumptions of the model described here, it is not generally the case that AV and PV derive from an underlying APV order.

## 3.4 General Discussion

Goldin-Meadow et al. (2008) proposed that APV (usually glossed as SOV in the silent gesture literature) is cognitively more basic than other word orders and is the default order adopted by all emerging communication systems. However, a growing body of literature has challenged this conclusion. Meir et al. (2017), for example, have argued that APV is no more cognitively basic than other orders, but reflects the relative salience of interacting entities: humans, which are typically agents, are more salient than inanimate objects and so tend to be mentioned first.

We had two main aims in the the present study. First, we sought to replicate the APV bias in a silent gesture task. Second, we set out to explore the role of salience in more detail. Specifically, we attempted to manipulate the salience of the agent in an event to investigate the hypothesis that the tendency to express the agent before the patient would be reduced for less salient agents, resulting in correspondingly more PAV responses.

Contrary to our first expectation, participants did not produce predominantly APV. In addition, we found no evidence that agent type influenced the relative ordering of the agent and patient. Nevertheless, we did find clear evidence of word order conditioning on agent type. In block 1, participants typically described character-agent events using AVP. However, this order was rare for generic-agent events. In addition, we found that participants in block 1 showed a strong tendency to omit generic agents from their descriptions. Of the orders in which at least one noun referent was expressed, PV occurred most frequently for this type of event. Responses in block 2 pointed to a strong self-priming effect whereby participants continued to produce the same ordering preferences they had established in block 1.

---

<sup>13</sup>Another possibility is that a two-gesture sequences may be further transformed, for example, from VP to PV, to avoid expressing the action in initial position. We explored this possibility by allowing the model to estimate the probability of reversing the order of V-initial incomplete sequences. This model performed exceptionally well, producing estimated surface distributions that exactly matched the observed distributions. However, this approach was post-hoc with little theoretical or empirical backing, hence, we do not discuss the details here.

Previous literature has made it obvious that in eliciting spontaneous, improvised utterances, it is hard to make participants include all information. As we discussed in §3.2.3, while adding a communicative component may have encouraged participants to be more informative, this approach was not appropriate for the present study. Moreover, previous studies suggest that improvisation may play out differently in a fully communicative setting (Hall et al., 2015; Schouwstra et al., 2020, preprint). In §3.3 we presented details of a computational model that avoided these methodological issues and inferred the word orders participants would have produced had they expressed all three constituents on every trial. This model provides an innovative way to deal with incomplete data sets, which we expect to be of potential use for others in the field. The model exploited the assumption that incomplete orders derive from an underlying complete order where one or more constituents have been omitted. The results of this analysis suggested that APV was the preferred order for describing generic-agent events in block 1, modulated by a strong tendency to omit the agent. In contrast, and consistent with our analysis of the observed data, AVP was the most common order for character-agent events.

### 3.4.1 Saliency and word order

#### 3.4.1.1 Generic-agent omissions

We saw in §3.2.5.2 that the rate of agent omissions in block 1 was significantly higher for generic agents than for character agents. We suggested that this reflected the relative saliency of the two types of agent. But why should less salient entities be omitted rather than expressed later in the sequence, as we predicted based on the saliency hypothesis proposed by Meir et al. (2017)? One possible explanation relates to the embodied nature of gestured descriptions. We observed that people typically enacted event actions, for example, miming the act of pushing using their own hands. In so doing, they in effect embodied the role of the agent while expressing the action. This phenomenon has been described in previous silent gesture studies (e.g., Hall et al., 2013; Kocab et al., 2018), and its presence is unsurprising given the performative nature of silent gesture.

One conclusion we can draw from this observation is that agents were not completely omitted from descriptions; rather, what was omitted was explicit reference to their physical attributes. More precisely, while some properties of the agent were encoded in the form of the action<sup>14</sup> (it is human, it acts volitionally, etc.), other properties such as gender or items of clothing were not expressed, presumably because they were not considered relevant in the context of the task. Such omissions were less likely to occur for character agents in block 1, we argue, because their physical characteristics were more salient and therefore more worthy of mention. It is also worth noting that in block 2, we saw a reduction in agent omissions across the board. We suggested that this may reflect a novelty effect: switching the set of agents may have made their individual attributes more salient and therefore more likely to be mentioned explicitly.

An alternative explanation for the tendency to omit generic agents in block 1 is that these were less easy to describe iconically than were character agents.<sup>15</sup> While this may form part of the explanation, the fact that the rate of generic-agent omissions was negligible in block 2

---

<sup>14</sup>A similar phenomenon has been described in sign languages. In so-called body-anchored verbs, the body is associated with the subject argument, and the form of the sign encodes some property of the subject, for example, that it has a mouth (Meir et al., 2007). Hall et al. (2013) discuss the relationship between body-as-agent in improvised gesture and body-as-subject in sign language.

<sup>15</sup>We thank an anonymous reviewer for this suggestion.

suggests that this cannot be the full story. In addition, as we noted in the Introduction, agent omission is widespread in sign languages and, consistent with our own interpretation, has been analysed as an agent backgrounding device (e.g., Kegl, 1990; Janzen et al., 2001; Barberà et al., 2018; Rissman et al., 2020).

### 3.4.1.2 AVP and (A)PV

An unexpected finding in block 1 was that agent type influenced the relative order of the patient and action: AVP was common for character-agent events, and (A)PV for generic-agent events. Here, we argue that the salience of the agent can influence constituent order indirectly by affecting the way participants construe events. This view on the role of salience differs from previous proposals (Meir et al., 2017), where salience is proposed as a direct influencer of constituent order.

We first consider the preference for expressing the action in final position when describing generic-agent events. While this pattern is consistent with previous proposals, for example, that concrete entities tend to be mentioned before abstract relations (Goldin-Meadow et al., 2008), an alternative explanation is that (A)PV, or PV more specifically, may reflect a patient-focused construal of an event. Where the identity of the agent is non-salient, the event may be more likely to be framed from the patient perspective. Consequently, the patient represents what Bock and Ferreira (2014) term the ‘aboutee’ and forms the starting point of the utterance (MacWhinney, 1977), while the agent is backgrounded through omission (Rissman et al., 2020).

The preference for describing character-agent events using AVP, on the other hand, is not only surprising in light of previous findings in the silent gesture literature, but also cannot be readily accommodated within any of the accounts discussed previously. Here, we offer a number of possible explanations. The first possibility is that AVP may have resulted from prolonged attentional focus on the agent. We observed that, in contrast to generic-agent descriptions, character-agent descriptions were often highly detailed.<sup>16</sup> In directing a large amount of attention to their physical attributes, participants may have more naturally proceeded to describing the action being performed by the character before turning their attention to the patient. If this explanation is correct, it raises important questions about the extent to which our findings reflect task-specific factors. For example, if participants were in some way restricted to providing the same amount of information about event agents, say, a single gesture, then this might potentially eliminate word order differences.

A related, task-agnostic explanation is that by analogy with the proposal that (A)PV reflects a patient-centered construal of an event, AVP may reflect an agent-centered construal. Accordingly, the agent is mentioned first, while the patient is expressed after the action reflecting its status as the background against which the agent performs the action. Thus, a highly agent-focused construal of an event could be glossed as ‘There is some character. This is the action they perform. This is the thing the action is directed towards.’

A third, perhaps more parsimonious explanation, is that both AVP and PV reflect native-language influence. AVP can be equated with SVO, while PV is analogous to the English passive construction where the *by*-phrase is not expressed, for example, *the clock is pushed (over)*.

---

<sup>16</sup>We also observed that people would sometimes include details that were not represented in the stimuli. For example, one participant consistently gestured a parrot sitting on his shoulder when describing the pirate, even though this was not depicted in the stimuli. We also observed examples of people enacting a stereotypical action associated with certain characters. For example, pantomiming cooking when describing the chef.

This interpretation of the data similarly leads to an explanation based on salience and event construal. The relatively high salience of character agents promotes an active, agent-focused construal, while events involving less salient, generic agents result in a passive, patient-focused construal. We return to the question of native-language influence below.

The proposal that salience influences event construal is, to our knowledge, new in the silent gesture literature. However, it is by no means new to the study of language production. Vogels et al. (2013), for example, argued that salience influences the global interpretation of a scene, which in turn affects structural choices. Similarly, Antón-Méndez (2017) proposed that event descriptions focus on what the more visually salient entity is doing or experiencing. In another study, as previously discussed, Rissman et al. (2018) argued that manipulating the salience of the agent affected whether participants provided an agent- or patient-focused construal of an event.

### 3.4.2 Silent gesture, native language, and word order

In the accounts outlined above, we suggested that AVP may result from increasing the salience of an event agent. However, although agent salience is typically not controlled for in the silent gesture literature, it is certainly not the case that previous studies have consistently used stimuli depicting ‘generic’ humans (Schouwstra and de Swart, 2014, for example, used events depicting witches, divers, and princesses, among others). Nevertheless, ours is the first study to find that events involving highly salient agents elicit AVP rather than APV. A possible source of this discrepancy might be the type of event used in the current study. Events in silent gesture studies often involve handling or manipulation of the patient (e.g., Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014; Christensen et al., 2016), for example, a witch eating a banana.<sup>17</sup> While the preference for APV has been attributed to the semantic relations between entities (e.g., Goldin-Meadow et al., 2008; Schouwstra, 2012) or structural iconicity (Christensen et al., 2016), an alternative interpretation lies in the observation that the form of the action gesture in handling events is likely to be influenced by the identity of the patient (e.g., a gesture that depicts eating a banana is likely to be different from one that depicts eating, say, a steak). Pertinent to this is a generalization noted by Napoli and Sutton-Spence (2014) in a review of 42 sign languages that if an argument affects the phonological shape of a verb, it typically precedes the verb (Generalization Two). This may in part explain why APV was less common in the current study where the form of the action was independent of the patient.

In the discussion above, we noted that word order preferences in block 1 may reflect influence from the participants’ native language. Despite the appealing simplicity of this suggestion, it is at odds with findings from previous silent gesture studies that have found no, or minimal, native language interference (e.g., Goldin-Meadow et al., 2008; Langus and Nespors, 2010; Futrell et al., 2015). However, these findings might also be explained by the tendency to use manipulation events. If APV is the natural order for representing such events, then, by extension, other orders may feel unnatural and may be avoided.

This argument notwithstanding, a native-language interpretation of our own data is not clear cut. We found, for example, that on moving from block 1 to block 2, participants in the

---

<sup>17</sup>We acknowledge that while some studies used only manipulation events, others used a mix of these and other types of event (e.g., Langus and Nespors, 2010; Hall et al., 2013; Gibson et al., 2013). Without a detailed breakdown of responses by event type, we can only speculate as to what, if any, the effect on word order might have been.

generic-first condition who had predominantly produced PV tended to shift to APV. However, if word order reflects native language, then we might expect a shift from a patient-focused, passive construal (PV), to an agent-focused, active construal, realized as AVP. A further complication, however, is that people tended to continue using the same ordering strategies established in block 1 when progressing to the second block. Thus, shifting from PV to APV may represent a tendency to persevere the order of the patient relative to the action while expressing the previously unseen agent in initial position.

The discussion presented here highlights the need for more research into how different event types influence structural choices in silent gesture and improvised communication (see also Schouwstra and de Swart, 2014; Christensen et al., 2016). In addition, more work is required to understand how these effects interact with influences from native language. Our findings also draw attention to an important methodological issue in the silent gesture literature, namely, how gesture sequences are analysed. Excluding incomplete orders, or categorizing them as consistent with an underlying complete order, could obscure important phenomena that might tell us something about the cognitive biases that shape structural choices during improvisation. Related to this is how improvised descriptions of events are interpreted. There is usually an implicit assumption in the literature that these can be mapped to a simple active clause. However, there are usually, if not always, multiple ways of representing the same event, and it would be surprising if this were not reflected in improvised communication.

### 3.5 Conclusion

The findings of this study support the hypothesis that word order in emerging communication systems reflects the relative salience of entities interacting in an event (Meir et al., 2017). However, rather than affecting word order directly, our results suggest that salience influences the perspective from which a producer frames an event, which in turn influences structural choices. Previous studies have demonstrated that word order in improvised communication is conditioned on certain properties of an event (e.g., Schouwstra and de Swart, 2014; Christensen et al., 2016; Hall et al., 2013; Gibson et al., 2013; Kocab et al., 2018), challenging the claim that APV (or SOV) is the default order. Our results add an additional layer to that argument: naturalness as it relates to constituent order is conditioned not only on the inherent properties of an event, but is mediated by the perspective of the producer and how they construe an event.



## Chapter 4

# Ordering Preferences in a Picture Selection Task: Investigating the Role of Modality

### 4.1 Introduction

Chapter 3 showed that manipulating the salience of the agent in a transitive event changed the word orders people used when describing events in a silent gesture task. We proposed that these structural variations derived from differences in the way people conceptualized and framed those events. However, our findings were inconclusive about the precise mechanisms underlying the specific word order preferences. We suggested three possible explanations. First, order of mention may directly reflect the way people mentally represent events. This accords with much of the silent gesture literature that assumes that word order reflects general cognitive biases that operate across modalities and communicative contexts (e.g., Gershkoff-Stowe and Goldin-Meadow, 2002; Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014; Christensen et al., 2016). Second, we suggested that factors specific to the gestural modality may have influenced the way different types of events were described. Other authors have similarly hypothesized that the affordances and constraints of the gestural modality could explain word order preferences when people describe certain types of event (e.g., Hall et al., 2013; Kocab et al., 2018). Third, word order may have been influenced by the participants' native language. While some studies have found that native language has little or no influence on structural choices in silent gesture (e.g. Goldin-Meadow et al., 2008; Langus and Nespors, 2010; Schouwstra and de Swart, 2014; Futrell et al., 2015), findings from other studies challenge the assumption that this is universally the case (e.g., Gibson et al., 2013; Meir et al., 2017).

In this study, I adopted a non-gestural, non-communicative paradigm to investigate in more detail the mechanisms underlying the word order variations found in Chapter 3. I conducted three experiments in which participants were first presented with a picture of an event, then asked to select images representing each of the event components (agent, patient, and action). The experimental approach used in this study was adapted from the paradigm adopted in a small number of studies that have investigated if word order in silent gesture reflects aspects of the manual modality, general properties of human cognition (Gershkoff-Stowe and Goldin-Meadow,

2002; Goldin-Meadow et al., 2008), or, alternatively, native language influences (Vastenius et al., 2016). I describe these studies in §4.1.1. In brief, two main findings are of interest to the present study. First, these studies found that ordering preferences in silent gesture could be replicated in a picture selection task. Second, the order in which people selected images depicting event components was both consistent and independent of the basic word order of the participants' native language. Based on these findings, the authors of these studies concluded that constituent order in silent gesture is independent of the manual modality and directly reflects the way people mentally represent events, consistent with the first of the three possibilities discussed above.

As well as enabling a comparison of ordering preferences across modalities, this non-gestural paradigm presents a number of additional advantages. First, it provides an opportunity to determine more clearly the preferred ordering of the three constituents. The results of the computational model described in Chapter 3 suggested that APV was the preferred order for describing generic-agent events, modulated by a strong tendency to omit the agent. Requiring participants to provide all and only three constituents will enable this finding to be verified. In addition, restricting participant responses in this way has the advantage of reducing the range of available ordering strategies compared with silent gesture. Although the flexibility of the gesture paradigm is generally considered one of its strengths, this same flexibility can also present challenges by introducing what might be considered 'noise' into the data. In Chapter 3, we noted that our data contained a large proportion of both incomplete responses and highly repetitious descriptions. In addition, on some trials participants expressed constituents simultaneously so that it was not possible to determine their relative ordering. We also found that people often included a gesture indicating the patient falling. Although we excluded these gestures from our analysis, it is conceivable that their presence could in some way have influenced the ordering of other the constituents.

#### 4.1.1 Picture reconstruction tasks

Gershkoff-Stowe and Goldin-Meadow (2002) report results from two studies: the first involving a silent gesture task in which English speaking participants described scenes using only gestures and no speech; and the second a picture reconstructing task where people were asked to stack transparencies, each depicting one element of a scene. The scenes depicted two objects, one of which was stationary while the other moved relative to the first, for example, a doll jumping into a hoop. In the silent gesture study, the authors investigated if manipulating aspects of the communicative context affected the extent to which participants used a consistent ordering strategy when describing scenes. They manipulated three factors: whether the signaller was sharing new knowledge with their communicative partner; whether the partner was permitted to provide feedback to indicate comprehension or request clarification; and whether the signaller and receiver swapped roles. The key finding from this experiment was that the communicative context had no effect either on the extent to which participants used a consistent order, or on the order used. Across all conditions, people strongly preferred to describe events using the order Stationary-Moving-Action. Importantly, this order does not reflect canonical English order.

In the second experiment, participants were asked to reconstruct the same scenes by stacking a set of transparencies depicting each of the event components, one on top of the other. Crucially, there was no requirement to place the cards in any particular order, since the final

picture had the same appearance regardless. The authors manipulated the temporal representation of the events such that reconstructed scenes depicted the moving object either in its initial state, prior to the action being performed, or in its final state after completion of the action. For example, a doll was depicted either alongside a hoop, with an arrow indicating the action about to be performed (e.g., jump), or inside the hoop with the same arrow indicating the action just completed. They also manipulated the communicative context. In the Self condition, participants were required to reconstruct the scene for themselves. Participants in the Other condition were told that another person would later describe each scene after viewing videos of the reconstruction process. In the Talk condition, participants were instructed to describe what they were doing to the experimenter while they stacked the transparencies.

There were three key findings from this experiment. First, participants were highly consistent in their ordering strategies, despite the fact that stacking order made no difference to the final appearance of the reconstructed scene. Second, when the moving object was depicted in its initial state, participants in both the Self and Other conditions stacked the event components in the same order as in the silent gesture experiment, that is, Stationary-Moving-Action. Third, the results demonstrated that the temporal representation of the event had a strong effect on the preferred ordering strategy. In all three conditions, Stationary-Moving-Action was less frequent when the object was depicted in its final state and there was a significant increase in the use of Stationary-Action-Moving.<sup>1</sup>

The authors propose that these ordering preferences reflect different ways of conceptualizing the same event. More specifically, they suggest that the order in which event components were stacked may reflect differing temporal perspectives on the events. When the moving object was depicted in its initial state, it was more likely to be introduced into the scene *before* the action. Conversely, when it was depicted in its final state, there was an increased probability of it being introduced *after* the action, echoing the temporal structure of the event. Moreover, they argue that these ordering strategies are independent of modality and of the communicative context. Instead, they reflect the way participants conceptualized the scenes depicted in these experiments.

In a later study, Goldin-Meadow et al. (2008) adopted the same multi-modal approach to investigate if ordering preferences in improvised descriptions of events resulted from modality-specific factors. This study involved speakers of four different languages: Chinese (SVO and SOV), English (SVO), Spanish (SVO), and Turkish (SOV). As in Gershkoff-Stowe and Goldin-Meadow (2002), participants in all language groups showed highly consistent ordering strategies both in the silent gesture task and in the picture reconstruction task. Moreover, the same order – Agent-Patient-Action – was used by all four groups, irrespective of modality, when describing transitive actions in place (e.g., a woman twisting a knob) and transitive actions crossing space (e.g., a man carrying a chicken to scaffolding).<sup>2</sup> Goldin-Meadow et al. (2008) proposed that Agent-Patient-Action reflects the ‘natural’ ordering of elements in an event. In addition, they

---

<sup>1</sup>Participants in the Talk condition were more likely to use Moving-Action-Stationary, which is analogous to the canonical English order (e.g, the doll jumps into the hoop). This is perhaps unsurprising given that they were instructed to describe what they were doing as they stacked the pictures. Participants in the Other condition also produced more Moving-Action-Stationary, but to a lesser extent.

<sup>2</sup>Included in the Goldin-Meadow et al. (2008) study were intransitive actions crossing space, for example, a man crawling towards a cat. This is the same class of event used by Gershkoff-Stowe and Goldin-Meadow (2002). Although Goldin-Meadow et al. (2008) discuss the positioning of end points (the stationary object) overall, they do not provide a detailed breakdown by event type (transitive or intransitive). It is therefore not possible to assess the extent to which their study replicates findings from Gershkoff-Stowe and Goldin-Meadow (2002).

conclude that this order reflects general cognitive processes that are modality independent and not specific to a linguistic context.

Both Gershkoff-Stowe and Goldin-Meadow (2002) and Goldin-Meadow et al. (2008) found that, in general, ordering preferences were not influenced by the canonical order of the participants' native language. In a more recent study, Vastenius et al. (2016) investigated in more detail the role of native language influence in a picture reconstruction task. They utilized the same class of events used in Goldin-Meadow et al. (2008) but adapted the procedure such that, rather than stacking transparencies, participants laid out pictures of event components vertically, from top to bottom. Thus, in contrast to the transparencies task, the final representation of the event differed according to the linear arrangement of the picture. The study involved speakers of Swedish (SVO) and Kurdish (SOV). Participants were assigned to one of two conditions: in one condition, they were instructed first to describe events verbally before arranging pictorial representations of event components; in the other, participants were required to arrange the event components without first providing a verbal description.

Of particular relevance to the present study is the finding that Kurdish participants expressed the patient before the action significantly more often than Swedish participants, in line with the respective ordering conventions of each language.<sup>3</sup> Nevertheless, in the non-verbalization condition, Swedish participants expressed the patient before the action in almost 60% of trials, contra the ordering conventions of their native language. The authors concluded that these results provide evidence for a shared preference for expressing the patient before the action, modulated, under certain circumstances, by an effect of native language. They further argued that the presence of native-language interference supports the proposal that the way we think about how to represent a scene is mediated through the structural conventions of the language we speak, or 'thinking-for-speaking' (Slobin, 1996). However, they also offer an alternative analysis suggesting that influence from native language conventions was rooted in task-specific factors. Specifically, the process of laying pictures in a row may have introduced an influence of native language due to its associations with writing.

This latter hypothesis is supported by findings from another study, conducted by Kline et al. (preprint), that compared the word orders produced by English-speaking participants in a silent gesture task with selection order in a task where participants clicked on non-iconic symbols representing event components. In this latter task, participants were asked to describe events to an alien avatar by clicking sequentially on the symbols corresponding to the event components. The task therefore differed from the other studies described in this section in that it was both explicitly communicative and language-like, involving sequential transmission of arbitrary written symbols to convey information about an event, as in a writing system. In contrast to those other studies, the order in which participants selected symbols did not match the ordering preferences found in the silent gesture task. Moreover, the preferred order in the picture selection task was SVO, the basic order of English. These findings suggest that native language may have a greater influence on ordering preferences when participants are engaged in a communicative, language-like task.

To summarize, the studies described in this section demonstrate that ordering strategies adopted in silent gesture can be replicated using a non-gestural modality in which participants select pictorial representations of event components. In addition, these strategies are broadly independent of the communicative context in which they are produced. The findings also

---

<sup>3</sup>The authors do not report on the positioning of the agent.

suggest that the particular orders that predominated in each of the studies may derive from non-linguistic cognitive processes that reflect the way people conceptualize events. Nevertheless, these studies also show that under certain conditions, native language influences can lead people to deviate from these putatively natural orders. In particular, introducing communicative or linguistic aspects into the task can increase the effect of native language ordering conventions.

## 4.2 Experiment 2: Picture selection

The main goal of this experiment was to investigate if the ordering preferences found in the first testing block<sup>4</sup> of the silent gesture experiment (Experiment 1) presented in Chapter 3 could be replicated using a non-gestural picture selection task similar to those described in the previous section. In the silent gesture study, participants typically described character-agent events using AVP. In contrast, this order was rare for generic-agent events. For this type of event, the findings suggested that APV was the preferred order, although participants tended to omit the agent from their descriptions. Among the trials in which at least one noun referent was expressed, PV was the most frequently used order, indicating a preference for expressing the patient before the action. If participants in the current experiment produced a consistent ordering strategy that replicated these findings, this would provide strong evidence that those ordering preferences directly reflect the way people mentally represented the events depicted in the study.

An important point to keep in mind is that in proposing that picture selection orders reflect the way people conceptualize events, Gershkoff-Stowe and Goldin-Meadow (2002) and Goldin-Meadow et al. (2008) make no specific reference to the picture reconstruction *process*. Rather, their hypothesis relates solely to the order in which pictures are selected. Indeed, the transparency task was specifically designed to ensure that selection order was not influenced by the process of reconstructing the scene: the resulting picture was the same irrespective of the order in which elements were placed and contained no information about how it was reconstructed. With this in mind, in this first experiment I did not include an event reconstruction procedure. Instead, participants were simply asked to select event components from an array of pictures. To eliminate, or reduce, any effects of native language on selection ordering preferences, the task did not involve any communicative aspects. In addition, the design of the experiment was intended to avoid any associations with the written modality. Details of the design and procedure are provided below.

### 4.2.1 Participants

Eighty participants were recruited through Amazon Mechanical Turk. Participants were at least 18 years old and were self-reported native English speakers. They were paid \$1.50 for their participation.

---

<sup>4</sup>I focus on block 1 responses in this study since there was strong evidence that ordering preferences established in this block were perseverated into block 2. In other words, word order preferences were not conditioned on agent type in the second block.

## 4.2.2 Materials

All experiments in this study were developed using the jsPsych library (de Leeuw, 2015) and were conducted online. The stimuli consisted of the same set of cartoon event images used in the silent gesture experiment described in Chapter 3. The set consisted of 80 character-agent and 80 generic-agent events. In addition, I produced separate images depicting the 10 humans (five ‘character’ and five ‘generic’), four objects and four actions. The design of these images was intended to address two potential problems with how event components were depicted in previous studies.

The first problem concerns the representation of event actions. In three of the studies described in §4.1.1 (Gershkoff-Stowe and Goldin-Meadow, 2002; Goldin-Meadow et al., 2008; Vastenius et al., 2016), actions were depicted as arrows. Analysing post-experiment debriefing responses, Vastenius et al. (2016) noted that, while there was a large degree of consistency across participants in how they interpreted pictures of agents, patients, and landmarks, there was much more variability in how action pictures were interpreted. Importantly, action pictures were not consistently interpreted as actions. Rather, as well as actions, they were also interpreted as motions or directions and labelled with verbs such as ‘take’, or with directional prepositions and adverbs. To ensure that participants interpreted action pictures as representing relations between interacting entities, the stimuli used in this experiment contained a greater degree of imagistic detail (Vastenius et al., 2016) such that they were more transparently identifiable with their referent. In addition, each action picture included an abstract representation of a patient. This was to ensure that pictures afforded both an agent-focused (i.e., a thing done *by an agent*) and a patient-focused construal (i.e., a thing done *to a patient*). Figure 4.1(a) shows an example action picture.

Second, in previous studies, some event components were depicted as they appeared while part of the event. For example, in the component images of the event *a man throws a ball into a basket* shown in Vastenius et al. (2016), the man was depicted with his hands raised above his head, while the ball was shown suspended in mid-air. Depicting agents mid-action is potentially problematic since it conflates the concept of the agent with that of the action into a single image. Consequently, one cannot be certain if the image will be interpreted as a representation of the agent, or as a composite representation of the agent and action. In addition, as Gershkoff-Stowe and Goldin-Meadow (2002) demonstrated, manipulating the way component images depict different perspectives of an event can influence the order in which those images are selected. In the example just given, depicting the man with his hands in the air (without the presence of the thrown object) may encourage a post-action perspective. Similarly, the suspended ball may be perceived as in-flight, again, encouraging a post-action construal of the event. It may therefore feel more natural to select the ball image after the man, since this representation aligns with the perceived temporal structure of the event. In other words, the ordering of the images may reflect a construal along the lines of *A man threw an object. The object (ball) flew through the air, towards a basket*. To avoid these potentially confounding issues, the event components were depicted independent of any event. Accordingly, humans and objects were presented in a static pose and not engaged in an interaction. Figure 4.1 shows three examples of component images (an action, an object, and a character human).

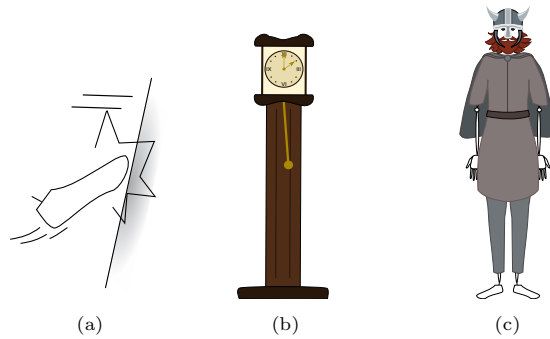


Figure 4.1: Three component images depicting (a) an action abstracted from a specific event and containing an abstract representation of a patient, (b) an object, and (c) a character human in a static pose, not engaged in an interaction.

### 4.2.3 Procedure

Participants were randomly assigned to either the character condition in which they were presented with events involving a character agent, or the generic condition where they were presented with generic-agent events. They were informed that they would be shown a series of scenes, each surrounded by nine images. They were further informed that their task was to click on the images that matched the scene. Before starting the experiment, participants were shown two example trials with the images matching the scene highlighted in green.

Participants were presented with 40 items, pseudorandomly ordered such that events in consecutive trials differed in all three constituents. Prior to each trial, a message was presented in the centre of the area where the scene would appear instructing the participant to click that area to view the scene. On clicking in the specified area, the participant was presented with an event image. After 0.5s a set of nine images depicting three humans, three objects, and three actions appeared arranged in a ring around the event image (see Figure 4.2). Three of these images depicted the event components – agent, patient, and action. The order of the nine images was randomized across trials, as was the left-right orientation of the event and component images.

When a participant clicked on one of the event components, its border changed from grey to green indicating that it had been selected. The image was also disabled so that it could not be selected again. Consequently, each image could only be selected once, and it was not possible to change the order of selection within a trial. If the participant clicked on an image that was not an event component, they received audio feedback in the form of an error tone. Progression to the next trial was automatic once the participant had selected all three event components.

### 4.2.4 Results

The data comprised a total of 3199 trials: 1600 were collected from 40 participants in the character condition and 1599 from 40 participants in the generic condition. One trial was lost due to a technical error.

The distribution across the six possible selection orders in each condition is shown in Figure 4.3. The first point to note is that, in contrast to previous picture selection studies (Gershkoff-Stowe and Goldin-Meadow, 2002; Goldin-Meadow et al., 2008), participants did not show a bias towards a single order. In addition, the orders produced in this experiment did not replicate



Figure 4.2: Experiment 2 example layout.

the ordering preferences found in Chapter 3. Whereas the results of that study showed a preference for describing character-agent events using AVP and generic-agent events using (A)PV, the predominant orders here were APV and PAV. These orders occurred in roughly equal proportions both between and within conditions. In the character condition, event components were selected in the order APV in 32.7% of trials, while PAV occurred in 35.5%. In the generic condition, APV accounted for 37.6% of trials and PAV for 30.4%.

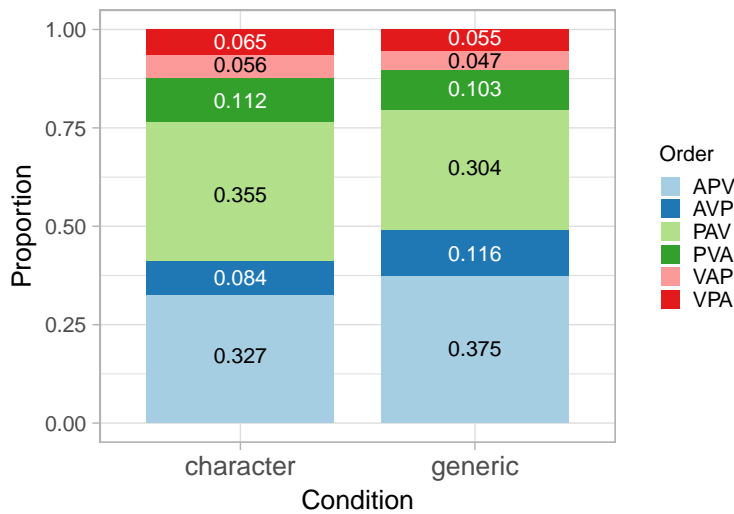


Figure 4.3: Experiment 2: Distribution of selection orders in each condition.

These findings point to a general tendency to select the action last, but suggest no preferred ordering of the agent and patient. In the character condition, participants selected the agent before the patient in 46.8% of trials. In the generic condition, the figure was 53.8% (see Figure 4.4). A mixed effects logistic regression analysis confirmed that the agent was selected before the patient at chance level, indicated by the non-significant intercept term (see Table 4.1).

These results are in contrast to the findings of the silent gesture study where there was a strong tendency to express the agent before the patient. The model also showed that there was no effect of condition, however, there was a significant main effect of image orientation such that participants were significantly less likely to select the agent first when it appeared on the right of the event image. I return to this observation in the discussion below.

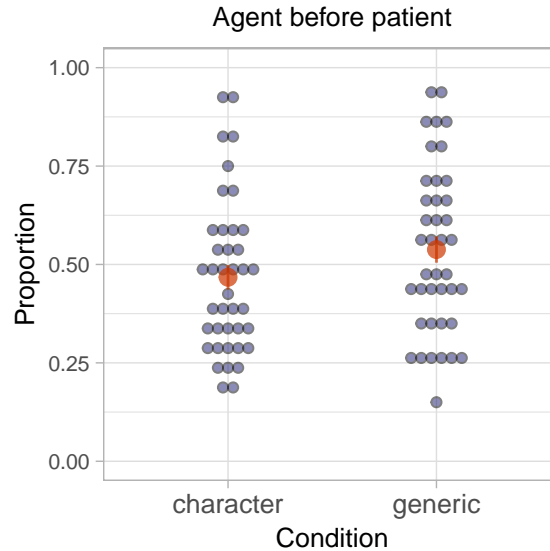


Figure 4.4: Experiment 2: Proportion of trials in each condition where the agent was selected before the patient. Small circles represent the proportion for each participant, and the large circles show the grand mean. Error bars represent the standard errors on the grand mean.

Table 4.1: Experiment 2: Mixed effects logistic regression analysis of agent-before-patient responses.

Predictor	$\beta$	$SE$	$p$
Intercept	0.063	0.128	0.624
condition	0.357	0.255	0.161
orientation <sup>a</sup>	-0.428	0.177	0.016*
condition:orientation	-0.235	0.354	0.506

Model:  $a\_before\_p \sim condition * orientation + (1 + orientation | participant) + (1 | item)^b$

<sup>a</sup>Orientation was a binary flag indicating if the agent appeared on the left (coded -0.5) or the right (coded 0.5) of the image.

<sup>b</sup>Binary inputs were deviation coded in all models in this study. Random effects in all models represent the maximal structures for which models converged without warnings.

I now investigate in more detail the occurrence of AVP and APV responses. In contrast to the findings of the silent gesture study, there was no evidence to suggest that AVP was the preferred order for describing character-agent events, or that it was more frequent for this type of event than for generic-agent events. In fact, there were *fewer* AVP responses in the character condition (8.4%) than in the generic condition (11.6%). However, a mixed effects logistic regression analysis showed that this difference was not significant (see Table 4.2). Consistent with the analysis of agent-before-patient responses, AVP was significantly less likely when the agent appeared on the right side of the event image.

Similarly, in contrast to the results of the silent gesture study, an analysis of APV responses found no evidence that this order was more common for generic-agent events than for character-agent events (see Table 4.3). Recall, however, that the evidence for an APV preference for

Table 4.2: Experiment 2: Mixed effects logistic regression analysis of AVP responses.

Predictor	$\beta$	$SE$	$p$
Intercept	-2.659	0.142	<0.001***
condition	0.255	0.267	0.340
orientation	-0.388	0.193	0.044*
condition:orientation	-0.214	0.326	0.511

Model:  $is\_avp \sim condition*orientation + (1 + orientation | participant)$

generic-agent events was indirect and came from the results of the computational model. Among the orders in which one or more noun referents were expressed at least once, PV was the most common order for this type of event, pointing to a preference for expressing the patient before the action. In the current experiment, participants selected the patient before the action in a majority of trials in both conditions (see Figure 4.5). In the character condition, patient-before-action responses accounted for 79.4% of trials, while in the generic condition the figure was 78.2%. This difference was not significant according to a mixed effects logistic regression analysis (see Table 4.4). This model did, however, indicate that participants were significantly more likely to select the patient before the action when it appeared on the left of the event image.

Table 4.3: Experiment 2: Mixed effects logistic regression analysis of APV responses.

Predictor	$\beta$	$SE$	$p$
Intercept	-0.813	0.143	<0.001***
condition	0.247	0.285	0.386
orientation	-0.252	0.163	0.122
condition:orientation	-0.229	0.324	0.480

Model:  $is\_avp \sim condition*orientation + (1 + orientation | participant) + (1 | item)$

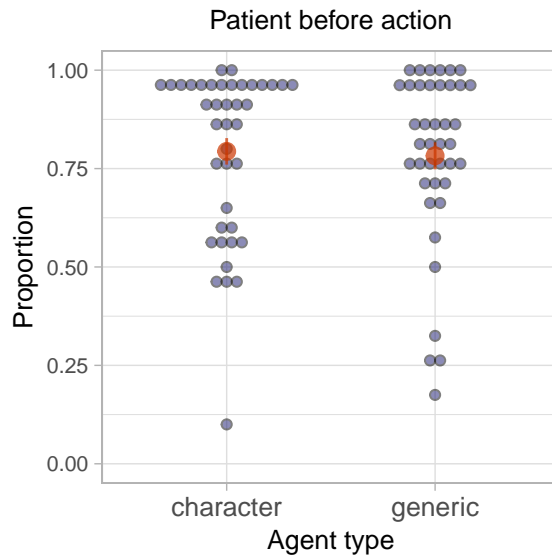


Figure 4.5: Experiment 2: Proportion of trials in each condition where the patient was selected before the action. Small circles represent the proportion for each participant, and the large circles show the grand mean. Error bars represent the standard errors on the grand mean.

Table 4.4: Experiment 2: Mixed effects logistic regression analysis of patient-before-action responses.

Predictor	$\beta$	$SE$	$p$
Intercept	1.874	0.185	<0.001
condition	-0.086	0.366	0.815
orientation	0.471	0.152	0.002**
condition:orientation	0.017	0.261	0.949

Model:  $p\_before\_v \sim condition*orientation + (1 + orientation | participant)$

#### 4.2.5 Discussion

The results of this experiment did not replicate the findings of the silent gesture study presented in Chapter 3. More specifically, there was no evidence for an AVP bias in the character condition, or for an APV preference in the generic condition. In addition, participants in the generic conditions were *not* more likely to select the patient before the action compared to those in the character condition. In fact, across all measures, there were no differences between conditions.

Another key finding was that APV and PAV occurred with equal frequency, a result that is at odds not only with the gesture study, but also with other studies that have used the same type of event (e.g., Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014) (i.e., extensional, non-reversible events). In addition, participants in this experiment showed no bias towards a single order, and, crucially, no preference for selecting agents before patients. Rather, the results showed that agents and patients were equally likely to be selected first. This was modulated by an effect of image orientation such that the patient was selected before the agent significantly more often when it appeared on the left in the event image. Image orientation also influenced the rate of AVP responses and, more broadly, the ordering of the patient relative to the action. The general conclusion from these results is that participants tended to select the patient earlier when it appeared on the left of the image.

Language production studies have similarly found an effect of image orientation on the order in which constituents are mentioned. Esaulova et al. (2019), for example, found that English-speaking participants produced more passive sentences when the patient was situated to the left of the agent. The same effect was reported by Pokhoday et al. (2019) in a study involving Russian speakers. One interpretation of these findings is that the left position is privileged for speakers of languages that use a left-to-right writing script (and the right for speakers of right-to-left languages). Consequently, the entity that appears on the left is more prominent and therefore more likely to be mentioned first.

Notwithstanding this effect of image orientation, the absence of an agent-first bias in the data warrants further investigation. The preference for expressing agents before patients is widespread in natural language, as evinced by the predominance of spoken languages that express the subject before the object (i.e., SOV, SVO and VSO) (Dryer, 2013).<sup>5</sup> Similarly, agent-before-patient is the preferred ordering strategy used in signed languages (Napoli and Sutton-Spence, 2014). Indeed, the ubiquity of agent-before-patient in both mature languages and in developing or degraded language systems has led Jackendoff (2002) to propose that Agent

<sup>5</sup>The terms *subject* and *object* are used to denote entities that are more agent-like and more patient-like, respectively. Dryer (2013) notes that ‘A language shown on the map as SOV could thus be equally well and perhaps more accurately described as APV.’

First is one of the earliest ordering principles adopted by pre-syntactic forms of language.

Also notable is that participants' ordering strategies showed no sensitivity to animacy distinctions. All of the event images used in this study depicted an animate agent acting on an inanimate patient, hence, the absence of an agent-before-patient bias can also be interpreted as an absence of an animate-first bias. A large body of literature attests to a bias for expressing animate entities before inanimate entities in natural language (e.g. Prat-Sala and Branigan, 2000; van Nice and Dietrich, 2003; Branigan et al., 2008; de Swart et al., 2008; Dennison, 2008; van de Velde et al., 2014; Esaulova et al., 2019). This bias has also been consistently reported in numerous silent gesture experiments (e.g., Goldin-Meadow et al., 2008; Hall et al., 2013; Schouwstra and de Swart, 2014; Christensen et al., 2016; Meir et al., 2017; Kocab et al., 2018) (see Chapter 5).

A possible explanation for these results is that the simple task of clicking on event components did not activate the process of formulating a mental representation of the event. Consequently, participants may not have attended to the semantic or conceptual properties of the entities represented (i.e., their role in the event or their animacy status), but rather to the visual properties of the stimuli. For example, it is possible that the images representing agents and patients were more salient because they were presented in colour. In contrast, actions were depicted with black-and-white line drawings; hence, they may have been less salient and therefore more likely to be selected last (e.g., Vogels et al., 2013; Clarke et al., 2015). Alternatively, or additionally, agent and patient images may have been more readily identified among the set of stimuli than were action images because they were more transparently related to their counterparts in the event scene.

Whatever the precise relationship between visual factors and selection order, one implication of this proposal is that reconstructing the scene, as in the transparencies task adopted by Gershkoff-Stowe and Goldin-Meadow (2002) and Goldin-Meadow et al. (2008), may have been key to eliciting consistent ordering strategies that reflected the way participants conceptualized events. In thinking about how to arrange event components, participants may have been more likely to attend to their conceptual properties and to the relationship between them (e.g., who was doing what to whom). Similarly, in a silent gesture task, participants must first apprehend the scene, formulating an understanding of the nature of the event and the entities involved.

The goal in Experiment 3 was to investigate this possibility by adapting the design of the clicking task such that participants progressively reconstructed the scene as they clicked on each of the event components. I anticipated two results from this design change. First, I predicted a significant increase in the occurrence of agent-before-patient responses. Second, if people's conceptualization of an event is reflected in the way they order event components, then encouraging participants to formulate a mental representation of the events should elicit more consistent ordering strategies.

## 4.3 Experiment 3: Scene reconstruction

### 4.3.1 Participants

Forty participants were recruited through Amazon Mechanical Turk. Participants were at least 18 years old and were self-reported native English speakers. They were paid \$1.50 for their participation.

### 4.3.2 Materials

The materials were the same as those used in Experiment 2.

### 4.3.3 Procedure

As in Experiment 2, participants were presented on each trial with an event surrounded by nine images and instructed to click on those that matched the scene. When they clicked on an event component, it appeared in the event reconstruction area on the right of the screen. Figure 4.6 shows an example trial in which the components were selected in the order APV.<sup>6</sup>

The interacting entities – agent and patient – were placed either on the right or the left of the event reconstruction area such that the left-right orientation of the final reconstructed scene matched that of the target scene. This design choice conferred two advantages. First, it ensured that the reconstructed scene exactly matched the target scene. Second, by restricting how participants could arrange the agent and patient, I hoped to minimize possible influences from their experience with written language. Studies have found that when people are asked to draw events freely, they do so in a way that is consistent with the writing system they have been exposed to. That is, speakers of left-to-right languages tend to proceed by first depicting agents on the left of the drawing while the reverse is found in speakers of right-to-left languages (Maass and Russo, 2003; Dobel et al., 2007).

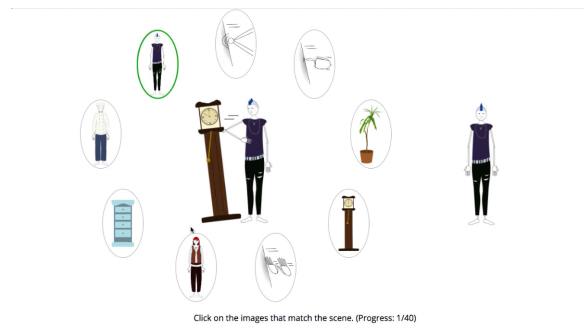
I noted in §4.2.2 that portraying event components performing their role in an event, as was done in previous picture reconstruction experiments (Goldin-Meadow et al., 2008; Vastenius et al., 2016), is potentially problematic. First, it conflates the concept of agent and action into a single image. Second, the way event components are represented can influence how people conceptualize and construe an event, which may in turn influence the order in which pictures are selected (Gershkoff-Stowe and Goldin-Meadow, 2002). The design of this experiment circumvented both of these problems by allowing the agent and patient to be introduced into the scene in an event-independent, neutral pose (see Figures 4.6(b) and (c)). Only when the action image was selected were the agent and/or patient ‘activated’ and depicted performing the relevant role in the event (see Figure 4.6(d)). If the action was selected before the agent, a generic ‘skeleton’ figure was shown performing the action in the position that would subsequently be occupied by the event agent (see, for example, Appendix B.1, Figure B.3).

---

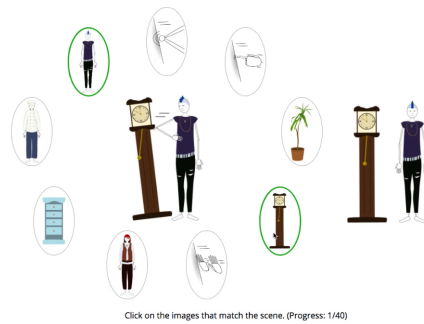
<sup>6</sup>The reconstruction steps for all other selection orders are shown in Appendix B.1.



(a) Initial trial layout



(b) Layout after agent selected



(c) Layout after patient selected



(d) Layout after action selected

Figure 4.6: Experiment 3: Example trial in which the event components were selected in the order APV.

### 4.3.4 Results

The analysis included 1600 trials from all 40 participants: 800 in the character condition and 800 in the generic condition.

Figure 4.7 shows the distribution of selection orders in each condition. What is immediately clear from these results is that, as in Experiment 2, participants did not show a preference for a single order. APV and PAV were again the most frequently produced orders, each accounting for roughly the same proportion of responses both between and within conditions. In the character condition, APV accounted for 35.4% of trials and PAV for 31.5%. The difference was greater in the generic condition: 37.6% of trials were APV and 28.9% PAV.

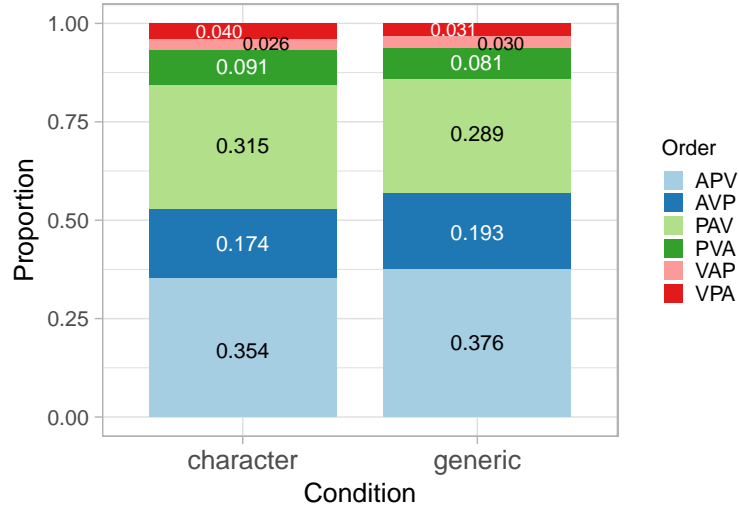


Figure 4.7: Experiment 3: Distribution of selection orders in each condition.

The manipulation was successful in eliciting more agent-before-patient responses. In the character condition, participants clicked on the agent before the patient in 55.4% of trials. In the generic condition, this figure was 59.9% (see Figure 4.8). A mixed effects logistic regression analysis indicated that the overall rate of agent-before-patient responses was greater than chance (see Table 4.5). The model indicated a marginal effect of image orientation such that participants were somewhat less likely to select the agent before the patient when it appeared on the right side of the event image.

Table 4.5: Experiment 3: Mixed effects logistic regression analysis of agent-before-patient responses.

Predictor	$\beta$	$SE$	$p$
Intercept	0.564	0.243	0.021*
condition	0.047	0.481	0.922
orientation	-0.474	0.286	0.097
condition:orientation	-0.389	0.566	0.493

Model:  $a\_before\_p \sim condition*orientation + (1 + orientation | participant) + (1 | item)$

I next compared agent-before-patient responses in this experiment with those from Experiment 2. Table 4.6 shows the results of a mixed effects logistic regression analysis of the combined data from the two experiments. The model confirmed that there were significantly

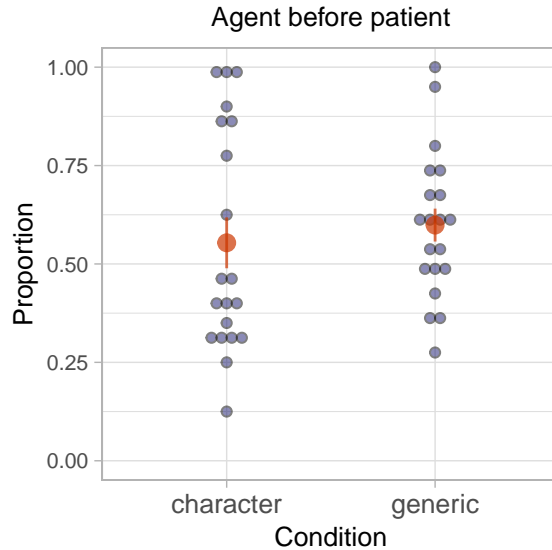


Figure 4.8: Experiment 3: Proportion of trials in each condition where the agent was selected before the patient. Small circles represent the proportion for each participant, and the large circles show the grand mean. Error bars represent the standard errors on the grand mean.

more agent-before-patient responses in Experiment 3 compared with Experiment 2. In addition, there was a significant main effect of image orientation and no interaction between orientation and experiment. This result indicates that although participants were less sensitive to image orientation in Experiment 3 compared with Experiment 2, this difference was not significant.

Table 4.6: Comparison of Experiments 2 and 3: Mixed effects logistic regression analysis of agent-before-patient responses.

Predictor	$\beta$	$SE$	$p$
Intercept	0.283	0.122	0.020*
condition	0.233	0.244	0.339
experiment	0.443	0.213	0.038*
orientation	-0.484	0.163	0.003**
condition:orientation	-0.222	0.323	0.493
condition:experiment	-0.282	0.426	0.508
experiment:orientation	-0.193	0.307	0.530
condition:experiment:orientation	0.247	0.611	0.686

Model: a.before.p ~ condition\*experiment\*orientation + (1 + orientation | participant) + (1 + orientation + experiment | item)

Consistent with the findings from the first experiment, there was no evidence for an AVP preference in the character condition (see Table 4.7). However, the use of this order increased in both conditions. In the character condition, event components were selected in the order AVP in 17.4% of trials (8.4% in Experiment 2). In the generic condition the figure was 19.3% (11.6% in Experiment 2). A mixed effects logistic regression analysis indicated that this increase was not statistically significant, however (see Table 4.8). This analysis also showed that across both experiments, AVP was less common when the agent appeared on the right of the event image. The analysis also indicated a marginal interaction between orientation and condition such that participants in the generic condition were slightly more sensitive to the relative positioning of

the agent and patient.

Table 4.7: Experiment 3: Mixed effects logistic regression analysis of AVP responses.

Predictor	$\beta$	$SE$	$p$
Intercept	-2.265	0.397	<0.001***
condition	-0.028	0.790	0.974
orientation	-0.450	0.304	0.138
condition:orientation	-0.811	0.593	0.171

Model:  $is\_avp \sim condition*orientation + (1 + orientation | participant)$

Table 4.8: Comparison of Experiments 2 and 3: Mixed effects logistic regression analysis of AVP responses.

Predictor	$\beta$	$SE$	$p$
Intercept	-2.558	0.165	<0.001***
condition	0.264	0.322	0.412
experiment	0.352	0.280	0.208
orientation	-0.390	0.171	0.023*
condition:orientation	-0.529	0.310	0.089
condition:experiment	0.472	0.561	0.400
experiment:orientation	-0.089	0.298	0.765
condition:experiment:orientation	-0.651	0.600	0.278

Model:  $is\_avp \sim condition*experiment*orientation + (1 + orientation | participant)$

Again, as in Experiment 2, there was no evidence that APV was the preferred order for generic agent events, or that it occurred more often in this condition than in the character condition (see Table 4.9). In addition, the patient was selected before the action in a majority of trials in both conditions (76.0% of trials in the character condition and 74.6% in the generic condition; see Figure 4.9), with no significant difference between the two (see Table 4.10). Unlike Experiment 2, the effect of image orientation on the rate of patient-before-action responses did not reach statistical significance.

Table 4.9: Experiment 3: Mixed effects logistic regression analysis of APV responses.

Predictor	$\beta$	$SE$	$p$
Intercept	-0.875	0.262	<0.001***
condition	0.056	0.521	0.914
orientation	0.004	0.208	0.983
condition:orientation	0.252	0.402	0.531

Model:  $is\_apv \sim condition*orientation + (1 + orientation | participant) + (1 | item)$

### 4.3.5 Discussion

In this experiment, participants reconstructed the target scene by clicking on each of the event components. The goal in introducing this procedure was to encourage participants to attend to the conceptual properties of event components and to the relationship between them, i.e., who

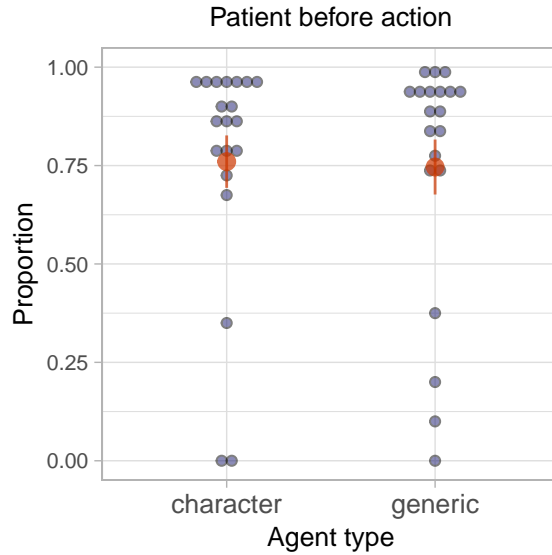


Figure 4.9: Experiment 3: Proportion of trials in each condition where the patient was selected before the action. Small circles represent the proportion for each participant, and the large circles show the grand mean. Error bars represent the standard errors on the grand mean.

Table 4.10: Experiment 3: Mixed effects logistic regression analysis of patient-before-action responses.

Predictor	$\beta$	$SE$	$p$
Intercept	1.591	0.373	<0.001
condition	0.144	0.745	0.846
orientation	0.385	0.244	0.115
condition:orientation	0.621	0.484	0.200

Model:  $p\_before\_v \sim condition * orientation + (1 + orientation | participant)$

was doing what to whom. The change was successful in eliciting significantly more agent-before-patient (or human-before-inanimate object) responses compared with Experiment 2. However, as in the previous experiment, participants did not produce a consistent ordering strategy. Moreover, the findings did not replicate those of the silent gesture experiment described in Chapter 3. In particular, the results showed that there was no effect of agent type on selection orders.

The results of this experiment suggest that introducing the reconstruction procedure was only partially successful in reducing the influence of the visual properties of the stimuli. This conclusion finds support in the observation that, although selection order was somewhat less sensitive to the left-right orientation of the target scene compared with Experiment 2, the difference was not significant. It is also worth noting that although the proportion of agent-before-patient responses in this experiment was greater than chance, it was nonetheless lower than might be expected given the strength and robustness of the agent-before-patient (or animate-before-inanimate) bias reported in natural language (e.g., Dryer, 2013; Jackendoff, 2002; Napoli and Sutton-Spence, 2014) and in silent gesture (e.g., Goldin-Meadow et al., 2008; Langus and Nespors, 2010; Schouwstra and de Swart, 2014). In the silent gesture study reported in Chapter 3, participants in the first block of testing expressed the agent before the patient in around 81% of character-agent trials and just under 88% of generic-agent trials.

In Experiment 4, I implemented an additional change to the task that was intended to further encourage participants to formulate a mental representation of the events, thereby reducing their reliance on the visual properties of the stimuli.

## 4.4 Experiment 4: Reconstructing scenes from memory

In the previous experiment, the target scene remained visible throughout the trial. Here, the target scene was removed from view while the participant completed the task. This design change was based on the premise that if selection order reflects the way people conceptualize events, then the effect should be strongest when people are required to recall an event from memory before reconstructing the target scene.

### 4.4.1 Participants

Forty participants were recruited through Amazon Mechanical Turk. Participants were at least 18 years old and were self-reported native English speakers. They were paid \$1.50 for their participation.

### 4.4.2 Materials

The materials were the same as those used in the previous two experiments.

### 4.4.3 Procedure

The procedure was the same as in Experiment 3, with the following modifications. The target scene was displayed on screen for 1.5 seconds. It was then removed from view and the array of nine images appeared in the same arrangement as in the previous two experiments. The participant was then prompted to select the images that matched the scene. As each event component was selected, the event was reconstructed in the location where the target scene had originally appeared (see Appendix B.2).

### 4.4.4 Results

The analysis included 1600 trials from all 40 participants: 800 in the character condition and 800 in the generic condition.

Figure 4.10 shows the distribution of selection orders in each condition. As is evident from these results, the manipulation was not successful in eliciting consistent ordering strategies. However, there were two notable differences compared with the previous two experiments. First, participants produced substantially fewer APV and PAV responses. In the character condition, APV accounted for 25.6% of trials, while PAV occurred in 24.1%. As in Experiment 3, the difference in the proportion of these two orders was greater in the generic condition: APV occurred in 29.3% of trials, and PAV in 18.9%. Second, AVP was the most common order in both conditions, accounting for 30.4% of trials in the character condition and 41.0% in the generic condition. I return to this second observation below.

Looking at the proportion of agent-before-patient responses, in line with expectations, both conditions saw an increase in the proportion of such trials. In the character condition, agent-before-patient responses accounted for 59.0% of trials (55.4% in Experiment 3). In the generic

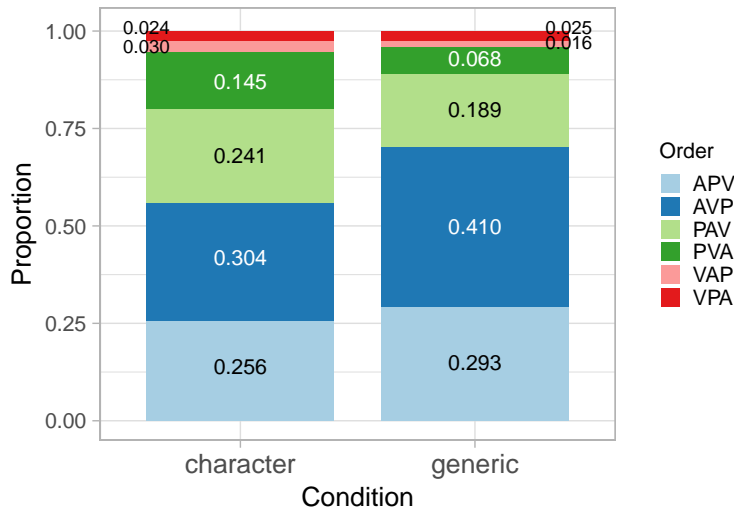


Figure 4.10: Experiment 4: Distribution of selection orders in each condition.

condition, the figure was higher at 71.9% (59.9% in Experiment 3). Figure 4.11 shows the proportion of trials in which the agent was selected before the patient. Nevertheless, a mixed effects logistical regression analysis of data from all three experiments indicated that this increase was not significant (see Table 4.11, contrast *experiment3-2*). Surprisingly, the model found that across all three experiments, there were significantly more agent-before-patient responses in the generic condition than in the character condition. This finding is the reverse of what was predicted under the salience hypothesis discussed in Chapter 3, and its cause is not clear.

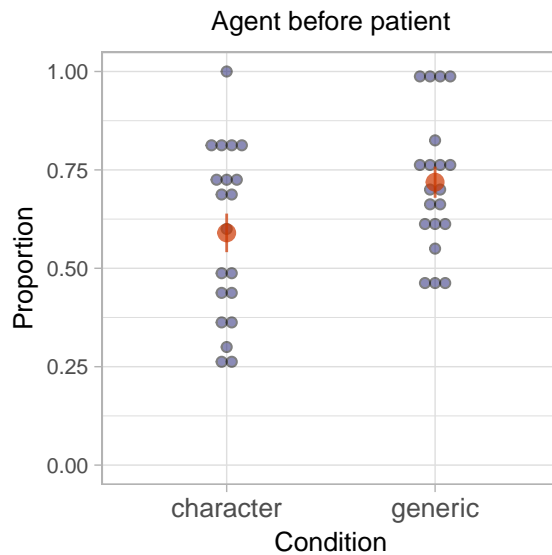


Figure 4.11: Experiment 4: Proportion of trials in each condition where the agent was selected before the patient. Small circles represent the proportion for each participant, and the large circles show the grand mean. Error bars represent the standard errors on the grand mean.

Turning to the proportion of AVP responses, as in the previous two experiments, there was no difference across conditions (see Table 4.12). I noted above that there was a substantial increase in this order compared with Experiment 3. This increase was statistically significant

Table 4.11: Comparison of Experiments 2, 3 and 4: Mixed effects logistic regression analysis of agent-before-patient responses.

Predictor	$\beta$	$SE$	$p$
Intercept	0.489	0.105	<0.001***
condition	0.424	0.209	0.043*
experiment2-1 <sup>a</sup>	0.411	0.208	0.048*
experiment3-2	0.399	0.256	0.119
orientation	-0.440	0.125	<0.001***
condition:orientation	-0.171	0.248	0.490
condition:experiment2-1	-0.298	0.417	0.475
condition:experiment3-2	0.666	0.511	0.193
experiment2-1:orientation	-0.088	0.273	0.747
experiment3-2:orientation	0.192	0.327	0.556
condition:experiment2-1:orientation	0.318	0.547	0.561
condition:experiment3-2:orientation	0.195	0.651	0.764

Model:  $a\_before\_p \sim condition*experiment*orientation + (1 + orientation | participant)$

<sup>a</sup>I used successive difference coding (also known as backward difference coding) for experiment; contrasts represent the difference between Experiments 2 and 1, and between Experiments 3 and 2.

according to a mixed effects logistic regression analysis of the data from all three experiments (see Table 4.13, contrast *experiment3-2*). The analysis also indicated a significant main effect of image orientation. Across the three experiments, this effect was significantly greater in the generic condition compared with the character condition. I return this observation in the Discussion below.

Table 4.12: Experiment 4: Mixed effects logistic regression analysis of AVP responses.

Predictor	$\beta$	$SE$	$p$
Intercept	-0.707	0.311	0.023*
condition	0.725	0.621	0.244
orientation	-0.350	0.183	0.055
condition:orientation	-0.495	0.374	0.186

Model:  $is\_avp \sim condition*orientation + (1 + orientation | participant)$

An analysis of APV responses showed that this selection order was similarly unaffected by condition (see Table 4.14). Looking at the proportion of trials in which the patient was selected before the action (see Figure 4.12), there were fewer such responses in this experiment compared with the previous two. In the character condition, the patient was selected before the patient in 64.3% of trials (79.4% in Experiment 2; 76.0% in Experiment 3). The proportion of such responses was lower in the generic condition, reaching only 54.9% (78.2% in Experiment 2; 74.6% in Experiment 3). This reduction was statistically significant and reflects the lower proportion of APV and PAV responses in this experiment, and the significant increase in AVP. As in the analysis of AVP responses, the model also indicated a significant main effect of image orientation and a significant interaction between orientation and condition, which I discuss below.

Table 4.13: Comparison of Experiments 2, 3 and 4: Mixed effects logistic regression analysis of AVP responses.

Predictor	$\beta$	$SE$	$p$
Intercept	-1.947	0.147	<0.001***
condition	0.427	0.291	0.143
experiment2-1	0.327	0.292	0.264
experiment3-2	1.680	0.367	<0.001***
orientation	-0.362	0.124	0.003**
condition:orientation	-0.528	0.234	0.024*
condition:experiment2-1	0.507	0.586	0.387
condition:experiment3-2	0.200	0.732	0.785
experiment2-1:orientation	-0.087	0.280	0.756
experiment3-2:orientation	0.080	0.308	0.795
condition:experiment2-1:orientation	-0.704	0.564	0.212
condition:experiment3-2:orientation	0.423	0.608	0.487

Model:  $is\_avp \sim condition * experiment * orientation + (1 + orientation | participant) + (1 | item)$

Table 4.14: Experiment 4: Mixed effects logistic regression analysis of APV responses.

Predictor	$\beta$	$SE$	$p$
Intercept	-1.203	0.188	<0.001***
condition	0.188	0.374	0.615
orientation	-0.147	0.124	0.236
condition:orientation	0.159	0.247	0.520

Model:  $is\_apv \sim condition * orientation + (1 | participant)$

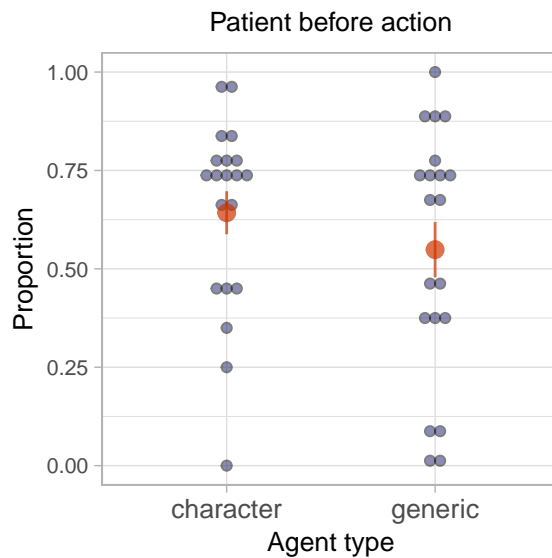


Figure 4.12: Experiment 4: Proportion of trials in each condition where the patient was selected before the action. Small circles represent the proportion for each participant, and the large circles show the grand mean. Error bars represent the standard errors on the grand mean.

#### 4.4.5 Discussion

As in the previous two experiments, participants in the present experiment did not produce consistent orders. In addition, the ordering preferences in this experiment did not replicate

Table 4.15: Comparison of Experiments 2, 3 and 4: Mixed effects logistic regression analysis of patient-before-action responses.

Predictor	$\beta$	$SE$	$p$
Intercept	1.321	0.154	<0.001***
condition	-0.235	0.308	0.445
experiment2-1	-0.135	0.301	0.654
experiment3-2	-1.320	0.383	<0.001***
orientation	0.339	0.100	<0.001***
condition:orientation	0.431	0.194	0.026*
condition:experiment2-1	-0.260	0.602	0.667
condition:experiment3-2	-0.445	0.766	0.561
experiment2-1:orientation	0.134	0.232	0.563
experiment3-2:orientation	-0.159	0.259	0.539
condition:experiment2-1:orientation	0.680	0.463	0.142
condition:experiment3-2:orientation	-0.132	0.513	0.797

Model:  $p\_before\_v \sim condition*experiment*orientation + (1 + orientation | participant) + (1 | item)$

those in the silent gesture study described in Chapter 3. The proportion of agent-before-patient responses was higher than in Experiment 3, however, the increase was not statistically significant. Moreover, although the agent was selected before the patient in almost 72% of trials in the generic condition, in the character condition, the figure was only 59%, far lower than might be expected.

In addition, the analyses showed that selection order was sensitive to the relative positioning of the agent and patient across all three experiments. Overall, participants were more likely to select left-positioned agents before right-positioned patients and to select left-positioned patients earlier (i.e., before the agent and/or action). This latter result was reflected in a decrease in AVP responses and a corresponding increase in patient-before-action responses for scenes depicting left-positioned patients. Taken together, these results suggest that the visual properties of the stimuli remained a significant factor in guiding the order in which people selected event components.

The results also showed that the effect of image orientation on the selection order of the patient was greater in the generic condition compared with the character condition. This result parallels the finding from the silent gesture experiment where participants had a slightly greater tendency to express patients when they appeared on the left of the image, an effect that was significantly greater when the events being described involved a generic agent (see Chapter 3). We suggested that this may have been because left-positioned patients are more prominent, particularly when agents are non-salient and therefore less likely to compete with the patient for the viewer’s attention. Thus, although participants in this study showed no systematic difference in selection orders across conditions, this findings hints at the possibility that agent salience in combination with the relative positioning of the agent and patient did have some effect on selection order.

Another key finding was that there was a significant increase in the proportion of AVP responses relative to the previous two experiments. This order is analogous to canonical English word order, SVO, suggesting that native language influences were stronger in this experiment compared with Experiments 2 and 3. One possible explanation for this finding is that partici-

pants may have recalled events by describing them internally.<sup>7</sup> The order in which they selected the event components may therefore reflect the order in which they mentally expressed them. This explanation echoes the account offered by Gershkoff-Stowe and Goldin-Meadow (2002) to explain why, in their study, participants who were asked to describe what they were doing as they arranged the event components were more likely to produce an order that was consistent with canonical English word order.

## 4.5 General discussion

I conducted three experiments designed to investigate in more detail the mechanisms underlying the ordering preferences found in the silent gesture study described in Chapter 3. I adopted a non-gestural picture selection paradigm based on the approach taken in a small number of studies that have investigated cross-modal word order preferences (Gershkoff-Stowe and Goldin-Meadow, 2002; Goldin-Meadow et al., 2008; Vastenius et al., 2016). Findings from these studies suggest that word order in non-linguistic, non-communicative tasks reflects the ‘natural’ sequencing of event components, that is to say, it directly reflects the way people mentally represent events. Under this view, finding consistent ordering strategies in the present study that replicated the findings from the silent gesture task would provide strong evidence that these orders were a direct reflection of the way people conceptualized the events.

In Experiment 2, participants were shown a series of character-agent or generic-agent events and asked to click on the images that matched the event, that is, the agent, patient, and action. Contrary to expectations, participants did not produce consistent ordering strategies. Analysis of the results suggested that the order in which participants selected the event components reflected visual properties of the stimuli. Two findings supported this conclusion. First, participants tended to select the action last, but showed no preferred ordering of the agent and patient. This finding is at odds with a substantial body of literature showing that there is a strong preference for expressing agents before patients, both in natural spoken and signed languages (e.g., Dryer, 2013; Jackendoff, 2002; Napoli and Sutton-Spence, 2014) and in silent gesture (e.g., Goldin-Meadow et al., 2008; Langus and Nespors, 2010; Schouwstra and de Swart, 2014). In addition, since all of the events presented in this experiment involved animate agents interacting with inanimate patients, the findings also run counter to a widely attested preference for expressing animate entities before inanimate entities (e.g., Branigan et al., 2008; Meir et al., 2017).

The second key finding was that selection order was sensitive to the relative positioning of the agent and patient in the event images. Participants were less likely to select agents before patients when they appeared on the right of the event image, and more likely to select left-positioned patients earlier, that is, before the agent and/or the action.

In the next experiment, the task was modified such that participants were required to reconstruct the target scene by clicking on each of the event components. The aim was to encourage people to attend more to the conceptual properties of the interacting entities and to the relationship between them. This manipulation was only partially successful. Although participants in Experiment 3 produced more agent-before-patient responses compared with Experiment 2, suggesting a decreased sensitivity to the visual properties of the stimuli, this difference did

---

<sup>7</sup>The experiment was conducted online, therefore one cannot rule out the possibility that participants may also have engaged in talking-out-loud to themselves.

not reach statistical significance. In addition, participants did not adopt a consistent ordering strategy, and selection order remained sensitive to the relative positioning of the agent and patient, although the effect was slightly weaker than in the first experiment.

In Experiment 4 I sought to increase the likelihood that participants would draw on their mental representation of the event by requiring them to reconstruct the target scene from memory after it was removed from view. Under the hypothesis that word order in a non-linguistic, non-communicative task reflects the way people conceptualize events (Gershkoff-Stowe and Goldin-Meadow, 2002; Goldin-Meadow et al., 2008), the effect should be strongest when they recall the target scene from memory. Nevertheless, the results of this experiment were broadly in line with findings from Experiments 2 and 3. However, they did differ in one key respect: there was a significant increase in the proportion of AVP responses, which was the most common order in both conditions. This may reflect native-language interference, possibly resulting from people describing the events to themselves during recall.

An analysis of the data from all three experiments revealed that ordering preferences overall were strongly influenced by the relative positioning of the agent and patient. Interestingly, the analysis showed that the effect of image orientation on the positioning of the patient was significantly greater in the generic condition compared with the character condition. This hints at the intriguing possibility that reducing the salience of the agent had the corresponding effect of making the patient more prominent. This effect was detectable in the current experiment through its interaction with the effects of image orientation.

In summary, across all three experiments participants did not produce consistent ordering strategies. Rather than reflecting the way people conceptualized the events, the findings suggested instead that the visual properties of the stimuli played a key role in determining the order in which participants selected event components. A corollary of this result is that selection orders did not replicate the findings of the silent gesture study where AVP was the preferred order for character-agent events and (A)PV for generic-agent events (see Chapter 3). Consequently, the results of the present study can tell us little about the mechanisms underlying the ordering strategies in silent gesture.

These findings raise an important question: why did participants in previous picture selection studies produce consistent orders that were, apparently, not influenced by visual properties of the stimuli? I suggest that the answer to this question lies in the way the stimuli were presented in the present study compared with previous picture selection studies. There were a number of differences that could potentially have influenced participant responses. First, whereas participants in previous studies were asked to arrange event components after viewing events presented as videos (Gershkoff-Stowe and Goldin-Meadow, 2002; Goldin-Meadow et al., 2008; Vastenius et al., 2016), in this study, events were depicted as static images, as is common in the silent gesture literature (e.g., Schouwstra and de Swart, 2014; Christensen et al., 2016). One possible consequence of this difference is that participants in the present study were less likely to conceptualize events as dynamic interactions between entities unfolding in space and time. In turn, this may have increased their reliance on presentational factors such as the relative positioning of entities. A second difference concerns the presentation of the event components. In the three experiments described here, participants were required to identify each of the components from a set of nine. The task therefore involved visually searching and matching to the target image, which may have resulted in the visual properties of the stimuli, for example their colour, being an important factor. In contrast, participants in previous picture selection

studies were not required to identify the components from an array, but were simply given the images and asked to arrange them.

A third difference between the present study and previous picture selection studies is the representation of event components. Recall from §4.2.2 that in previous studies these were depicted as they appeared in the event, for example, a man presented with his hands raised in a throwing event. I argued that this approach was problematic for two reasons: first, it conflates the concepts of the agent and the action; second, and more relevant to the present discussion, I noted that it may encourage a particular perspective on an event, which may in turn influence selection order (Gershkoff-Stowe and Goldin-Meadow, 2002). However, the findings of the experiments presented here suggest an alternative interpretation: rather than encouraging people to adopt a particular perspective on the event, this perspective may be inherent to the pictures themselves. Consider the example event discussed in §4.2.2 in which a man throws a ball. I suggested that depicting the man in a post-action pose, that is, with his hands in the air with the ball not present, may encourage a viewer to formulate a perspective in which the man first throws an object (the ball), which subsequently flies through the air. Alternatively, it could be argued that a viewer need not adopt this perspective at all, or any other; they need simply arrange pictures in a way that is consistent with the structural and temporal properties of the event as conveyed by the component images themselves. Staying with the same example, if the pictures were arranged such that the ball was placed before the man, this might feel unnatural, since it may be considered more consistent with an event in which a ball first flies through the air and is then caught by the man. In the present study, such considerations would not have been in effect since the event components were presented in a neutral pose, independent of any particular event.

It is important to note, however, that these explanations cannot, of course, explain why both Gershkoff-Stowe and Goldin-Meadow (2002) and Goldin-Meadow et al. (2008) found that participants produced the *same* consistent order in a gesture task and a picture selection task. Further research is required to address this question.

## 4.6 Conclusion

The silent gesture paradigm has proven to be a valuable resource for investigating how people convey information about events in the absence of a conventionalized communication system. Nevertheless, questions remain about the mechanisms underlying the ordering preferences found in silent gesture. In particular, there is an ongoing debate over the extent to which word order reflects aspects of general cognition that can be generalized to other modalities such as spoken language, or factors that are specific to the gestural modality (e.g., Kocab et al., 2018). In the present study, I set out to address this question using a picture selection task adapted from previous studies (Gershkoff-Stowe and Goldin-Meadow, 2002; Goldin-Meadow et al., 2008) that have found that, in the absence of communicative pressures or language-like factors, ordering preferences are free from the influence of native language word order. In addition, these preferences have been found to match those produced when people describe the same events using improvised silent gesture. However, the results of this study cast doubt on the claim that the orders produced in these experiments directly reflect the way people mentally represent events. Rather, the findings point to the possibility that depicting event components at a particular point in an event, for example, in a post-action pose, may encourage people to

arrange pictures in a way that is consistent with that representation. When event components are presented in a more neutral manner, as in the current study, low-level visual properties may play a stronger role in directing ordering preferences.



## Chapter 5

# Cognitive Saliency vs Communicative Pressure: Investigating the Effects of Animacy on Word Order using Artificial Language Learning

In the previous two chapters, I looked in more detail at the saliency hypothesis proposed by Meir et al. (2017) to account for word order variation in new and emerging communication systems. Specifically, In Chapter 3 I asked if manipulating the saliency of a human agent interacting with an inanimate object could modulate the tendency to express animate entities before inanimate entities in a silent gesture task. The results of this study suggested that manipulating the saliency of the agent can influence word order indirectly by affecting the way participants construe events. However, the results were inconclusive about the exact mechanisms underlying the specific word order preferences. In Chapter 4, I pursued this question further and investigated the role of modality using a series of non-gestural picture selection tasks. Again, the results of this study were inconclusive about the precise relationship between saliency, modality, and word order. They nevertheless highlighted important methodological issues that warrant further investigation before firm conclusions can be drawn about the nature of word order variation in such tasks.

In this chapter, I return to the question of how animacy and, relatedly, semantic reversibility influence word order in emerging communication systems. As discussed in Chapter 2, the saliency hypothesis is only one of a number of hypotheses that have been put forward in the silent literature to account for animacy-based word order variation. An open question concerns the extent to which modality and communicative pressures also play a role. Here I focus on the former and test the saliency hypothesis against the noisy channel hypothesis (Gibson et al., 2013).

The material presented in this chapter forms the text of a manuscript submitted to *Cognitive Science* on 7 August 2020 following peer review and an invitation to revise and resubmit,

received from the journal on 20 July 2020 (original submission data: 10 January 2020). The text has been formatted to be consistent with the rest of the thesis. The experiments described in this chapter were conceived in cooperation with all four of my supervisors – Simon Kirby, Kenny Smith, Jenny Culbertson, and Marieke Schouwstra – who are also co-authors on the paper. I designed and developed the experiment, collected the data, performed the analysis and wrote the first draft of the paper myself. All co-authors provided advice and guidance on the work presented in this chapter and contributed to editing the paper.

## 5.1 Introduction

The relationship between animacy and linguistic structure has been well documented in typology and the psycholinguistics literature (see de Swart et al., 2008, for a review). For example, cross-linguistic experimental work provides strong evidence that animate entities tend to be assigned to more prominent syntactic roles than inanimate entities, or tend to be expressed earlier in an utterance (Prat-Sala and Branigan, 2000; van Nice and Dietrich, 2003; Branigan et al., 2008; Dennison, 2008; van de Velde et al., 2014; Esaulova et al., 2019). A number of recent studies have demonstrated that word order variation in newly emerging sign languages is also sensitive to animacy (e.g., Meir et al., 2017), suggesting that such effects are operational from the very earliest stages in the development of a language and may play a key role in shaping the evolution of word order conventions.

To understand the relationship between word order and animacy in newly developing languages that lack linguistic conventions, a number of studies have adopted the silent gesture paradigm (e.g., Gibson et al., 2013; Hall et al., 2013; Kocab et al., 2018). In these studies, participants produce gestured, non-verbal descriptions of *reversible* events, so called because both interacting entities – typically human – could plausibly be either the semantic agent or patient, and *non-reversible* events – typically involving an animate agent and inanimate patient – where there is only one plausible interpretation. A common finding from these studies is that word order choices are sensitive to the relative animacy of the interacting entities.

In this study, we conducted a series of artificial language learning experiments using a non-gestural paradigm to test two competing hypotheses that have emerged from this work. In brief, the first hypothesis proposes that word order variations derive from communicative pressures arising from the potential ambiguity of reversible events (Gibson et al., 2013). The second proposes that word order is conditioned on the relative salience of entities interacting in an event (Meir et al., 2017). We describe these hypotheses in more detail below. First, we provide an overview of how animacy affects linguistic structure in newly emerging sign languages.

### 5.1.1 Animacy and linguistic structure in emerging sign languages

Unlike spoken languages whose lineages stretch back thousands of years, emerging sign languages have developed independently within the last century. They therefore offer a rich seam of information on the biases and pressures that shape newly developing communication systems. Studies of emerging sign languages suggest that animacy influences structural choices from the earliest stages in the development of a new language. In an early study of Nicaraguan Sign Language (NSL), Senghas et al. (1997) found clear conditioning of word order on event type. Non-reversible events were typically described using constructions in which the action was ex-

pressed last while the two NPs appeared in no preferred order. In contrast, reversible events were always described with two verbs, one of which described the action performed by the agent, and the other the resulting action by the patient, for example, MAN PUSH, WOMAN FALL. In a more recent study of NSL, Flaherty (2014) found that these so-called paired verb constructions are used for both types of event, but appear more often with reversible events. Meir (2010) has suggested that this construction may enable people to accurately convey information about participant roles without the use of explicit role-marking devices such as conventionalized word order or case marking.

Flaherty (2014) also found that NSL signers tended to describe non-reversible events using SOV, while SOV and OSV occurred in roughly equal proportions for reversible events. Similar findings have been reported for Israeli Sign Language (Meir, 2010), Al-Sayyid Bedouin Sign Language (Meir, 2010; Meir et al., 2017), and Central Taurus Sign Language (Ergin et al., 2018). To account for these findings, Meir et al. (2017) have proposed that word order variations reflect the relative salience of interacting entities (see also Branigan et al., 2008). Human entities are more salient than inanimate entities, resulting in S-before-O for non-reversible events; where both the subject and object are human, there is no preferred ordering of these constituents. We discuss this hypothesis in more detail below.

### 5.1.2 Animacy and word order in silent gesture

A number of studies have sought to address the question of how animacy influences constituent order in the absence of linguistic conventions using the silent gesture paradigm. This paradigm enables participants to describe stimuli without being restricted by the conventions or norms of their native language. It is therefore becoming increasingly popular as a means of investigating the cognitive biases that shape language structure in emerging communication systems (Kirby, 2017).

Echoing findings from emerging sign languages, silent gesture studies have found that people tend to vary their choice of word order as a function of event type. When describing non-reversible events, people typically use SOV, or orders that are classed as consistent with SOV. For reversible events, these orders are usually less common. However, across studies, findings in relation to this category of event have been considerably less consistent than those reported for non-reversible events. Gibson et al. (2013), for example, found that English-speaking participants showed a preference for SVO when describing reversible events. Futrell et al. (2015) replicated this result in a study involving Irish, Russian and Tagalog speakers. Hall et al. (2013), on the other hand, found that English-speaking participants tended to describe reversible events using orders in which the object did not appear immediately before the verb. These included SVO, but also OSV, OSVO and SOSV. Similar results were reported by Kocab et al. (2018). In another study, Meir et al. (2017) found that the distribution of word orders for reversible events varied according to the language experience of the participants. A summary of the relevant findings from these studies is presented in Appendix C.1.

To explain the preference for describing reversible events using SVO, Gibson et al. (2013) proposed the *noisy channel hypothesis*, according to which SVO is more robust than SOV against information loss when communication takes place in the presence of noise. The authors argue that when a producer communicates about an event, there is some possibility that part of the message may be lost in transmission. Suppose, for example, the producer sends the message

*boy man kick*. If part of this message is lost such that the comprehender receives, say, *man kick*, they will not be able to determine if *man* refers to the subject or the object, because both entities appear before the verb. If, on the other hand, the producer sends *boy kick man* and the comprehender receives *kick man*, they will be able to recover at least part of the message – in this case, the role of *man* – because the subject and object appear on different sides of the verb. In support of this proposal, Gibson et al. (2013) note that SOV languages are more likely to mark argument structure morphologically than are SVO languages (Sinnemäki, 2010), suggesting that the latter is more communicatively transparent. The observation that many mature sign languages exhibit a preference for SVO for reversible events (Napoli and Sutton-Spence, 2014) has similarly been taken as evidence that this order conveys some communicative advantage.

Hall et al. (2013), however, note that the noisy channel hypothesis cannot explain the increased use of orders such as OSV for reversible events. They proposed an alternative account – termed the *role conflict hypothesis* – based on modality-specific production constraints. They noted that when participants in their study produced gestures to describe human entities, they typically used their own body, for example, flexing their muscles to indicate *man*. Similarly, when describing actions, people implicitly embodied the role of the subject by physically enacting the action, for example, motioning with their hands to express *push*. The authors suggested that, for reversible events, it may feel unnatural to embody the role of an animate object immediately before enacting an action, since this might feel like a conflict of roles. Consequently, they argue, people avoid OV clusters for events of this kind.

However, neither the noisy channel nor the role conflict hypothesis satisfactorily explains the results reported by Meir et al. (2017). This study tested groups of participants representing three spoken languages and three emerging sign languages. The authors reported that only participants with significant exposure to an SVO language preferred this order for reversible events (Hebrew speakers and Israeli Sign Language signers who were literate in Hebrew). A key finding from this study was that in all three signing groups, non-literate participants described reversible events using OSV and SOV in roughly equal proportions. Both of these orders are inconsistent with the noisy channel hypothesis, while SOV runs counter to the predictions of the role conflict hypothesis.

To explain why SVO was used among some participant groups, Meir et al. (2017) proposed that this reflected interference from another language. Specifically, they suggested that people may fall back on a convention with which they are familiar when there is no straightforward mechanism for indicating who is doing what to whom. To explain the roughly equal proportion of SOV and OSV for reversible events among the non-literate signers, they proposed an account based on a ‘human-first’ principle and a broader appeal to the notion of salience.<sup>1</sup> According to their hypothesis human entities are more salient than inanimate entities and are therefore expressed first (see also Prat-Sala and Branigan, 2000; van Nice and Dietrich, 2003; Branigan et al., 2008; Dennison, 2008; van de Velde et al., 2014; Esaulova et al., 2019). This results in S-before-O when the subject is human and the object is inanimate, and no preferred ordering of S and O when both are human. To explain the preference for V-final orders, Meir et al. (2017)

---

<sup>1</sup>In the scene literature, ‘salience’ refers to measurable, low-level visual properties such as luminance and size. Elsewhere, the term has been applied more loosely to a range of phenomena that can be broadly summarized as referring to factors that make an entity more prominent, important, or interesting, and therefore more likely to attract the attention of the viewer (Ferreira and Rehrig, 2019). For Meir et al. (2017), the salience of human entities derives from the central importance of conspecifics to human cognition.

draw on previous proposals that concrete entities are more cognitively basic than abstract entities (Gentner and Boroditsky, 2001) and so tend to be mentioned first (Goldin-Meadow et al., 2008).<sup>2</sup> Meir et al. (2017) situate much of their argument around a broad concept of salience, but also appeal to the notion of conceptual accessibility. This refers to the ease with which concepts are represented in thought and/or retrieved from memory (Bock and Warren, 1985) and has itself been equated in the literature with salience (e.g., Prat-Sala and Branigan, 2000; Gleitman et al., 2007; Myachykov et al., 2012). To synthesize these notions under one umbrella term, and as a notational convenience, for the remainder of this paper we refer to this account as the *cognitive salience hypothesis*.

While each of the three hypotheses described above – noisy channel, role conflict and cognitive salience – is consistent, at least in part, with the results on which it is based, other studies have reported results that run counter to all three. Kocab et al. (2018), for example, found that when people described reversible events involving two *inanimate* entities, such as a car crashing into a truck, verb medial orders were rarely used, contrary to the predictions of the noisy channel hypothesis. In addition, Kocab et al. (2018) found that, as for animate-animate events, people tended to avoid orders containing an OV cluster when describing inanimate-inanimate events. This finding cannot be explained by the role conflict hypothesis, since there was no tendency to embody inanimate entities or to physically enact events in which an inanimate entity was the subject. Lastly, Kocab et al. (2018) found that OSV was the most common order for inanimate-inanimate events, with SOV accounting for a minority of responses. Similar results were reported in an earlier study by Gershkoff-Stowe and Goldin-Meadow (2002). This result is inconsistent with the cognitive salience hypothesis, which predicts roughly equal proportions of SOV and OSV since the interacting entities occupy the same level on the animacy hierarchy and are therefore equally salient. Another study, by Kline et al. (preprint), sought to test the noisy channel hypothesis against the role conflict hypothesis by modifying the silent gesture paradigm such that participants explicitly marked the locations of entities in the event using one hand for each. This modification was intended to act as a case-marking system, thereby reducing the pressure to use word order to disambiguate participant roles. Accordingly, Kline et al. (preprint) predicted that under the noisy channel hypothesis there would be no conditioning of word order on event type. The role conflict hypothesis, on the hand, predicted that animacy/reversibility-conditioned word order would not be affected by case marking. Consistent with the noisy channel hypothesis, participants produced predominantly SOV-like orders for both reversible and non-reversible events. However, an analysis of the gestures used in this task revealed that the use of spatial marking was correlated with a reduction in the use of body-based gestures, thereby reducing the potential for role conflict. The findings from this study were therefore inconclusive about the mechanism underlying the distribution of word orders in this task.

As the preceding discussion demonstrates, each of the three hypotheses can account for some of the results reported in the silent gesture literature, but falls short of providing a full explanation. Table 5.1 provides a summary of the three hypotheses and the evidence for and against each.

---

<sup>2</sup>Meir et al. (2017) use the terms ‘salient’ rather than ‘basic’.

Table 5.1: Summary of the three main hypotheses that have been proposed in the silent gesture literature to explain the effects of animacy on word order variation.

Hypothesis	Main claim	Evidence for	Evidence against
Noisy channel (Gibson et al., 2013)	SVO preferred for reversible events; word order used to disambiguate participant roles.	<ul style="list-style-type: none"> <li>• SVO preferred for reversible events in some silent gesture experiments involving speakers of SVO and non-SVO languages (e.g., Futrell et al., 2015)</li> <li>• SVO preferred for reversible events in some sign languages (Napoli and Sutton-Spence, 2014)</li> <li>• More SOV languages use case marking than do SVO languages (Sinnemäki, 2010)</li> </ul>	<ul style="list-style-type: none"> <li>• Some gesture studies, including Gibson et al. (2013), do not find SVO preference for reversible events in speakers of non-SVO languages (e.g., Meir et al., 2017)</li> <li>• SOV and OSV common for reversible events in some studies (e.g., Hall et al., 2013; Meir et al., 2017)</li> <li>• Reversible events involving inanimate entities do not elicit SVO (Kocab et al., 2018; Gershkoff-Stowe and Goldin-Meadow, 2002)</li> <li>• No evidence that SVO is preferred for reversible events in emerging sign languages (e.g., Flaherty, 2014)</li> </ul>
Role conflict (Hall et al., 2013)	OV clusters avoided for reversible events; word order reflects production constraints specific to the gestural modality.	<ul style="list-style-type: none"> <li>• Avoidance of OV clusters for reversible events (e.g. Hall et al., 2013; Kocab et al., 2018)</li> </ul>	<ul style="list-style-type: none"> <li>• Avoidance of OV clusters for reversible events in the absence of embodiment (Kocab et al., 2018)</li> <li>• SOV common for reversible events in some studies (e.g., Gibson et al., 2013; Meir et al., 2017)</li> </ul>
Cognitive salience (Meir et al., 2017)	Word order in emerging communication systems reflects the relative salience of entities interacting in an event.	<ul style="list-style-type: none"> <li>• Language production studies demonstrate that animate entities tend to be expressed before inanimate entities (e.g. Prat-Sala and Branigan, 2000)</li> <li>• Studies of emerging sign languages demonstrate SOV and OSV equally preferred for reversible events (e.g., Flaherty, 2014; Ergin et al., 2018)</li> </ul>	<ul style="list-style-type: none"> <li>• Reversible events involving inanimate entities elicit few SOV responses (Kocab et al., 2018)</li> <li>• Studies showing speakers of non-SVO languages prefer SVO for reversible events (e.g., Futrell et al., 2015)</li> </ul>

## 5.2 Testing the noisy channel and cognitive salience hypotheses

While the role conflict hypothesis appeals to production constraints specific to the gestural modality, the noisy channel and cognitive salience hypotheses are modality independent. Further, as discussed in the previous section, each of these two hypotheses finds some support from spoken-language typology and language production studies. For example, the suggestion that SVO is communicatively more transparent than SOV, as claimed by the noisy channel hypothesis, is supported by the observation that SOV languages are more likely to explicitly mark for case than are SVO languages (Sinnemäki, 2010). The cognitive salience hypothesis,

on the other hand, finds support in spoken-language production studies showing that animate entities tend to be mentioned earlier than inanimate entities (e.g., Prat-Sala and Branigan, 2000; van Nice and Dietrich, 2003; Branigan et al., 2008; Dennison, 2008; van de Velde et al., 2014; Esaulova et al., 2019), and is grounded in hypotheses concerning cognition (Gentner and Boroditsky, 2001) and language processing (Bock and Warren, 1985). Moreover, each of the two hypotheses has been presented as an explanatory framework for the distribution of word orders found in *spoken* rather than sign languages. Therefore, while there are undoubtedly additional modality-specific factors which might differentiate how word order patterns emerge in sign and spoken languages, for example, the availability of embodiment and iconicity in the manual modality (Kocab et al., 2018), it is important to test the validity of the specific hypotheses targeted here in a non-gestural modality. In this study, we adopted an artificial language learning paradigm to test and compare the two hypotheses. We conducted three experiments in which participants were trained on a novel language and asked to describe reversible and non-reversible events. Each experiment was preregistered prior to data collection. Supporting material is available on the Open Science Framework (OSF) at <https://osf.io/hvw2d>.

The approach adopted in this study offers a number of advantages over the silent gesture paradigm. First, as mentioned, it allows us to test if word order conditioning on animacy can be replicated in a non-gestural modality. Replicating findings from silent gesture would provide strong evidence that animacy effects represent general, modality-independent biases and pressures that shape linguistic structure across modalities. In addition, this would provide evidence against the role conflict hypothesis, or at least show that the phenomenon it seeks to explain can be accounted for by other modality-general processes. In contrast, failure to find any conditioning on animacy might indicate that such effects are modality specific.

Second, the methodology adopted in this study allows for greater control over the responses that participants can provide, and over the specific constructions available to them. By design, silent gesture studies typically impose few restrictions on how people respond to stimuli. Consequently, data are often extremely noisy, with responses ranging from single-constituent gestures to highly repetitious sequences in which the same constituents appear multiple times. In addition, in the gestural modality constituents may be expressed simultaneously. These factors can complicate the interpretation of results and reduce the amount of data that can be included in the analysis. Moreover, there is little consistency across the gesture literature as to how responses should be coded. For example, where one study may code an agent-action-patient-action response as SVOV, another might analyze the same sequence as two separate clauses (SV and OV). The paradigm adopted in this study restricts the types of responses available to participants, which results in fewer coding decisions and therefore restricts experimenter degrees of freedom.

Despite these advantages, there are nevertheless a number of key differences between silent gesture and the artificial language learning paradigm adopted in this study that may bear on the interpretation of the results. We have already noted the possible effects of the manual modality in driving word order preferences in silent gesture and emerging sign languages. Another possible factor concerns the explicitly language-like character of an artificial language learning task. This may influence, for example, the extent to which participants draw on previous linguistic experience (see §5.2.1). Another difference concerns learning and recall, which are not typically a concern in silent gesture studies. Specifically, some words may be more easily learned and/or retrieved than others and therefore more likely to be mentioned earlier. It is worth noting

however that a similar phenomenon may also operate in silent gesture. That is, word order may reflect how easily a concept can be conveyed using gesture. However, to our knowledge, there are no studies investigating this question making it difficult to formulate clear predictions about incrementality effects across modalities.

In the next section, we detail the predictions that each hypothesis makes about how people should describe events as a function of animacy. Before we do this, a comment on notation: It is common practice in silent gesture and artificial language learning studies to equate agents with syntactic subjects and patients with objects, ostensibly as a notational convenience. We do not follow this convention for two reasons. First, we make no assumptions about the syntactic status of event constituents in our study. Second, and more importantly, using this convention may obscure or distort possible relationships between the orders people use in the experimental setting and constructions available in their native language. For example, a patient-action-agent utterance produced by an English speaker could be considered analogous to the passive construction. Coding the utterance OVS obscures this relationship and could potentially lead one to erroneously conclude that word order choices do not conform to the rules or conventions of the participant's native language. To avoid these problems we code words according to their semantic role. We use 'A' for agent, 'P' for patient, and, to avoid confusion with agent, 'V' for action.

### 5.2.1 Predictions

The noisy channel and cognitive salience hypotheses make different predictions about how animacy influences the way people describe events given certain restrictions on the word orders available to them. In this study, participants were asked to describe a series of reversible and non-reversible events using an artificial language in which one constituent always appeared in a fixed position and the positioning of the other two constituents was variable. There were two language types: one in which the patient was always expressed in initial position and the ordering of the agent and action was flexible (P-first); and one in which the action was always expressed last and the agent and patient were flexibly ordered (V-final). The P-first language permitted both patient-agent-action (PAV) and patient-action-agent (PVA). The V-final language allowed agent-patient-action (APV) and patient-agent-action (PAV). Participants were assigned to one of these two language types. Since both language types permitted PAV, we describe the predictions in terms of this order. Fig. 5.1 shows a schematic illustration of the predictions, which we describe below for each language type.

#### 5.2.1.1 P-first language

According to the noisy channel hypothesis, verb-medial orders should be preferred for reversible events. Thus, when presented with the choice of PAV or PVA, people should produce a higher proportion of PVA relative to PAV. For non-reversible events, the noisy channel hypothesis makes no direct prediction. We would therefore expect people to either probability match, that is, to produce each order in the same proportion as it appeared in the training input (e.g., Hudson Kam and Newport, 2005; Smith and Wonnacott, 2010), or if no training is provided, to alternate randomly between the two orders. Note that this prediction assumes that no other pressures or biases are operational. Nevertheless, it is important to acknowledge that the participants in this study may bring prior expectations to the task. For example, based on

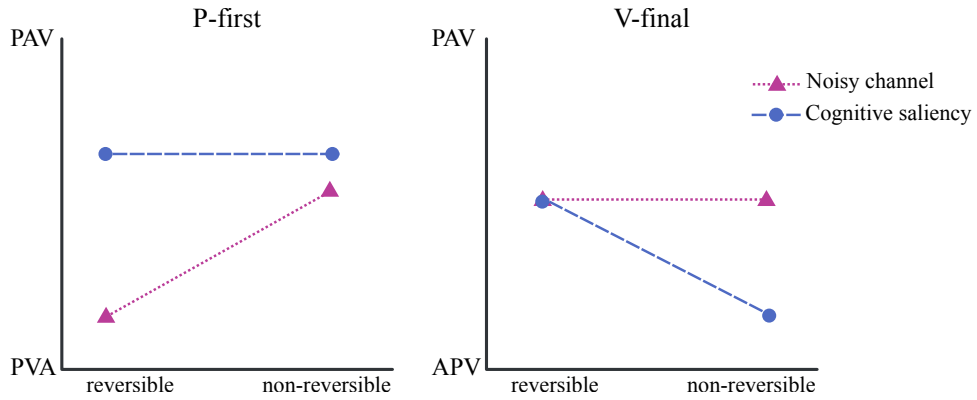


Figure 5.1: Schematic illustrating the overall pattern of responses predicted by the noisy channel hypothesis and the cognitive saliency hypothesis for a P-first language (left) and a V-final language (right). Moving up the y-axis indicates increasing use of PAV and a corresponding decrease in the other available order.

their linguistic experience, they may expect that one word order should be more frequent than another. The effect of this would be to shift the overall proportion of PAV up or down, but would not be predicted to influence conditioning on event type.<sup>3</sup>

The cognitive saliency hypothesis states that order of mention reflects the relative cognitive saliency of entities in an event. Under this hypothesis, we would expect to see an overall preference for PAV over PVA since relations are less cognitively salient, or basic, than concrete entities and so tend to be mentioned last (see also Goldin-Meadow et al., 2008). The hypothesis predicts no conditioning on event type since the language does not permit flexible ordering of the agent and patient.

### 5.2.1.2 V-final language

Given a choice between APV and PAV, the cognitive saliency hypothesis predicts that APV should be preferred over PAV for describing non-reversible events, since it places the more salient human agent before the less salient inanimate patient. For reversible events, APV and PAV should appear in roughly equal proportions reflecting the equal saliency of agent and patient.

Under the noisy channel hypothesis, both APV and PAV are sub-optimal for communicating about reversible events. We would therefore expect no conditioning on event type, and, in the absence of any other pressures or biases such as an effect of prior linguistic experience, probability matching or alternating randomly between the two orders.

## 5.3 Experiment 5: Free production task

Participants were trained on a novel lexicon and a grammar. The grammar of the language conformed to either the P-first or V-final rule. Preregistration files can be obtained at <https://osf.io/b2xuj>.

<sup>3</sup>We thank an anonymous reviewer for highlighting this point.

### 5.3.1 Method

#### 5.3.1.1 Participants

We recruited 60 participants through Amazon Mechanical Turk. Participants were at least 18 years old and were self-reported native English speakers. They were paid \$5.00 for their participation.

#### 5.3.1.2 Materials

All experiments in this study were developed using the jsPsych library (de Leeuw, 2015). Elicitation stimuli consisted of a set of cartoon images depicting transitive events in which a human agent acted on either a human or an inanimate patient. The stimuli depicted four humans, which appeared as both agents and patients, three inanimate objects, and four actions, resulting in a set of 11 individual concepts. The inanimate objects were designed to be similar in size and scale to the human entities to control for the possibility that more visually prominent entities may be preferentially expressed before less prominent entities (Myachykov et al., 2011). Fig. 5.2 shows an example of each type of concept (human, object or action). Combining these concepts resulted in 96 transitive events evenly distributed across reversible and non-reversible events (the human referents in reversible events always differed from each other). Fig. 5.3 shows two examples of the event images used to elicit event descriptions.

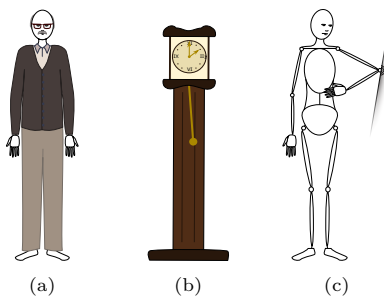


Figure 5.2: Experiment 5: Examples of images presented during vocabulary training. The images depicted (a) humans, (b) inanimate objects, and (c) representations of actions.

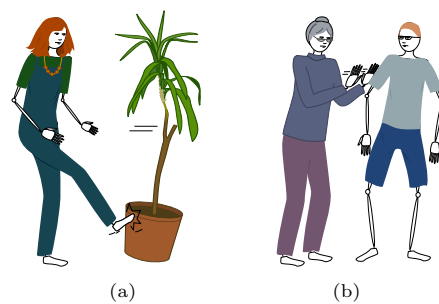


Figure 5.3: Experiment 5: Example stimuli showing (a) a non-reversible event in which a girl is kicking a potted plant, and (b) a reversible event in which a woman is pushing a boy.

Each language was constructed by randomly associating one of 11 one- or two-syllable words with one of the 11 concepts. A different set of concept-label associations was generated for each participant. Event descriptions consisted of three uninflected constituents ordered according to the appropriate language type rule.

### 5.3.1.3 Procedure

Participants were randomly assigned to one of the two language types (P-first or V-final). They were told that they would learn a new language called Shebrish. They were further advised that the experiment was divided into different phases and that they would be given full instructions on how to complete the task before each phase. Participants were given feedback on their responses during training. No feedback was provided in the critical testing phase. Progression through trials in each phase was self-paced with no time limits imposed.

The experiment consisted of two lexical training and four grammar training phases. This multi-phase approach was intended to add interest to the task and facilitate learning. Screenshots from each phase and a demo video of the experiment are provided on the OSF repository.

**Phase 1 – Vocabulary Exposure:** This phase introduced participants to the vocabulary of the novel language. Each item (human, object or action) was presented along with its label in Shebrish for 2.5s. Participants were then asked to identify the corresponding word for the item by clicking on one of three options. The full set of items was shown three times in random order.

**Phase 2 – Vocabulary Comprehension:** Participants were shown pairs of items of the same type (for example, two humans) along with a single label. They were asked to click on the item that matched the label. The linear arrangement of the target item and distractor was randomized across trials. The full set of labels was presented three times in random order.

**Phase 3 – Sentence Exposure:** On each trial, an event image was presented along with its description in the novel language. Participants were asked to click on the word corresponding to one of the three constituents, for example, the one doing the action. They were shown 16 events such that each action was shown 4 times, twice as part of a reversible event and twice as part of a non-reversible event. Each of the two permitted word orders appeared an equal number of times. Trials were pseudo-randomly ordered such that consecutive events differed in all three constituents. Across trials, the orientation of the event image was randomized such that the agent appeared either to the left or to the right of the patient.

**Phase 4 – Vocabulary Production (1):** Participants were presented with 14 previously unseen events (7 reversible and 7 non-reversible), pseudo-randomly ordered across trials. On each trial, the randomly oriented event was shown with a partially completed description. The missing constituent was replaced with a text box. Participants were instructed to type the word for the target (missing) constituent (for example, the action). Each of the two permitted word orders appeared an equal number of times.

**Phase 5 – Sentence Selection:** This phase consisted of 16 pseudo-randomly ordered trials in which previously unseen, randomly oriented events (8 reversible and 8 non-reversible) were presented along with two descriptions. Each description contained the same words, but in a different order. Only one description constituted a valid word order for a given language type. Participants were instructed to click on the correct description. Each of the two permitted word orders appeared as a target description an equal number of times. The order of the distractor description was chosen at random.

**Phase 6 – Vocabulary Production (2):** This phase was the same as Phase 4 – Vocabulary Production (1) and provided another opportunity to practice typing the newly learned vocabulary before the critical testing phase. Participants were presented with 14 previously unseen

events.

**Phase 7 – Sentence Production:** This was the main testing phase. Participants saw 36 previously unseen, randomly oriented events (18 reversible and 18 non-reversible) presented in pseudo-random order. On each trial they were instructed to type the description of the scene using the novel language. No feedback was provided in this phase.

### 5.3.1.4 Coding responses

Word order on each sentence production trial was determined by comparing the description supplied by the participant with the event constituent labels. If the description did not contain exactly three words, the trial was marked ‘ambiguous’. To accommodate spelling or lexical errors, we attempted to identify the intended constituent by comparing the supplied word to each of the words in the lexicon. Details of this procedure are provided in Appendix C.2. Responses that did not contain a single occurrence of each constituent in an order consistent with the language type were treated as ‘invalid’.

### 5.3.1.5 Analysis

Data from each language type were analysed separately using a mixed effects logistic regression model. All statistical analyses were carried out using the *lme4* package (Bates et al., 2015b) in R (R Core Team, 2017). The preregistered model included event type (reversible or non-reversible) as the fixed effect, and participant (intercept and slope) and event (intercept) as random effects. The dependent variable was a binary flag indicating if the participant used PAV word order on a trial. We used dummy coding for event type. The model specification in *lme4* format was  $is\_pav \sim event\_type + (1 + event\_type|workerId) + (1|event)$ .

As the schematic in Fig. 5.1 illustrates, within each language type, the noisy channel and cognitive salience hypotheses make different predictions about the proportion of PAV responses for a given event type, and about the direction of the effect. Given the nature of the predictions, it was convenient to use different reference levels for event type for each language type. Table 5.2 specifies the event type reference level for each language type and summarizes the predicted outcome of the statistical analysis.

Table 5.2: Experiment 5: Summary of predictions of the statistical analysis

Language type	Event type reference level	Predictions	
		Cognitive salience	Noisy channel
P-first	non-reversible	Majority PAV: significant (+ve) intercept No effect of event type: non-significant slope	No preferred order: non-significant intercept ↑ PVA for reversible: significant (-ve) slope
V-final	reversible	No preferred order: non-significant intercept ↑ APV for non-reversible: significant (-ve) slope	No preferred order: non-significant intercept No effect of event type: non-significant slope

## 5.3.2 Results

As per the pre-registered plan, participants who scored below 70% lexical accuracy in the critical trials were excluded from the analysis. Eight participants were excluded on this basis.

In addition, trials coded ‘ambiguous’ or where the word order was invalid were excluded. Three participants provided an invalid or ambiguous order on every trial and were therefore excluded. This left 49 participants, 26 in the P-first group and 23 in the V-final group. Of the remaining data set, 168 trials (17.9%) were excluded from the P-first group on the basis that they were either invalid or ambiguous. A total of 17 (2.1%) were excluded from the V-final group. This is considerably fewer than in the P-first group, an observation that we return to below. This left 768 trials in the P-first group (382 non-reversible and 386 reversible) and 811 in the V-final group (405 non-reversible and 406 reversible).

### 5.3.2.1 Proportion of PAV responses

Fig. 5.4 shows the proportion of responses coded as PAV plotted by event type for each language.

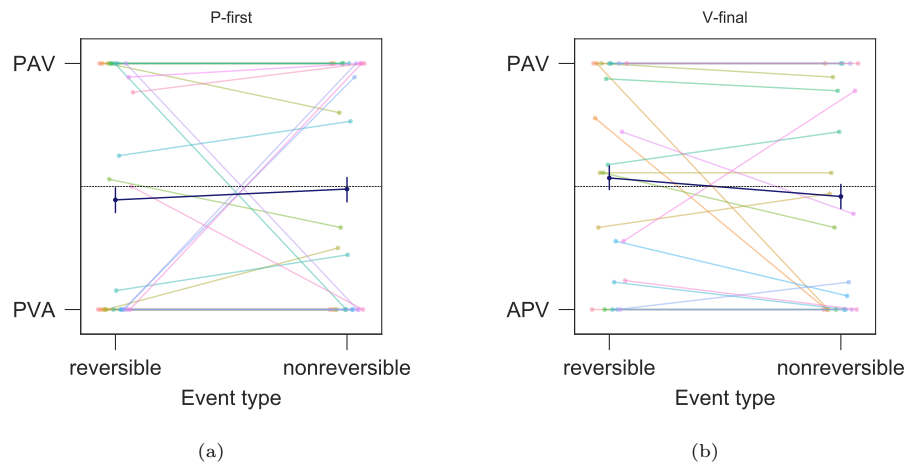


Figure 5.4: Experiment 5: The proportion of PAV responses by event type for (a) the P-first group and (b) the V-final group. The dark lines indicate the averaged responses. Colored lines represent the averaged responses for an individual participant. In the P-first group, there was an overall tendency to produce fewer PAV (hence more PVA) responses when describing reversible events, consistent with the noisy channel hypotheses. In the V-final group, participants on average produced fewer PAV (hence more APV) responses when describing non-reversible events compared with reversible events, as predicted by the cognitive salience hypothesis. These differences were not statistically significant.

As predicted by the cognitive salience hypothesis, APV and PAV occurred in roughly equal proportions for reversible events in the V-final group (estimate for intercept: 53.4% PAV;  $\beta=0.604$ ,  $SE=0.996$ ,  $p=0.544$ ).<sup>4</sup> Non-reversible events elicited fewer PAV responses (45.9%) than APV, however, the difference was not significant at the  $p<0.05$  level (estimate for slope:  $\beta=-1.196$ ,  $SE=0.672$ ,  $p=0.075$ ). These results therefore did not provide sufficient evidence for an effect of event type consistent with the cognitive salience hypothesis.

Similarly, we found no evidence for an effect of event type in the P-first group, contra the noisy channel hypothesis. People were slightly less likely to produce PAV (hence, more likely to produce PVA) for reversible events (44.6%) compared with non-reversible events (49.0%). However, this difference was not significant ( $\beta=-1.4292$ ,  $SE=1.482$ ,  $p=0.335$ ). In addition, contrary to the predictions of the cognitive salience hypothesis, PAV was not preferred over

<sup>4</sup>The pre-registered model reported a convergence warning for both language types. Simplifying the random effects structure removed the convergence warning, but had no substantial effect on the coefficient estimates. We therefore retained the full, pre-registered random effects structure in all of our analyses.

PVA (estimate for the intercept:  $\beta=-0.619$ ,  $SE=1.384$ ,  $p=0.655$ ).

In both language types, the averaged responses appeared consistent with probability matching, that is, each word order variant was produced in the same proportion as it appeared in the training data. However, analysis of individual-level responses indicated that people’s responses were much more categorical than would be expected under probability matching. Exactly half of the participants in the P-first group used a single order on all analysed trials. Of these, five consistently used PAV and eight used PVA. The tendency to regularize was somewhat less pronounced in the V-final group. Here, 39% of participants used a single order. APV was used consistently by four participants, and five people used PAV.

What might account for these findings? One possible explanation is that people had difficulty learning the grammar of the language on which they were trained, failing to produce its variable word order. It is notable that both the proportion of trial exclusions and the tendency to produce a single order were greater in the P-first group than in the V-final group, suggesting a possible correlation between a failure to learn the language rule and the tendency to regularize towards a single order. Previous studies have found that regularization of variable input decreases as the learning task becomes easier (e.g., Ferdinand et al., 2019; Samara et al., 2017). Alternatively, it is possible that participants *did* learn that the language permitted variable word order, but adopted a strategy of using only one order in the critical phase. Ferdinand et al. (2019), for example, found that people tended to produce more regular output even when they had learned the distribution of variants in the input. In the present study, this tendency may have resulted from a bias deriving from the participants’ previous linguistic experience, as we noted in §5.2.1. Either way, participants in the present experiment may have settled on a single word order as a means of reducing cognitive load during recall of newly learned lexical items.

### 5.3.3 Discussion

Results from the V-final group were qualitatively in line with the cognitive salience hypothesis. Participants showed a slight preference for APV when describing non-reversible events, while producing APV and PAV in roughly equal proportions for reversible events. However, the effect of event type was at best marginal. Moreover, there was no evidence that participants in the P-first group had an overall preference for PAV over PVA. We therefore cannot claim that these data support the cognitive salience hypothesis. Equally, analysis of the P-first data provided no evidence to support the noisy channel hypothesis.

Although the averaged responses in both language types appeared to be consistent with probability matching, we found that a substantial proportion of participants used a single order on all trials. This may have been due to the cognitive demands of the task (e.g. Ferdinand et al., 2019), which required participants to learn then reproduce both a vocabulary and a grammar in a relatively short space of time. To investigate this possibility, we ran a second experiment intended to make the learning task easier. We predicted that decreasing cognitive load in the critical trials would reduce the tendency to use a single order and therefore allow us to detect conditioning of word order on event type.

## 5.4 Experiment 6: Restricted production task

In this experiment we implemented two changes to the design of Experiment 5 aimed at reducing the tendency to regularize towards a single word order by facilitating learning and lessening cognitive load in the critical trials. First, studies have found that regularization of variable input is reduced when the vocabulary is more familiar, or more easily retrieved from memory (e.g., Samara et al., 2017). We therefore changed the lexicon of the novel language to make it more English-like. For example, some labels were phonetically similar to their English counterparts (e.g. *kerla* for *girl*), or contained some element that had an association with the referent (e.g. *legip* for *kick*).

Second, we eliminated grammar training and only trained participants on the novel vocabulary. In the test trials, participants were presented with an event image accompanied by a partially completed description consisting of a single constituent. In the P-first language, the patient label was always provided and appeared in sentence-initial position; in the V-final language, the action label was provided in sentence-final position. Participants were asked to complete the description by providing the missing constituents. Thus, instead of requiring participants to learn and reproduce the grammar of the language, the rule was made explicit through the provision of the relevant fixed-position constituent on every trial. In addition to reducing the learning demands placed on participants, eliminating grammar training had the added advantage of making the task more faithful to the improvisational character of silent gesture where word order is not modeled prior to testing.

This experiment was preregistered prior to data collection and following collection and initial analysis of data for Experiment 5. Preregistration files can be obtained at <https://osf.io/vph48>.

### 5.4.1 Method

#### 5.4.1.1 Participants

The duration of this experiment was shorter than Experiment 5 due to the absence of grammar training and reduced vocabulary exposure. We therefore increased the number of participants relative to Experiment 5, recruiting 80 people through Amazon Mechanical Turk. We also implemented a filtering procedure so that only people who learned the novel vocabulary to a determined level proceeded to the final testing phase. Participants were paid between \$1.70 and \$2.50 for their participation, depending on how far they progressed through the experiment. They were at least 18 years old and were self-reported native English speakers.

#### 5.4.1.2 Materials

The set of visual stimuli was the same as in Experiment 5. A new lexicon was generated, as described above. Since each word was intended to have some association with its referent, item-label mappings were the same for all participants. These are shown in Table 5.3.

#### 5.4.1.3 Procedure

As in Experiment 5, participants were randomly assigned to one of the two language types. Progression was self-paced and feedback was provided in all but the critical testing phase. Screenshots from each phase and a demo video are provided on the OSF repository.

Table 5.3: Experiment 6: List of item-label mappings.

Item	Label
girl	kerla
boy	poyod
old woman	nanoo
old man	papee
potted plant	potri
bird feeder	burfu
clock	tiktu
kick	legip
prod	pinga
push	shufa
elbow	nokor

**Phase 1 – Vocabulary Exposure:** This phase introduced participants to the novel vocabulary. Each item (human, object or action) was presented along with its corresponding label in the novel language for 2.5s. Participants were then asked to identify the correct word for the item by clicking on one of three options. The participant was forced to provide the correct response before progressing to the next trial. The full set of items was shown twice, in random order.

**Phase 2 – Vocabulary Exposure - typed response:** As in Phase 1 – Vocabulary Exposure, participants were presented with an item and its corresponding label for 2.5s. The label then disappeared and participants were asked to provide the word for the item by typing. The full set of items was presented twice, in random order.

**Phase 3 – Vocabulary Testing:** In this phase, items were presented without their corresponding label. Participants were instructed to provide the label by typing. The full set of items was presented three times, in random order. Participants were informed that if they learned the language well enough, they would proceed to the bonus phase. The minimum score required to progress to the final phase was 24 (out of 33). A running score was visible to the participant throughout this phase. However, they were not informed of the target score.

**Phase 4 – Sentence Completion:** This was the main testing phase. Participants were informed that they would be shown pictures of scenes and a partially completed description. On each trial, a single constituent was provided, as described above. Participants were instructed to complete the description by typing in the spaces provided. As in Experiment 5, events were shown in pseudo-random order and the orientation of each image was randomized across trials. No feedback was given. There were a total of 32 trials (16 reversible and 16 non-reversible events).

#### 5.4.1.4 Coding responses

The coding procedure was largely the same as in Experiment 5, with very minor changes to reflect the fact that participants were required to provide two constituents instead of three.

#### 5.4.1.5 Analysis

We used the same mixed effects logistic regression model as in Experiment 5 (§5.3.1.5).

## 5.4.2 Results

Eighty participants completed 32 trials each, except one, for whom three trials were dropped due to a technical error. This resulted in a total of 2557 trials. One hundred trials were excluded because the supplied order was invalid. All but one of these occurred in the P-first group, accounting for around 8% of responses in this group. The final set of analysed trials consisted of 1178 from the P-first group (598 reversible and 580 non-reversible) and 1279 from the V-final group (639 reversible and 640 non-reversible). Data from all 80 participants were included in the analysis. We first discuss the extent to which individual participants varied their word order, then look at the proportion of PAV responses in each group.

### 5.4.2.1 Word order variability

The design changes in this experiment were intended to reduce regularization by making the learning task easier. This manipulation was not successful. In fact, a much higher proportion of participants produced a single order in this experiment compared with Experiment 5. In the P-first group, only nine out of 40 participants varied their responses. In the V-final group, 20 participants used a single order and 20 varied their responses. In both groups, and in contrast to Experiment 5, there was a clear preference for one order over the other among those who did not vary their responses. Of the 31 participants who consistently used a single order in the P-first group, 28 used PVA. In the V-final group, 19 out of 20 participants used APV on all trials. We suggest that these preferences can be attributed to an effect of the participants' native language. We return to this in §5.4.3.

### 5.4.2.2 Proportion of PAV responses

The proportion of responses coded as PAV in each language group is shown in Fig. 5.5. Overall, participants in both groups showed a strong tendency to avoid PAV. In the V-final group, PAV was produced in only 14.6% of trials. Reversible events were described with PAV in 15.3% of trials. This was significantly less than the 50% predicted by the cognitive salience hypothesis (estimate for the intercept:  $\beta=-3.731$ ,  $SE=0.696$ ,  $p<0.001$ ). Non-reversible events elicited fewer PAV responses (13.9%), however, the difference was not significant ( $\beta=-0.935$ ,  $SE=0.660$ ,  $p=0.156$ ). Similarly, in the P-first group we found no evidence for conditioning on event type: PAV accounted for 15.9% of reversible events compared with 16.2% of non-reversible events ( $\beta=-0.987$ ,  $SE=1.883$ ,  $p=0.6$ ). Contra the cognitive salience hypothesis, the overall proportion of PAV was low at only 16.0%.

## 5.4.3 Discussion

As in Experiment 5, we found no evidence to support either the cognitive salience or the noisy channel hypothesis. Contrary to both accounts, participants did not condition their responses on event type. In addition, despite our efforts to ease the learning task, the tendency to regularize was considerably more pronounced in this experiment compared with Experiment 5. Moreover, unlike the first experiment, there was a clear preference for a particular order both overall and among participants who used a single order on all trials: PVA in the P-first group and APV in the V-final group.

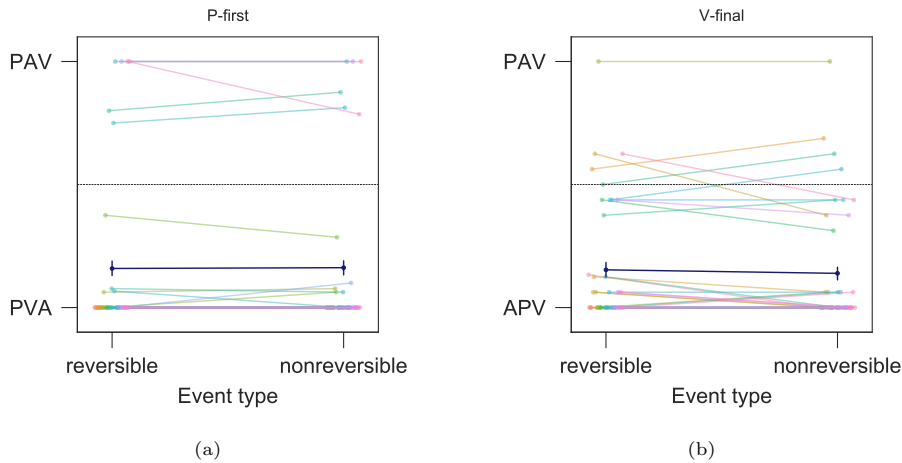


Figure 5.5: Experiment 6: The proportion of PAV responses by event type in (a) the P-first group and (b) the V-final group. Dark lines indicate averaged responses. Colored lines represent averaged responses for an individual participant. In both groups, participants showed a strong tendency to produce a single order. There was also a strong bias against PAV. These results indicate an overall bias in favor of PVA in the P-first group and a preference for APV in the V-final group.

One possible explanation for the latter observation is that participants in this experiment relied more heavily on their native language because they did not receive explicit modeling of the two possible orders. In the P-first group, the preferred order – PVA – is partially consistent with SVO in that it is verb-medial. Moreover, PVA could be associated with the English passive construction. In contrast, PAV is inconsistent with the canonical active English word order in that it is both patient-first and verb-final. In the V-final group, APV may have been preferred over PAV because, like canonical English utterances, it is agent-first. Alternatively, the preference for APV may reflect a more universal agent-first bias (Jackendoff and Wittenberg, 2017). This bias may derive from the salience of agents relative to patients (Rissman et al., 2018). It was not possible in the current study to differentiate between these two alternatives.

What can account for the greater tendency to use a single order in this experiment compared with Experiment 5, particularly given the semi-improvisational nature of the task? One possibility is that the set of expectations people bring to a task differs across modalities. Whereas the unrestricted nature of silent gesture may result in a high degree of variability, a written task may elicit a stronger association with a formal linguistic system. Hence, based on prior linguistic experience, participants may have assumed that there should be a single ‘correct’ order in which to express constituents. In the absence of an explicit grammatical model, participants may have drawn on their native language when trying to determine what that order should be.

These effects – native language and task-specific associations – may have masked or even overridden the influence of animacy on word order. In the next section, we report on a post-hoc exploratory analysis in which we investigated another possibly confounding effect: word order conditioning on image orientation. To pre-empt the findings, we found a strong effect of image orientation in the V-final group in both Experiment 5 and Experiment 6 reflecting a tendency to express the patient before the agent more often when it appeared on the left side of the image compared with when it appeared on the right.

## 5.5 Image orientation and word order in Experiments 5-6

Previous studies have found evidence that word order choices may be influenced by image orientation (e.g., Ferreira, 1994; Esaulova et al., 2019). In the two experiments reported above, the orientation of the event images was randomized across trials to mitigate against any possible systematic effects on constituents order. However, the pre-registered statistical model did not include orientation as a predictor, hence, any such effect was not controlled for statistically. In this section, we address this issue and investigate if participants were sensitive to the orientation of event images.

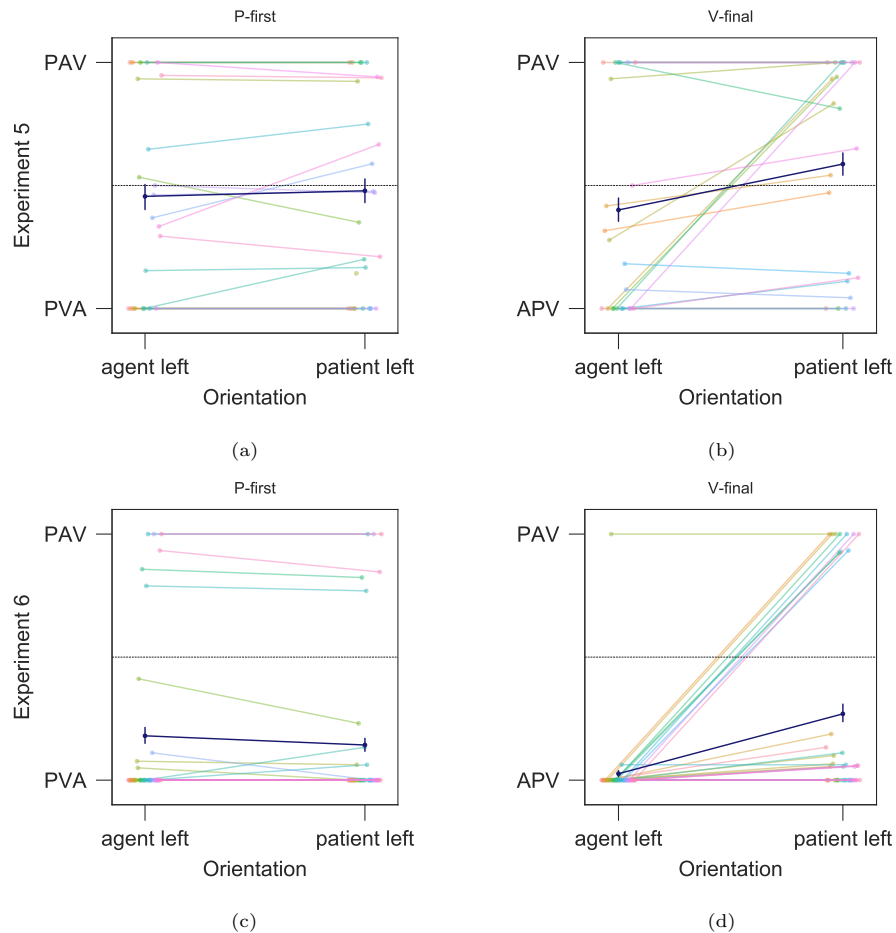


Figure 5.6: Experiments 5 and 6: The proportion of PAV responses by image orientation in the P-first group – (a) and (c) – and the V-final group – (b) and (d). Dark lines show the averaged responses. Colored lines indicate the means for each participant. In the V-final group people were significantly more likely to produce PAV when the patient appeared on the left of the image. There was no effect of image orientation in the P-first group.

Figs. 5.6a and 5.6b show the proportion of PAV responses by image orientation in Experiment 5 in the P-first group and the V-final group, respectively. Figs. 5.6c and 5.6d plot the results from Experiment 6. In both experiments, participants in the V-final group were significantly more likely to produce PAV when the patient appeared on the left of the image. In Experiment 5, participants produced PAV on 58.7% of trials when the patient appeared on the left of the image, compared with 40.1% when the agent appeared on the left ( $\beta=4.229$ ,  $SE=1.730$ ,  $p=0.015$ ). In Experiment 6, PAV accounted for 26.9% of V-final trials when the

patient appeared on the left, compared with only 2.6% when the agent appeared on the left ( $\beta=10.593$ ,  $SE=1.677$ ,  $p<0.001$ ). Details of the statistical analysis are provided in Appendix C.3.1.

In the P-first group, we found no effect of orientation in either Experiment 5 ( $\beta=-0.064$ ,  $SE=0.369$ ,  $p=0.862$ ) or Experiment 6 ( $\beta=-0.363$ ,  $SE=0.406$ ,  $p=0.371$ ).<sup>5</sup> Further exploratory analysis, however, suggested a possible effect of image orientation on the tendency to produce invalid orders in the P-first group (see Appendix C.3.2). Fig. 5.7 shows the proportion of invalid orders in each experiment plotted by image orientation.

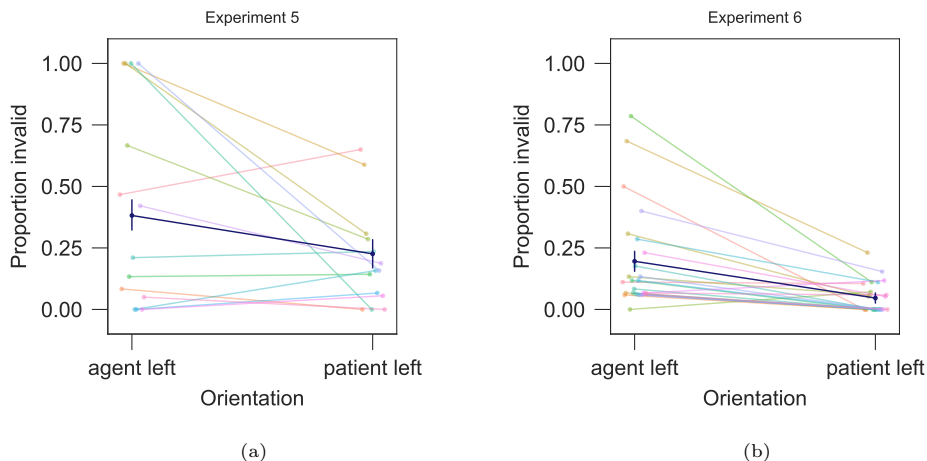


Figure 5.7: Experiments 5 and 6: The proportion of invalid responses in the P-first group by image orientation in (a) Experiment 5 and (b) Experiment 6. Dark lines indicate averaged responses. Colored lines represent the averaged responses for individual participants. In both experiments, there was a greater proportion of invalid orders when the agent appeared on the left of the image.

In Experiment 5, there was a greater tendency in the P-first group to produce an invalid order when the agent appeared on the left (38.2%) than when it appeared on the right (22.6%). However, these differences were not significant ( $\beta=-1.232$ ,  $SE=1.185$ ,  $p=0.298$ ). In Experiment 6, we found a significant effect of image orientation ( $\beta=-2.000$ ,  $SE=0.345$ ,  $p<0.001$ ) such that participants were more likely to provide an invalid order when the agent appeared on the left of the image (19.6% of trials included in the analysis) compared with when it appeared on the right (4.6%).

What can we conclude from this analysis of invalid orders? Although the effect of image orientation on the error rate was not significant in Experiment 5, it is nevertheless notable that in both experiments, errors were proportionally greater when the agent appeared on the left. Interestingly, although the most frequently produced invalid orders differed across the two experiments – AVP in Experiment 5 where participants provided all three constituents (84.3% of invalid orders) and PVP in Experiment 6 where participants completed a description in which the patient constituent was already provided (91.1%) – they have in common the final VP cluster. We hypothesize that this pattern of errors points to two underlying mechanisms. First, there is a tendency to produce orders that are consistent with one’s native language. Hence, VP occurred because it is consistent with SVO. Overlaying this is a tendency to express entities in the order that they appear on screen, moving from left to right. In the P-first group

<sup>5</sup>There was, however, a significant interaction between event type and orientation in Experiment 5 ( $\beta=2.126$ ,  $SE=0.845$ ,  $p=0.012$ ). It is not clear what the causal relationship underlying this effect might be.

in Experiment 6, where the patient constituent was provided, participants' native language experience may have led them to expect that the referent of the first NP was the agent. This expectation may have been reinforced when it aligned with the linear arrangement of entities on screen, that is, when the agent appeared to the left of the patient.

### 5.5.1 Discussion

Post-hoc analysis of responses in the V-final group suggested that word order was sensitive to the relative positioning of the agent and the patient in the scene. Specifically, people were more likely to express the patient before the agent (producing PAV) when the former appeared on the left side of the image. In the P-first group, we found suggestive evidence that word order errors were sensitive to the linear arrangement of the agent and patient. In both experiments, invalid orders, typically ending in VP, were more common when the patient appeared to the right of the agent. We hypothesize that this latter finding may have resulted from native-language effects modulated by a tendency to read the scene from left to right.

Previous studies have found that the effect of conceptual properties, such as animacy, on structural choices may be overridden by contextual factors, such as discourse prominence (Prat-Sala and Branigan, 2000) or visual salience (Rissman et al., 2018). It is possible that in the present study, a combination of native language interference and the influence of the visual layout of the scenes exerted a stronger influence on word order than did the conceptual properties of the event. In the next experiment, we attempted to eliminate the effect of image orientation and encourage participants to formulate a more holistic mental conceptualization of the event.

## 5.6 Experiment 7: Eliminating the influence of image orientation

In the previous two experiments, participants were presented with a static event image in the critical trials that remained on screen while they input the description. In this experiment, we implemented two changes that were designed to eliminate the possibility of 'reading' the scene from left to right, and to reduce the influence of image configuration on word order. First, the image was visible for only 4s, after which participants were asked to complete the description. Second, the image was presented rotating around the y-axis. To better understand how this looked, one can imagine a transparency on which an image is printed suspended from a piece of string and rotating around the vertical axis. As the transparency rotates, the image is mirrored from the original presentation on each half (180°) rotation. A number of parameters relating to the rotation of the image were randomized across trials: the direction of rotation (clockwise or anti-clockwise), the initial left-right orientation, and the initial angle of the image relative to y-axis. Screenshots of this presentation are provided on the OSF repository.

This experiment was preregistered prior to data collection and following collection and analysis of data from Experiment 5 and Experiment 6. Preregistration files can be obtained at <https://osf.io/wxz49>.

## 5.6.1 Method

### 5.6.1.1 Participants

We recruited 80 participants through Amazon Mechanical Turk. We used the same filtering procedure as in Experiment 6 and participants were paid between \$1.70 and \$2.50 for their participation. They were at least 18 years old and were self-reported native English speakers.

### 5.6.1.2 Materials

All materials were the same as in Experiment 6.

### 5.6.1.3 Procedure

The procedure was the same as in Experiment 6, except for the changes made to the critical trials described above. A video demonstrating each phase is provided on the OSF repository.

### 5.6.1.4 Coding responses

We used the same coding procedure as in Experiment 6 (§5.4.1.4).

### 5.6.1.5 Analysis

We analysed the data using the same logistic regression model described in §5.3.1.5.

## 5.6.2 Results

Four participants completed fewer than 32 trials in the critical testing phase due to technical error. One participant completed 31 trials, two completed 29 and one completed 26. This resulted in a total of 2547 trials. Of these, 27 (2.1%) were excluded from the P-first group and 10 (0.8%) from the V-final group because the order provided was not permitted by the language type or could not be unambiguously identified. The remaining data set contained 1247 P-first trials (622 reversible, 625 non-reversible) and 1263 V-final trials (632 reversible, 631 non-reversible). Data from all 80 participants were included in the analysis.

### 5.6.2.1 Proportion of PAV responses

As in Experiment 6, participants in both groups showed a strong tendency to use a single order across all trials and to avoid PAV. In the P-first group, 29 participants used a single order, of whom 24 used PVA on all trials. In the V-final group, 24 participants used a single order. All of these individuals used APV. The proportion of PAV responses was therefore consistent with Experiment 6 in being low overall in both language groups (see Fig. 5.8).

Looking first at the V-final group, we found that PAV accounted for 11% of trials overall. Reversible events elicited PAV on 12.2% of trials, which was significantly less than the 50% predicted by the cognitive salience hypothesis (intercept estimate:  $\beta=-4.799$ ,  $SE=1.152$ ,  $p<0.001$ ). For non-reversible events, PAV accounted for 9.8% of responses. Unlike the previous two experiments, this difference was significant ( $\beta=-4.615$ ,  $SE=2.093$ ,  $p=0.027$ ). The overall pattern was echoed at the individual level. Of the 16 participants who varied their responses, 12 used PAV more often when describing reversible events compared with non-reversible events. Three participants showed the opposite pattern and one produced PAV on exactly 50% of trials.

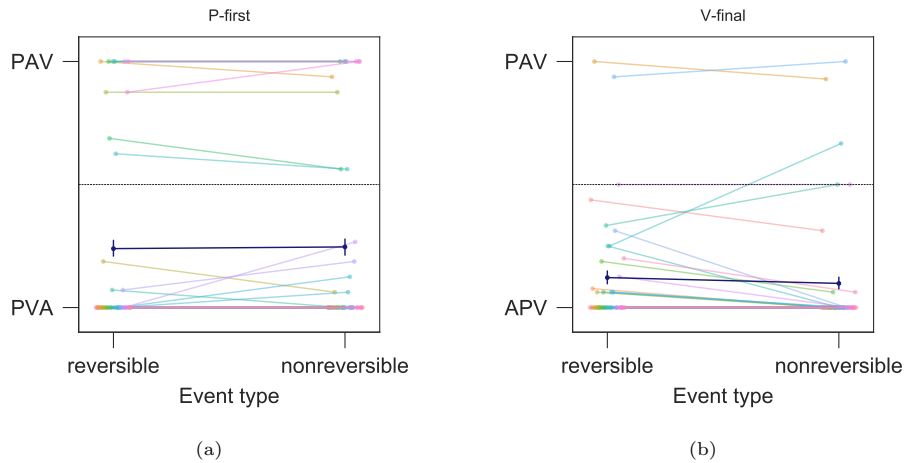


Figure 5.8: Experiment 7: The proportion of PAV responses by event type in (a) the P-first group and (b) the V-final group. Dark lines indicate averaged responses. Colored lines represent averaged responses for individual participants. As in Experiment 6, participants showed a strong tendency to produce a single order and an overall bias against PAV. Of the 16 participants in the V-final group who varied their responses, 12 used PAV more for reversible than for non-reversible events. In the P-first group, five out of 11 participants used more PVA for reversible than for non-reversible events.

Consistent with the two previous experiments, and contra the noisy channel hypothesis, we found no evidence for an effect of event type in the P-first group. Reversible events were described with PAV on 24.0% of trials compared with 24.6% of non-reversible trials ( $\beta=-1.687$ ,  $SE=1.916$ ,  $p=0.378$ ). Contrary to the cognitive salience hypothesis, participants tended to avoid PAV, which accounted for only 24.3% of responses overall (intercept estimate:  $\beta=-10.072$ ,  $SE=1.559$ ,  $p<0.001$ ). Unlike the V-final group, there was no consistent pattern of responses as a function of event type across participants: of the 11 participants who varied their choice of word order, 5 used PVA more for reversible events than for non-reversible events and 6 produced the opposite pattern.

### 5.6.3 Discussion

In this experiment, as in the previous two, we found no evidence to support the noisy channel hypothesis. Participants in the P-first group were not sensitive to the animacy of the patient, showing no tendency to prefer PVA over PAV for reversible events. In contrast to the previous two experiments, however, we did find partial support for the cognitive salience hypothesis. Although participants in the P-first group did not show an overall preference for PAV, contra the hypothesis, those in the V-final group were significantly more likely to produce PAV for reversible compared with non-reversible events, albeit at a lower rate than the predicted 50%. This effect was small, but was nevertheless consistent across participants who varied their word order choices. We discuss these deviations from predicted results in more detail in §5.8.3.

## 5.7 Combined analysis

In this section, we present an exploratory analysis of the combined data from all three experiments to investigate if there was a detectable effect of event type in each language type across the bigger sample.

Fig. 5.9 shows the proportion of responses coded as PAV plotted by event type across all three experiments. Across the three V-final groups, PAV accounted for 23.4% of reversible trials and 20.1% of non-reversible trials. Participants across the P-first groups produced PAV on 26.0% of reversible trials compared with 27.4% of non-reversible trials.

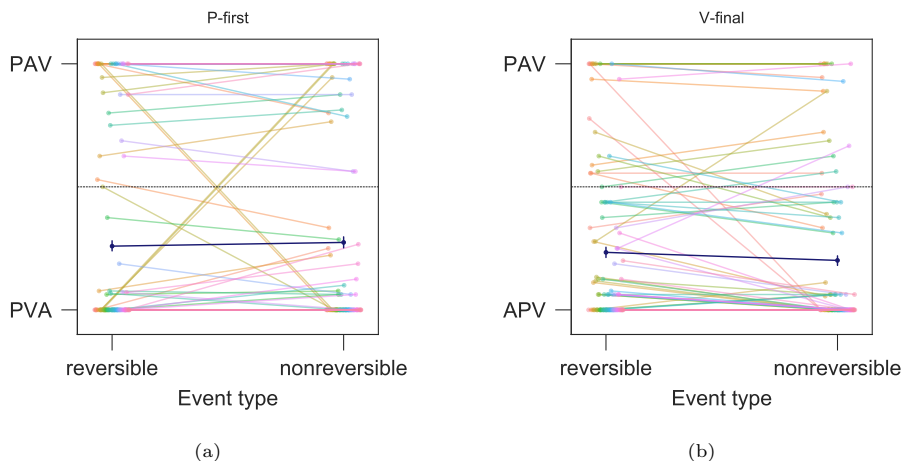


Figure 5.9: Experiments 5, 6, and 7: The proportion of PAV responses by event type across all three experiments in (a) the P-first group and (b) the V-final group. Dark lines indicate averaged responses. Colored lines represent averaged responses for individual participants. In the V-final group, participants produced significantly more PAV when describing reversible events compared with non-reversible events. There was no effect of event type in the P-first group.

### 5.7.1 V-final

We submitted the data to a mixed effects logistic regression analysis that included as fixed effects experiment, event type, the entity that appeared on the left of the image (‘entity left’), and their interactions. The maximal random effects structure that converged without warnings included a by-participant intercept and slopes for event type, entity left and their interaction.<sup>6</sup> We used deviation coding for the binary predictors event type and entity left and Helmert coding for the categorical variable experiment. The model specification in *lme4* format was:  $is\_pav \sim exp * entity\_type * entity\_left + (1 + event\_type * entity\_left | workerId)$ .

The analysis indicated a small but significant main effect of event type ( $\beta=2.634$ ,  $SE=1.339$ ,  $p=0.049$ ). This suggests that, across all three experiments, participants in the V-final group were significantly more likely to produce PAV when describing reversible events compared with non-reversible events. Looking at the interactions, we found only a marginally greater effect of event type in Experiment 7 compared with the averaged responses from Experiment 5 and Experiment 6 ( $\beta=0.952$ ,  $SE=0.564$ ,  $p=0.091$ ). There was no difference in the effect of event type in Experiment 6 compared with Experiment 5 ( $\beta=0.426$ ,  $SE=1.238$ ,  $p=0.731$ ). In other words, the effect of event type was largely the same across all three experiments. We were unable to detect this effect in Experiment 5 and Experiment 6 most likely due to insufficient data and the effect of image orientation.

<sup>6</sup>By design, Experiment 7 eliminated orientation as a factor. For the purpose of this analysis, we selected a flag indicating the entity on the left of the image at the end of the presentation as a proxy for entity left. We conducted an exploratory analysis investigating a number of different factors (the entity on the left at the start of the presentation, the entity on the left the end of the presentation, the entity moving ‘towards’ the viewer at the start of the presentation, and the entity moving ‘towards’ the viewer at the end) and found no effect on the proportion of PAV responses. The choice of factor for the present analysis was therefore somewhat arbitrary.

In addition, we found a strong overall preference for APV over PAV ( $\beta=-6.644$ ,  $SE=1.084$ ,  $p<0.001$ ). The model also confirmed that the proportion of PAV responses in Experiment 6 was significantly less than in Experiment 5 ( $\beta=-5.134$ ,  $SE=1.287$ ,  $p<0.001$ ). Recall that in Experiment 5, the mean proportions of APV and PAV responses were roughly equal, as in the training input. In addition, among the participants who used a single order on all trials, approximately half used APV and half used PAV. However, when participants were not trained on the grammar of the language, as was the case in Experiment 6, there was a shift towards APV both among those who used a single order and those who varied their responses. This suggests that, in the absence of explicit grammar training, participants were more likely to be influenced by their native language, resulting in a preference for expressing the agent before the patient.

Our analysis also indicated a main effect of image orientation ( $\beta=4.239$ ,  $SE=1.489$ ,  $p=0.004$ ). In addition, the model confirmed that our manipulation of the experimental design in Experiment 7 was successful. Compared with the averaged results across Experiment 5 and Experiment 6, the effect of image orientation was significantly smaller in Experiment 7 ( $\beta=-2.555$ ,  $SE=0.582$ ,  $p<0.001$ ). Interestingly, the model also indicated a significant interaction between experiment and entity left such that participants in Experiment 6 were significantly more likely to produce PAV when the patient was on the left of the image compared with Experiment 5 ( $\beta=3.526$ ,  $SE=1.435$ ,  $p=0.014$ ). A possible explanation for this could be that participants in Experiment 5 were somewhat less reliant on image orientation as a result of having been trained on a grammar. In the absence of a language model, participants in Experiment 6 may have been more likely to exploit perceptual properties of the scenes when formulating descriptions.

### 5.7.2 P-first

The model used for this analysis included a by-participant intercept and a slope for event type. The model was  $is\_pav \sim exp * entity\_type * entity\_left + (1 + event\_type | workerId)$ . Consistent with the individual analyses of the three experiments, we found no effect of event type ( $\beta=-0.656$ ,  $SE=1.487$ ,  $p=0.659$ ) and no effect of image orientation ( $\beta=-0.138$ ,  $SE=0.223$ ,  $p=0.537$ ). The results therefore did not support the noisy channel hypothesis. In addition, contra the cognitive salience hypothesis, we found that across all three experiments, participants tended to avoid PAV in favor of PVA ( $\beta=-7.871$ ,  $SE=1.485$ ,  $p<0.001$ ). The model also confirmed the observation that participants in Experiment 6 produced PAV significantly less often than those in Experiment 5 ( $\beta=-3.810$ ,  $SE=1.754$ ,  $p=0.030$ ).<sup>7</sup>

### 5.7.3 Discussion

Post-hoc analysis of results pooled from all three experiments suggested that, in line with the cognitive salience hypothesis, participants in the V-final groups were significantly more likely to produce PAV when describing reversible events compared with non-reversible events. However, we did not find support for the more specific prediction that PAV and APV should be produced in roughly equal proportions for reversible events. In addition, in the P-first groups, participants did not show a preference for PAV over PVA. Our results therefore offer only partial support

<sup>7</sup>We also found a significant three-way interaction between experiment, event type and image orientation ( $\beta=1.503$ ,  $SE=0.564$ ,  $p=0.008$ ) indicating that, compared with Experiment 5, in Experiment 6, the effect of image orientation was greater for reversible events compared with non-reversible events such that PAV was more common when the patient appeared on the left. The causal mechanism underlying this effect is unclear.

for the cognitive salience hypothesis. In contrast, we found no evidence to support the noisy channel hypothesis. Pooled data from the three experiments indicated an overall preference for PVA for both reversible and non-reversible events, with no effect of event type.

## 5.8 General Discussion

### 5.8.1 Cognitive salience hypothesis

According to the cognitive salience hypothesis, word order reflects the relative salience of referents in an event such that animate entities are expressed before inanimate entities and concrete entities are expressed before abstract relations. Our findings provided partial support for this hypothesis. Consistent with predictions, we found that participants in the V-final group produced more PAV when describing reversible events compared with non-reversible events. In Experiments 5 and 6, this difference did not reach statistical significance. However, in Experiment 7, where we successfully eliminated the potentially confounding effect of image orientation, the results indicated a small but significant effect of event type. A further post-hoc analysis of the combined results from all three experiments also found a main effect of event type such that PAV was more frequent for reversible events compared with non-reversible events.

However, we did not find that APV and PAV were equally preferred for reversible events. Although results in Experiment 5 were consistent with this prediction, we found the same overall pattern for non-reversible events. Hence, we cannot conclude that this finding supports the cognitive salience hypothesis. In addition, we did not find the overall preference for PAV predicted by the cognitive salience hypothesis in the P-first group. Instead, in Experiments 6 and 7 participants avoided PAV, showing a strong preference for PVA. In Experiment 5, results overall indicated no preference for one order over the other (PAV or PVA).

### 5.8.2 Noisy channel hypothesis

The noisy channel hypothesis predicts that people should avoid orders in which the agent and patient appear on the same side of the action when describing reversible events. Hence, in a P-first language, PVA should be the preferred order for such events. Across the three experiments and in the combined analysis, however, this prediction was not borne out. Our results therefore do not support the noisy channel hypothesis.

A possible counter to this conclusion could be that participants in the current study were not subjected to a communicative pressure, the absence of which might obviate the need to mitigate against information loss. However, the silent gesture experiment reported by Gibson et al. (2013) similarly did not involve explicit communication. The authors acknowledge this fact and suggest that the noisy channel hypothesis need not be restricted to communicative contexts. Accordingly, individuals encode messages so as to maximize meaning recoverability of event representations from memory. Under this memory-based version of the hypothesis, we might expect that the effects of event reversibility would be evident, or more evident, when people are asked to recall an event rather than describe one that is visible to them. Although our own study was not designed to explicitly test this prediction, we note that in the combined analysis, we found that participants in Experiment 7, where they were asked to describe events from memory, were no more likely to use PVA for reversible events than were people in Experiments 5 and 6.

### 5.8.3 Deviations from predicted results: native language and other influences

The predictions outlined in §5.2.1 in relation to the noisy channel hypothesis assumed that participants would probability match (Experiment 5) or randomize between the two available orders (Experiments 6 and 7) in the V-final group and in the P-first group when describing non-reversible events. The cognitive salience hypothesis, on the other hand, predicted that participants would produce APV and PAV in roughly equal proportions when describing reversible events. We also noted, however, that participants' prior linguistic experience might lead them to produce one dominant order over another. The results of this study suggest that native language did indeed have a strong influence on word order choices.

First, we note that APV is partially consistent with canonical English word order in that it expresses the agent first. This may explain the overall preference for this order in the V-final groups in Experiments 6 and 7. The deviation from the predicted equal occurrence of APV and PAV for reversible events in the V-final group under the cognitive salience hypothesis could therefore reflect participants' prior expectations that a language should have a single 'correct' order, and a preference for an order that is consistent with their native language. Similarly, PVA is partially consistent with English word order in being both verb medial and analogous to the English passive construction. In contrast, PAV differs markedly from canonical English word order, which may be why participants strongly dispreferred this order in the P-first group, *contra* the cognitive salience hypothesis.

Further evidence for native language interference comes from the distribution of invalid orders across the two language groups. In all three experiments, there were substantially more trial exclusions from the P-first group than from the V-final group. Some authors have suggested that patient/object-initial languages may be less easily learned than agent/subject-initial languages (e.g., Tily et al., 2011; Tabullo et al., 2012). However, given that participants were not required to learn a grammar in Experiments 6 and 7, a more likely explanation for the uneven distribution of exclusions in the present study is that participants were biased towards an expectation that agents are expressed in initial position, possibly as a result of native language influence. This conclusion is supported by the observation that the majority of invalid orders in the P-first group ended in a VP cluster, consistent with SVO, even when the patient constituent was provided in initial position. In addition, AVP was the most common invalid order in both language groups in Experiment 5 where there were no constraints on the order in which constituents could be expressed.

While the evidence for native language interference is compelling, there are other explanations that could account for the ordering preferences in each language group. In the V-final group, the preference for expressing the agent first may reflect a universal, cross-linguistic bias. 'Agent First' has been proposed as a fundamental ordering principle prefiguring the development of syntactic rules in the earliest stages of language evolution. Jackendoff (2002) proposed that people adopt this ordering strategy as a means of unambiguously identifying participants roles. Alternatively, or additionally, it may derive from a non-communicative cognitive preference for expressing the more salient agent before the less salient patient (Rissman et al., 2018). This interpretation reflects a broader formulation of the cognitive salience hypothesis that goes beyond the animacy-based explanation to include contributions from other properties of referents. Determining the relative contributions of animacy-based salience and that deriving from

other factors such as participant roles remains an area for further research.

Turning to the P-first group, a possible explanation for the PVA preference is that this resulted from the interaction between a preference for using a single order and the effects proposed by the noisy channel hypothesis. That is, if PVA was preferred for reversible events, as predicted by the hypothesis, this order may therefore have been adopted overall reflecting a bias in favour of using a single, consistent order. A summary of the possible explanations for the ordering preferences found in Experiments 6 and 7 is provided in Appendix C.4.

Our results also point to a potentially complex set of interactions between native language, task specific effects, and the perceptual properties of stimuli. In contrast to Experiments 6 and 7, in Experiment 5 there was no overall bias in favor of APV in the V-final group or PVA in the P-first group. This may indicate that participants were less likely to draw on native language knowledge when explicit grammar training was provided. Further, the effect of image orientation in the V-final group was significantly greater in Experiment 6 than in Experiment 5, possibly reflecting an increased reliance on the perceptual properties of the stimuli. The specific nature of this effect, namely, the tendency to mention entities as they appear from left to right, may in turn reflect participants' experience with written language. In Experiment 7, eliminating the effects of image orientation increased the effect of event type in the V-final language groups, suggesting that animacy-based salience is subtle, and may be overridden by other effects (Prat-Sala and Branigan, 2000; Rissman et al., 2018).

In summary, the effects of native language and participants' prior linguistic experience may explain the deviation from the pattern of results predicted by both the cognitive salience and the noisy channel hypothesis. Our results also indicate that other factors, such as the visual layout of the stimuli, may account for the weak effect of animacy on word order in the V-final groups.

#### 5.8.4 Outstanding questions

One outstanding question concerns the extent to which animacy-based word order variation in silent gesture is specific to the gestural modality, as suggested by the role conflict hypothesis (Hall et al., 2013). Kocab et al. (2018) have suggested that the affordances of the manual modality – specifically, the availability of embodiment and iconic gestures – play a key role, raising questions about the extent to which results from silent gesture generalize to spoken language. Although the current study was not designed to test these issues directly, we noted previously that a failure to find an effect of event type would provide evidence in favor of modality-based accounts. Our combined analysis *did* find a significant, albeit small, effect of event type in the V-final groups, but not in the P-first groups. Our results are therefore inconclusive with respect to this question and highlight the need to better understand how modality-specific factors influence word order choices in the absence of linguistic conventions.

Our study also leaves open the question of how communicative pressures influence the evolution of word order conventions in newly emerging languages. Although we found no evidence to support the noisy channel hypothesis specifically, we do not conclude from this that communication plays no role. As previously noted, reversible events are inherently different to non-reversible events due to their potential ambiguity. Further work is therefore required to understand how people negotiate this ambiguity in the absence of linguistic conventions.

## 5.9 Conclusion

A large body of literature attests to the influence of animacy on linguistic structure. Evidence from emerging sign languages suggests that animacy distinctions influence word order variation from the very earliest stages in the emergence of a new language. Further evidence from typology and psycholinguistics demonstrates that such influences remain operational throughout the lifetime of many languages. Numerous studies have investigated the mechanisms underlying the relationship between animacy and word order variation using the silent gesture paradigm. Three main hypotheses have emerged from this work, each of which can account for some of the evidence, but cannot satisfactorily explain all. In the current study, we tested two of these hypotheses – the noisy channel hypothesis, proposed by Gibson et al. (2013), and the cognitive salience hypothesis put forward by Meir et al. (2017). We found partial support for the cognitive salience hypothesis, suggesting that word order variation may reflect the relative salience of interacting entities. This finding is consistent with evidence from emerging sign languages (e.g., Flaherty, 2014) and from language production studies (e.g., Prat-Sala and Branigan, 2000). In contrast, we found no evidence to support the noisy channel hypothesis. Nevertheless, our results pointed to a range of factors that may explain the deviations from the results predicted by both of these hypotheses, in particular, the effects of native language and of participants’ prior linguistic experience. Despite the absence of direct evidence in favour of the noisy channel hypothesis, we do not conclude that communicative pressures associated with animacy distinctions play no role in shaping the evolution of word order conventions. We suggest that further work is needed to investigate this question. Our findings also underscore the need for future work to better understand how task-specific factors, particularly in relation to modality, influence word order choices in the absence of linguistic conventions.



## Chapter 6

# Summary and Conclusion

### 6.1 Aims and contributions

Word order is one of the most basic structuring devices in language (Heine and Kuteva, 2002; Jackendoff, 2002). A vast body of research has been dedicated to understanding why some word orders are more common than others (e.g., Newmeyer, 2000; Lupyan and Christiansen, 2002; Maurits and Griffiths, 2014; Roberts and Levinson, 2017). These studies can tell us a great deal about why languages converge over generations on one dominant word order rather than another. Nevertheless, I have argued that focusing on statistically dominant word orders masks a considerable degree of cross-linguistic and language-internal variation in how people structure information about events. Understanding the underlying causes of this variation is key to uncovering the biases and pressures that shape linguistic structure in emerging communication systems.

Evidence from restricted communication systems in the spoken modality suggests that in the absence of grammatical rules and conventions, producers structure utterances according to two high-level ordering principles: Agent First and Focus Last (Jackendoff, 2002). However, studies of restricted and emerging communication systems in the manual modality suggest that word order is more flexible than these overarching principles might imply. Further, evidence from typology and language production studies have uncovered various factors that influence structural choices in fully developed languages. In §1.3 I argued that these findings can provide valuable insight into the factors that influence word order in emerging communication systems. Much of this work has focused on conceptual accessibility (Bock and Warren, 1985), which has been linked to inherent properties of referents, such as animacy (e.g., Branigan et al., 2008), and to context-dependent properties such as discourse salience (e.g., Prat-Sala and Branigan, 2000), and visual prominence (e.g., Gleitman et al., 2007; Myachykov and Tomlin, 2008).

Taking as its starting point findings from these studies and from restricted and emerging communication systems in the manual modality, the work presented in this thesis has investigated how properties of referents influence word order in emerging languages. In particular, I have focused on the relationship between word order and the animacy status of referents. The effects of animacy on linguistic structure are numerous and wide-ranging (e.g., Yamamoto, 1999). In addition, these effects are evident from the very earliest stages in the development of a communication system (e.g., Senghas et al., 1997; Coppola, 2002; Meir et al., 2017; Ergin et al., 2018). Moreover, as I have argued, the relationship between animacy and word order

has particular relevance to the study of language evolution. Not only is the animacy distinction fundamental to human cognition (Opfer and Gelman, 2011), but there is also an intimate relationship between animacy and the potential to fulfil certain semantic roles (Dahl, 2008). In addition, the animacy of entities interacting in an event, combined with the semantic reversibility of verbs, has implications for how a receiver might infer the meaning of a potentially ambiguous utterance in the absence of conventionalized mechanisms for marking participant roles.

The experiments described in this thesis build on previous work in the silent gesture literature. This paradigm has been adopted by an increasing number of researchers to investigate how people convey information about concepts in the absence of linguistic conventions. In one of the earliest and most cited silent gesture studies, Goldin-Meadow et al. (2008) reported that participants used the order APV to describe events, irrespective of their native language. On the basis of their findings, and drawing a parallel between agents and syntactic subjects and patients and objects, Goldin-Meadow et al. (2008) concluded that SOV is the default order used by all newly developing communication systems. However, as I argued in Chapter 2, evidence from other silent gesture studies reveals a more complex picture. Early work suggested that word order reflects the way people mentally represent events (Goldin-Meadow et al., 1996; Gershkoff-Stowe and Goldin-Meadow, 2002). More recent studies have added to this picture and shown that semantic, structural, and temporal properties of events influence word order preferences (Schouwstra, 2012; Schouwstra and de Swart, 2014; Christensen et al., 2016).

Of particular relevance to this thesis is a body of work that has found that word order in silent gesture is conditioned on the animacy of referents and/or the semantic reversibility of events. The precise nature of this conditioning is unclear, however. Three main hypotheses have been proposed, each of which is consistent with some of the data but falls short of providing a full explanation. The noisy channel hypothesis (Gibson et al., 2013) emphasizes the potential ambiguity of semantically reversible events and proposes that language producers structure utterances to maximize message recoverability in the presence of noise. The role conflict hypothesis (Hall et al., 2013), on the other hand, is motivated by the observation that people typically embody the role of animate entities and also enact actions in silent gesture. Consequently, they avoid orders in which the patient is expressed immediately before the action. A third hypothesis, proposed by Meir et al. (2017), highlights the importance of animacy to human cognition. According to this hypothesis, human entities are more salient than inanimate entities and so tend to be mentioned first. Meir et al. (2017) combined this human-first principle with the proposal that concrete entities tend to precede more abstract relations (Goldin-Meadow et al., 2008) to explain the predominance of SOV. I have termed this the salience, or cognitive salience, hypothesis.

In the experiments presented in this thesis and summarized below, I have: investigated the salience hypothesis in more detail and asked whether the tendency to express human entities before inanimate objects can be modulated by contextually derived salience – the ‘interestingness’ – of agents in a silent gesture task (Chapter 3); explored the role of modality in determining word order preferences (Chapter 4); and tested the salience hypothesis against the communication-based noisy channel hypothesis (Chapter 5).

## 6.2 Summary of experiments

Experiment 1 (Chapter 3) took as its starting point findings from language production studies showing that salience-based word order preferences derive not only from inherent properties of referents such as animacy (Branigan et al., 2008), but also from context-dependent salience, such as discourse prominence (e.g., Prat-Sala and Branigan, 2000), or visual salience (e.g., Gleitman et al., 2007; Myachykov and Tomlin, 2008; Clarke et al., 2015). This study had two main aims. The first was to replicate the APV preference for describing extensional, semantically non-reversible events in a silent gesture task. The second aim was to investigate if manipulating the salience of the event agent would influence the preference for expressing animate entities before inanimate entities. The specific prediction was that events involving non-salient generic agents would elicit fewer APV responses and correspondingly more PAV responses than those involving more salient character agents.

Contrary to expectations, participants did not produce predominantly APV. In addition, manipulating the salience of the agent had no influence on the relative ordering of the agent and patient. Nevertheless, the results of this experiment showed a clear relationship between agent salience and word order. In the first testing block, participants typically described character-agent events using AVP. However, when they described generic-agent events, they showed a strong tendency to omit the agent. In addition, of the orders containing at least one noun, PV occurred most often for events of this kind. Some previous studies have classified such incomplete orders as consistent with APV (e.g., Goldin-Meadow et al., 2008) based on an assumption that the agent would have preceded the patient had it been expressed. Avoiding this a priori assumption, I developed a computational model that inferred the word orders participants would have produced had they expressed all three constituents on every trial. This model exploited the weaker assumption that incomplete orders derive from any consistently ordered underlying complete order where one or more constituents have been omitted. The results of this analysis suggested that APV was the preferred order for generic-agent events, modulated by a strong tendency to omit the agent. In contrast, and consistent with the empirically derived data, AVP was the most common order for character-agent events.

I proposed that the results of this study point to the possibility that, rather than influencing word order directly as suggested by Meir et al. (2017), the salience of referents has a more indirect influence on word order through its effect on how people distribute their attention and subsequently construe an event, that is, on their perspective on an event. This proposal is new in the silent gesture literature, but is by no means novel in the study of language production. A number of authors have previously proposed that salience influences a speaker's global interpretation of a scene (Vogels et al., 2013) and the perspective from which they construe an event (Antón-Méndez, 2017; Rissman et al., 2018).

Nevertheless, the findings of this experiment were inconclusive about the precise mechanisms underlying the specific word order preferences for each agent type. I suggested three possible explanations. First, order of mention may be a direct outgrowth of the way people mentally represent events. Under this view, word order reflects general cognitive biases that operate across modalities and communicative contexts (Gershkoff-Stowe and Goldin-Meadow, 2002; Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014; Christensen et al., 2016). Second, ordering preferences may have been driven by modality-specific factors, as has been proposed in previous silent gesture studies (e.g., Hall et al., 2013; Kocab et al., 2018). The third possibility

is that word order in silent gesture is influenced by native language. I discuss this possibility in more detail below.

In Chapter 4 I presented details of three experiments in which I explored the role of modality in determining word order. In these experiments, I investigated if the ordering preferences found in Experiment 1 would be replicated in a non-gestural task. I adopted a picture selection paradigm first introduced by Gershkoff-Stowe and Goldin-Meadow (2002) and later used by Goldin-Meadow et al. (2008) and, in a modified form, by Vastenius et al. (2016). I adapted this paradigm to address some issues identified with previous studies. First, to ensure that participants interpreted action pictures as representing relations between entities, I designed the stimuli such that they contained a high degree of imagistic detail. In addition, the action pictures included an abstract depiction of a patient so that they afforded both an agent- and patient-focused construal of the event. Finally, I designed the pictures of humans and objects such that they were depicted independent of any particular event.

In all three experiments described in Chapter 4, participants were shown the same character- and generic-agent events used in Experiment 1. In the first experiment (Experiment 2), their task was to click on images depicting the event agent, patient, and action. The main finding was that participants did not produce the same ordering preferences found in Experiment 1. In addition, in contrast to previous picture-selection studies, participants did not use a consistent ordering strategy across items. Further analysis of the results indicated that picture selection order reflected the visual properties of the stimuli. Two findings supported this conclusion. First, people tended to select action images last, which may have resulted from their being less visually salient than agent and patient images and/or less readily associated with their intended referent. Second, the analysis showed that selection order was sensitive to the relative positioning of the agent and patient in the event images.

To reduce the influence of these visual factors and encourage participants to attend to the conceptual and relational properties of the interacting entities, in Experiment 3 I modified the task such that participants were required to reconstruct the target scene. This manipulation was partially successful. Participants produced more agent-before-patient responses compared with Experiment 2, however, selection order remained sensitive to the relative positioning of the agent and patient, albeit to a lesser extent. In Experiment 4, I introduced a further change such that participants were required to reconstruct the target scene from memory. As in the previous two experiments, participants did not adopt a consistent ordering strategy and did not condition their responses on agent type. Crucially, however, this experiment saw a significant increase in the proportion of AVP responses compared with the previous two. As in Experiment 1, I suggested that this may have been due to native-language interference.

In summary, the word orders found in Experiment 1 were not replicated in any of the three experiments presented in Chapter 4. Importantly, selection order was not conditioned on agent type, but was sensitive to the visual properties of the stimuli. When the effect of image orientation was eliminated, the results suggested an increased influence of native language. The findings were therefore inconclusive about the mechanisms underlying the ordering preferences found in the silent gesture task (Experiment 1). In addition, the influence of stimuli properties and native language cast some doubt over the claim that the orders produced in picture-selection experiments directly reflect the way people mentally represent events.

In Chapter 5, I turned my focus to the relationship between animacy, semantic reversibility, and word order in silent gesture. I reported three artificial language learning experiments in

which I tested the salience hypothesis proposed by Meir et al. (2017) against the noisy channel hypothesis put forward by Gibson et al. (2013). In the first of these experiments (Experiment 5), participants were trained on one of two artificial languages: a P-first language in which events could be described using either PAV or PVA word order; or a V-final language where APV and PAV were permitted. Following language training, participants were shown a series of reversible and non-reversible events and asked to provide a written description of those events using the language on which they had been trained. Each of the two hypotheses made different predictions about the orders that people would use to describe each event type.

The overall results of this experiment suggested that participants in both language groups produced each of the permitted orders in equal proportions, with no conditioning on event type. However, further analysis revealed that a substantial proportion of participants in fact used a single order on all trials. I suggested that this finding might reflect the high cognitive demands of the task, resulting in a tendency to regularize to a single order (Ferdinand et al., 2019). Experiment 6 was designed to investigate this possibility. I implemented two design changes intended to reduce the cognitive demands of the task. First, I made the vocabulary more English-like (Samara et al., 2017). Second, participants were trained on the vocabulary only. During the critical testing phase, they were presented with an event image and a partially completed description consisting of a single constituent: either the patient label, which always appeared in initial position; or the action label in final position. Participants were required to complete the description of the event using the vocabulary on which they had been trained. This design change was not successful in reducing regularization. In fact, a higher proportion of participants regularized to a single order. In addition, unlike Experiment 5, regularizing participants in both language groups showed a clear preference for one order over the other. In the P-first group, PVA was the preferred order, while APV predominated in the V-final group. I proposed that these preferences were driven by native language interference, which I return to below.

Echoing findings from the picture-selection experiments described in Chapter 4, post-hoc analysis of results from both Experiment 5 and 6 showed that responses in the V-final group were sensitive to the relative positioning of the agent and patient such that PAV was more common when the patient appeared on the left side of the image. An analysis of word order errors in the P-first group similarly suggested an effect of image orientation: invalid orders, typically ending in VP, were more common when the patient appeared on the right of the image. I argued that this latter finding might also point to native language interference.

In Experiment 7, I introduced two changes designed to eliminate the possibility of ‘reading’ the scene from left to right. First, the target event image was made visible for only 4s so that participants were required to describe the scene from memory. Second, the image was presented rotating around the y-axis such that the relative positioning of the agent and patient varied continuously. Echoing results from Experiment 6, participants in both language groups showed a strong tendency to use a single order: PVA in the P-first group and APV in the V-final group. Unlike Experiments 5 and 6, however, responses in the V-final group showed some sensitivity to agent type consistent with the salience hypothesis. More specifically, the results showed that participants were significantly more likely to produce PAV for reversible events compared with non-reversible events. Nevertheless, the specific prediction that APV and PAV should appear in equal proportions for reversible events was not borne out. In addition, as mentioned, there was no overall preference for PAV in the P-first group, contra the salience hypothesis. In terms

of the noisy channel hypothesis, the data provided no evidence to support this proposal.

A post-hoc combined analysis of the data from all three experiments supported these conclusions. The results of the three experiments presented in Chapter 5 therefore offered tentative support for the salience hypothesis, but were not consistent with the noisy channel hypothesis. I concluded that these experiments demonstrated that the properties of the referents (i.e., animate or inanimate) rather than the semantic reversibility of events influenced word order choices. Nevertheless, I argued that these findings do not rule out the possibility that communicative factors driven by semantic reversibility play a key role in shaping word order preferences in emerging communication systems.

### 6.3 Outstanding questions and areas for further research

In the discussion above, I drew attention to a number of areas where further research is required to provide a fuller understanding of word order preferences in the experiments described here and elsewhere, particularly in the silent gesture literature. Of crucial importance to this field is understanding the role of modality. In Chapter 3 I highlighted two ways in which word order in Experiment 1 may have been influenced by the manual modality. First, I noted an apparent difference in the way people described each agent type: while descriptions of generic agents tended to be quite minimal, descriptions of character agents were typically highly detailed. I suggested a possible relationship between the tendency to provide detailed descriptions of character agents and the preference for expressing actions before patients. Specifically, I proposed that increased attentional focus on the agent may have heightened the salience, or perceived contextual relevance of the action relative to the patient. If correct, then eliminating the difference between character- and generic-agent descriptions could in turn eliminate word order differences, resulting in an overall preference for patient-before-action preferences in line with previous silent gesture studies (e.g., Goldin-Meadow et al., 2008). I further suggested that this could be tested by restricting the amount of information that participants can provide about an entity to, say, a single gesture.<sup>1</sup>

The second finding for which I proposed a possible effect of modality was the tendency to omit generic agents in the first testing block of Experiment 1. I argued that this phenomenon might be more accurately characterized as omission of explicit reference to the physical attributes of the agent rather than of the agent itself. The manual modality affords such a strategy because it enables participants to simultaneously embody the role of the agent while enacting the action (see also Meir et al., 2007; Hall et al., 2013). An important outstanding question, therefore, is whether incomplete orders are surface manifestations of complete orders with omissions, as assumed in the computational model I developed and described in Chapter 3, or instead represent utterances with simultaneous expressions of multiple constituents. If the latter, then this raises questions about the generalizability of findings from silent gesture to the spoken modality (see also Kocab et al., 2018).

Also relevant to the question of modality is event type and, relatedly, properties of referents.

---

<sup>1</sup>It is interesting to note that this phenomenon need not be exclusive to the gestural modality, but may arise in any context where people describe rather than label event constituents. To test this, one could devise an experiment in which participants describe events involving unfamiliar entities (e.g., aliens interacting with novel objects) using an artificial language with words for attributes and actions but not whole entities. One could then investigate whether manipulating the number of attributes associated with and used to describe an agent had a systematic effect on the positioning of the action relative to the patient.

As I noted in Chapter 3, the elicitation stimuli used in silent gesture studies often depict events in which agents manually manipulate patients. I highlighted the possibility that such events may naturally elicit APV in the manual modality, noting a general characteristic of sign languages whereby if an argument affects the phonological shape of a verb, it typically precedes the verb (Napoli and Sutton-Spence, 2014). It is notable that in Experiment 1, where the form of the action was independent of the patient, the proportion of APV responses was smaller than expected based on previous silent gesture studies. Moreover, where event type has been systematically explored such that manipulation events are compared with creation or other intensional events, APV has been preferred for the former but not the latter (Schouwstra, 2012; Schouwstra and de Swart, 2014; Christensen et al., 2016; Napoli et al., 2017). The accounts that have been proposed in the silent gesture literature to explain the preference for APV appeal to modality-independent preferences for conveying information about events. As this discussion demonstrates, however, more research is required to understand the relative contribution of modality-specific factors and those rooted in general cognition.

Another key question that remains to be answered concerns the role of communicative pressures in shaping word order in silent gesture, particularly in relation to reversible events. Previous attempts to test the noisy channel hypothesis (Gibson et al., 2013) against the role conflict hypothesis (Hall et al., 2013) have been inconclusive (Kocab et al., 2018; Kline et al., preprint), while the three experiments described in Chapter 5 found no evidence in support of the noisy channel hypothesis. Nevertheless, I argued that such results cannot be taken as evidence that communicative pressures play no role. The increased use of spatial indexing for reversible events observed by Gibson et al. (2013) and Hall et al. (2013), for example, suggests that such events do impose a pressure to explicitly mark participant roles. In addition, I noted in Chapter 2 that findings from both silent gesture and sign language studies indicate that word order is more variable for reversible compared with non-reversible events (silent gesture: Hall et al., 2013; Meir et al., 2017; Kocab et al., 2018; sign languages: Johnston et al., 2007; Meir, 2010; Meir et al., 2017; Ergin et al., 2018). I suggested that this apparently counterintuitive finding might indicate that people explore a greater range of strategies when communicating about reversible events precisely because of the pressure to clearly mark who is doing what to whom. Further research into this phenomenon will provide valuable insights into the relationship between communication and linguistic structure.

The work presented in this thesis also highlights the need to better understand how native language influences the way people describe events in experimental settings. In Chapter 3 I suggested that native language may have been a factor in determining word order choices in Experiment 1. Specifically, I noted that a parallel could be drawn between the orders AVP and PV and the English active and passive constructions, respectively. Native language influence was also evident in the picture-selection task in Experiment 4 (Chapter 4) where participants were required to recall the event image from memory. Similarly, in Experiments 6 and 7 (Chapter 5), where participants were not trained on a grammar prior to completing the task, I argued that native language influence could explain the preference for PVA in the P-first language group and APV in the V-final group.

Nevertheless, a native-language interpretation of these results is by no means clear cut in all cases. In relation to Experiment 1 I noted that this interpretation was complicated by the finding that participants who had predominantly produce PV in block 1 shifted to APV in block 2. If native language were the main determiner of word order, we would expect participants to

shift to either AVP or PVA. Further, in relation to Experiments 6 and 7, the preference for APV in the V-final group might alternatively reflect a more universal agent-first bias (Jackendoff, 2002).

Numerous silent gesture studies have found that native language does not influence the orders people use to describe events (e.g. Goldin-Meadow et al., 2008; Langus and Nespors, 2010; Schouwstra and de Swart, 2014; Futrell et al., 2015). However, findings from some studies challenge the assumption that this is universally the case (e.g., Gibson et al., 2013; Meir et al., 2017). Meir et al. (2017), for example, suggested that the preferences among some groups for AVP (or SVO) when describing reversible events could be attributed to an effect of native language. Specifically, they suggested that individuals lack a straightforward strategy for ordering the agent and patient in such events and therefore fall back on their native language. Although this explanation is compelling, it leaves a number of questions unanswered. First, it cannot explain the finding reported by Futrell et al. (2015) that Irish and Tagalog speakers preferred AVP for reversible events. Second, the native-language account cannot explain the pattern of responses found by Kocab et al. (2018) and Gershkoff-Stowe and Goldin-Meadow (2002) in relation to inanimate-inanimate events, which should present the same difficulties as animate-animate events. Third, if speakers are more likely to use their native language for reversible events, then speakers of an SOV language should use APV *more* often than for non-reversible events. In fact, Meir et al. (2017) found the opposite result.

What this discussion demonstrates is that the role of native language in shaping structural choices in silent gesture and other experimental paradigms remains poorly understood. While there is evidence for native language effects in some circumstances, it is not always possible to rule out influences from alternative candidate factors. In addition, evidence for native language effects is apparent under some but not all circumstances. Understanding the role of native language is crucial to our ability to draw firm conclusions about the factors that shape word order in emerging communication systems. Further, understanding how and under what circumstances people rely on or are influenced by the conventions of their native language will shed light on the biases and pressures that operate when people convey information about events using an alternative communication system.

Lastly, the work presented in this thesis has uncovered some important methodological issues in the silent gesture literature. In Chapter 3 I drew attention to the fact that some studies have claimed a bias for APV based on the proportion of gesture responses classed as consistent with this order. In Gibson et al. (2013), these included all responses in which the patient was expressed before the action, while in Goldin-Meadow et al. (2008), they included the incomplete orders AV and PV. This highlights two broad issues with the way gesture strings are analysed and interpreted across the literature. First, in relation to incomplete orders, I argued that categorizing these as consistent with some other order, or excluding them from the analysis as in some studies, could overlook important phenomena that tell us something about the cognitive biases that shape linguistic structure. Second, the inconsistent approaches to dealing with incomplete orders limits the extent to which results across different studies can be compared. The same problem arises with repetitious orders. For example, where one study may code an agent-action-patient-action response as a single string (e.g., Meir et al., 2017), another might analyse the same sequence as two separate clauses (AV and PV) (e.g., Langus and Nespors, 2010).

Coding strategies in the silent gesture literature more generally are not consistent across

studies, and are sometimes not fully explained or justified. For example, Langus and Nespors (2010) included in their analysis the gesture string in each trial that was produced last and ignored failed attempts. However, they did not make clear what constituted a string or how they identified a failed attempt. Christensen et al. (2016), on the other hand, included only the *first* spontaneously produced string; strings containing repeated sequences or self-repair were excluded from the analysis. Yet another approach was taken by Gibson et al. (2013) who excluded all trials in which the patient was omitted, or the patient or action were expressed more than once. As an increasing number of studies use this paradigm, it is becoming ever-more important for researchers to adopt consistent and transparent coding strategies. Without this, it is not possible to draw clear comparisons between studies.

## 6.4 General conclusions

Early work in the silent gesture literature suggested that emerging communication systems may adopt a single, default word order – SOV (Goldin-Meadow et al., 2008). A growing body of evidence has challenged this view, revealing a range of factors that influence word order in the absence of linguistic conventions. The work presented in this thesis supports the view that word order is more variable and is subject to a range of influences. A number of studies have highlighted the relationship between word order and certain properties of events (e.g., Schouwstra, 2012; Christensen et al., 2016; Gibson et al., 2013). The experiments described in this thesis provide evidence that properties of the referents themselves also influence word order choices. Specifically, these experiments support the hypothesis that word order in emerging communication systems is influenced by the relative salience of entities in an event. I have demonstrated that word order is influenced not only by inherent properties of referents – animacy in particular (Meir et al., 2017) – but also by contextually derived salience. Findings from language production studies demonstrate that contextually derived salience influences linguistic structure in fully developed languages (e.g., Prat-Sala and Branigan, 2000; Gleitman et al., 2007). The work presented in this thesis is the first to show that the same effects operate in restricted communication systems.

This work has also drawn attention to a number of areas for future study. Of particular importance is understanding the relationship between modality and word order preferences. While the silent gesture paradigm has proven a valuable resource for investigating how people convey information in the absence of linguistic conventions, questions remain about the extent to which findings from these studies can be generalized to the spoken modality. Another important question concerns the role of native language. Gaining a better understanding of how and under what circumstances people rely on, or are influenced by, their native language in experimental settings is crucial.

Finally, although the work presented here did not support the noisy channel hypothesis specifically (Gibson et al., 2013), I have argued that this cannot be taken as evidence that communicative pressures play no role in shaping linguistic structure. On the contrary, there is ample evidence that semantic reversibility, in combination with the animacy status of referents, imposes a pressure to elucidate who is doing what to whom in an event. Future work will help uncover how this pressure affects word order in emerging communication systems.



# Bibliography

- Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21(3):435–483.
- Antón-Méndez, I. (2017). Visual salience effects on speaker choices: Direct or indirect influences on linguistic processing? *Applied Psycholinguistics*, 38(3):601–631.
- Barberà, G., Cabredo Hofherr, P., and Quer, J. (2018). Agent-backgrounding in Catalan Sign Language (LSC). *Sign Language and Linguistics*, 21(2):334–349.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015a). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Becker, M. (2014). *The Acquisition of Syntactic Structure: Animacy and Thematic Alignment*. Cambridge University Press.
- Bock, K. and Ferreira, V. S. (2014). Syntactically Speaking. In Goldrick, M., Ferreira, V. S., and Miozzo, M., editors, *Oxford Handbook of Language Production*, chapter 2, pages 21–47. Oxford University Press.
- Bock, K. J. and Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21(1):47–67.
- Botha, R. (2007). On homesign systems as a potential window on language evolution. *Language and Communication*, 27(1):41–53.
- Branigan, H. P. and Feleki, E. (1999). Conceptual accessibility and serial order in Greek speech production. In Hahn, M. and Stoness, S. C., editors, *21st Annual Conference of the Cognitive-Science-Society*, pages 96–101.
- Branigan, H. P., Pickering, M. J., and Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, 118(2):172–189.
- Campbell, L. and Poser, W. J. (2008). *Language Classification: History and Method*. Cambridge University Press.
- Carrigan, E. M. and Coppola, M. (2017). Successful communication does not drive language development: Evidence from adult homesign. *Cognition*, 158:10–27.

- Christensen, P., Fusaroli, R., and Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, 146:67–80.
- Christianson, K. and Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, 98(2):105–135.
- Clarke, A. D. F., Elsner, M., and Rohde, H. (2015). Giving Good Directions: Order of Mention Reflects Visual Salience. *Frontiers in Psychology*, 6.
- Coco, M. I., Malcolm, G. L., and Keller, F. (2014). The Interplay of Bottom-Up and Top-Down Mechanisms in Visual Guidance during Object Naming. *Quarterly Journal of Experimental Psychology*, 67(6):1096–1120.
- Coppola, M. (2002). *The Emergence of Grammatical Categories in Home Sign: Evidence from Family-based Gesture Systems in Nicaragua*. PhD thesis, University of Rochester.
- Coppola, M. and Newport, E. L. (2005). Grammatical Subjects in home sign: Abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences*, 102(52):19249–19253.
- Culbertson, J., Schouwstra, M., and Kirby, S. (2020). From the world to word order: deriving biases in noun phrase order from statistical properties of the world”. *Language*.
- Dahl, Ö. (2008). Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua*, 118(2):141–150.
- de Hoop, H. and Lamers, M. (2006). Incremental distinguishability of subject and object. In Kulikov, L., Malchukov, A., and de Swart, P., editors, *Case, Valency and Transitivity*, chapter 12, pages 269–287. John Benjamins.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1):1–12.
- de Swart, P., Lamers, M., and Lestrade, S. (2008). Animacy, argument structure, and argument encoding. *Lingua*, 118(2):131–140.
- de Vos, C. and Pfau, R. (2015). Sign Language Typology: The Contribution of Rural Sign Languages. *Annual Review of Linguistics*, 1(1):265–288.
- Dennison, H. Y. (2008). Universal versus language-specific conceptual effects on shifted word-order production in Korean: evidence from bilinguals. *Working Papers in Linguistics: University of Hawaii at Manoa*, 39(2):1–16.
- Dobel, C., Diesendruck, G., and Bölte, J. (2007). How Writing System and Age Influence Spatial Representations of Actions. *Psychological Science*, 18(6):487–491.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3):547–619.
- Dryer, M. S. (2013). Order of subject, object and verb. In Dryer, M. S. and Haspelmath, M., editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology.

- Ergin, R., Meir, I., Aran, D. I., Padden, C., and Jackendoff, R. (2018). The Development of Argument Structure in Central Taurus Sign Language. *Sign Language Studies*, 18(4):612–639.
- Esaulova, Y., Penke, M., and Dolscheid, S. (2019). Describing Events: Changes in Eye Movements and Language Production Due to Visual and Conceptual Properties of Scenes. *Frontiers in Psychology*, 10.
- Evans, N. and Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(5):429–448.
- Ferdinand, V., Kirby, S., and Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, 184:53–68.
- Ferreira, F. (1994). Choice of Passive Voice is Affected by Verb Type and Animacy. *Journal of Memory and Language*, 33(6):715–736.
- Ferreira, F. and Rehrig, G. (2019). Linearisation during language production: evidence from scene meaning and saliency maps. *Language, Cognition and Neuroscience*, pages 1–11.
- Flaherty, M. E. (2014). *The Emergence of Argument Structural Devices in Nicaraguan Sign Language*. University of Chicago, Division of the Social Sciences, Department of Psychology.
- Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., and Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*.
- Gell-Mann, M. and Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences of the United States of America*, 108(42):17290–17295.
- Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A Weakly Informative Default Prior Distribution for Logistic and other Regression Models. *The Annals of Applied Statistics*, 2(4):1360–1383.
- Gelman, R. and Spelke, E. S. (1981). *The development of thoughts about animate and inanimate objects: Implications for research in social cognition*. Cambridge University Press.
- Gentner, D. and Boroditsky, L. (2001). Individuation, relativity, and early word learning. In Bowerman, M. and Levinson, S. C., editors, *Language acquisition and conceptual development*, chapter 8, pages 215–256. Cambridge University Press.
- Gershkoff-Stowe, L. and Goldin-Meadow, S. (2002). Is there a natural order for expressing semantic relations? *Cognitive Psychology*, 45(3):375–412.
- Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., and Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, 24(7):1079–1088.
- Gleitman, L. R., January, D., Nappa, R., and Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57(4):544–569.

- Goldin-Meadow, S. (1985). Language development under atypical learning conditions: Replication and implications of a study of deaf children of hearing parents. In Nelson, K. E., editor, *Children's Language (Vol. 5)*, chapter 7, pages 197–245. Lawrence Erlbaum Associates.
- Goldin-Meadow, S., McNeill, D., and Singleton, J. (1996). Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. *Psychological review*, 103(1):34–55.
- Goldin-Meadow, S. and Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, 391(6664):279–281.
- Goldin-Meadow, S., Özyürek, A., Sancar, B., and Mylander, C. (2009). Making Language Around the Globe: A Crosslinguistic Study of Homesign in the United States, China, and Turkey. In Guo, J., Lieven, E., Budwig, N., and Ervin-Tripp, S., editors, *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*, chapter 2, pages 27–39. Taylor and Francis.
- Goldin-Meadow, S., So, W. C., Ozyürek, A., and Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, 105(27):9163–9168.
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Language*, 2:73–113.
- Hall, M. L., Ahn, Y. D., Mayberry, R. I., and Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of Memory and Language*, 81:16–33.
- Hall, M. L., Ferreira, V. S., and Mayberry, R. I. (2014). Investigating Constituent Order Change With Elicited Pantomime: A Functional Account of SVO Emergence. *Cognitive Science*, 38(5):943–972.
- Hall, M. L., Mayberry, R. I., and Ferreira, V. S. (2013). Cognitive constraints on constituent order: evidence from elicited pantomime. *Cognition*, 129(1):1–17.
- Heine, B. and Kuteva, T. (2002). On the Evolution of Grammatical Forms. In Wray, A., editor, *The Transition to Language*, volume 2, pages 376–397. Oxford University Press.
- Hudson Kam, C. L. and Newport, E. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, 1(2):151–195.
- Hurford, J. R. (2012). *The Origins of Grammar: Language in the Light of Evolution*. Oxford University Press.
- Hwang, H. (2017). The role of thematic role accessibility in production: evidence from Korean. *Lang. Cogn. Neurosci.*, 32(1):117–128.
- Hwang, H. and Kaiser, E. (2015). Accessibility effects on production vary cross-linguistically: Evidence from English and Korean. *Journal of Memory and Language*, 84:190–204.

- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.
- Jackendoff, R. and Wittenberg, E. (2017). Linear grammar as a possible stepping-stone in the evolution of language. *Psychonomic Bulletin and Review*, 24(1):219–224.
- Jaeger, T. F. and Norcliffe, E. J. (2009). The Cross-linguistic Study of Sentence Production. *Language and Linguistics Compass*, 3(4):866–887.
- Janzen, T., O’Dea, B., and Shaffer, B. (2001). The Construal of Events: Passives in American Sign Language. *Sign Language Studies*, 1(3):281–310.
- Johnston, T., Vermeerbergen, M., Schembri, A., and Leeson, L. (2007). ‘Real data are messy’: Considering the cross-linguistic analysis of constituent ordering in Australian Sign Language (Auslan), Vlaamse Gebarentaal (VGT) and Irish Sign Language (ISL). In Perniss, P. M., Pfau, R., and Steinbach, M., editors, *Visible variation: Comparative studies on sign language structure*, pages 163–206. Mouton de Gruyter.
- Kegl, J. (1990). Predicate argument structure and verb-class organization in the ASL lexicon. In Lucas, C., editor, *Sign Language Research: Theoretical Issues*, pages 149–175. Galludet University Press.
- Kimball, A. E., Shantz, K., Eager, C., and Roy, J. (2019). Confronting Quasi-Separation in Logistic Mixed Effects for Linguistic Data: A Bayesian Approach. *Journal of Quantitative Linguistics*, 26(3):231–255.
- Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin and Review*, 24(1):118–137.
- Klein, W. and Perdue, C. (1997). The Basic Variety (or: Couldn’t natural languages be much simpler?). *Second Language Research*, 13(4):301–347.
- Kline, M., Salinas, M. A., Lim, E., Fedorenko, E., and Gibson, E. (2017). Preprint – Word order patterns in gesture are sensitive to modality-specific production constraints. PsyArXiv. <https://psyarxiv.com/62puh/>.
- Kocab, A., Lam, H., and Snedeker, J. (2018). When Cars Hit Trucks and Girls Hug Boys: The Effect of Animacy on Word Order in Gestural Language Creation. *Cognitive Science*, 42(3):918–938.
- Langus, A. and Nespors, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, 60(4):291–318.
- Lupyan, G. and Christiansen, M. H. (2002). Case, Word Order, and Language Learnability: Insights from Connectionist Modeling. In Gray, W. D. and Schunn, C. D., editors, *Proceedings of 24th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates.
- Maass, A. and Russo, A. (2003). Directional Bias in the Mental Representation of Spatial Events. *Psychological Science*, 14(4):296–301.
- MacWhinney, B. (1977). Starting Points. *Language*, 53(1):152–168.

- Mandler, J. M. (1992). The foundations of conceptual thought in infancy. *Cognitive Development*, 7(3):273–285.
- Marno, H., Langus, A., Omidbeigi, M., Asaadi, S., Seyed-Allaei, S., and Nespors, M. (2015). A new perspective on word order preferences: the availability of a lexicon triggers the use of SVO word order. *Frontiers in Psychology*, 6.
- Maurits, L. and Griffiths, T. L. (2014). Tracing the roots of syntax with Bayesian phylogenetics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(37):13576–13581.
- Meir, I. (2010). The Emergence of Argument Structure in Two New Sign Languages. In Rappaport Hovav, M., Doron, E., and Sichel, I., editors, *Lexical Semantics, Syntax, and Event Structure*, chapter 6, pages 101–123. Oxford University Press.
- Meir, I., Aronoff, M., Börstell, C., Hwang, S.-O., Ilkbasaran, D., Kastner, I., Lopic, R., Lifshitz Ben-Basat, A., Padden, C., and Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, 158:189–207.
- Meir, I., Padden, C. A., Aronoff, M., and Sandler, W. (2007). Body as subject. *Journal of Linguistics*, 43(3):531–563.
- Mithun, M. (1992). Is basic word order universal? In Payne, D., editor, *Pragmatics of Word Order Flexibility*, pages 15–61. John Benjamins, typologica edition.
- Motamedi, Y. (2017). *Artificial sign language learning: a method for evolutionary linguistics*. PhD thesis, University of Edinburgh.
- Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., and Kirby, S. (2018). The emergence of spatial modulation in artificial sign languages. Retrieved from [osf.io/p6zy4](https://osf.io/p6zy4).
- Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., and Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, 192:103964.
- Myachykov, A., Garrod, S., and Scheepers, C. (2009). Attention and syntax in sentence production: A critical review. *Discours*.
- Myachykov, A., Garrod, S., and Scheepers, C. (2012). Determinants of structural choice in visually situated sentence production. *Acta Psychologica*, 141(3):304–315.
- Myachykov, A., Thompson, D., Scheepers, C., and Garrod, S. (2011). Visual Attention and Structural Choice in Sentence Production Across Languages. *Language and Linguistics Compass*, 5(2):95–107.
- Myachykov, A. and Tomlin, R. (2008). Perceptual priming and structural choice in Russian sentence production. *Journal of Cognitive Science*, 9(1):31–48.
- Napoli, D. J., Spence, R. S., and de Quadros, R. M. (2017). Influence of predicate sense on word order in sign languages: Intensional and extensional verbs. *Language*, 93(3):641–670.

- Napoli, D. J. and Sutton-Spence, R. (2014). Order of the major constituents in sign languages: implications for all language. *Frontiers in Psychology*, 5.
- Newmeyer, F. J. (2000). On the Reconstruction of ‘Proto-World’ Word Order. In Knight, C., Studdert-Kennedy, M., and Hurford, J., editors, *The Evolutionary Emergence of Language*, pages 372–388. Cambridge University Press.
- Nölle, J., Staib, M., Fusaroli, R., and Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, 181:93–104.
- Opfer, J. E. and Gelman, S. A. (2011). Development of the Animate–Inanimate Distinction. In Goswami, U., editor, *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, chapter 8, pages 213–238. Wiley-Blackwell, second edition.
- Özçalışkan, Ş., Lucero, C., and Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition*, 148:10–18.
- Padden, C. A., Meir, I., Sandler, W., and Aronoff, M. (2010). Against all expectations: Encoding subjects and objects in a new language. In Gerdts, D., Moore, J., and Polinsky, M., editors, *Hypothesis A/Hypothesis B: Linguistic Explorations in Honor of David M. Perlmutter*, pages 383–400. MIT Press.
- Peirce, J. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2:10.
- Pokhoday, M., Shtyrov, Y., and Myachykov, A. (2019). Effects of Visual Priming and Event Orientation on Word Order Choice in Russian Sentence Production. *Frontiers in Psychology*, 10(JULY):1661.
- Prat-Sala, M. and Branigan, H. P. (2000). Discourse Constraints on Syntactic Processing in Language Production: A Cross-Linguistic Study in English and Spanish. *Journal of Memory and Language*, 42(2):168–182.
- Primus, B. (2012). Animacy, Generalized Semantic Roles, and Differential Object Marking. In Lamers, M. and de Swart, P., editors, *Case, Word Order and Prominence. Studies in Theoretical Psycholinguistics*, volume 40. Springer.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Rissman, L. and Goldin-Meadow, S. (2017). The Development of Causal Structure without a Language Model. *Language Learning and Development*, 13(3):286–299.
- Rissman, L., Horton, L., Flaherty, M., Senghas, A., Coppola, M., Brentari, D., and Goldin-Meadow, S. (2020). The communicative importance of agent-backgrounding: Evidence from homesign and Nicaraguan Sign Language. *Cognition*, 203:104332.
- Rissman, L., Woodward, A., and Goldin-Meadow, S. (2018). Occluding the face diminishes the conceptual accessibility of an animate agent. *Language, Cognition and Neuroscience*, pages 1–16.

- Roberts, S. G. and Levinson, S. C. (2017). Conversation, cognition and cultural evolution. *Interaction Studies*, 18(3):402–442.
- Samara, A., Smith, K., Brown, H., and Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, 94:85–114.
- Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences*, 102(7):2661–2665.
- Schouwstra, M. (2012). *Semantic Structures, Communicative Strategies and the Emergence of Language*. LOT.
- Schouwstra, M. (2017). Temporal Structure in Emerging Language: From Natural Data to Silent Gesture. *Cognitive Science*, 41:928–940.
- Schouwstra, M. and de Swart, H. (2014). The semantic origins of word order. *Cognition*, 131(3):431–6.
- Schouwstra, M., de Swart, H., and Thompson, B. (2019). Interpreting Silent Gesture: Cognitive Biases and Rational Inference in Emerging Language Systems. *Cognitive Science*, 43(7).
- Schouwstra, M., Kirby, S., and Culbertson, J. (2017). Silent gesture and noun phrase universals. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pages 3095–3100.
- Schouwstra, M., Smith, K., and Kirby, S. (2020). The Emergence of Word Order Conventions: improvisation, interaction and transmission. PsyArXiv. <https://psyarxiv.com/wdfu2/>.
- Senghas, A., Coppola, M., Newport, E. L., and Supalla, T. (1997). Argument Structure in Nicaraguan Sign Language: The Emergence of Grammatical Devices. In Hughes, E., Hughes, M., and Greenhill, A., editors, *Proceedings of the Boston University Conference on Language Development*, pages 550–561. Cascadilla Press.
- Sinnemäki, K. (2010). Word order in zero-marking languages. *Studies in Language*, 34(4):869–912.
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In Gumperz, J. J. and Levinson, S. C., editors, *Rethinking linguistic relativity*, chapter 3, pages 70–96. Cambridge University Press.
- Smith, K. and Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3):444–449.
- Tabullo, Á., Arismendi, M., Wainelboim, A., Primero, G., Vernis, S., Segura, E., Zanutto, S., and Yorio, A. (2012). On the Learnability of Frequent and Infrequent Word Orders: An Artificial Language Learning Study. *Quarterly Journal of Experimental Psychology*, 65(9):1848–1863.
- Tanaka, M. N., Branigan, H. P., McLean, J. F., and Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, 65(3):318–330.

- Tily, H. J., Frank, M. C., and Jaeger, T. F. (2011). The learnability of constructed languages reflects typological patterns. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 1364–1369.
- van de Velde, M., Meyer, A. S., and Konopka, A. E. (2014). Message formulation and structural assembly: Describing “easy” and “hard” events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, 71(1):124–144.
- van Nice, K. Y. and Dietrich, R. (2003). Task sensitivity of animacy effects: evidence from German picture descriptions. *Linguistics*, 41(5).
- Vastenius, A., van de Weijer, J., and Zlatev, J. (2016). The influence of native language word order and cognitive biases in pictorial event representations. *Cognitive Semiotics*, 9(1):45–77.
- Vogels, J., Krahmer, E., and Maes, A. (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes*, 28(9):1323–1349.
- Yamamoto, M. (1999). *Animacy and Reference : A Cognitive Approach to Corpus Linguistics*. Studies in Language Companion Series. John Benjamins Publishing Co.



# Appendix A

## A.1 Chapter 3: List of items

Table A.1: Set of humans (agents), inanimate objects (patients) and actions.

Item	Category
girl	generic human
boy	generic human
teenager	generic human
woman	generic human
man	generic human
chef	character human
king	character human
pirate	character human
punk	character human
viking	character human
bird table	object
clock	object
drawers	object
plant	object
push	action
kick	action
elbow	action
poke	action

## A.2 Chapter 3: Patient-before-action in block 1

This supplementary analysis considered the proportion of trials in which the patient was expressed before the action in block 1 of testing. The data included all trials in which it was possible to determine the order of the first-mentioned patient with respect to the first-mentioned action. There were 410 such trials, representing 229 descriptions of character-agent events and 181 generic-agent events. Participants were significantly more likely to express the patient before the action when describing generic-agent events.

Table A.2: Mixed effects logistic regression analysis of patient-before-action responses in block 1.

Predictor	$\beta$	$SE$	$p$
Intercept	4.818	1.985	0.015*
agent_type	6.257	2.989	0.036*
orientation	0.284	0.671	0.672
agent_type:orientation	0.736	1.359	0.588

Model:  $p\_before\_act \sim agent\_type*orientation + (1 | participant) + (1 | item)$

## A.3 Chapter 3: Model supplementary data

### A.3.1 Generic-agent events

Table A.3: Model: Bootstrap means and 95% CIs for each word order in the observed surface distribution, best-fit surface distribution, and inferred underlying distribution based on block 1 descriptions of generic-agent events.

	Observed distribution			Best-fit surface distribution			Inferred underlying distribution		
	lower 95% CI	mean	upper 95% CI	lower 95% CI	mean	upper 95% CI	lower 95% CI	mean	upper 95% CI
APV	0.168	0.220	0.276	0.172	0.223	0.276	0.645	0.746	0.842
AVP	0.034	0.065	0.099	0.008	0.026	0.047	0.026	0.087	0.158
PAV	0.017	0.039	0.065	0.017	0.042	0.069	0.066	0.141	0.224
PVA	0	0.004	0.013	0	0.008	0.022	0	0.027	0.066
VAP	0	0	0	0	0	0	0	0	0
VPA	0	0	0	0	0	0	0	0	0
AV	0	0	0	0	0	0	NA	NA	NA
PV	0.276	0.337	0.397	0.272	0.333	0.392	NA	NA	NA
VA	0	0	0	0	0	0	NA	NA	NA
VP	0	0	0	0.013	0.032	0.056	NA	NA	NA
V	0.276	0.336	0.397	0.276	0.336	0.397	NA	NA	NA

The best-fit model had omission probabilities as free parameters. Observed and best-fit surface distribution data were generated by drawing 10,000 samples of  $n = 232$  trials from each distribution. The inferred underlying distribution data were generated by drawing 10,000 samples of  $n = 76$  trials.

### A.3.2 Character-agent events

Table A.4: Model: Bootstrap means and 95% CIs for each word order in the observed surface distribution, best-fit surface distribution, and inferred underlying distribution based on block 1 descriptions of character-agent events.

	Observed distribution			Best-fit surface distribution			Inferred underlying distribution		
	lower 95% CI	mean	upper 95% CI	lower 95% CI	mean	upper 95% CI	lower 95% CI	mean	upper 95% CI
APV	0.119	0.163	0.211	0.132	0.179	0.229	0.172	0.234	0.301
AVP	0.383	0.445	0.511	0.383	0.448	0.511	0.509	0.584	0.656
PAV	0.004	0.018	0.035	0.013	0.034	0.062	0.018	0.044	0.080
PVA	0.053	0.088	0.128	0.066	0.102	0.141	0.086	0.132	0.184
VAP	0	0.004	0.013	0	0.004	0.013	0	0.006	0.018
VPA	0	0	0	0	0	0	0	0	0
AV	0.048	0.080	0.115	0.031	0.059	0.093	NA	NA	NA
PV	0.066	0.106	0.145	0.013	0.036	0.062	NA	NA	NA
VA	0	0	0	0	0.009	0.022	NA	NA	NA
VP	0	0	0	0.022	0.051	0.079	NA	NA	NA
V	0.062	0.097	0.137	0.048	0.079	0.115	NA	NA	NA

The best-fit model had omission probabilities as fixed parameters. Observed and best-fit surface distribution data were generated by drawing 10,000 samples of  $n = 227$  trials from each distribution. The inferred underlying distribution data were generated by drawing 10,000 samples of  $n = 163$  trials.

# Appendix B

## B.1 Chapter 4: Experiment 3 reconstruction steps

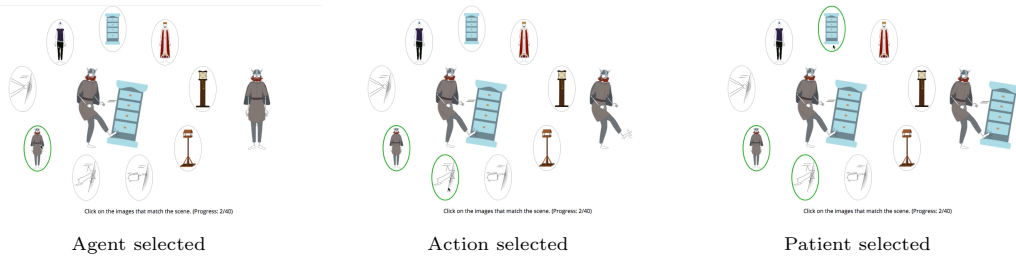


Figure B.1: Reconstruction steps – AVP



Figure B.2: Reconstruction steps – PAV

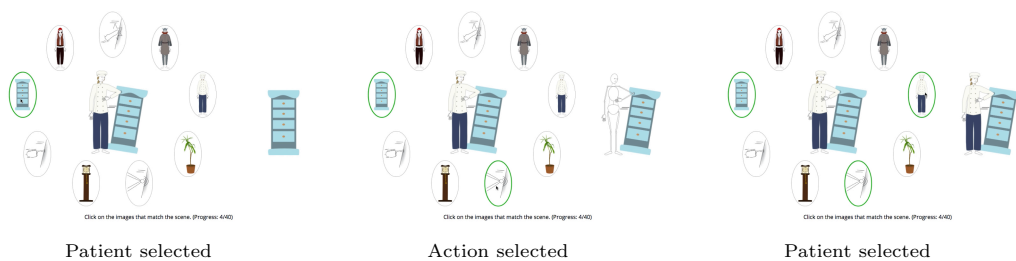
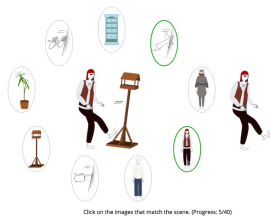


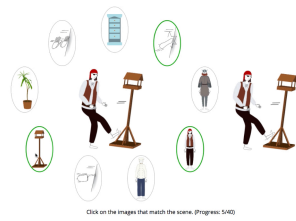
Figure B.3: Reconstruction steps – PVA



Action selected

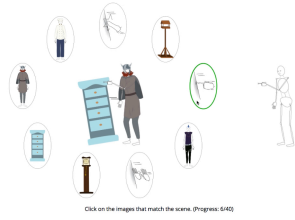


Agent selected

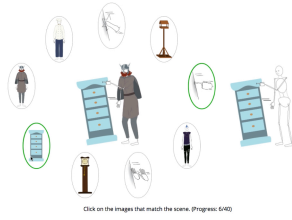


Patient selected

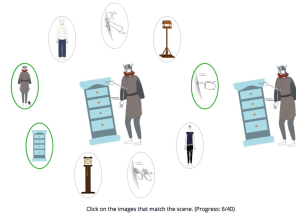
Figure B.4: Reconstruction steps – VAP



Action selected



Patient selected



Agent selected

Figure B.5: Reconstruction steps – VPA





# Appendix C

## C.1 Chapter 5: Summary of silent gesture results

The table below summarizes findings from studies investigating the effects of animacy on word order in silent gesture and signing tasks.

Table C.1

Study	Type	Native Language	Other language experience	Preferred order		Additional observations	
				Non-reversible	Reversible		
Gibson et al. (2013) <sup>a</sup>	Silent gesture	English (SVO)	N/A	SOV-like	SVO-like	SOV and SVO occurred in roughly equal proportions for non-reversible events	
		Korean (SOV)	N/A	SOV-like	SOV-like		
		Japanese (SOV)	N/A	SOV-like	SOV-like		
		Irish (VSO)	English (SVO)	SOV and SVO	SVO		
Futrell et al. (2015)	Silent gesture	Russian (SVO)	N/A	SOV	SVO		
		Tagalog (VSO)	N/A	SOV	SVO		
		English (SVO)	N/A	SOV-like	SVO		
Hall et al. (2013) <sup>b</sup>	Silent gesture	English (SVO)	N/A	SOV-like	SVO	Word orders for reversible events were less likely to contain an OV cluster compared with non-reversible events.	
Kocab et al. (2018) <sup>c</sup>	Silent gesture	English (SVO)	N/A	SOV	No preferred order	Word orders for reversible events were less likely to contain an OV cluster compared with non-reversible events.	
Meair et al. (2017) <sup>d</sup>	Silent gesture	Arabic (spoken: SVO, written: VSO)	Hebrew (SVO)	SOV	SOV and SVO	OSV also increased for reversible events.	
		Hebrew (SVO)	N/A	SOV	SVO		
		Turkish (SOV)	N/A	SOV	SOV		
		ABSL (iterate)	Arabic (VSO), ISL	SOV	SOV, SVO, and OSV		
		ISL (iterate)	Hebrew (SVO)	SOV	SVO		
		ABSL (non-iterate)	N/A	SOV	OSV		
		ISL (non-iterate)	N/A	SOV	SOV and OSV		
		KQSL (non-iterate)	N/A	SOV	SOV and OSV		

<sup>a</sup> SOV-like: orders in which the object appeared before the verb; SVO-like: orders in which the verb appeared before the object.

<sup>b</sup> SOV-like: included SOV, SV and OV. Results refer to Experiment 1.

<sup>c</sup> Results for reversible events refer to the animate-animate condition.

<sup>d</sup> All hearing participants were literate.

## C.2 Chapter 5: Word order coding procedure

On trials where it was not possible to match a word to one of the event constituents, we attempted to identify the intended word or referent using the procedure illustrated in Fig. C.1. The first step was to calculate the normalized Levenshtein distance between the word and all other words in the lexicon to identify a candidate match.

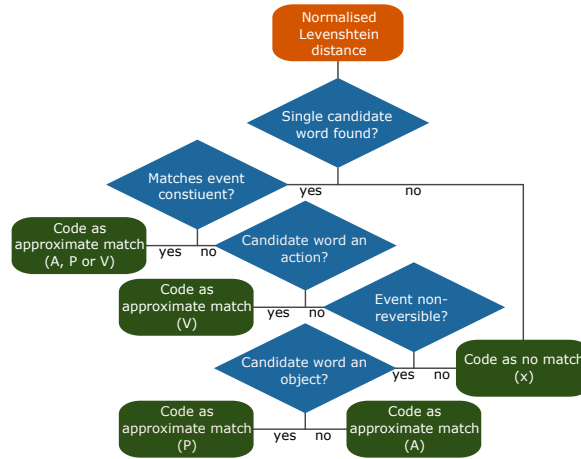


Figure C.1: Coding procedure used when a word supplied by the participant did not match any of the event constituents.

Following this procedure, if the coded word order contained duplicate constituents, exact matches (coded with lowercase 'a', 'p' or 'v') were retained and approximate word matches were replaced with 'x'. For example, aAv, which contains both an exact and an approximate match to the event agent ('a' and 'A', respectively), would be recoded as axv. For the analysis, exact matches were treated the same as approximate matches. For example, both 'apv' and 'Apv' were analyzed as an agent-patient-action response.

## C.3 Chapter 5: Image orientation: Experiments 5-6

### C.3.1 Word order

We investigated the effects of image orientation on the distribution of (valid) word orders in Experiment 5 and 6. We analysed the data using a mixed effects logistic regression with event type, image orientation and their interaction as fixed effects. Image orientation was coded as a binary flag labeled ‘mirror’ indicating if the agent appeared on the left of the image. We used deviation coding for both event type and orientation. The random effects structure varied due to convergence issues with some of the analyses.<sup>1</sup> Details of the analysis performed for each language type are detailed in Table C.2.

Table C.2: Details of the analyses performed for each language type in Experiment 5 and 6 investigating if word order was sensitive to image orientation.

Language type	Exp.	Model	Predictor	$\beta$	<i>SE</i>	<i>p</i> -value
V-final	Exp. 5	$is\_pav \sim event\_type * mirror + (1 + event\_type + mirror   workerId) + (1   event)$	<i>event\_type</i>	1.806	1.155	0.118
			<i>mirror</i>	4.229	1.730	0.015
			<i>event\_type</i> × <i>mirror</i>	0.004	0.975	0.997
	Exp. 6	$is\_pav \sim event\_type * mirror + (1   workerId)$	<i>event\_type</i>	-0.339	1.118	0.762
			<i>mirror</i>	10.593	1.677	<0.001
			<i>event\_type</i> × <i>mirror</i>	2.243	2.250	0.319
P-first	Exp. 5	$is\_pav \sim event\_type * mirror + (0 + event\_type   workerId)$	<i>event\_type</i>	1.259	2.010	0.531
			<i>mirror</i>	-0.064	0.369	0.862
			<i>event\_type</i> × <i>mirror</i>	-0.726	0.734	0.323
	Exp. 6	$is\_pav \sim event\_type * mirror + (1   workerId)$	<i>event\_type</i>	0.140	0.403	0.729
			<i>mirror</i>	-0.363	0.406	0.371
			<i>event\_type</i> × <i>mirror</i>	2.126	0.845	0.012

### C.3.2 Word order errors

We investigated the effect of image orientation on the tendency to produce an invalid order. For this analysis, we included only those responses where the word order could be unambiguously identified. We therefore excluded trials marked as ‘ambiguous’ or where one of the constituents was coded as ‘x’, since these represented lexical rather than grammatical errors. In addition, since we were only interested in responses from participants who produced invalid orders, we excluded individuals who produced a valid order on every trial.

A total of 7 participants from the V-final group and 14 from the P-first group were included in the analysis from Experiment 5. The most common word order error was AVP, which

<sup>1</sup>Convergence problems frequently occur in logistic regression models when there is (quasi)-complete separation in the data. An alternative to simplifying the model, as we did here, is to use Bayesian inference (Gelman et al., 2008; Kimball et al., 2019). To check the reliability of our simplified models, we conducted a separate Bayesian analysis that included the maximal random effects structure. In all cases, the results were in line with those obtained using the simplified maximum likelihood model. We do not report the details of the Bayesian models here. Instead, we refer the interested reader to the supplementary material.

accounted for 84.3% of invalid trials overall (97.5% of invalid V-final trials and 72.5% of invalid P-first trials). Twenty-four participants were included in the analysis from Experiment 6. All of these participants were in the P-first group. The overwhelming majority of invalid trials were PVP (91.1%).

Table C.3 lists the details of the statistical analyses. Binary predictors were deviation coded in all models.

Table C.3: Details of the analyses performed for each language type in Experiments 5 and 6 investigating the effect of image orientation on the tendency to produce an invalid order.

Language type	Exp.	Model	Predictor	$\beta$	<i>SE</i>	<i>p</i> -value
V-final	Exp. 5	$is\_invalid\_order \sim event\_type * mirror + (1 workerId)$	<i>event.type</i>	0.019	0.850	0.983
			<i>mirror</i>	-0.682	0.863	0.429
			$event.type \times mirror$	-0.015	1.721	0.993
	Exp. 6	NA	<i>event.type</i>	NA	NA	NA
			<i>mirror</i>	NA	NA	NA
			$event.type \times mirror$	NA	NA	NA
P-first	Exp. 5	$is\_invalid\_order \sim event.type * mirror + (1 + event.type + mirror workerId)$	<i>event.type</i>	0.863	0.813	0.288
			<i>mirror</i>	-1.232	1.185	0.298
			$event.type \times mirror$	-1.460	0.867	0.092
	Exp. 6	$is\_invalid\_order \sim event.type * mirror + (1 + event.type workerId)$	<i>event.type</i>	-0.553	0.500	0.269
			<i>mirror</i>	-2.000	0.345	<0.001
			$event.type \times mirror$	-0.992	0.690	0.151

## C.4 Chapter 5: Word order preferences in Experiments 6 and 7: candidate explanations

The table below lists the possible factors that may have resulted in an overall preference in Experiments 6 and 7 for APV in the V-final group and PVA in the P-first group.

Table C.4: Candidate explanations for the APV preference in the V-final group and the PVA preference in the P-first group in Experiments 6 and 7.

<b>V-final: APV &gt; PAV</b>		<b>P-first: PVA &gt; PAV</b>	
Explanation	Comment	Explanation	Comment
Native language interference	Agent-first consistent with canonical English word order.	Native language interference	Verb-medial consistent with canonical English word order; PVA analogous to English passive construction.
Salience of agents relative to patients	This explanation is broadly consistent with the cognitive salience hypothesis (more salient entities are expressed first).	Noisy-channel + previous linguistic experience	A preference for PVA for reversible events, as predicted by the noisy-channel hypothesis, combined with a preference for using a single order, deriving from previous linguistic experience, resulting in predominantly PVA responses.
Communicative pressure + agent-first preference	A consistent ordering strategy may serve to unambiguously identify participant roles. The agent-first preference may derive from previous linguistic experience and/or may reflect the salience of agents relative to patients.		