Conjunction, disjunction, and scalar implicatures in British Sign Language

Matthew Simon Brown
University College London
PhD Linguistics
May 2025

I, Matthew Simon Brown (also known as Matt Brown), confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. The ethics for this project were approved by the UCL Ethics Committee (project reference LING-2019-10-07).

Abstract

This project explores conjunction and disjunction in British Sign Language (BSL), and the effects of linguistic context and language acquisition on the relative strength of exclusivity inferences from disjunction in BSL (a type of scalar implicature).

Part 1 investigates BSL conjunction and disjunction via an analysis of spontaneous conversation data from the BSL Corpus, examining manual, non-manual, and asyndetic articulation. Descriptive findings provide relative frequencies of various ways of expressing these coordinations. A hierarchical clustering analysis indicates that contrary to assertions made for some other sign languages, there is no robust association between specific forms of co-occurring non-manual feature and coordination types in the sample, with the qualified exception of English mouthings. It is argued that such forms of non-manual marking are prosodic.

Part 2 documents a scalar implicature judgement task experiment with a continuous rating scale paradigm. Deaf adult BSL users were exposed to short descriptions of picture stimuli given by a video “guessing game partner”. Test conditions elicited exclusivity inferences from disjunctive statements; control conditions were designed to verify semantic understanding and reinforce the accessibility of the stronger scalar alternatives. The lexical structure of the coordinations and contrastive non-manual prosody over the coordinates were manipulated per trial item; the participants were grouped by age of BSL acquisition and English reading accuracy score. Bayesian mixed effects modelling indicates that for participants with the most language experience, where the lexical context is minimal, contrastive non-manual prosody has a tendency to increase the strength of exclusivity inferences, as has been demonstrated in analogous spoken language experiments. For those with less language experience, lexical context is more influential, but readings are generally more inclusive. These findings are argued to be compatible with a “probabilistic pragmatics” account of implicatures, and to further support the characterisation of non-manual marking of BSL coordination as prosodic.

Impact statement

Potential benefits of the research presented in this thesis are as follows.

1. **Language development in deaf children**. This thesis is not a direct study of language development, but the scalar implicature study presented in chapter 5 identifies some statistically credible differences in a specific area of pragmatic ability between adult participants with differing BSL and English language acquisition experiences. This has direct relevance to recent academic "calls to action" (briefly highlighted in §6.2) in the context of the social and cognitive benefits of pragmatic competence for deaf children and adults (with general implications for well-being).
2. **BSL teaching, and interpreter training**. The usage-based BSL Corpus study presented in chapter 3, and the study of exclusivity inferences in chapter 5, were both undertaken towards addressing gaps in existing documentation of conjunctive and disjunctive coordination in BSL (in plain terms, ways of expressing equivalents of "and" and "or"). The chapter 3 findings may be useful and relevant for BSL teachers and curriculum designers. BSL/English interpreters and their trainers may additionally find the discussions and new evidence presented in chapters 4 and 5 to be relevant in the context of working with intended meaning. The general topic area (inferential communication in the context of diverse language acquisition experiences) is well-suited to Continuous Professional Development workshops for interpreters.
3. **Statistical methods**. In experimental pragmatics, the judgement task rating paradigm and statistical approach presented in chapter 5 (Bayesian mixed effects modelling of continuous measurement responses based on an ordered beta distribution) was felt to be relatively novel and may be of interest. Separately, sign language corpus linguists may find the demonstration of a hierarchical clustering analysis in chapter 3 to be a useful method for approaching categorical annotation data (if they were not previously aware of it).
4. **Further research**. To the best knowledge of the author and his supervisors, this thesis was the first study of scalar implicatures in BSL, and it is hoped that it will not be the last. Very little is known about scalar implicatures in any sign language, so the findings and methods may also be of international interest in sign language linguistics. The final chapter summarises a number of directions for future study in sign language corpus linguistics, experimental sign language pragmatics, and the development of pragmatic competence in deaf children and adults.

Acknowledgements

- I am privileged in that this research project was supported by a studentship from the Arts and Humanities Research Council via the London Arts and Humanities Partnership, grant number AH/L503873/1 515892.
- I sincerely thank the deaf friends, colleagues, and strangers who have supported, included, and taught me over the years, not least those who contributed to this project's experimental data. This BSL research belongs to the BSL community.
- Thanks to Anne Leahy, who tracked down the Hutton (1855) reference to an old sign reminiscent of BSL's present day AND, and to Adam Schembri for asking her.
- Gratitude to Calum Medlock and Afshan Jamil, DCAL administrative officers, for their support with participant recruitment; also to Richard Jardine, Executive Officer at UCL Linguistics, for support and advice over the last several years, particularly through the Covid pandemic. Nothing gets done without a good administrator.
- I greatly valued the professional services of Alexandra Shaw, Marcel Hirshman, and Lathan Sivasubramaniam at Remark Access in producing BSL video stimuli and translations.
- Thank you to Kate Rowley, Mairead MacSweeney, and Bencie Woll for constructive questions and comments on a work-in-progress presentation of results.
- I'm forever grateful to Heidi Proctor, fellow traveller, long-suffering ear to bend and shoulder to cry on, for all of her support and practical suggestions.
- My academic life would not have even started without Peter Llewellyn-Jones, my former supervisor at Leeds, who saw something in me I didn't know was there.
- I am indebted to my examiners, Nausicaa Pouscoulous and Nick Palfreyman, for their constructive and thoughtful recommendations and comments.
- Much appreciation to Richard Breheny, my secondary supervisor, whose theoretical insights and experimental pragmatics expertise were greatly valued.
- Lastly, to Kearsy Cormier, my inspirational primary supervisor, without whose unfailing encouragement, tolerance, patience, generosity, and belief in me above and beyond the call of duty, this project would never have been completed.

Dedications

- To Kathy Gray, my Level 1 Deaf Community & Culture evening class tutor in 2003. Teachers generally value students that engage and ask questions, but there are limits. Kathy's solution to my repeated derailments of her lesson plan was to lend me her personal copy of Sutton-Spence & Woll (1999), a BSL linguistics textbook years ahead of the course's curriculum. I didn't understand a lot of it the first time through, but read it cover to cover and then bought my own copy. A couple of decades later, I was performing my own linguistics research. We sometimes forget that small acts of kindness can change the course of a life.
- To my parents, Barrie and Dianne Brown, to whom I owe everything.
- To my husband Markus, who is my entire horizon.

Table of contents

Abstract ... iii
Impact statement ... iv
Acknowledgements ... v
Dedications ... vi
Table of contents ... vii
Abbreviations ... xi
Glossing conventions ... xii
A note on terminology ... xii

1 General introduction ... 1
1.1 About this thesis ... 1
1.2 Conjunction and disjunction ... 2
1.3 Scalar implicatures ... 6
1.4 British Sign Language (BSL) ... 11
1.4.1 The sociolinguistic context of BSL ... 11
1.4.2 Phonology, and manuals versus non-manuals ... 14
1.4.3 BSL fingerspelling and borrowings from English ... 16

2 Review: conjunction and disjunction in BSL and other sign languages ... 18
2.1 Introduction ... 18
2.2 Corpus-attested conjunction and disjunction in BSL ... 19
2.2.1 Conjunction ... 19
2.2.2 Disjunction ... 21
2.2.3 List buoys ... 22
2.3 Manual forms and asyndesis ... 24
2.4 Non-manual contributions ... 28
2.5 Conclusions ... 32

3 Usage-based study: conjunction and disjunction in the BSL Corpus ... 33
3.1 Introduction ... 33
3.2 Method ... 34
3.2.1 Participants and conversational data ... 34
3.2.2 Coding and procedure ... 34
3.2.3 Analysis ... 36

3.3 Results 36
3.3.1 Descriptive statistics 36
3.3.1.1 Token types per coordination type 36
3.3.1.2 Tokens per coordination type for syndetic coordination 37
3.3.1.3 Surrounding context for asyndetic coordination 38
3.3.1.4 Brow movements per coordination type 39
3.3.1.5 Body lean per coordination type 39
3.3.1.6 Mouthings per coordination type 40
3.3.1.7 List buoy episodes 41
3.3.2 Quantitative analysis 42
3.4 Discussion 45
3.4.1 Questions addressed 45
3.4.2 Asyndetic disjunction 47
3.4.3 Disjunctive list buoys 50
3.4.4 Non-manual features 52
3.4.5 Limitations and future directions 53
3.5 Conclusions 53
4 Review: scalar implicatures and judgement tasks 55
4.1 Theoretical accounts 56
4.1.1 Gricean and “traditional” accounts 56
4.1.2 Relevance Theory 59
4.1.3 Probabilistic accounts 60
4.1.3.1 The “constraint-based” account 61
4.1.3.2 The Rational Speech Act framework 62
4.1.3.3 Recursive reasoning and theory of mind 64
4.2 Experimental approaches 66
4.2.1 Judgement task paradigms 67
4.2.2 The consequences of rating scale granularity 71
4.2.3 The acquisition of pragmatic competence 75
4.2.4 Scalar implicatures in American Sign Language (ASL) 81
4.2.5 Language background issues and grammaticality judgements in BSL 88
4.3 Conclusions 90

5 Experimental study: the effects of linguistic context and language acquisition on the strength of exclusivity inferences in BSL 94
5.1 Introduction 94
5.1.1 Conjunction, disjunction, and stress in BSL and other sign languages 94
5.1.2 Age of BSL acquisition and English reading ability 95
5.1.3 Judgement task design 96
5.1.4 Research questions and predictions 98
5.2 Method 99
5.2.1 Participants and exclusion criteria 99
5.2.2 Materials and procedure 100
5.2.2.1 BSL sentence/picture task 101
5.2.2.2 English reading accuracy task 108
5.2.3 Analysis 108
5.3 Results 110
5.3.1 Descriptive statistics 110
5.3.1.1 Participants 110
5.3.1.2 Age of BSL acquisition 110
5.3.1.3 Reading accuracy scores 111
5.3.1.4 Test condition responses 111
5.3.2 Quantitative analysis 115
5.3.2.1 Summary 115
5.3.2.2 Marginal effects of item factors 117
5.3.2.3 Marginal effects of participant factors 117
5.3.2.4 Marginal effects of factor interactions 118
5.4 Discussion 120
5.4.1 Research questions addressed 120
5.4.2 Supported accounts of scalar implicature, and future directions 125
5.4.3 Limitations 129
5.4.3.1 Sample size and participant balance 129
5.4.3.2 Participant reactions to the task 130
5.4.3.3 Remote participation 131
5.5 Conclusions 132

6 General conclusions 134
6.1 Summary 134
6.2 Future directions 138
References 140
Appendix A: Information sheet 152
Appendix B: Main task instructions 155
Appendix C: Main task trial item combinations 157
Appendix D: English reading task instructions 158
Appendix E: English reading task questionnaire 159
Appendix F: Tables of group average marginal effects 162

Abbreviations

AoA	age of acquisition
ASD	Autistic Spectrum Disorder
ASL	American Sign Language
Auslan	Australian Sign Language
BSL	British Sign Language
CLU	clause-like unit
DCAL	Deafness, Cognition, and Language Centre, University College London
DGS	Deutsche Gebärdensprache (German Sign Language)
DTS	Dansk tegnsprog (Danish Sign Language)
GCI	generalised conversational implicature
HCPC	Hierarchical Clustering of Principal Components
HKSL	Hong Kong Sign Language
HZJ	Hrvatski znakovni jezik (Croatian Sign Language)
ISL	Irish Sign Language
Libras	Língua Brasileira de Sinais (Brazilian Sign Language)
LSF	Langue des signes française (French Sign Language)
MCA	Multiple Correspondence Analysis
NGT	Nederlandse Gebarentaal (Sign Language of the Netherlands)
NZSL	New Zealand Sign Language
PCI	particularised conversational implicature
RA	reading accuracy
RSA	the Rational Speech Act framework
RSL	Русский жестовый язык (Russian Sign Language)
SSE	Sign Supported English
STS	Svenskt teckenspråk (Swedish Sign Language)

Glossing conventions

Written glosses of sign language forms are represented by small capitals (e.g. SIGN). Glosses containing hyphens always refer to a single sign (e.g. NOT-HAVE).

When referring to BSL, especially for citation forms and annotated video data from the BSL Corpus, the gloss used matches an "ID gloss" following BSL Corpus conventions (Cormier, Fenlon, Gulamani, & Smith, 2017) unless indicated otherwise. These glosses are the index to a lemmatised BSL SignBank database entry (e.g. DEPEND04). Glosses with a dotted underline are hyperlinked to their SignBank page (e.g. AND). ID glosses may also have a colon-separated prefix which denotes a type or function identifier (such as FS: for a fingerspelling sequence, G: for a gesture, PT: for a pointing sign, CA: for constructed action). BSL SignBank is a public resource, but some of the data is restricted: researchers can request full access by visiting the website (https://bslsignbank.ucl.ac.uk/) and registering (see About > For Researchers).

For BSL glosses cited from other literature, and for the glossing of any other sign language, the glosses echo those given by the cited source.

Conventionalised manual gestures found across spoken and signed languages are identified in italics (e.g. *palm-up*).

A note on terminology

The pragmatics literature often refers to a hypothetical "speaker" and "listener" when discussing conversational implicatures. In this thesis, I prefer to use "producer" and "receiver" or "addressee", since they are neutral terms with regards to language modality (signed versus spoken forms). However, there are also cases where I echo the terms used by the literature under discussion. The occasional use of terms like "speaker" and "listener" (and verbs like "say") is always intended to include BSL and other sign languages.

1 General introduction

1.1 About this thesis

The seeds of this project were planted during my former studies and practice in the field of British Sign Language (BSL) / English community interpreting. People are generally aware that what we say is not always what we mean, but for professional interpreters, the difference between literal and intended meaning is daily bread and butter. It was translation and interpreting studies which first introduced me to pragmatics, the study of context and the relationship between meaning and user. That led me to language sciences and sign language corpus linguistics, and from there, research. I no longer practice as an interpreter, and this thesis is not about that kind of interpreting, but I still think about how interpreters should handle implicit meaning and the setting-specific balance between justifiable explicitation and over-interpreting.

The chapters that follow are ultimately concerned with an extremely specific kind of implicit meaning: scalar implicatures derived from the contrast between conjunction and disjunction. We will get to the nitty gritty in good time, but it is first important to note that implicature is generally assumed to be “language universal”, a human thing, a product of cognition. Scalar implicatures specifically have been described as “the *drosophila* of experimental pragmatics” (Degen & Tanenhaus, 2019; Mazzarella, Trouche, Mercier, & Noveck, 2018). This is in reference to the fruit fly, the tiny insect favoured in genetics research because of the speed with which it reproduces (when it isn’t hanging around your used teabags). The analogy alludes to the fact that there have been a great many experimental studies of scalar implicature in spoken and written forms of language, some of them produced by early career researchers like myself. Yet at the time this project was first proposed, to my knowledge there had been a total of one study of scalar implicatures in any sign language (Davidson, 2011), and it was American Sign Language (ASL), which does not belong to the same language family as BSL. If scalar implicatures really are “language universal”, I felt it to be scientifically and politically important that more than one sign language has had the opportunity to contribute evidence. For those who might find the *drosophila* comparison unflattering, it should be noted that studies of fruit fly genetics have led to important breakthroughs in medicine. Obviously I make no such claim here.

The remainder of this chapter serves as a general introduction to the corpus-based and experimental studies that follow, establishing some necessary background and

technical terms. For the current chapter only, when first defining a technical term that will be frequently used throughout the chapters that follow, it will be marked with **bold text**.

In chapter 2, existing documentation and studies of conjunction and disjunction in BSL and other sign languages are examined, establishing expectations and research questions for the corpus-based study that follows in chapter 3. This discusses findings from an original analysis of a conversational data sample from the BSL Corpus, including the relative frequencies of five distinct clusters of conjunctive and disjunctive coordination structures and some details of their associations with form and context.

In chapter 4, we will explore theoretical and methodological issues around the experimental study of scalar implicatures, with the aim of informing the design and research questions for such a study in BSL. That follows in chapter 5, which discusses a web-based sentence/picture judgement task experiment with a continuous rating paradigm, designed to capture variable strengths of exclusivity inference made by deaf users of BSL in response to pragmatically infelicitous disjunctive statements (based in part on conclusions drawn in previous chapters). Interactions between linguistic context (levels of prosodic and lexical context in the coordination structure) and participant language background (levels of age of BSL acquisition and English reading accuracy) are analysed, and the results are discussed in terms of contrasting accounts of implicature.

Finally, in chapter 6, the conclusions of the previous chapters are drawn together and suggestions are made for future study.

1.2 Conjunction and disjunction

In linguistics, conjunction and disjunction are forms of **coordination**, a syntactic structure in which two or more units of the same type are linked in such a way that they seem to behave as one (Haspelmath, 2004, 2007). This is believed to be possible in all languages, but there is considerable cross-linguistic variation in how coordination is constructed. Examples in English are given below, with the unit structure indicated by nested square brackets. The sub-units might be noun phrases (example 1), verbs (2), or clauses (3), and can be referred to as **coordinates**. If there is a lexical item which delimits the coordinates (or a bound morpheme in some languages), it is a **coordinator** (marked with an underline), but this is not always necessary, as we will later see.

1. I'll have [[beans on toast] <u>and</u> [a cuppa]], please.
2. They [[worship] <u>and</u> [adore]] her.
3. She was [[checking her watch] <u>and</u> [planning her exit]].

In each example above, the coordinated unit takes on the same grammatical role as the uncoordinated sub-units. For example, the coordinated noun phrases in example 1 (“beans on toast and a cuppa”) have the role of object in the subject-verb-object structure: if they were replaced by a much simpler uncoordinated noun phrase (e.g. “that”), the sentence would still be grammatically conventional.

There are distinct types of coordination. If the coordination has an additive or cumulative sense, such that all of the coordinates are equally relevant or possible or required, it is referred to as **conjunctive** (in English, the most common lexical conjunction is “and”, examples 1-3). But if the coordinator marks a sense of contingent alternatives, or items of unequal relevance or likelihood, it is **disjunctive** (in English, the primary disjunction is “or”, example 4). If the coordinates have a sense of contrast or unexpectedness, the coordination is **adversative** (in English, “but”, “yet”, “however”; example 5). If there is a sense of an inferential relationship between the coordinates, the coordination is **illative** (in English, “so”, “thus”, “for”, “therefore”; example 6).

4. You can have [[cake] or [death]].
5. He is [[very nice] but [a bit dim]].
6. [[The lecture was boring] so [I jumped out of the window]].

The chapters that follow are almost entirely focused on conjunction and disjunction, but adversative coordination may pop up in passing. It is worth mentioning that there is an inconsistency in the use of the term “conjunction” across reference sources: in some contexts, by some authors, it is used to mean a lexical coordinator of any type. However, for the remainder of this thesis, **conjunction** is only used to refer to additive coordination (“and”), and **disjunction** is used to mean alternative coordination (“or”).

One of the hallmarks of coordination is that it is symmetrical or commutative: changing the order of the coordinates does not strongly affect the meaning (example 7). In some cases the conventional choice of coordinator may differ with order depending on the coordination type, the specific language, and the semantics involved (example 8). If a linked clause structure denotes an asymmetric relationship where one clause has a dependency on the other (e.g. causal, conditional, purposeful), it is technically **subordination** rather than coordination (in English, often marked by subordinators such as “because” or “when”, as in example 9). For subordination, swapping the order of the clauses will almost certainly change the meaning.

7. I do not like [[green eggs] and [ham]].
   ≈ I do not like [[ham] and [green eggs]].
8. [[It was raining], so [we stayed at home]].
   ≈ [[We stayed at home], for [it was raining]].
9. [[I was late] because [I stopped for coffee]].
   ≠ [[I stopped for coffee] because [I was late]].

This thesis is mostly unconcerned with subordination, but there is a related issue worth noting. The difficulty for curious language users is that it is not always a straightforward matter to distinguish coordination from subordination by syntax alone. In examples 10-13 below, each sentence has the same general structure: at first glance, the units all seem to be linked by the conjunction “and”. However, in semantic and pragmatic terms (on which more follows later), we might infer simultaneous “while” (10), sequential “and then” (11), cause/effect “and as a result” (12), and purposeful “in order to” (13) meanings, the latter two of which are technically subordination rather than coordination. None of these meanings are guaranteed by the presence of “and” alone: as a semantic and pragmatic phenomenon, the sense comes from the context. In pragmatics, this is sometimes described as **conjunction enrichment**. In grammatical terms, some might describe it as syndetic parataxis (to be clarified below).

10. Matt [[worked on his thesis] and [shed bitter tears]].
11. Matt [[took off his boots] and [jumped into bed]].
12. Matt [[fell down a hole] and [broke his leg]].
13. Matt [[went to the shop] and [bought a cake]].

When coordination is overtly marked by a lexical coordinator, this is referred to as **syndesis.** In English, when three or more items are linked, a coordinator is typically inserted just once, before the final item (**monosyndesis**, example 14), but this is not compulsory: coordinators can also delimit each and every listed item (**polysyndesis**, example 15, Shakespeare). It is also possible for the coordinator to be entirely absent (**asyndesis**, example 16, Conrad). Polysyndetic and asyndetic coordination are somewhat associated with emphatic, rhetorical, poetic, or stylistic purposes in English, and otherwise monosyndetic coordination is more typical, but this association cannot be assumed for other languages (more detail on this point follows in §2.1, §2.3). The chapters that follow will frequently refer to syndetic and asyndetic coordination.

14. I'll have [[eggs], [beans], and [chips]].
15. [[Poison], or [fire], or [suffocating streams]], I'll not endure it.
16. [[An empty stream], [a great silence], [an impenetrable forest]].

As a sidebar, note that asyndetic coordination has a related concept in subordination. When an asymmetric dependency between adjacent clauses can be inferred despite the absence of an explicit subordinator, this is referred to as **parataxis**; it can be **syndetic parataxis**, which (in English) can be marked with what looks like a coordinating conjunction (examples 12-13 above), or **asyndetic parataxis**, which is not marked with any linking item (also sometimes referred to as **juxtaposition**). If subordinated clauses are explicitly marked, that is **hypotaxis** (example 9 above). Examples 17-20 below are versions of the "enriched conjunction" structures in examples 10-13 above, but this time omitting the conjunction so that the clauses are simply juxtaposed. It does not make much of a difference: the same intended meanings (simultaneity, sequence, causality, purpose) can still be pragmatically inferred. They can be classed as asyndetic coordination (17-18) and asyndetic parataxis (19-20) or otherwise, depending on your preferred theoretical framework, but regardless, the syntax alone does not distinguish the function or the sense.

17. [[Matt worked on his thesis]. [He shed bitter tears]].
18. [[Matt took off his boots]. [He jumped into bed]].
19. [[Matt fell down a hole]. [He broke his leg]].
20. [[Matt went to the shop]. [He bought a cake]].

Finally, in the interests of tying this section to the next, it should be noted that conjunction and disjunction have analogues in formal logic (but experience in that field is not required to understand this thesis). Conjunction is an operator in Boolean algebra and propositional logic (related systems of logical operations on binary truth values). The proposition "A and B" can only evaluate as true if proposition A and proposition B are both individually true. However, in natural languages, there is an inherent ambiguity as to whether "or" signifies "either or both" or "either but not both". In logic, the former is an **inclusive disjunction**, where "A or B" is true as long as either one or both of A and B are true; the latter is an **exclusive disjunction**, where "A xor B" is only true if A is true and B is false, or A is false and B is true. Boolean logic makes a symbolic distinction between these three operators (∧ and, ∨ inclusive or, ⊕ exclusive or): see Table 1. They also have corresponding operations in set theory (∩ intersection, ∪ union, △ symmetric difference), another system of logic. But interestingly, and most relevant of all for our purposes, while

natural human languages tend to make some kind of marked distinction between conjunction and disjunction, no known natural language makes a fully lexicalised distinction between exclusive and inclusive disjunction (Haspelmath, 2007). Context is often needed to establish the intended meaning of "or", requiring **pragmatic competence** (the ability to convey and infer intentions and meanings beyond literal semantics).

A	B	A ∧ B	A ∨ B	A ⊕ B
TRUE	TRUE	TRUE	TRUE	FALSE
TRUE	FALSE	FALSE	TRUE	TRUE
FALSE	TRUE	FALSE	TRUE	TRUE
FALSE	FALSE	FALSE	FALSE	FALSE

Table 1: Boolean truth table for ∧ conjunction, ∨ inclusive disjunction, and ⊕ exclusive disjunction

Regarding how any of the above might play out in BSL, there has been relatively little research on the topic in comparison to studies of spoken and written forms of language. Chapters 2 and 3 explore how conjunction and disjunction can be expressed in BSL, consulting existing sign language documentation and performing an original usage-based exploration of a sample of conversation data from the BSL Corpus.

1.3 Scalar implicatures

We will not encounter an actual study of scalar implicatures until the latter half of this thesis, but there will be references to that goal throughout. For those who are less familiar with implicatures and pragmatics, the following serves as an introduction in relatively relaxed terms; some of what follows will be covered more formally in chapter 4.

In linguistics, **implicatures** (utterances for which the intended meaning is not made explicit) typically fall under the heading of **pragmatics**, the study of the relationship between linguistic form and user: how context, common ground, situation, and conversational goals affect linguistic choices and mutual understanding. In this field, communication is often characterised as a purposeful activity, with goals or intentions. "Context" is typically understood by the layperson to mean the choice of words used in the same sentence (**lexical** and **grammatical context**), or topics and information already established during the conversation (**discourse context**). What is less widely understood is that there is also **prosodic context** (influences from intonation or modulation choices), **epistemic context** (unstated "world knowledge" or prior beliefs, conversational goals), **social context** (relationships between those conversing, e.g. perceptions of relative

status, alignments), and **situational context** (the physical environment or communication medium). Any of these contextual influences might change how a statement is understood. Pragmatics is therefore a thoroughly human area of study with enormous interdisciplinary potential, testing and blurring the imaginary borders between linguistics, psychology, cognitive science, and philosophy. Lexical, prosodic, and epistemic context will be especially important in the chapters to follow.

One particular contextual influence on the interpretation of intended meaning may come from the receiver's consideration of alternative statements which were not explicitly expressed by the producer: the implication that unstated contrasting alternatives are not true, or to put it more precisely, that the producer believes that what they chose to express is more true than what they chose *not* to express. This may seem blatantly obvious on first reading, but there are subtle and sometimes powerful consequences to the idea, especially when you take the producer's *intentions* into account. An early observation of the phenomenon is often attributed to John Stuart Mill:

> If I say to any one, "I saw some of your children to-day", he might be justified in inferring that I did not see them all, **not because the words mean it**, but because, if I had seen them all, it is most likely that I should have said so: though even this cannot be presumed unless it is presupposed that I must have known whether the children I saw were all or not.
>
> *– Mill (1865), in Horn (2009), Geurts (2010); emphasis mine*

In other words, meaning can come not just from what was said, but from what the producer and receiver *know* or *believe* (**presuppositions**), as well as what was *not* said but might have been. It took another century for Western academia to produce a fuller framework for the implications derived from unstated alternatives. Grice (1968, 1975), in a lecture series and publications which shaped the study of pragmatics up to the present day, characterised the phenomenon as falling under "conversational implicature", said to be governed by a "co-operative principle" and a set of "conversational maxims" under which people tend to operate. We will discuss this "Gricean" framework (and alternatives) in more detail later (§4.1), but for those unfamiliar with implicatures, consider the following exchange between Abir and Barbara, who are good friends:

> ABIR: Carlos is handsome, clever, and generous.
> BARBARA: He's *generous*.

Assuming that Barbara is being “co-operative” (she is not trying to mislead Abir by providing information which she secretly believes to be false), and given that she has echoed only one of the three attributes which Abir has proposed as true descriptions of Carlos, it would be reasonable to think that Barbara is deliberately trying to imply that she does not think the other two descriptions are accurate and true, even though she has not said so explicitly. This is a key concept in understanding implicature, worth drumming home once more: we don’t simply derive information from the surface semantics and literal meaning of an utterance, we also attribute communicative *intent* to the linguistic choices of others, and we have practically simultaneous access to many different kinds of supporting context. That includes alternative and contrasting utterance choices that might have been made instead. If Barbara had completely agreed with Abir, she could have said “Yes, he certainly is”, or echoed all three of the adjectives. The fact that she did not seems to suggest that she does not agree that Carlos is handsome or clever and that her intention is to communicate this disagreement. Her intonation and manner (the prosodic context, more difficult to convey in written form) may further support a reading of disagreement, adding focus or emphasis to the constituents that are agreed on as opposed to those which are not. Barbara had many other options around *explicitly* indicating that she disagrees, such as flat-out contradiction (“No, he’s neither handsome nor clever” or “No, he’s ugly and stupid”), but she might have had conscious or unconscious motivations to avoid doing so, such as wishing to avoid face-threatening behaviour, or as part of a general drive towards economical communication, among many others.

A more complex example is as follows:

ABIR: Is Carlos joining us later?
BARBARA: He’s out with a man tonight.

There is potentially more than one inference to be drawn from Barbara’s response. On similar lines to the previous example, a reasonable reading of her general intention is to convey “No, I don’t think Carlos is joining us”, even though she has not explicitly said so. If she thought it were true that Carlos will join them, she could have answered the question unambiguously and economically (“Yes”); instead she has offered an alternative account of what she believes Carlos will be doing. But it might also be possible to derive a further intention from Barbara’s choice of saying “a man”, possibly dependent on shared knowledge, but also involving logic and semantics. If Carlos were meeting his brother, or his father, or his boyfriend (all conventionally understood as being men), it would have been almost equally economic and more informative for Barbara to have said so (“He’s out

with his boyfriend tonight", or "He's out with Dmitri tonight", if Barbara knows that Abir knows that Dmitri is Carlos's boyfriend). The fact that she did not say any of those things implies that she does not believe them to be true, and that she intends for Abir to understand that it is some other man that Carlos is seeing (not a brother or father or boyfriend or any man that Barbara knows that Abir knows about), despite the fact that it would still be *literally* true that "Carlos is out with a man" if it were in fact any of those.

Speaking of literal semantics: in some cases, the unstated alternative is a *thoroughly* logical one. This leads us, at last, to our working definition of scalar implicatures, a term originally coined by Horn (1972) in an exploration of the Gricean framework, but the concept was also known for a time as "quantity implicatures" or "Horn scales":

> **Scalar implicature**: when the use of an informatively weaker statement is understood as implicating that any stronger alternative on the same scale is not true, even when the stronger is not logically excluded by the weaker.

A "scale" in this context is a logical sequence of semantically related but contrasting terms, ordered by increasing "informativity" (in the sense of precision or certainty). Scales can be composed of several members, but for simplicity, we will limit examples to two-member scales. The scale most often cited and studied is <*some, all*>, which features in Mill's example quoted above, the relevant part repeated here:

JOHN: I saw some of your children today.

This kind of statement is often understood as meaning "some but not all of your children", despite that not being the literal meaning. Following the definition of scalar implicatures above, "I saw all of your children today" would have been "informatively stronger", because "all" is absolute and precise: it only refers to one possibility, the one in which each and every one of your children was seen by John. In comparison, "some" is informatively weaker, essentially vaguer, in that for the same set of children there may be a large number of possible combinations for which of them were seen or not, and the use of "some" does not tell you which of those possibilities it was. Crucially, in purely logical terms, "all" is also included in the possibilities covered by "some", since if John had seen all of them, it would still be literally true that he had also seen some of them. We will unpack this in a little more detail later (§4.1.1). But you might think that by deliberately not saying "all", John must not believe that "all" was true; in which case by saying "some" he must intend for you to understand "some but not all". The range of possibilities has been narrowed: you have derived a scalar implicature.

It is here that we can finally tie in disjunction and conjunction, which can be said to form a scale *<or, and>*. As described near the end of the previous section, these coordinators might be viewed in terms of logical operations on the truth value of the coordinated items. We also established that “or” is typically ambiguous between “either or both” and “either but not both”. Consider this example:

ABIR: Where did Carlos go?
BARBARA: He went to the library or the pub.

Following the argument given above for the *<some, all>* scale, “and” is more informative than “or”. If Barbara thought it were definitely true that Carlos went to both the library and the pub, she would be expected to say “the library and the pub”. If the proposition “A and B” is true, there is only one logical combination of possible truth values for A and B respectively, that A is true and B is also true: “and” is therefore the maximally informative member of the *<or, and>* scale. However, referring back to Table 1, for the alternative construction “A or B” to evaluate as true, there are at least two possible combinations of truth values, and three if the disjunction is understood as inclusive rather than exclusive (again, to be picked up in §4.1.1); therefore, “or” is informatively weaker. But if Abir were to think that by using “or”, Barbara was implying that “both” should be rejected (because if she thought that “both” was a truthful description, she would have used “and”) then he might believe that in this case Barbara intended for “or” to be understood as exclusive, i.e. that Carlos went to the library, or the pub, but not both. In which case, Abir would have derived a scalar implicature.

To reiterate, this kind of inference is not at all guaranteed and is dependent on context as well as logic. A different situation and context might not be equally likely to carry the same kind of exclusivity inference:

BARBARA: I’m hungry.
ABIR: You can have a biscuit or a cake.

In this example, does Abir want Barbara to understand that she can have a biscuit or a cake but not both? Without more context, it is hard to say. The inference might be influenced by the syntactic context of the modal qualifier “can”, suggesting that general permission is being given. It might depend on the intonation contours applied: for example, we will explore in chapters to follow how stressing a disjunction can contribute to a reading of exclusivity in English (“You can have a biscuit *or* a cake”), and what that might look like in BSL. It might also depend on their situation and goals, the details of their relationship, or

their shared experiences, such as Barbara's knowledge of Abir's views on the propriety of snacking between meals, or any other information thought to be relevant.

Adult humans manage to resolve these ambiguities fairly effectively most of the time, almost always without noticing, using several cognitive and linguistic facilities. The field of study required to explain the phenomenon is quite broad. However, as we will see in chapter 4, one of the aims of the study of scalar implicatures is to explore and identify the contextual and cognitive influences on the strength of belief that the unstated stronger alternative was intended to be rejected. Chapter 5 of this thesis is such a study, concerning the interpretation by a range of BSL users of disjunctive statements with systematically manipulated prosodic and lexical context.

1.4 British Sign Language (BSL)

This section establishes some background for BSL, as well as terminology and key concepts from the sign linguistics field that are repeatedly referenced across the chapters that follow. For one-off references to more specific technical terms, footnotes will be used.

1.4.1 The sociolinguistic context of BSL

BSL is the language of the deaf signing community in the United Kingdom, and plays a central role in defining a sense of cultural and linguistic identity (Ladd, 2003; Rowley & Cormier, 2023). The **modality** of BSL is visual-gestural, in contrast with the oral-auditory modality of English, the main language of the UK's hearing majority. BSL has a different grammar and vocabulary to English, and it is also not necessarily related to sign languages in other countries that count English as a majority language. For example, American Sign Language (ASL) and BSL can be fairly described as mutually unintelligible (Schembri, Rowley, & Leeson, 2024). However, BSL does have a historical relationship with other sign languages that are cognate to an extent, such as Australian Sign Language (Auslan) and New Zealand Sign Language (Schembri et al., 2010). It is also related, perhaps less closely, to varieties of South African Sign Language, Maltese Sign Language, and Maritime Sign Language in parts of Canada (Johnston & Schembri, 2007), the latter of which is thought to be moribund (Yoel, 2009). BSL does have contact with English via mouthings and the manual alphabet (both described further below). Deaf users of BSL are generally bilingual in spoken and/or written English to some extent, although that extent varies. The number of BSL users across the UK is not known, but recent estimations have claimed it may be as high as 125,000 (Schembri et al., 2024) against a total population of

almost 70 million; this is likely to include some hearing children of deaf adults and other hearing BSL users.

The history of sign language in the British Isles is difficult to trace (Kyle & Woll, 1985), but there is evidence of signing being used in the fourteenth century, and plausible links to modern BSL dating from the seventeenth century (Sutton-Spence & Woll, 1999). The term “British Sign Language” is a relatively recent coinage in comparison, first appearing in published literature in the 1970s (Leahy & Brown, 2020), but it is now in widespread use. After decades of campaigning, BSL was only recently recognised in law as an official language of the UK (*British Sign Language Act*, 2022). A secondary school qualification in BSL was also recently announced (GOV.UK, 2023).

In terms of language transmission, a minority of deaf children are born to at least one deaf parent. The proportion is typically claimed to be 5-10%, but it is difficult to find an authoritative source for a recent UK estimate. Such statistics have usually been based on indications from surveys conducted by third sector or charitable organisations, or extrapolations from public health data on congenital deafness. In the United States, a figure for the proportion around the turn of the millennium was estimated at closer to 5% of deaf people (Mitchell & Karchmer, 2004); there is no particular reason to think that it would be very different in the UK, but this is essentially speculation. Whatever the exact proportion, it sets a small upper limit on the numbers of deaf BSL users who, as infants, had access to deaf signing parents as language models, given that not all deaf parents use BSL or count it as their preferred language. This entails that unlike the largely vertical transmission of spoken languages across generations (parent to child), the transmission of sign languages tends towards the horizontal (peer to peer). On the whole, deaf children and adolescents are much more likely to learn BSL at school or in the community than from their parents, and their age of first language acquisition is more likely to be delayed in comparison to their hearing peers. They may also initially acquire BSL from deaf or hearing parents who have themselves only recently started to learn it. Even in the twenty-first century, language deprivation is still a reality for a proportion of deaf children in the UK and elsewhere, with all of the linguistic, cognitive, developmental, and social consequences that might entail (e.g. Buyle & Crollen, 2022; Cheng, Roth, Halgren, & Mayberry, 2019; Hall, Eigsti, Bortfeld, & Lillo-Martin, 2017; Hall, Hall, & Caselli, 2019; Hall, 2017; Hecht, 2020; Matthews & Kelly, 2022; Szarkowski, Young, Matthews, & Meinzen-Derr, 2020; Twomey, Price, Waters, & MacSweeney, 2020).

However, the lack of early language models is not the only reason why schools have played such a prominent historical role in shaping the transmission and sociolinguistic

variation of BSL (Quinn, 2010). Following British industrialisation in the eighteenth and nineteenth centuries, mass migration to urban centres contributed to the establishment of specialist schools for deaf children, including residential boarding schools. By 1870, there were 22 deaf schools across Britain, creating regional focuses for BSL transmission (Kyle & Woll, 1985). But at the Congress of Milan in 1880, European educators called for a ban on the use of sign languages in the classroom in favour of oral methods intended to reinforce spoken languages and lip-reading, with the now thoroughly discredited justification that acquiring a sign language actively interferes with the acquisition of spoken and written language. This ban was enforced to varying extents across British deaf schools over the following decades (tailing off over the second half of the twentieth century). Regardless, many deaf children persisted in signing outside of the classroom.

By the 1980s, many of the residential deaf schools had been closed and were replaced with mainstreaming education practices, in which deaf children were integrated into classes of hearing children or attended a specialist unit attached to a state school. In recent times, deaf education policy has typically favoured a “total communication” approach, a mix of communication strategies intended to be tailored to the needs of individual pupils (e.g. BSL, spoken English, Sign Supported English, fingerspelling, lip-reading, symbolic/visual communication aids). The closure of deaf schools, and the attitudes towards sign language of some educators and policy-makers, have been perceived by some members of the BSL community as a direct attack on the status and transmission of BSL as a natural language (Ladd, 2003; Rowley & Cormier, 2023). Specialised deaf schools still exist, although some have recently broadened their remit to include children with “SLCN” (speech, language and communication needs not necessarily related to deafness).

Another potential source of transmission change is the reported closure of deaf clubs over recent years (local community organisations with a variety of origins, some with long histories, that focus on face-to-face social activities and mutual support). In this case the driving factors might be part of wider social changes, not least the technological impact of relatively affordable and portable access to video calls and social media, as well as local government and third sector funding issues and the general economic background. However, deaf grassroots organisations continue to be active and adaptive.

Deaf education policies and practices, as well as broader social norms, have therefore varied across time and place, suggesting that age and geographic region are likely to be predictors of BSL variation. Several usage-based studies in the last decade have found such effects (Börstell, Crasborn, & Schembri, 2024; Brown & Cormier, 2017;

Proctor & Cormier, 2022; Stamp, Schembri, Fenlon, & Rentelis, 2015; Stamp et al., 2014), although not all such studies have identified sociolinguistic factors as interacting with linguistic phenomena (Fenlon, Schembri, & Cormier, 2018).

1.4.2 Phonology, and manuals versus non-manuals

The term **phonology**, the study of the fundamental constituent parts of a language, is frequently used in sign language linguistics despite its etymological roots in Ancient Greek φωνή (phōnḗ), "voice" or "sound", and λόγος (lógos), "word" or "speech". This usage is intended to highlight parallels between the studies of signed and spoken constituents and to give them parity, reinforcing that sign languages are true human languages and not simply gesture or pantomime. Historically, alternative sign-centric terms have been coined from time to time, but they are barely used in the current literature (if at all).

A common perception of sign languages is that they are "languages of the hands", but this is not the whole story (Sutton-Spence & Woll, 1999). It is more accurate to say that the hands and arms are important articulators among many. Historically, when researchers have set out to specify **manual** sign formation in terms of phonological distinction (e.g. Battison, 1978; Fenlon, Cormier, & Schembri, 2015; Stokoe, 1960), the four most frequently used parameters relate to **handshape** (the form produced by articulators in the fingers); **orientation** (rotation of the hands via articulation of the forearm and wrist); **location** (where the sign is produced in signing space relative to the torso or head, or where the hands make contact with each other or another part of the body); and **movement** (changes in location during the course of the sign's production). Some signs only specify one hand, while others are typically produced with two. Almost all signers favour one hand over the other, whether left or right: the **dominant** hand is typically more active in terms of movement and will usually be the hand that produces one-handed signs, while the **non-dominant** hand is more likely to have a fixed location in two-handed production. Manual sign formation has a number of phonological constraints on acceptable combinations of handshape, location and movement (Battison, 1978; Brentari, 1998; Fenlon, Cormier, & Brentari, 2019; Sandler, 1989, 1993; van der Kooij & Crasborn, 2016). Furthermore, every sign language has a specific phonemic inventory: for example, there are handshapes in the core lexicon of BSL that are not used in ASL, and vice versa.

However, lexical distinction, grammatical function, and other kinds of qualifying information can be conveyed by articulators other than the hands: collectively these elements can be referred to as **non-manual features**. The mouth (including the lips, cheeks, and tongue) provides important non-manual contrasts, broadly falling into two

categories. **Mouthings** are visible movements of the lips that voluntarily co-occur with the production of a sign, and are derived from the lip and tongue movements of words from spoken languages. In BSL, mouthings tend to co-occur sporadically with individual signs or short phrases: it is not possible to continuously produce English on the lips and sign in BSL at the same time as their grammars are very different (in the UK, the production of signs in English word order with continuous English mouth patterns is often referred to as SSE, Sign Supported English). The use of mouthing varies with the individual and may be frequent or rare, although for any given sign language, there may be trends in the presence or absence of mouthings per sign that correlate with grammatical role (Proctor & Cormier, 2022). Meanwhile, **mouth gestures** are distinct from mouthings, in that they are defined as being unrelated to any spoken word forms (Pfau & Quer, 2010). In some accounts, when they consistently co-occur with specific signs, they have been presented as if they were part of the sign's phonological specification. More conservatively, they might be described as conventionally associated with the sign to a greater or lesser extent (Johnston, van Roekel, & Schembri, 2016). For example, the BSL sign TRUE has been associated with a mouth gesture resembling "um"; the sign REALLY can be accompanied by a mouth gesture resembling "fan" or "fant"; the sign NOT-HAVE is quite often marked by a "boo" mouth gesture with a plosive exhalation[1].

Non-manual features, including mouth gestures and other forms of facial articulation, can also be conventional for "adverbial" meanings (Sutton-Spence & Woll, 1999). Specific articulations can convey information about manner and degree (e.g. size, extent, speed, continuity, repetition, cessation, emotional affect), as can modifications to manual production. For example, in the citation form of the BSL sign WORK, the dominant hand acts with a repeated "chopping" movement on the radial edge of the non-dominant hand: in combination with conventionalised facial gestures for extent or intensity or attitude, this movement can be faster or slower, and the contact between the hands can be exaggerated or abbreviated, to indicate varying levels of effort or continuity as well as feelings about the work. Non-manuals have also sometimes been said to distinguish homophones, such as for the manual form of the sign BOSS, which can also mean "god". These two meanings are essentially identical in terms of their manual articulation, but the

[1] Note that the BSL SignBank citation forms hyperlinked here do not display the mouth gestures referred to, due to a principled decision by the authors to omit all non-manual features, analogous to the typical representation of audio pronunciations in online spoken language dictionaries with neutral modulation.

latter might be marked with an upward eye gaze. However, this latter kind of marking is arguably much less conventional and there are not very many examples in BSL.

Lastly, there are non-manual features that have been associated with grammatical and discourse functions in BSL and other sign languages (Brennan, 1992; Pfau & Quer, 2010; Sutton-Spence & Woll, 1999), although for many sign languages (including BSL) there has been relatively little usage-based research in this area. The textbook example is negation: this can be marked in BSL by a shaking of the head throughout the negated phrase constituents, but this has been shown to be non-obligatory in Auslan (closely related to BSL), and lexicalised manual negators (such as variants of NO, NOT, and NEVER) can also be used with or without a head shake (Johnston, 2018). Brow movements (raising or furrowing movements of the eyebrows and forehead) have been linked to specific structures such as conditionals, question forms, and topic marking in a number of sign languages (Zeshan, 2004), but the presence of specific brow articulations is not necessarily predictive of specific grammatical phenomena in BSL (e.g. Hodge, Manrique, Winter, & Cormier, 2025). Head nods have been associated with the marking of phrase boundaries, among several other things (Sutton-Spence & Woll, 1999). Body turns and leans have been associated with emphasis, contrast, and constituent focus, although explanations of their function have differed (van der Kooij, Crasborn, & Emmerik, 2006; Wilbur & Patschke, 1998). This is a complex area and there has not been a total consensus in the field, particularly with regard to questions around whether or not non-manual contributions to grammar are required or even dominant for any particular sign language. Specific examples and issues relevant to the current project are covered in more detail in the next chapter.

1.4.3 BSL fingerspelling and borrowings from English

Fingerspelling, or the **manual alphabet**, is not part of the core BSL lexicon: it originates in an invented system that represents English orthography via sign-like hand arrangements corresponding to the 26 letters of the alphabet (Brennan, 2001). This can be used to spell out names and words for which there is no equivalent sign. However, there are some signs in BSL which are derived from fingerspelling sequences or single letters, which can be thought of as loans or borrowings from English that have become to some extent lexicalised (in the sense that they have taken on a conventionalised meaning within the community that is not necessarily predictable from the original manual alphabet form). Manual alphabet loans may also undergo “nativisation”, in which the production of the letter sequence is reduced or blended in order to conform more closely to the core

lexicon's phonological constraints, especially if the sign was originally formed from a sequence of two or more letters. The present day BSL manual alphabet is relatively unusual in world terms, in that it is formed via two-handed production (with one exception, the letter C, which is one-handed); in contrast, the majority of the world's sign languages have entirely one-handed manual alphabets. This difference in BSL has some phonological consequences for the nativisation process and the typical forms of manual alphabet loan signs (Brown & Cormier, 2017; Cormier, Schembri, & Tyrone, 2008).

The frequency and character of use of fingerspelling by the deaf BSL community is diverse: it can depend on situation, topic, and register, and it has also been shown to vary with geographical region and with age, most likely due in part to changes in deaf education ideology and practice over the decades (Brown & Cormier, 2017; Sutton-Spence, 1994; Sutton-Spence, Woll, & Allsop, 1990).

2 Review: conjunction and disjunction in BSL and other sign languages

2.1 Introduction

In the general introduction, we defined conjunction, disjunction, and related terms, illustrated by examples in English (§1.2). The different articulations and uses of manual and non-manual forms in BSL were also briefly described (§1.4.2). We now turn to the question of how conjunctions and disjunctions are expressed in BSL.

It has been suggested that the written form of a language can provide a stimulus for the development of morphosyntactic devices for clause combination, as opposed to the asyndetic or intonational coordination that is more frequently found in spoken forms (Mithun, 1988). The argument proposes that the written modality typically lacks prosody (intonational modulations of pitch, volume, and rhythm) and the shared epistemic context of bilateral discourse: these deficiencies may drive the development of relatively complex and fixed grammatical structures for coordination in the written form, which can then in turn influence conventions in the original spoken form. Haspelmath (2007) further suggests that languages with no written form (of which there are many) “often lack indigenous coordinators and now use coordinators borrowed from prestige languages”. Sign languages that have evolved naturally within deaf communities are thought to be typically younger than the spoken languages of the hearing majority, and no sign language is known to have a commonly accepted written form (Hopkins, 2008). The “prestige” languages in these cases are the spoken/written languages used by the hearing community; for BSL, that would primarily be English. In terms of the prosodic contributions to coordination that written forms can lack, spoken languages are produced by the voice and are sometimes said to be linear in nature, but prosody and co-speech gesture can be thought of as adding another dimension. However, sign languages naturally exploit three-dimensional space and can be produced with multiple articulators simultaneously, and also feature co-sign gesture. It is certainly not impossible that prosody or gesture might become conventionalised or grammaticalised over time for coordination (or other structures) in a signed or spoken language, even without any pressure from a written form which lacks them. That said, the extent to which we can currently generalise the development of conventional gestures and prosody into grammatical sign language structures is a matter of debate, and further research is suggested (Heine & Kuteva, 2007; van Loon, Roland, & Markus, 2014).

The logical next question for the purposes of the current project, then, is how is coordination actually expressed in BSL, and conjunction and disjunction in particular? To be more specific, are they marked with lexical signs produced only with the hands, or is coordination also associated with non-manual features, and what forms might they take? Does asyndetic coordination happen, and if so, how frequent is it in comparison to syndetic coordination? Is it possible that coordination might be conventionally signalled by non-manual features alone, or are non-manuals in coordination structures appropriately comparable to intonation in spoken languages? Has English had any influence?

We begin by exploring existing literature and documentation on coordination in sign languages (and BSL in particular wherever possible), with the aim of establishing expectations for an exploratory corpus study to follow in the next chapter. Starting with a description of lexical database items attested by the BSL Corpus project, we then move on to publications that have described or categorised the manual and non-manual forms associated with signed conjunction and disjunction, and discuss some of issues raised.

2.2 Corpus-attested conjunction and disjunction in BSL

BSL SignBank (Fenlon, Cormier, et al., 2014) is an online research- and usage-led dictionary hosted by University College London. It was based initially on the Dictionary of British Sign Language (Brien, 1992), the preceding paper-based reference source for BSL vocabulary, but has been updated and expanded with attested signs from the BSL Corpus (Schembri, Fenlon, Rentelis, Reynolds, & Cormier, 2013). In addition to serving as a community resource, SignBank is the lexical database reference source for the “ID gloss” annotation tiers in the Corpus (unique identifiers for the citation forms of BSL lemmas), as well as providing phonological descriptors and other metadata. Several entries in SignBank have keywords which are associated with conjunction or disjunction, briefly described below.

2.2.1 Conjunction

BSL SignBank lists AND (Figure 1) as a corpus-attested sign with keywords “and” and “then”; however, it does not appear in Brien (1992). In comparison to spoken English corpora, in which (depending on genre and register) variants of “and” can account for something like 2.5% of tokens (Leech, Raysen, & Wilson, 2001), AND is likely to be an order of magnitude less frequent, representing only 0.2% of tokens in a sample of 25,000 from conversation and narrative data in the BSL Corpus (Fenlon, Schembri, Rentelis,

Vinson, & Cormier, 2014). It has been suggested to have an origin in Signed English[2], and a manual for schools from that period does illustrate the form (Working Party on Signed English, 1984), but since Signed English borrowed extensively from the BSL lexicon, this does not in itself constitute evidence that AND originates there. There is, for example, inconclusive evidence of a somewhat similar form with the same meaning which pre-dates Signed English by more than a century (Hutton 1855); but the origins are essentially unknown and it is not currently possible to definitively establish whether or not AND is a product of contact with English. Of note in passing is AND04 (Figure 8), which is fully distinct from AND phonologically and most likely a loan from ISL or ASL, vanishingly rare in BSL Corpus conversational data (attested only twice, both instances in the Belfast region).

Figure 1: BSL SignBank citation form of AND

Figure 2: BSL SignBank citation form of PLUS

PLUS (Figure 2) is a visually motivated representation of the mathematical plus symbol (+), with a similar additive connotation to “plus” in English, but also the mathematical connotation. It is infrequent: an ad hoc query of dominant hand ID gloss tokens from conversation data in the BSL Corpus suggests that it accounts for approximately 0.08% of tokens.

NEXT (Figure 3) can have the temporally sequential conjunctive adverb sense of “after” or “next” in a series, or the sense of “(and) then”. It may also be used by some individuals to coordinate items with the non-temporal sense of “and”. The token is of moderate frequency, appearing 55th on a list of the 100 most frequent signs in conversational data, constituting 0.3% of tokens (Fenlon, Schembri, et al., 2014), but the proportion that represents the non-temporal additive conjunction usage is not known.

Finally, SAME (Figure 4) is primarily associated with the adjectival sense of “same (as)” or adverbial “also”, but it is additionally described by Brien (1992) and BSL SignBank

[2] Signed English was a pedagogical system intended to support deaf children’s English acquisition, invented in the late 1970s but not used in UK schools today. It borrowed from BSL vocabulary, but had something resembling English syntax, employing fabricated signs to represent English parts of speech and grammatical morphemes that are not found in BSL. Not to be confused with Sign Supported English (see §1.4.2).

as having the sense of the additive conjunction phrase “as well as”. The token is very common, the 9th most frequent in conversation data, representing 1.0% of the sample (Fenlon, Schembri, et al., 2014), but again, it is not currently known how frequently the sign is used in the “as well as” additive context.

Figure 3: BSL SignBank citation form of NEXT

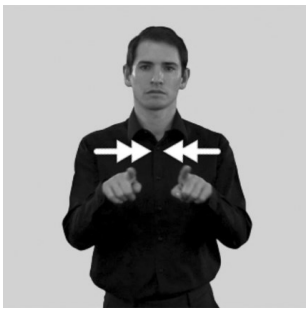


Figure 4: BSL SignBank citation form of SAME

2.2.2 Disjunction

OR is a loan sign from English via fingerspelling. The citation form (Figure 5) is essentially the fingerspelled sequence -O-R-, but like other fingerspelling loans that have become at least partly lexicalised, it may undergo reduction or blending in natural usage, conforming more closely with phonological constraints on lexical form (Brown & Cormier, 2017). It is not listed in Brien (1992), perhaps due to its origin. An ad hoc query of conversation data in the BSL Corpus suggests that it is rare, accounting for 0.05% of dominant hand tokens.

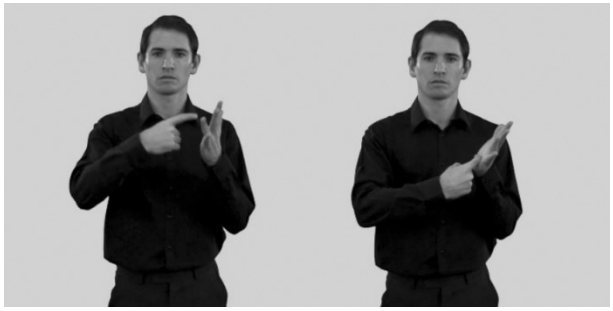


Figure 5: BSL SignBank citation form of OR

The other candidate for a disjunctive coordinator is G:WELL (Figure 6), the basic form of which is widely discussed in literature as part of a family of “palm-up open-hand” discourse markers and gestures (hence the G: for “gesture” prefix on the SignBank ID gloss), and is often referred to as *palm-up*. The form has been documented in the context of many spoken and signed languages (Conlin, Neidle, & Hagstrom, 2003; Cooperrider, Abner, & Goldin-Meadow, 2018; Engberg-Pedersen, 2002; Kendon, 2004; McKee & Wallingford, 2011; Müller, 2004; Neidle & Lee, 2008; van Loon et al., 2014; Zeshan, 2004): we will examine some of these accounts in more detail in discussions to follow.

Determining *palm-up*'s status as lexicalised or conventionalised in BSL presents several challenges, as it exhibits moderate phonological variation and has also been associated with a fairly wide variety of functions and meanings, each of which may vary in terms of lexicalisation extent. We will return to this issue shortly (§2.3), and it will also be directly relevant in the latter half of this thesis, when we discuss specific theoretical accounts of scalar implicature which require contrasting terms to be equally lexicalised (§4.2.4, §5.4.2). An additional difficulty is that the form itself is *very* common, the second most frequent of all tokens in conversational data within the BSL Corpus, making up approximately 5.5% of ID glosses in a sample (Fenlon, Schembri, et al., 2014). Five of its six English keywords in BSL SignBank are connectives, including the disjunction "or", three adversative coordinators ("however", "nevertheless", "but") and a subordinator ("although"). The remaining keyword is "well", in reference to its gestural usage in marking epistemic uncertainty, indefiniteness, or hesitancy (analogous to English "Well, …"). A similar form is listed in Brien (1992, fig. 1382), but interestingly, while the adversative functions are attributed there, the disjunctive "or" meaning is not. In several signed and spoken languages, the form has also been associated with discourse regulation and turn-taking, in the sense of yielding a turn or inviting a contribution. The citation form is two-handed but it can be produced with one hand (on the dominant hand, or on the non-dominant hand as simultaneous articulation with other one-handed signs on the dominant hand).

Figure 6: BSL SignBank citation form of G:WELL (*palm-up*)

Figure 7: Illustrative snapshot of a BSL list buoy form

2.2.3 List buoys

List buoy forms are moderately frequent in BSL Corpus conversational data (Fenlon, Schembri, et al., 2014), although they resemble a form of polysyndetic coordination to some extent, and this increases their apparent frequency. They are listed here separately from the previous sections on conjunction and disjunction as there have been occasional suggestions that they might be associated with both, although elsewhere they have been described only as additive (discussed in §2.3, §3.4.3). Like *palm-up*, list buoys are not

exactly “lexical items” in a very strict sense of the term, and characterising them as coordinators is not entirely straightforward: their dominant hand actions are comparable to pronominal pointing signs and pointing-based deictic reference-tracking strategies, exploiting the spatial and simultaneous features of two-handed production in sign languages. However, list buoys differ in that the dominant and non-dominant handshapes are a fairly consistent and finite set of distinct and highly specified phonological forms, so list buoys are in that sense highly conventionalised if not fully lexicalised. Furthermore, they can be coordinative, in the sense that they delimit a series of items such that they might potentially function as a single grammatical unit (§1.2): this does not in and of itself prevent list buoys from also having the deictic reference potential. The following description is adapted from Liddell (2003) for ASL, as it also describes BSL list buoys well: there is evidence to suggest that list buoys are articulated similarly across different sign languages with no familial connection, perhaps with some variation around acceptable usage of the thumb (Asada, 2019; Liddell, Vogt-Svendsen, & Bergman, 2007).

To coordinate each item, the non-dominant hand forms a number-like “buoy”, with a handshape similar to the cardinal number signs for the numerals one to five (see Figure 7 above for an example); however, actual cardinal numbers themselves are typically produced on the *dominant* hand with a different location and orientation. The individual digits of these non-dominant “buoy numerals” can then be sequentially pointed to with the dominant hand, delimiting or prefacing a series of coordinated items: in the current study we will refer to this series as a “list buoy episode”. The buoy hand can be held in place while the dominant hand is actively forming other sign sequences, or it can be temporarily dropped and re-established later if both hands are required to form one of the coordinated units (hence the “buoy” metaphor, referencing something that disappears from view and then bobs back up). Once any item has been enumerated and established within the “list”, the corresponding buoy numeral can be referred back to later in the discourse with a pointing sign, serving as a point of anaphoric reference, but this may or may not happen in natural conversation. Despite the visual similarity to numbers, the coordinating buoy items do not necessarily have the ordinal adverb sense of “Firstly, secondly, thirdly” constructions; they can, but they may also represent a non-numeric sequence, or a list with no inherent order.

A conventionalised derivation of the general list buoy form also exists in BSL: the sign OTHERS, which sweeps the pointing dominant hand across multiple digits on the non-dominant hand, conveying a sense of indefinite quantities analogous to “and so on”, “*et cetera*”, or “several”. This sign can be used to terminate a list buoy episode, or

independently from one. The sign OTHER (“other”, “another”, “otherwise”, “else”) is also quite possibly a lexicalised derivation of a list buoy form.

2.3 Manual forms and asyndesis

We now turn away from the BSL Corpus to other relevant literature. The only empirical study of syndetic and asyndetic coordination in BSL prior to the current project seems to be Waters & Sutton-Spence (2005), a study of “connectives” (conjunctions, disjunctions, and other forms of coordination and subordination) and accompanying patterns of mouthing. 102 minutes of BSL footage produced by 27 deaf adults were analysed, taken from public television broadcasts and narrative/instructive video data from University of Bristol archives. 238 instances of connectives (both syndetic and asyndetic) were identified, categorised and described in terms of form, relative frequency and genre variation. The findings indicated that asyndetic conjunction is moderately common, and the authors suggest that asyndesis in BSL does not have the stylistic or emphatic qualities typically ascribed to it in English. For syndetic coordination, there are differences in the data between frequencies of specific connective use per discourse genre; while this interaction is not analysed for statistical significance, there is an apparent descriptive difference between the highest (televised news) and lowest (personal narrative) rates of lexicalised connectives. It seems plausible that in personal narratives, other non-lexical cohesion strategies that exist in BSL (e.g. the use of deictic reference) may be more favoured than the relatively complex lexical/syntactic constructions required by “bursts” of detailed information such as televised news, which are typically produced under time pressure and translated from English source materials. Furthermore, personal narratives are more likely to be *implicitly* additive or progressive, built up as they typically are from a sequence of events, thereby requiring fewer explicit connectives.

Asyndetic coordination in BSL is described by Waters & Sutton-Spence as potentially being either additive or alternative, and examples of both were apparently found in the sample; however there is no clear discussion of how the two might be distinguished from each other in the absence of a syndetic coordinator. List buoys are also discussed in the role of coordination; it is asserted here that they are additive only. Nothing resembling AND was apparently identified in the data, but PLUS was described as a rare connective of addition and NEXT was associated with temporally ordered lists. OR was confirmed as a fingerspelling loan derived from English that likely undergoes some degree of phonological reduction, but it was noted that it was relatively infrequent. This was attributed to signers having access to a non-derived alternative, a sign whose description and illustration very

closely resemble that of the *palm-up* form indexed in BSL SignBank as G:WELL, presented in this study as a fully lexicalised form directly equivalent in meaning to English “or” without reference to any other coordinating or discourse-regulative functions (§2.2.2). In the case of discourse regulation, the omission might be explained by a lack of relevance to the study, but the *palm-up* form had also been previously associated with adversative senses in Brien (1992) and the study did not seem to capture any examples of it being used as adversative coordination (while other “connectives of contrast” were identified).

A much richer picture of *palm-up* from another branch of the BSL language family can be found in a NZSL corpus-based study (McKee & Wallingford, 2011) where it is described as both sign and gesture, and is associated with a remarkable range of functions and meanings. In a sample of 5,134 sign tokens from conversation and interview data with 120 signers, *palm-up* was identified as 261 of the tokens, making it once again the second most frequent token; the rate was just under 5.1%, close to a finding of 5.5% for BSL (Fenlon, Schembri, et al., 2014). The keywords attributed to *palm-up* in NZSL do include disjunctive and adversative coordinating functions, matching all of the equivalent BSL SignBank connective keywords; however, the study also stirs conjunctive (“as well as”) and temporal-sequential (“and then”) meanings into the pot. As in other signed and spoken languages that use a similar semi-conventionalised gesture (§2.2.2), it is further identified as signalling uncertainty or indefiniteness. In terms of grammatical role, the study associates the form with an extremely broad array of functions (connective, determiner, pronoun, preposition, adverb, verb, phrase, negator, interjection, and interrogative). Following corpus analysis, these roles are re-cast within four functional categories: cohesive (e.g. as a within-sentence coordinator), modal (i.e. related to epistemic uncertainty), interactive (e.g. discourse level cohesion and turn management), and finally as a “frame” for English mouthings. Case examples of all four categories are discussed in some depth. Of primary relevance to the current study is the finding that a full third of the NZSL *palm-up* tokens were classed as representing cohesive tokens, with examples given of disjunctive, sequential, causative and elaborative coordination. A more recent study of BSL Corpus conversational data in a similar paradigm suggests that a comparable proportion of *palm-up* tokens have a cohesive function (Arnold, 2019).

Moving away from the BSL language family, Engberg-Pedersen (2002) reviews the usage of *palm-up* (referred to as “the presentational gesture”) in Danish Sign Language (DTS), where it serves as an example of the terminological difficulties in distinguishing so-called “non-linguistic” (a proxy for “gestural”) from so-called “linguistic” (lexicalised) items in sign linguistics research. A variety of discourse-regulative and connective meanings are

ascribed to *palm-up* in DTS (e.g. as a marker linking sentences and clauses together), all based around a central theme of “presenting something for inspection”, but it is not explicitly discussed in terms of any potential as a lexicalised coordinator.

In a general literature review, Pfau (2016) suggests that sign languages typically employ few lexicalised coordinators. A sign meaning “but” is said to be common in NGT, but another identified as meaning “and” is infrequent and is considered an influence from spoken Dutch. NGT and other sign languages also use a sign similar to BSL’s PLUS (Figure 2) but this is again attributed to language contact. It is claimed here that conjunctive coordination is instead more often marked by a non-manual body lean (§2.4). Conjunction in sign languages more generally is said to be most often realised through what is referred to here as “ellipsis”, i.e. asyndetic coordination, but some sign languages may be exceptional. For example, in DGS, a lexical sign described as meaning “both” can reinforce conjunction; alternatively a coordinated verb can be realised twice.

Tang and Lau (2012) discuss the syntax of coordination and subordination in Hong Kong Sign Language (HKSL) in some depth. It is asserted that lexicalised signs with the same meanings as “and”, “or” and “but” are all to be found in HKSL, but that their use usually marks a switch to Chinese word order; “juxtaposition” (asyndetic coordination or parataxis) is asserted to be common in sign languages, with examples given from ASL.

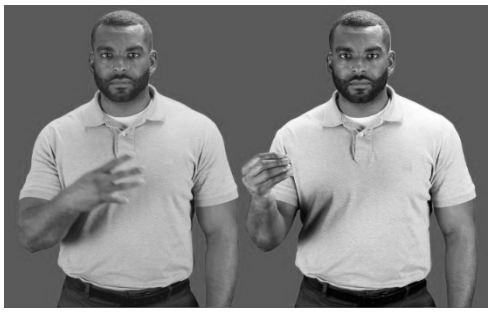


Figure 8: Stills from the ASL SignBank citation form of AND
(Hochgesang, Crasborn, & Lillo-Martin, 2020)

ASL is described by Davidson (2011) as having a small set of lexicalised disjunctions available, including a fingerspelled loan from English “or” (not to be confused with the BSL fingerspelled loan OR, as the ASL and BSL manual alphabets, and the respective mechanisms of incorporating fingerspelling as loans, are very different). The study’s informants (a small number of deaf ASL users and linguists) suggest that this loan is also available for use within list buoy episodes to make a disjunctive “list of alternatives” reading more explicit. A lexicalised ASL disjunction, OR-WHICH, is described as being homophonous with a sign meaning “which”; some of the informants reported a sense that

this sign is awkward in alternative questions (where it may conflict with the reading of "which?"), while others reported a feeling that it was "out of date" and less likely to be used by younger signers. For conjunction in ASL, there is the previously-mentioned lexical item AND (Figure 8). It is suggested that AND is much less common in ASL than "and" is in English; it is further described as being used for "emphasis" and more likely to feature in direct translations of English. As a side-note, Zimmer (1989) also found that the frequency of ASL's lexical AND is higher in some registers (e.g. university lectures), possibly reflecting the increased situational language contact with English. Finally, Davidson also suggests that ASL has two different forms of "general use coordinator" available, which can function either conjunctively or disjunctively depending on non-manual context. One of these is list buoys (glossed by Davidson as "COORD-L"); the other is a completely non-manual "body gestural" form of spatial contrast ("COORD-SHIFT"). Two of the three experimental sections of the thesis manipulate COORD-SHIFT in experimental studies of scalar implicature in ASL, the results of which were later revised and published separately (Davidson, 2013; Davidson & Mayberry, 2015): aspects of this work will be examined in some detail in a later chapter (§4.2.4), especially with regard to the lexicalisation extent of COORD-SHIFT.

None of the ASL studies reviewed made explicit reference to the *palm-up* form as a candidate for a lexical coordinator, which suggests that it is less closely (or not at all) associated with disjunction in ASL, in contrast to BSL and some other sign languages. Something phonologically similar has been referred to as a grammaticalised "indefinite focus particle" in ASL (Conlin et al., 2003; Neidle & Lee, 2008), usually sentence-final, indicating uncertainty, hypothesis or "domain-widening" (i.e. indefiniteness), which is consistent with the gestural discourse functions it has been associated with in many signed and spoken languages (see §2.2.2). In another account (Hoza, 2011), the form is also attributed with various discourse-regulating roles in ASL such as turn regulation, hedging, and pause-filling (this time as a lexicalised sign WELL); it is also described as a face-saving marker of politeness.

Further regarding indefiniteness and uncertainty, the grammaticalisation potential of the *palm-up* gesture in sign languages is discussed as a case study in van Loon et al. (2014), citing some of the literature mentioned above on DTS, ASL, and NZSL, with additional examples from NGT. The discourse regulation and epistemic uncertainty functions of *palm-up* are described in detail, but the description of its within-sentence connective potential is briefer and only uses examples from the same NZSL study discussed earlier (McKee & Wallingford, 2011). A theoretical "grammaticalisation pathway" is proposed, with the discourse-connective and epistemic roles emerging from the

discourse regulation role: use of the gesture as a grammaticalised connective within single-speaker utterances develops from use as a presentational marker of turn-yielding that "connects" different speakers' utterances. The authors emphasise the need for further empirical research to test this theory against.

Finally, in our initial introduction, we also briefly looked at examples of structures for which it is impossible to distinguish the nature of the relationship between coordinated units through syntax alone, illustrated by cases in English where paratactic subordination can look very like asyndetic or syndetic conjunction, with semantics, pragmatics, and discourse-contextual considerations being required to identify the intended function and meaning (§1.2). There has been some documentation and discussion of parataxis in sign languages, all taking a functionalist approach (see Halliday & Matthiessen, 2004), which tends to make less of a distinction between coordination and subordination in any case. Most recently, Carneiro, El Khouri, Santos, & Ludwig (2023) investigated a number of data sources in Libras (Brazilian Sign Language), including spontaneous conversation data taken from social media and a corpus-in-progress. Examples of juxtaposition (asyndesis), manual coordinators/subordinators, buoy forms, and non-manual contributions (described in the next section) were found; the approach is descriptive (rather than quantitative or statistically inferential). For BSL, Rudge (2022) briefly mentions parataxis in an overview of a systemic functional approach to analysing sign language structure. That school of analysis is beyond the scope of this thesis, but the suggestion in this case is only that items such as BUT and *palm-up* (identified as an apparently lexical sign glossed as OR) can have paratactic and hypotactic functions in BSL. Hodge (2014) also mentions parataxis during a study of clause structure in the Auslan Corpus, mostly in terms of contrast with hypotaxis, affirming the functionalist view that clause relationships are better defined and identified by their discourse function than the labelling of any particular linking item as a "coordinator" or "subordinator".

2.4 Non-manual contributions

No relevant data or discussion could be located in existing literature on BSL to any interaction between non-manual features (facial expression, body lean/positioning, brow movement, head movements, mouth gestures, direction of eye gaze, or eye contact between interlocutors) and the syntax of within-sentence coordination, with the occasional exception of mouthings. However, there has been some discussion on the topic for other sign languages, and occasionally empirical data.

Pfau (2016) remarks that conjunctions are sometimes accompanied by raised brows or head tilts, and it is further suggested that in some sign languages these elements may be obligatory, but no specific languages are identified. Tang and Lau (2012), however, cite the view expressed by Padden (1988) that lexical coordinators in ASL such as BUT may only be considered "true" coordinators if they are accompanied by a pause or a head shake. They also describe non-manual contributions to asyndetic coordination in HKSL: where a lexical coordinator is absent, non-manuals are said to provide the distinction between conjunctive (clause-marking head nods only), disjunctive (body turns) and adversative coordination (head turns and body leans). However, the evidence base for these claims is not clearly identified. The authors do note that identifying coordination in sign languages is not straightforward, due to the difficulty of identifying clause boundaries in the first place: morphosyntactic devices that mark such boundaries may exist for any given sign language, but they are often not obligatory.

In the NZSL study of *palm-up* mentioned previously (McKee & Wallingford, 2011), non-manuals accompanying the form were also analysed. Older signers were more likely than younger to use the gesture with affective non-manuals (non-syntactic emotional expressions), and were also more likely to accompany *palm-up* with an English mouthing. 39% of those *palm-up* tokens identified as having a cohesive or connective function were found to have a range of accompanying head or body movements: this is partly explained by the authors in terms of the role of non-manuals in marking clause and discourse boundaries. However, while specific body, head and facial non-manuals are listed against broad functional categories, no analysis of specific non-manuals associated with specific types of coordinating function (e.g. conjunctive, disjunctive, adversative) is described.

We also previously noted (§2.3) an investigation of clause linking in Libras conversational data (Carneiro et al., 2023), which took a functionalist approach. For non-manual contributions, various forms of articulation were identified in the context of both paratactic and hypotactic constructions, including eyebrow raises, chin lifting, head nodding, mouthings, and pauses. The discussion does not directly describe the extent to which these contributions might be considered conventional, but does characterise some of them as prosodic. Correlations between specific non-manuals and specific functions are not quantitatively analysed, but the authors note that conditionally linked clauses seem to be associated with raised brows, and purposeful subordinations with slight chin elevations and narrowed eyes. There do not appear to be any associations noted for symmetric coordinations such as conjunction and disjunction. Regarding body turns and leans, the authors point out that the range of movement of the Libras corpus contributors is limited by

the fact that they are sitting down; it is worth noting for the following chapter that this is also the case for BSL Corpus conversational data.

Continuing on the topic of prosody, the use of prosodic contributions to mark constituent focus or contrastive stress was previously alluded to in the introductory chapter (§1.3). We saw an example of how a disjunction can be contrastively stressed to reinforce an exclusivity inference in English ("You can have a biscuit *or* a cake" might be more likely understood as implying "but not both" than the unstressed version). "Constituent focus", on the other hand, is more often described as differences in meaning from emphasising sentence elements such as subjects, objects, and verbs. "*Barbara* ate the cake" reinforces that it was Barbara who ate the cake and not someone else; "Barbara ate the *cake*" reinforces that it was the cake that Barbara ate and not something else; "Barbara *ate* the cake" reinforces that the cake was eaten rather than meeting with some other destiny at Barbara's hands. The details of prosodic stress marking can vary between languages; for many spoken languages it is typically expected to be intonational contours of pitch, volume, and rhythm. While the phenomenon in sign languages is not yet as thoroughly researched, there have been some useful investigations. For example, Wilbur (2012) presents a cross-linguistic review of literature on sign language information structures, largely with regard to ASL but with additional examples from HZJ, NGT and Libras. Specific co-occurring non-manuals are described as playing a role in stressing lexical items within focused constituents and are further analysed in terms of a descriptive categorical system of focus proposed by Dik (1989). For contrastively marking conjunctive, disjunctive and adversative coordination in ASL (under what is termed by Dik as "parallel focus"), left/right body leans are said to form a non-manual realisation of stress that can accompany list buoy coordination structures; furthermore, the same form of stress can mark asyndetic coordination for a pair of items (Wilbur & Patschke, 1998). The given description of these body leans is highly reminiscent of the COORD-SHIFT "general use coordinator" proposed by Davidson (2011), previously described in §2.3; we will return to this comparison in §4.2.4.

However, other studies have been more reluctant to associate specific non-manual features with specific functional categories of focal stress, and they have tended to take a more usage-driven approach than some of the documentation mentioned previously. In one study of NGT data, van der Kooij, Crasborn & Emmerik (2006) found that while some individual uses of forward and backward body leans in NGT are comparable to fairly tightly defined accounts of their functional associations in ASL, others are better explained in terms of pragmatics and the specific conversational interaction. In a later NGT study,

Crasborn & van der Kooij (2013) found a surprising range of non-manual features co-occurring with various types of focus and stress (eye contact, raised eyebrows, raised head, head nods, left/right body leans), and suggest that non-manuals can vary for any particular focus type. On similar lines, Schlenker et al. (2016), in a discussion of inferences taken from variously stressed disjunction forms in ASL and LSF, also suggests that there is at least some variation and cross-over in the prosodic means used to mark different categories of focus. They further concur with the conclusions of Crasborn & van der Kooij (2013), which argues that increased amplitude and acceleration in the movement of signs (as opposed to any specific non-manual articulation) are a key part of marking prosodic contrast, and that the extent of markedness is more salient than the precise form. An investigation of related phenomena in RSL (Kimmelman, 2014) takes a similar view. In later related work, Kimmelman (2019) also explicitly argues that this kind of non-manual contribution is optional, at least in NGT and RSL.

On that last point, it may be worth summarising the backdrop of some historical differences in viewpoint within the sign linguistics field on the general role and status of non-manuals in syntax, as summarised by Pfau and Quer (2010) in the context of a proposed scheme for categorising and sub-categorising non-manual features. Some researchers have contended that “grammatical” non-manuals are morphosyntactic, as there is often no clear division that can be drawn between morphemic and syntactic marking. Non-manual marking of this kind has been asserted to take dedicated, well-defined and obligatory positions in signed sentence structures, and the spread of specific non-manuals over specific phrasal domains has been positioned as providing evidence for hierarchical constituents in a generative framework, particularly in the earlier ASL literature (e.g. Neidle et al. 2000; Wilbur & Patschke 1998). However, there are often cross-overs in the ways that grammatical markers and prosodic domain markers are articulated by signers, such that there are cases where the two different functions are difficult or impossible to distinguish. Alternative arguments have been made, which hold that so-called “grammatical” non-manuals are very unlikely to be obligatory (citing observed examples of variation or absence) and that they are essentially prosodic in nature. In this latter view, it is not disputed that non-manuals can have a complex relationship with syntax, but the two are said to operate independently (e.g. Johnston & Schembri 2007; Sandler 2010; Sandler & Lillo-Martin 2006).

2.5 Conclusions

The consensus from literature appears to be that asyndetic coordination is a common but by no means exclusive strategy for conjunction in sign languages. While lexicalised conjunctions and disjunctions are sometimes borrowings from contact with other languages, non-loan lexical coordinators also exist, potentially (but not necessarily) through a process of grammaticalisation of gestural forms (for example, *palm-up*, and perhaps list buoy constructions). BSL has been documented as featuring examples of all of these strategies: asyndesis, the lexical conjunctions AND, NEXT, and PLUS, the English-loaned lexical disjunction OR, the polysyndetic list buoy construction that seems to share some provenance with deictic reference forms, and the semi-lexicalised disjunctive reading of *palm-up* (although for this latter, the extent to which the form is conventionalised for within-sentence disjunctive coordination in BSL could still use some clarification).

However, there has not been a historical consensus in the field on whether non-manual features that co-occur with coordination are obligatory or not, and whether they are better understood as morphosyntactic or prosodic. It is possible that this might vary between sign languages. Very little has been said in print on the issue for BSL specifically, and existing documentation of BSL coordination is relatively sparse in any case.

Therefore, following this literature search and informed by it, an empirical investigation into conjunctive and disjunctive utterances in spontaneous conversation data in the BSL Corpus was undertaken, partly with the aim of clarifying the picture above but also in the interests of designing a further study of exclusivity inferences from disjunction in BSL. This initial exploratory study is presented in the next chapter.

3 Usage-based study: conjunction and disjunction in the BSL Corpus

3.1 Introduction

In the previous chapter, we examined documentation and discussion of coordinators with conjunctive and disjunctive functions in BSL and other sign languages, including some empirical data (but relatively little from spontaneous conversation data), as well as a number of more general or theoretical positions. What is still lacking is machine-readable data on BSL conjunctions and disjunctions in conversational usage, and data about non-manual contributions to coordination in particular. Given that the final goal of this project is an experimental study of contextual and cognitive influences on the strength of exclusivity inferences from disjunction in BSL, and given that inferential communication is co-constructed in bilateral discourse via multiple linguistic and epistemic contexts, a usage-led snapshot of BSL conjunctions and disjunctions from a peer-to-peer conversational setting has the potential to inform that experimental design.

The BSL Corpus (Schembri, Fenlon, Rentelis, & Cormier, 2014) is a resource uniquely positioned to provide such a snapshot. At the time of the current study, over 70,000 sign tokens had been identified and annotated either with “ID glosses”, unique identifiers that correspond to a lemmatised database entry in the BSL SignBank lexicon (Fenlon, Cormier, et al., 2014) or to a function/form identifier (Cormier et al., 2017). Many additional annotation tiers relating to grammatical function, non-manual features, and translations (among other things) have been added to differing subsets of task data during the lifespans of several supplementary projects. The video data of BSL usage was collected from pairs of participants for four task areas, one of which was designed to elicit spontaneous conversation data. In this chapter, I present an exploratory quantitative study of a sample of that data, focusing on sentence-level conjunctive and disjunctive coordination, with the following research questions in mind:

- How frequent is asyndetic coordination in BSL?
- For syndetic coordination, which tokens take on the role of conjunctive and disjunctive coordinators, and of what type, and with what frequencies?
- What kind of non-manual features are statistically associated with conjunction and disjunction, if any, and in what contexts?

3.2 Method

3.2.1 Participants and conversational data

Participants in the original BSL Corpus project were all deaf users of BSL as their main or preferred language, recruited and filmed between 2008 and 2010 with the aim of creating a broadly representative sample in terms of age, gender, ethnicity, geographical region and family language background (i.e. deaf or hearing parents). Members of each conversational pair were acquainted but were not supposed to have a close or familial relationship. For the conversational data task, each pair was given up to 30 minutes to talk about whatever they wished (although most conversations were shorter).

For the current study, a sample of 40 participants across four regions were selected from conversational data where "clause-like unit" (CLU) boundary annotations were available (5,956 in total), as this assisted to some extent with differentiating between discourse-level and sentence-level coordination. The participant sample was largely balanced for region, gender, age group and language background as shown in Table 2.

The conversational task data took the form of digital video files in HD 720p resolution, with annotation data stored in an XML format produced by ELAN, a media annotation tool (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). The first 500 tokens per participant (~20,000 tokens in total) all had ID glosses, CLU boundaries, and translations completed as part of previous BSL Corpus projects.

	Gender		**Family**		**Age group**			
Region	**Male**	**Female**	**Deaf**	**Hearing**	**16-40**	**41-65**	**65+**	**Subtotal**
Birmingham	5	5	5	5	5	3	2	**10**
Bristol	5	5	4	6	3	4	3	**10**
London	5	5	4	6	3	5	2	**10**
Manchester	5	5	3	7	4	2	4	**10**
Subtotal	**20**	**20**	**16**	**24**	**15**	**14**	**11**	**40**

Table 2: Distribution of participants across social categories

3.2.2 Coding and procedure

Mouthing, facial expression and grammatical function annotations were already available for some participants, but not across the entire sample. A new annotation scheme was therefore designed to capture the linguistic parameters listed in Table 3. These values

were packed into single annotations on a single ELAN tier in the form of a delimited string, suitable for unpacking into separate columns in the data cleaning and analysis stage.

As this study was being conducted in the context of a broader project concerning pragmatic reasoning, utterances with syntactic constructions known to modify or cancel the derivation of scalar implicatures were excluded: coordination embedded within questions and negation, and those modified by existential modal qualifiers (e.g. CAN, SHOULD, MUST). Discourse-regulatory ("between-sentence") coordinations were also not included, using CLU boundaries as the primary basis for making this distinction. To support the aggregation of findings, for pointing and other deictic signs that rely on context and reference, only the main functional identifier of the token was duplicated: for example PT:PRO3SG (a form analogous to a third person singular pronoun) was recorded in the scheme as PT:PRO, while instances of the various non-dominant handshapes for list buoys (LBUOY-TWO, LBUOY-THREE etc.) were all recorded as LBUOY.

Parameter	Allowed values	Description
coordination	conjunction, disjunction	The type of coordination.
token type	lexical, gestural, list, asyndetic	Whether the coordinator produced on the hands was lexicalised (with a lemma listed in SignBank); or a gestural/pointing annotation with a G: or PT: prefix; or a list buoy episode; or absent (asyndetic).
ID gloss	<any ID gloss>, none	The ID gloss of the identified coordinator, where applicable.
mouthing	<any word>, none	The perceived English mouth pattern observed to co-occur with the coordinator, if any.
brow movement	neutral, up, down	Whether observed brow movement co-occurring with the coordinates moved upwards or downwards, or remained relatively neutral.
body lean	neutral, forward, back, side	Whether the upper body moved forwards, backwards or side-to-side during the coordinated phrase, or remained relatively neutral.
context	<any ID gloss>, <notes>, none	Note-like annotations of any other linguistic features within scope, which were judged to potentially affect a conjunctive or disjunctive interpretation.

Table 3: Annotation scheme

The annotation process was undertaken in a series of passes. For familiarisation, each segment of video in the sample was first watched from start to finish without annotating, consulting English translations only in cases where regional or idiolectic variation made initial comprehension difficult. The sample was then worked through again, creating annotations for all structures of potential interest, i.e. utterances which appeared at first

glance to have at least some sense of conjunction or disjunction without any disqualifying discourse or grammatical context (see above). The token type, coordination, and ID gloss parameters were completed in this first pass. Brow movement, body lean, mouthing and context parameters were completed in a second pass. All annotations were then checked for consistency and accuracy. While a formal inter-annotator agreement process was not carried out, deaf signing colleagues from the BSL Corpus team were consulted wherever there was uncertainty.

At the end of this process, 162 within-sentence coordinated utterances were available for statistical analysis.

3.2.3 Analysis

On completion of the annotation and auditing process, the annotations were exported and unpacked into a table of factors. Descriptive statistics for frequency and distribution were produced, and inspected for any immediately obvious patterns.

Multiple Correspondence Analysis (MCA) and Hierarchical Clustering on Principal Components (HCPC) were chosen as the quantitative statistical analysis methods. MCA is well-suited to exploring the associations between categorical factors in a dataset and produces measures of the significance and variance of those associations across the set, forming a predictive model. This model can then be re-structured by HCPC, producing a set of "clusters" of similar data points with details on the strength and variance of factors within and between each cluster. Analysis was performed using R Statistical Software 3.6.1 (R Core Team, 2024) and the packages *FactoMineR* 2.11 (Husson, Josse, & Lê, 2008) and *factoextra* 1.0.7 (Kassambara & Mundt, 2020). Data visualisation was created with *ggplot2* 3.5.1 (Wickham, 2016) and *ComplexHeatmap* 2.21.1 (Gu, 2022).

3.3 Results

3.3.1 Descriptive statistics

3.3.1.1 Token types per coordination type

Table 4 shows that conjunction was found to be more frequent than disjunction within the conversation data. Asyndesis was the most common form of conjunction. Proportionally speaking (Figure 9), asyndesis was found to be almost equally common for conjunction and disjunction: this was unexpected and is explored in the discussion section. In terms of syndetic coordination, nearly a third of all conjunctions were list buoys, with relatively few

using a lexicalised coordinator or gestural tokens. The most common type of syndetic disjunction, almost half of all disjunctions, was gestural tokens (always *palm-up*), while lexicalised disjunctions and disjunctive list buoys were much rarer.

	Asyndetic	Gestural	Lexical	List buoy	Total
Conjunction	43	9	13	31	**96**
Disjunction	28	31	4	3	**66**
Total	**71**	**40**	**17**	**34**	**162**

Table 4: Frequencies of token types per coordination type

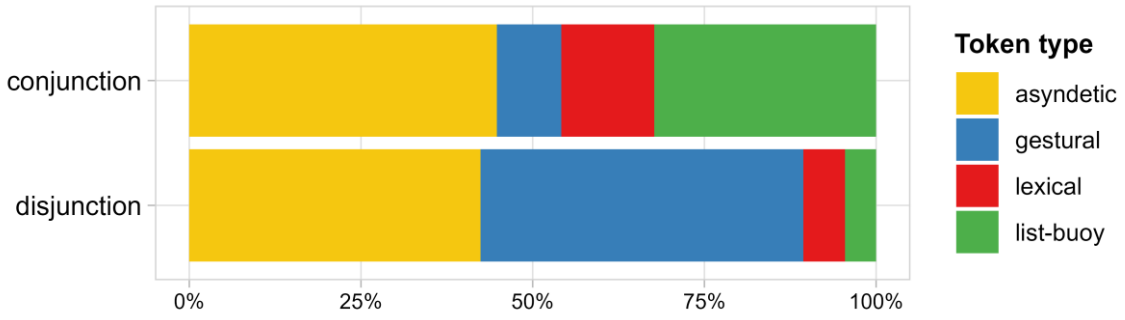


Figure 9: Proportional frequencies of token types per coordination type

3.3.1.2 Tokens per coordination type for syndetic coordination

Putting asyndetic coordination to one side, Table 5 shows the ID glosses or functional annotations associated with coordination type for the 91 identified examples of syndetic coordination. As expected from Table 4, list buoys are a common strategy, accounting for more than half of syndetic conjunctions. Lexicalised coordinators such as AND, PLUS and OR were relatively rare. NEXT was found to be even rarer with just two identified examples functioning as non-temporal conjunction. There was occasional use of pointing signs (determiners and pronominals) and a single use of a depicting sign[3] to give a sense of conjunction, using spatial contrast to create a list-like structure. SAME was not observed (§2.2.1). For syndetic disjunction, it is clear that the fingerspelling-derived loan sign OR and disjunctive list buoy episodes are rarely used in comparison with the *palm-up* G:WELL sign.

[3] Depicting constructions/handshapes identify classes of entities or handled objects. These have in the past been compared to classifiers in spoken languages, but the comparison is problematic (Cormier, Quinto-Pozos, Sevcikova, & Schembri, 2012).

ID gloss	Conjunction	Disjunction	Total
AND	4	-	**4**
depicting signs	1	-	**1**
G:WELL (*palm-up*)	-	31	**31**
list buoy(s)	31	3	**34**
NEXT	2	-	**2**
OR	-	4	**4**
PLUS	6	-	**6**
pointing signs	8	-	**8**
production error[4]	1	-	**1**
TOTAL	**53**	**38**	**91**

Table 5: ID glosses associated with syndetic coordination

3.3.1.3 Surrounding context for asyndetic coordination

The qualitative notes entered into the "context" field for the 71 asyndetic coordination episodes can be briefly summarised as follows. The overwhelming majority of asyndetic conjunctions (42 out of 43) had no identifiable context that contributed to the conjunctive reading, i.e. the phrases were not identified as containing any additional signs, non-manual markers not already coded as brow movements or body leans or mouthings, or any other overt context that contributed. Even the exception is debatable: it contained PT:PRO3PL-TWO within scope, a specific kind of third person plural pronominal sign equivalent to "(both of) the two of them", but the utterance would likely not be understood very differently if the sign were removed. However, the picture for asyndetic disjunction was very different: none of the 28 examples were deemed to have no contributing context (see §3.4.2 for discussion).

[4] This single case was a sign originally annotated as HAVE, but the production was highly atypical for that sign, and it had a clear mouthing of "and". The context was an English name, the "Game and Country Fair". It was felt that this token more likely represents a production error or possibly an idiolectic sign. Given that it had both a mouthing and a manual articulation, it was left in as an example of syndetic conjunction.

3.3.1.4 Brow movements per coordination type

Table 6 and Figure 10 describe the absolute and proportional frequencies of brow movement per coordination type. Literature has suggested (§2.4) that in some other sign languages, specific brow movements can contribute to, or are expected to co-occur with, various forms of coordination. In this BSL sample, for conjunction, the majority (over 70%) had a neutral brow, with upward and downward brow movement split equally among the remainder. For disjunction, the three brow positions are distributed somewhat more evenly; there is more non-neutral brow movement than is found for conjunction, and upward movement is proportionally more common than downward, but half of brow positions still remain neutral.

	Down	Neutral	Up	Total
Conjunction	14	68	14	**96**
Disjunction	13	32	21	**66**
Total	**27**	**100**	**35**	**162**

Table 6: Frequencies of brow movement per coordination type

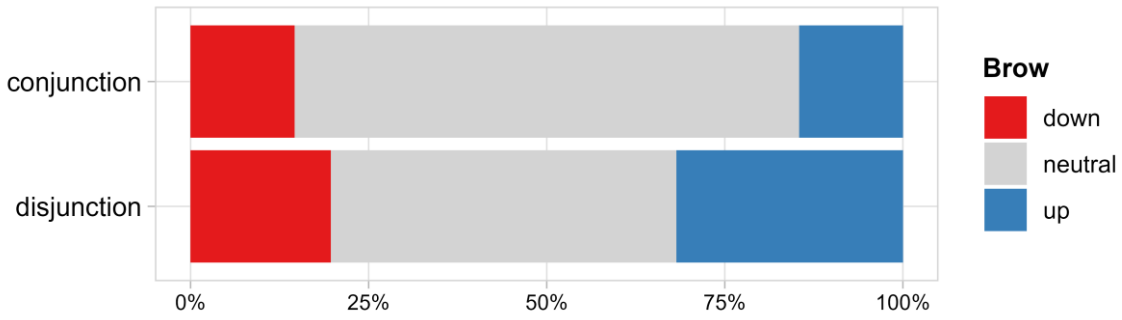


Figure 10: Proportional frequencies of brow movement per coordination type

3.3.1.5 Body lean per coordination type

Table 7 and Figure 11 show the absolute and relative frequencies of body lean per coordination type. Relatively few examples were observed, with forward and backward leans being even less likely than side-to-side. Neutral body positions dominated, although disjunctions were slightly more often marked with a body lean than conjunctions.

	Backwards	Forwards	Neutral	Sideways	Total
Conjunction	0	2	90	4	**96**
Disjunction	2	0	58	6	**66**
Total	**2**	**2**	**148**	**10**	**162**

Table 7: Frequencies of body lean per coordination type

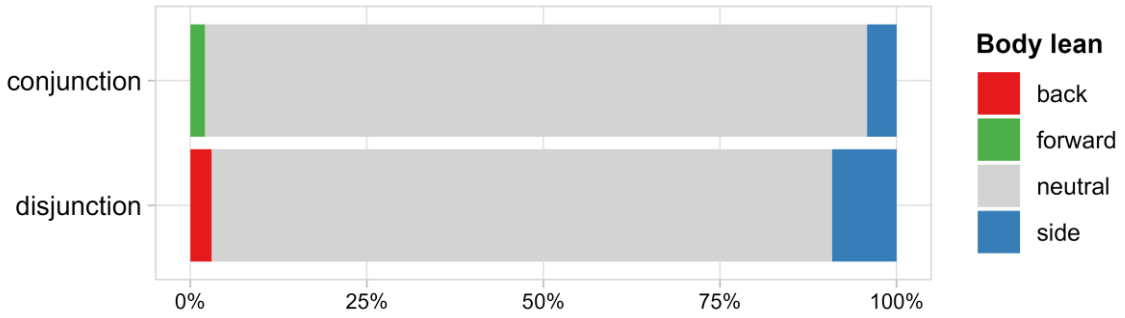


Figure 11: Proportional frequencies of body lean per coordination type

3.3.1.6 Mouthings per coordination type

Table 8 gives the frequencies of observed English mouthings that occurred simultaneously or near-simultaneously with conjunctions and disjunctions; Figure 12 shows the proportions of mouthings per coordination type. For conjunction, a large majority of the coordinations (79%) had no mouthing, and the mouthings that did occur are somewhat varied. The examples of cardinal numbers and days of the week all co-occurred with list buoy pointers. In comparison, half of the disjunctions are marked with a mouthing, almost always “or”.

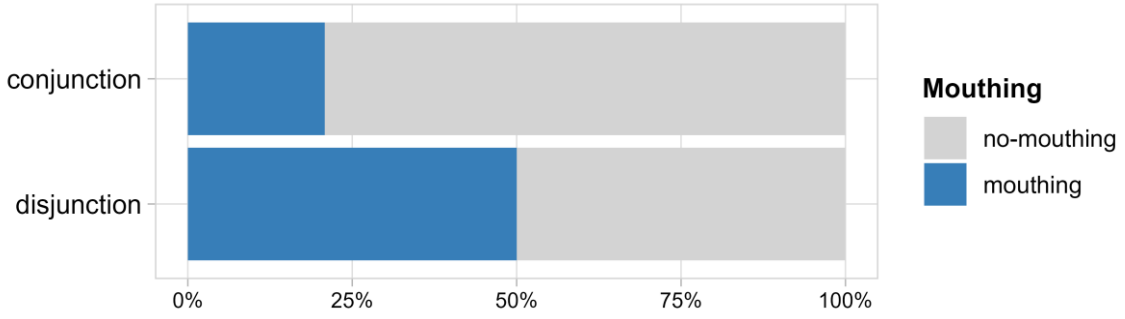


Figure 12: Proportional frequencies of mouthing per coordination type

Mouthing	Conjunction	Disjunction	Total
“and”	8	-	**8**
“both”	2	-	**2**
“next”	1	-	**1**
“or”	1	32	**33**
“or what”	-	1	**1**
“plus”	3	-	**3**
cardinal numbers	2	-	**2**
days of the week	3	-	**3**
no mouthing	76	33	**109**
TOTAL	**96**	**66**	**162**

Table 8: Mouthings per coordination type

3.3.1.7 List buoy episodes

The 34 list buoy episodes were inspected for the frequencies of constituent properties, noting the number of buoys per episode, the number of coordinates, the sequence of list buoy numerals used, and the number of tokens per coordinate based on actions of the dominant hand (for this latter, the following token types were excluded: the dominant and non-dominant list buoy tokens themselves, any variations on the G:ERM hesitancy gesture, ambiguous/vague pointing PT: tokens, unknown and indecipherable tokens, and false starts). Two ambiguous or fragmentary episodes are not included here due to difficulties with determining their overall structure. Due to the small sub-sample, the findings are not tabulated here or analysed in depth, but a few key observations are made as pointers towards any future study of list buoy constructions.

More than a quarter (9 of 32) of the episodes omitted the buoy before the first coordinate. More than half (19 of 32) had two coordinates; the mean number of coordinates per episode was 2.3. For episodes with only two coordinates, the mean of token counts per coordinate was 1.8; the mean for episodes of all lengths was 1.7. More than a quarter of episodes (9 of 32) had one or fewer mean tokens per coordinate (less than one was possible in cases of termination with buoy forms closely resembling the derived lexical sign OTHERS (§2.2.3), with a sense of “and so on”, such that the final coordinate effectively had zero non-buoy tokens and the preceding coordinates all had only one token each). Buoy numerals were always in a rising sequence, although they did

not necessarily start with 1, and in a few cases repeated a numeral, e.g. 1-1-2-3. None of the episodes “back-tracked” to a previously enumerated item (§2.2.3): this was not surprising given that these utterances were all examples of within-sentence coordination. Future study might investigate list buoy back-tracking at the discourse level.

3.3.2 Quantitative analysis

Hierarchical Clustering of Principal Components (HCPC) was chosen as the main analytical tool. This method organises data into “clusters” (groups of data points that share a degree of similarity), painting a picture of the factors that commonly combine within the data set, and providing details of the variance and strength of associations between variables within and between clusters. For purely categorical data, HCPC takes the output from a Multiple Correspondence Analysis (MCA) as input.

The data for the 162 instances of within-sentence coordination were first formatted for MCA. “Token type” (four levels) and “coordination type” (two levels) were set up as the active variables, with “brow”, “body lean” and “mouthing” as supplementary qualitative variables. In theory, MCA allows for analyses with any number of categorical variables, each with multiple levels; however, in practice, greatly increasing the total product of factor level combinations (too many variables, or too many levels, or both) can result in an excessive number of dimensions of correspondence, risking a loss of explanatory power. For this analysis, the “mouthing” variable was collapsed into the binary values “no-mouthing” (representing the “none” value) and “mouthing” (all other values). The “body lean” variable was originally coded with four levels, but descriptive statistics (§3.3.1.5) revealed that the large majority of observations (91%) had a value of “neutral”, with the “forward, “backward”, and “side” values making up the remainder. The “body lean” variable was therefore reduced to the binary values “neutral” (representing no body lean) and “leaning” (any direction of lean). The three values of the “brow” variable (“neutral”, “up”, “down”) were found to be somewhat more evenly balanced and were left intact.

The MCA analysis determined that four dimensions of correspondence were required to explain all of the variation in the data. HCPC was then run on the MCA results, using Ward’s minimum variance method to partition the data into distinctive clusters. The question of how many clusters to create is an open choice for the researcher, and depends on correspondences within the data. Too many partitions risks overfitting, essentially magnifying noise; too few, and you risk the loss of informative distinctions. One useful criterion that can assist in judging the most effective balance is the gain of inertia (a measure of internal coherence) from making each partition. In this case, eight clusters

would account for all of the inertia, corresponding to all possible combinations of the two active variables, but on examination, the three least effective partitions only produced a minor net gain of inertia. Another estimation is the “elbow” method, in which the total variance (the within-cluster sum of squares) per cluster added is plotted, revealing the diminishing returns of making each partition; any visible “elbow” in the resulting curve (i.e. where it begins to flatten out more rapidly) suggests an informative number of clusters. Based on consideration of both of these estimates (Figure 13) and then comparing HCPC results with four, five and six clusters, partitioning for a total of five clusters was felt to produce the most informative analysis overall.

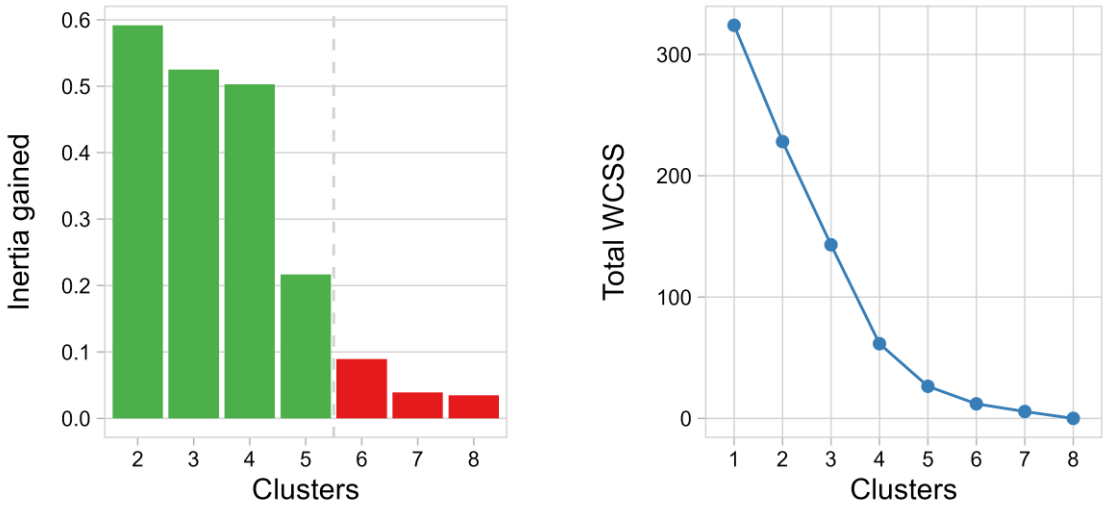


Figure 13: Inertia gained per partition (left), within-cluster sum of squares per cluster (right)

Chi-squared statistics were produced for each variable, indicating their relative significance in differentiating between clusters. Token type, coordination type and mouthing were significant factors ($p < 0.001$ in each case). Brow movement was not significant for the sample as a whole, but had an influence within one specific cluster (reported below). Body lean was not significant.

Figure 14 illustrates the sample as both a dendrogram (indicating how the clusters were partitioned, with the height of each split being proportional to the coherence gained by making it, echoing Figure 13) and a map (a visual impression of the relative sizes of each cluster and the distribution of variables within them). The cluster number in the top row refers to the numbered list of descriptions given below. It can be clearly seen that instances of body leans and brow movements are scattered across the entire sample, while mouthing is somewhat more coherently associated with some clusters than others, and conjunction and disjunction are quite strongly attached to certain token types for the most part.

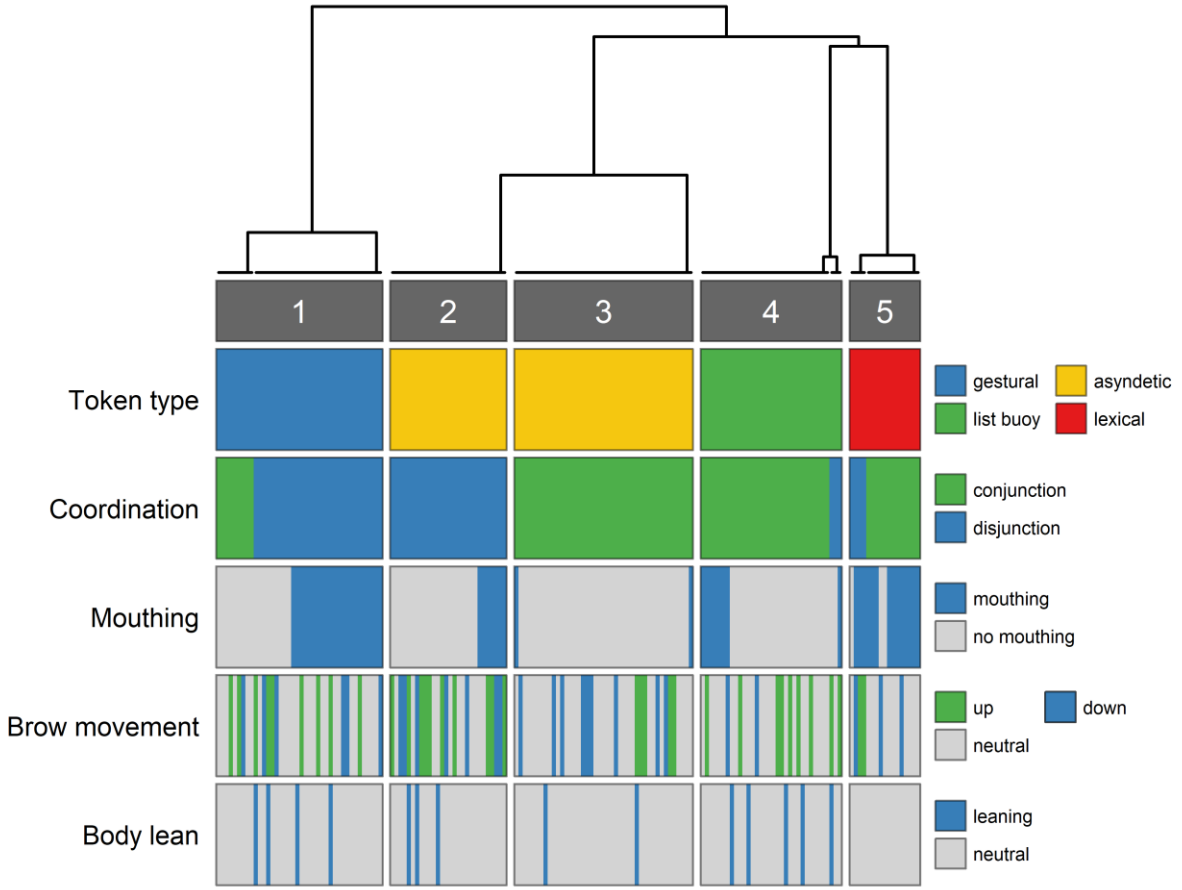


Figure 14: HCPC cluster dendrogram and map of variables

Each of the five clusters is quantitatively described below, labelled in terms of its relationship to the active variables. For given associations, the *p* value represents the significance of the variable in explaining the cluster in terms of an analysis of variance. The *v* value represents a comparison of the variable's representation within the cluster to its representation within the entire sample: the sign of the value (i.e. positive or negative) indicates whether the variable's representation in the cluster is greater (if positive) or smaller (if negative) than in the whole sample, and an absolute magnitude of over 1.96 suggests a significant difference comparable to a 95% confidence interval (Husson, Josse, & Pagès, 2010).

1. **Gestural tokens** (mostly *palm-up*, some pointing signs). There is a significant association with disjunction ($p < 0.001$, $v \approx 5.40$, 77.5% of the cluster). There is also a significant association with mouthing ($p < 0.01$, $v \approx 3.34$, 55.0% of the cluster). There is no significant association with brow movement, body lean, or their absence.
2. **Asyndetic disjunction**. There is a significant negative association against neutral brow movement ($p < 0.01$, $v \approx -2.60$, 39.3% of the cluster), but no significant association with either upward or downward movements. In terms of effect size, the

difference between the representation of neutral brows within the cluster and the entire sample is smaller than the size of effects found in other clusters. There is no significant association with mouthing, body lean, or their absence.

3. **Asyndetic conjunction**. There is a significant association with the absence of mouthing ($p < 0.001$, $v \approx 4.95$, 95.3% of the cluster). There is no significant association with brow movement or body lean, or their absence.
4. **List buoys**. There is a significant association with conjunction ($p < 0.001$, $v \approx 4.48$, 91.2% of the cluster). There were no significant associations with mouthing, brow movement, body lean, or their absence.
5. **Lexicalised coordinators**. There is a significant association with mouthing ($p < 0.001$, $v \approx 4.35$, 82.3% of the cluster). There is no significant association with coordinator type (conjunction or disjunction); both are represented within the cluster. There is also no significant association with brow movement or body lean or their absence. This cluster is the smallest, about a tenth of the entire sample.

3.4 Discussion

3.4.1 Questions addressed

Returning to our original questions of how conjunction and disjunction are expressed in the BSL Corpus in terms of the frequencies of asyndesis and lexical/gestural tokens, and how non-manuals might be associated with them, the results of the descriptive and inferential analyses above suggest that:

- Lexicalised coordinators (such as AND, NEXT, PLUS, OR) are attested but relatively rare, for both conjunction and disjunction. They are likely to feature mouthing.
- Asyndetic conjunction is the most frequent form of conjunction and is very unlikely to feature mouthing.
- Asyndetic disjunction can be tentatively linked with non-neutral brow movements, but not with either upward or downward brow movement specifically. Further examination of these items suggested an association with specific contexts (discussed below).
- Syndetic conjunction is much more likely to feature a pointing-based list buoy construction than a lexicalised coordinator. List buoys have no particular mouthing or non-mouthing association. While the large majority of list buoys were conjunctive, a small number of disjunctive list buoys were attested (discussed below). A small number of other pointing-based reference strategies were observed for conjunction.

- Syndetic disjunction is the most frequent form of disjunction, and is much more likely to feature the *palm-up* G:WELL sign than the fingerspelling loan OR.
- Body lean is not significantly associated with any of the coordination strategies studied here: it does happen (fairly infrequently), but not with any predictability with regard to the other categorical factors used in this analysis.

Some of these results were expected, confirming empirical findings from previous studies or matching existing documentation, particularly with regards to asyndesis and manual coordinator forms (§2.2, §2.3). Asyndesis appears to be a common strategy for conjunction, as asserted previously for BSL and other sign languages (Pfau, 2016; Waters & Sutton-Spence, 2005). For lexical coordinators, although relatively infrequent in comparison to asyndesis and list buoys, all of the lexical items documented in BSL SignBank with keywords relating to conjunction and disjunction (§2.2) were found at least once, with one exception: SAME with any of its conjunctive keyword senses, which it inherited from Brien (1992). This item was also not apparently found in any coordinative context by Waters & Sutton-Spence (2005). A speculative explanation for this is that it is so rarely used with this sense that it did not happen to appear in this sample (conversation data from 40 of the 249 BSL Corpus participants); the token can easily be found in the sample but only with its other senses (e.g. “same” or “similar”). It is also possible that it is more closely associated with discourse-level coordination, which was not examined by this study. Further corpus study would be required to establish whether these speculations are supportable; if not, the keywords may be questionable.

With regards to syndetic disjunction forms, the findings are again consistent with Waters & Sutton-Spence (2005) to the extent that the loan sign OR is attested but rare, and *palm-up* (G:WELL) is much more frequent. In fact, *palm-up* appears to be the preferred manual disjunction, which suggests that it is at least somewhat lexicalised for that function. This is not necessarily the case for the other discourse functions it has also been connected with (Arnold, 2019; McKee & Wallingford, 2011; van Loon et al., 2014).

Turning to non-manuals, asyndetic disjunction was also found (unexpectedly often) and it had a minor disassociation from neutral brow positions (discussed in more detail below), making it the only form of coordination in the sample which had any kind of tentative link with non-manual marking other than mouthing. Body leans were relatively infrequent across the entire sample, and had no detectable associations with any kind of coordination. No other forms (such as head nods) were observed to overtly mark coordination during the annotation process. These findings suggest that in contrast to

some general suggestions in the literature about non-manual contributions to coordination in other sign languages (Padden, 1988; Pfau, 2016; Tang & Lau, 2012), specific brow movements and body leans are certainly not obligatory markers of coordination in BSL and are not even conventional.

It should be noted, however, that while this analysis does suggest some reasonably clear trends for manual and non-manual contributions to coordination, the overall picture of conjunction and disjunction in BSL is one of some diversity. In the following sections, we will discuss some of the more variable or less expected cases in more depth.

3.4.2 Asyndetic disjunction

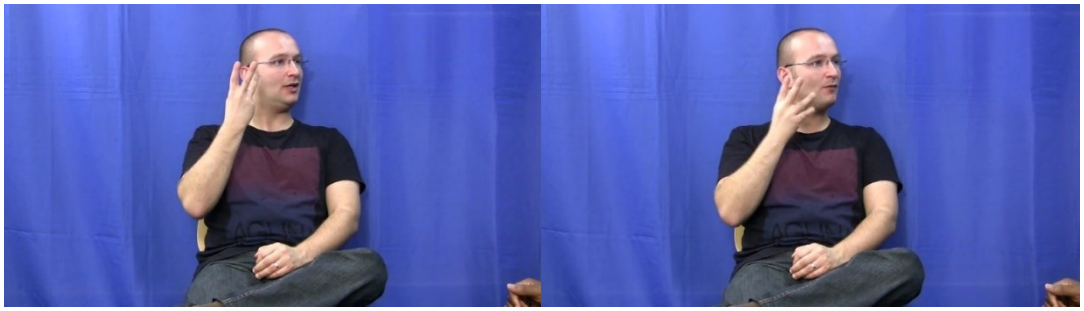
Within the sample, 28 of the 162 cases were annotated as being examples of asyndetic disjunction, which was enough for the analysis method to identify them as a distinctive cluster of associations. Proportionally speaking, the rate of asyndesis within all disjunctive coordinations was comparable to that found within all conjunctions (see Figure 9): this was not expected. Since disjunction tends to be more overtly marked than conjunction, it is also interesting that asyndetic disjunction is the only occasion that a significant association with the brow movement variable was detected, although it was a negative correlation with neutral brows (i.e. no particular preference for upward or downward movement) and the size of the effect (in terms of the cluster's contrast with the entire sample) was fairly mild. Just over a third of asyndetic disjunctions retained neutral brows.

Given that this demonstrates that asyndetic disjunction is not guaranteed to be non-manually marked in BSL, and that by definition asyndesis has no lexical coordinator, how is it distinguishable from conjunction? Waters & Sutton-Spence (2005) also previously suggested that either type of coordination can be asyndetic in BSL, but did not explicitly discuss what might set them apart. It seemed likely that semantic or discourse context might be making some kind of consistent contribution, and on further inspection of every case, this intuition was validated. A clear majority of 20 of the 28 examples had a "numeric hedging" context around quantities or time, such as examples 1 and 2 below; other clearly contributing contexts were identified for the rest. In comparison, only 2 of 38 examples of *syndetic* disjunction had this kind of hedging. Both used *palm-up* between or immediately before the numbers, as in example 3, which co-articulates *palm-up* on the non-dominant hand and also contains the gesture G:ERM (another open hand gesture but with wiggling fingers, used in BSL as a pause-filler or a marker of deliberation/uncertainty). These three specific examples are not marked with brow movement.

Example 1. **Participant BM06, conversation data, 871-874 seconds [video]**

ID glosses	FOUR02 FIVE PT:PRO1PL-FIVE
Literal gloss	four five five-of-us
Translation	Four or five of us.

Example 2. **Participant LN23, conversation data, 296-300 seconds [video]**



ID glosses	BEEN PAST REGULAR WAKE O'CLOCK-THREE02 O'CLOCK-FOUR02 WAKE
Literal gloss	been recent regular wake three-o'clock four-o'clock wake
Translation	Recently I've been waking up regularly at three or four o'clock.

Example 3. **Participant BL08, conversation data, 419-424 seconds [video]**

ID glosses	PERHAPS OTHER G:ERM EIGHT MONTH G:WELL SIX MONTH GO
Literal gloss	perhaps another erm eight months or six months to-go
Translation	… perhaps with another, uh, eight months or six months to go.

This prevalence of numeric hedging in asyndetic disjunction can be explained in terms of the pragmatic construal of intended meaning (more follows on this general topic in the next chapter). There has been much discussion in the pragmatics literature around the indefiniteness of numbers in certain contexts (e.g. Breheny, 2008; Carston, 1998; Geurts, 2006; Levinson, 2000); as Geurts put it, “there is no consensus on the meaning of number words like five”. A number does not necessarily describe an exact quantity: it can also imply a lower bound (“at least N”), an upper bound (“at most N”), or a range (“about N”) without requiring any explicit qualification, depending on the context. For example, the intended meaning of “Three years of experience are required” in the context of a job advert is likely at least three, not precisely three. On similar lines, an asyndetic list of numbers can be construed as setting out the bounds of the most probable range of possibilities, especially when describing a singular entity or occurrence, such that an explicit lexicalised disjunction is not required (and therefore dispreferred according to communicative principles of economy). As an illustrative example in English, an utterance such as “The cake arrives in three, four days” does not easily license an interpretation of “The cake arrives in three days and in four days”, even though asyndesis is often understood as conjunction. The sense may be influenced by grammatical context (the definite article “the” implying a specific or unique cake), previous discourse context around the purchasing of a cake for a specific purpose, prosodic context in the form of intonational and rhythmic modulations that are conventional for signalling uncertainty, and any other relevant context. If the cake cannot arrive in both three days and four days, the intended sense must be an exclusive disjunction (three or four but not both). In the BSL sample, all of the examples of numeric hedging had this kind of context and an exclusive sense.

Examining the remaining 8 occurrences of asyndetic disjunction that did not involve numeric hedging, all of them had other contextual reasons of one kind or another for an exclusive reading. Two had a lexical item within scope giving additional linguistic context for uncertainty between alternatives (ANYTHING in the sense of “any of them” or “whichever”, and this sign also bears a phonological resemblance to a two-handed *palm-up*; the other was DEPEND04 with the sense in that context of “depends which”). Two had an affective facial expression giving an clear sense of uncertainty. One listed logically opposed alternatives, licensing an exclusive reading (“retire” versus “continue working”). One also involved points in time but of a non-numeric relative nature (“last week” versus “this week”). The remaining three examples all had an English mouthing of “or” in the expected coordinator position between the manually articulated coordinates, as well as co-occurring non-neutral brow movements. It can be argued that these last three “mouthed

coordinators” are not true examples of asyndetic coordination. In the annotation scheme designed for this study, syndetic coordination was originally defined in terms of the occurrence of manually signed production, i.e. mouthing alone was not considered to warrant a classification of syndetic coordination. With hindsight, this decision bears re-examining in the context of code-switching and bilingualism.

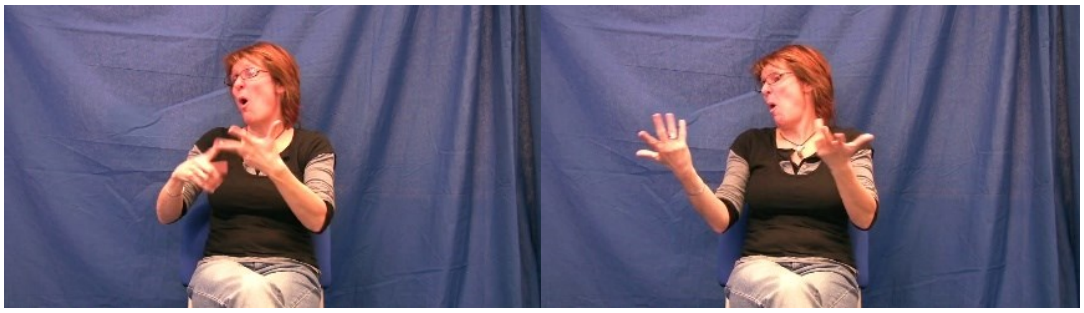
The conclusion to be drawn is that in BSL, where a sense of possible alternatives is clearly afforded by context (e.g. a speculative range of discrete quantities attributed to a single entity, a pair of diametrically opposed alternatives which cannot simultaneously be true, prosodic or semantic or discourse context contributing to uncertainty etc.), asyndetic disjunction is possible, but not necessarily preferred, and it is strongly associated with exclusivity. For numeric hedging specifically, a preference for asyndetic disjunction is indicated. If the discourse context is not sufficient to guarantee the intended reading, asyndetic disjunction tends to be additionally marked, perhaps via non-coordinating lexical context that explicitly indicates a sense of alternatives, or perhaps non-manually (with English mouthings or other non-manual features that are not especially conventional). Otherwise, disjunction is syndetic and uses a conventionalised lexical coordinator: the data suggests that the *palm-up* (G:WELL) sign can be regarded as such a form for that function. In the absence of any of the above, a conjunctive reading is afforded.

3.4.3 Disjunctive list buoys

As mentioned (§2.2.3), BSL and Auslan literature has occasionally asserted in passing (Johnston & Schembri, 2007; Waters & Sutton-Spence, 2005) that the function of list buoys is to effect “additive coordination”, which is synonymous with conjunction; the “additive” label does not usually imply disjunction, which is “alternative coordination”. However, in the current study, three examples from the BSL Corpus were identified where the list buoy episode has a clear sense of uncertain alternatives; the *palm-up* token G:WELL also appears within the context of all three. Two are shown in examples 4 and 5. In example 4, the second and third items in the list (supporting children, being a foster carer) are marked with a preceding G:WELL immediately before the list buoy pointers, and the brows are raised across all three items. Each coordinate is therefore marked with a sense of uncertainty or possibility, giving the entire construction a disjunctive “list of alternatives” reading. In example 5, a similar mixture of list buoy pointers and *palm-up* mark food and drink as conceptual possibilities, giving the listed pair a clear sense of non-exclusive alternatives. Only one list buoy marker clearly occurs within the phrase on this occasion, but the second G:WELL token before EAT is articulated rapidly and somewhat ambiguously

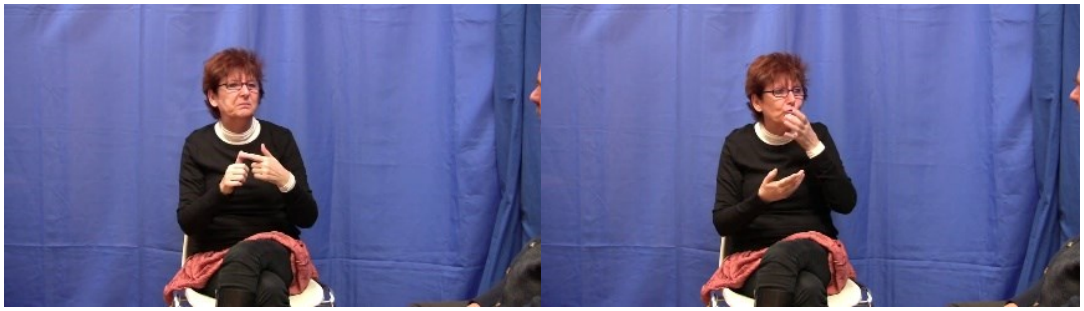
and it may also include a blended buoy. There is a third and final G:WELL marker near the end of the utterance. The brows remain neutral.

Example 4. **Participant BL10, conversation data, 179-186 seconds [video]**



ID glosses	G:WELL CAN WORK TEACH02 SCHOOL02 G:WELL DEAF SCHOOL02 PT:LBUOY G:WELL SUPPORT CHILD PT:LBUOY G:WELL LOOK-AFTER
Literal gloss	well possible work teach school or deaf school [list-item] or support child [list-item] or foster
Translation	Well, they could work as a teacher in a school or a deaf school, or otherwise supporting the children, or as foster carers.

Example 5. **Participant BL29, conversation data, 516-522 seconds [video]**



ID glosses	WANT TRUE PT:POSS3SG PT:LBUOY G:WELL DRINK PT:POSS3SG G:WELL EAT PT:POSS3SG G:WELL CA:DISAPPOINTMENT
Literal gloss	want true their [list-item] or drink their or food their well [disappointed]
Translation	I wanted to try their real cuisine or drinks. I was disappointed.

It appears, then, that while BSL list buoys effect an additive/conjunctive reading "by default", they do not intrinsically *prevent* a disjunctive interpretation of a list of alternatives. Overt marking (whether manual or non-manual) seems to be required for a reading of disjunction, but as previously suggested, that generally appears to hold for the contrast between conjunction and disjunction in any case. While it is the *palm-up* form that explicitly marks or carries the sense of disjunction in the list buoy episodes identified above, the buoys are still involved in delimiting the coordination structure. In contrast with previous characterisations of list buoys as purely additive itemisation (Johnston & Schembri, 2007;

Waters & Sutton-Spence, 2005), we might prefer to say that while list buoys conventionally mark out conjunctive coordination structures (in addition to their discourse-level deictic reference-tracking role), they can also be involved in forming disjunctive structures. Similar arguments have been made previously for ASL list buoys, although the additional marking differs from BSL (Davidson, 2013; Wilbur, 2012).

3.4.4 Non-manual features

As discussed previously (§2.4), Pfau (2016) suggests in very general terms that specific non-manual features (such as raised brows and head tilts) may be obligatory markers of coordination in some sign languages. Tang & Lau (2012) claims that for HKSL, asyndetic conjunction is marked with specific forms of non-manual feature, and asyndetic disjunctions “require” a body turn. However, in this analysis of a sample of BSL conversational data, no predictable contribution was found for body turns/leans to either conjunction or disjunction, regardless of lexical status or token type. For brow movements, the analysis did suggest a significant negative association against neutral brow movement for asyndetic disjunction, but with neither upward nor downward brow expressions being associated and more or less equally likely; the effect size was mild in comparison to the entire sample, i.e. neutral brows did still occur quite often. Brow movements were not significant for the other four clusters of coordination identified by the analytical method, and no other forms of non-manual feature were identified as repeatedly co-occurring with coordinators or coordinates during analysis. The participants were seated, which does slightly limit the range of possible body movements, and this might (or might not) contribute to the low frequency of body leans, as also noted in the Libras study of clause structures described in the previous chapter (Carneiro et al., 2023). However, sitting down is an unremarkable situation. Even if body leans could be demonstrated to be a little more frequent or more overtly marked when standing, it would not necessarily change the conclusions drawn.

Therefore, it cannot be concluded from this study of corpus-attested spontaneous conversation data that either conjunctive or disjunctive coordination have any particular obligatory or even typical form of non-manual marking in BSL, whether the coordination is syndetic or asyndetic. It seems more likely that brow movements and body leans occur sporadically in this BSL data because they are not an intrinsic part of the expression of coordination but are instead voluntary and prosodic in nature, used for a specific communicative purpose; based on some of the cross-linguistic literature on the use of body leans (§2.4), we might speculate that this purpose is contrastive stress or focus.

3.4.5 Limitations and future directions

The data sample used in this analysis was only of moderate size (162 within-sentence coordination episodes captured from a 20,000 token sample across 40 conversational partners), but this was felt to be more than sufficient for the statistical analysis method chosen. The decision to restrict the sample to conversational data which had clause-like boundary annotations previously completed by a previous project entailed that the four regions represented were all in England, so the sample does not reflect any usage from Scotland, Wales, or Northern Ireland. The "naturalness" of the conversation data might also be mildly criticised, given that it was elicited by researchers in a contrived setting and produced under the subtle pressure of a video camera. It is therefore important to clarify here that I am not claiming that these findings are necessarily representative of everyday usage across the BSL community. These things said, the BSL Corpus is (at the time of writing) the largest and most thoroughly annotated data set available for BSL usage. Furthermore, principled design and procedures underpinning the original collection of the corpus data did aim to encourage natural, free and relaxed interaction.

It does also bear repeating that this study deliberately excluded examples of coordination in negated, modally qualified, and question-form contexts, and also excluded coordination at the discourse level. The former are grammatical structures known to change or cancel the particular kind of pragmatic reasoning this thesis is ultimately concerned with (scalar implicatures), and discourse coordination is also likely to be beyond the scope of that kind of inference. Future studies not bound by these requirements might shed light on similarities and differences between sentence and discourse level coordination in BSL for a broader range of grammatical contexts.

3.5 Conclusions

This exploratory study of spontaneous conversation data from the BSL Corpus has demonstrated that while there is some diversity in the expression of conjunction and disjunction in BSL, some constructions are more frequently employed than others, and associations are detectable. While BSL has a number of attested lexical items associated with conjunctive and disjunctive functions (such as AND, PLUS, NEXT, OR), their use is relatively infrequent. Asyndetic coordination is the most common strategy for conjunction; asyndesis is also attested for disjunction, but in that case tends to pragmatically rely on prosodic or epistemic context, and heavily favours exclusive readings ("either but not both"). List buoys are the next most frequently employed strategy to represent conjunction;

they are also further attested (much less frequently) to have a disjunctive sense when additional context is available, such as the *palm-up* sign indexed in BSL SignBank as G:WELL. In fact, syndetic disjunction is most frequently denoted by *palm-up* in any case, which therefore seems to be at least somewhat conventionalised for a disjunctive function if not fully lexicalised. In comparison, the fingerspelling loan OR is used very infrequently.

Non-manual marking of coordination was attested to sometimes occur in the form of brow movements and spatial contrast via body leans, but the study found no predictable association between non-manuals and either type of coordination, with the sole exception of brow movements for asyndetic disjunction. However, even in that case, neither upward nor downward brow movements were significantly favoured, with the only predictable tendency being against a neutral brow position; the size of the effect was low-to-moderate, with over a third of asyndetic disjunctions retaining a neutral brow position. English mouthings were associated with fully lexicalised coordinators and *palm-up*, but in any other case their use was infrequent; a significant association with the absence of mouthing was indicated for asyndetic conjunction. There was no pattern of association for body leans, and no other repeated form of non-manual marking was noted.

Taken together, we can infer from these findings that disjunction tends to be more overtly marked than conjunction in BSL, as is likely the case in most languages. However, the occasional assertion in the literature for other sign languages that either type of coordination might *require* specific non-manual markings, of a morphosyntactic nature or otherwise, is not at all supported for BSL by this analysis. The conclusion preferred, concurring with Crasborn & van der Kooij (2013) for NGT (see §2.4), is that non-manual marking of coordination in BSL is prosodic and voluntary, functioning as emphasis or contrastive stress, and that there is no single conventionalised manner of marking: the amplitude or overtness of the marking is more of a pragmatic contributor to intended meaning than the precise form.

Assuming that this is the case, it is possible that this kind of prosodic marking could contribute to the pragmatic reading of manual disjunctions as either exclusive or inclusive: this is one of the conjectures that the remainder of this thesis sets out to test.

4 Review: scalar implicatures and judgement tasks

We have now established a picture of what conjunction and disjunction can look like in conversational BSL, supported by an original analysis of a sample from the BSL Corpus. One of the primary goals of that investigation was to inform the design of the experimental scalar implicature study that follows in chapter 5. Further to that end, this chapter explores theoretical and experimental issues from the pragmatics literature that are also likely to be informative, as well as further learning points from sign language linguistics.

The preamble to this thesis gave a brief introduction to scalar implicatures (§1.3), pointing out in particular exclusivity inferences arising from the contrast between conjunction and disjunction (whereby “either or both” and “either but not both” are both possible to be construed from “or”, depending on context). However, competing theoretical accounts of implicature may be shaped by deeper assumptions about the cognitive basis of language. How is the nature of belief in intended meaning best characterised? Are there “stages” or “modules” (semantic, pragmatic) involved in producing and understanding language? If so, what is the architecture and flow of the connections or “interfaces” between them? Is the ability to reason pragmatically innate, or is it acquired from childhood exposure to a linguistic environment, or a combination of both? Does language modality matter (whether the language is spoken or signed)? How might any of this be empirically investigated?

We will bear such questions in mind during the following reviews, first establishing some contrasting theoretical accounts of scalar implicature (§4.1), and then evaluating a number of theoretical and methodological lessons from experimental approaches to their study (§4.2). The role of language acquisition in developing pragmatic competence is likely to be of special interest to sign language researchers, given the atypical acquisition experience of most deaf users of sign languages (§1.4.1). However, experimental studies of scalar implicature derivation in a sign language have so far been exceptionally rare. We will discuss one such study in some detail, but spoken language studies exploring issues around childhood performance with scalar implicatures may also provide relevant points around the acquisition of pragmatic competence. We will also very briefly cover some issues around age of sign language acquisition, and discuss an example of a paper on BSL grammaticality judgements with an inclusive approach in that regard.

4.1 Theoretical accounts

As mentioned very early in this thesis, Grice (1968, 1975) is widely credited as giving a seminally important description of the mechanisms of conversational implicature, fuelling decades of discussion and research. In the Gricean framework, conversational norms are characterised first and foremost as arising from a “co-operative principle”. Conversation is assumed to be co-operative by default, in that it happens for a purpose or purposes: people are generally rational, and what they say and understand tends to be tailored towards facilitating such purposes. This main principle is further elaborated by four conversational maxims and sub-maxims. The maxim of Quantity concerns the extent to which co-operative speakers should be informative (enough, but not too much); the maxim of Quality concerns truthfulness (say things that are known to be true or supportable); the maxim of Manner reinforces clarity and brevity; and the maxim of Relation merely requires utterances to “be relevant” (which Grice describes as the most difficult of these concepts despite the very terse wording). In their original formulation, the principle and maxims can read like requirements or directives, but they are only intended to be descriptive, setting out tendencies within the underlying reasoning of typical humans in everyday conversation. The maxims can be violated, on purpose or by accident, and ostensively or covertly (they are maxims, not rules or laws): people can and do lie, or make mistakes, or talk too much, but these are all less than optimally co-operative exchanges by definition.

Over the following half century, theorists and researchers have attempted to refine, extend, or replace the Gricean framework, not least in terms of accounting for the cognitive processes involved in deriving meaning (Grice was essentially a philosopher, and did not set out to investigate or speculate on the cognitive architecture that might be required). As a co-operative agent myself, I hesitate to be overly informative, given that others have already provided thorough reviews and summaries of the historical state of the art (Breheny, 2019; Degen, 2023; Khorsheed, Price, & van Tiel, 2022; Papafragou & Skordos, 2016). However, in particular, there are three influential theories (or groups of theories which share key commonalities) that are contextually informative for the experimental pragmatics reviews to follow.

4.1.1 Gricean and “traditional” accounts

Towards understanding the logical basis for scalar implicatures, consider again the example of a quantity-based statement from Mill (1865), first mentioned in §1.3, this time with explicitly elaborated alternatives:

1. I saw some of your children today.
2. I saw some but not all of your children today.
3. I saw all of your children today.

Grice proposed that the co-operative principle and conversational maxims can explain how, when people hear or see example 1 produced, they often take away the understanding that example 2 was meant, because they take the producer's decision not to produce example 3 into account, even though example 1 does not actually exclude example 3 (in purely literal terms). To give a slightly simplified version of the Gricean chain of reasoning: given that you believe me to be rational and co-operative (in accordance with the co-operative principle), and also believe me to be informative enough but appropriately economical while still giving a relevant and true account (in accordance with the maxims of conversation), if I truly believed that I had seen all of your children, you would expect me to have said so; because I did not say so, you are led to believe that I must not believe it to be true that I had seen them all; therefore, by "some", you believe that I must have intended for you to understand "some but not all". The same chain of reasoning can be applied to items coordinated with a disjunction: the production of example 4 may be understood as implying that example 5 was meant given that example 6 was not produced instead, despite example 4 not necessarily excluding example 6 in purely literal terms:

4. There is a biscuit or a cake in the cupboard.
5. There is either a biscuit or a cake in the cupboard, but not both.
6. There is a biscuit and a cake in the cupboard.

However, it is important to note that the derivation of the "pragmatic" meaning is definitely not guaranteed: implicatures are described by Grice as being explicitly or contextually cancellable, and later works by others characterise them as defeasible (capable of being cancelled or annulled or blocked). To give a trivial example, if I explicitly added "In fact, I saw them all" to the end of example 1, the "some but not all" implicature would not be licensed. Much discussion has ensued about the linguistic, sensory, and epistemic contributions that might constitute pertinent context, as well as what various interpretations of "defeasibility" might imply for the temporal order of processing literal and intended meanings, as we will see.

Grice called this general kind of reasoning from unstated alternatives "conversational implicature": the term "scalar implicature" was later coined by Horn (1972), in reference to the idea that for at least some implicatures, the key terms can be said to form a "scale" of

increasing informativity (in the sense of precision or certainty). Quantifier and coordinator scales are popular choices for illustration and study because semantically, they correspond directly to fundamental logical operators in set theory and Boolean algebra respectively (§1.2). In the quantifier <*some, all*> scale, “all” is more strongly informative and certain than “some”, because for all possible states of a set containing elements that may or not have a property (e.g. a collection of children, each of which may or may not have been seen), there is only one possible combination in which all of the elements have the property, so the meaning is certain; but there are many possible combinations in which only some of them have it and the utterance “some” does not specify which. Likewise, with the coordinator <*or, and*> scale, for a pair of coordinated propositions, each with a logical truth value (true or false), conjunctive “and” is more informative than disjunctive “or”. With a conjunction, there is only one possible way in which the entire proposition can evaluate as true, which is when both coordinates are true (biscuit in cupboard is true and cake in cupboard is true). In comparison, disjunction is informatively weaker, because either coordinate (but not both) can be true if the disjunction is read as exclusive (two possibilities), and either coordinate (or both) can be true if the disjunction is deemed to be inclusive (three possibilities); recall that no known natural language makes a lexicalised distinction between inclusive and exclusive disjunction (Haspelmath, 2007). There are very many other possible scales (<*might, must*>, <*clever, brilliant*>, and so on) that depend on other kinds of semantic contrast beyond logical operations. Scales are also not limited to two members, but binary contrasts are easier to describe and study.

Over the following decades, a number of theorists attempted to modify the original Gricean account, often towards addressing perceived problems or gaps, but sometimes with the aim of accounting for proposals on the nature of the actual cognitive processes involved in deriving pragmatic meanings. Theories in this neighbourhood have sometimes been loosely or inconsistently bundled together under the labels “neo-Gricean”, “post-Gricean” or “traditional”. One of the most debated and studied has been the “default” account proposed by Levinson (2000), perhaps due to its experimental tractability. This model is relatively close to the original Gricean framework in that it does not particularly challenge the assumptions of rationality, co-operation and purposefulness, but it has more to say about the sequence of reasoning involved and the lexical status of scalar terms.

In the “default” account, implicatures as a whole are generally classed as either “generalised conversational implicatures” (GCIs) or “particularised conversational implicatures” (PCIs). Scalar implicatures of the kind discussed in the numbered examples above are said to be GCIs: they are “generalised” in the sense that their derived pragmatic

meaning is proposed to be carried or triggered automatically by certain forms of words, i.e. the scale members are all (to some extent) lexicalised. Otherwise, lacking a conventional formula, an implicature is treated as "particularised". It is further assumed that there is some form of staged cognitive process involved in arriving at an intended meaning. For GCIs specifically, the intended pragmatic meaning of a scalar implicature (a rejection of the unstated stronger alternative) is proposed to "go through" automatically by default (hence the "default" model), and additional time and processing costs are only invoked if contributing context warrants the cancellation of the implicature, in which case the literal semantic meaning is derived instead. This hypothesis can be contrasted directly with the reversed alternative, a "literal first" hypothesis, in which the unenriched literal meaning of the utterance is always derived in the first instance, and it is the derivation of the pragmatic meaning that invokes additional cognitive costs. It has occasionally been assumed that the original Gricean account of implicature aligns with a "literal first" hypothesis, but Breheny (2019) and others have pointed out that this is not actually prescribed by Grice's work. Regardless, it was realised that these opposing hypotheses on the temporal sequence and costs of processing implicatures should be amenable to experimental online detection ("online" in this context meaning taking measurements while responses are being processed and then made). This might be done via measuring response times (where slower responses are taken as indicative of increased processing costs), eye-tracking (where the sequence of visual focus on contextual elements is taken as indicative of the chain of reasoning), or brain activity/imaging techniques (where detected activations are taken as indicative of the sequence and costs of the process), or combinations thereof. Several studies of the period explored this question experimentally (e.g. Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; De Neys & Schaeken, 2007; Huang & Snedeker, 2009). Evidence has tended to suggest an increased cost of processing implicatures in comparison to literal controls, which is clearly contrary to the predictions of the "default" model. Rejecting either the "default" or "literal first" hypotheses does not necessarily challenge assumptions that the nature of processing an implicature is sequential or modular, although some researchers have explicitly done so.

4.1.2 Relevance Theory

Pulling on that last thread, another influential and competing work was Relevance Theory (Sperber & Wilson, 1986, 1995), sometimes labelled by others as a "contextual" account; for a relatively digestible summary see Wilson & Sperber (2002). This departs significantly from several Gricean notions in that it challenges the adequacy of maxim- or norm-based

accounts of conversation, as well as rejecting any “code model” of communication which assumes the automatic sequential calculation of propositional truth values. An over-arching inferential theory of cognition based on “relevance” is proposed instead, aiming to accommodate all forms of pragmatic communication, including implicatures as well as loose or figurative speech: human cognition itself is proposed to be geared towards maximising this communicative relevance. An interpretation of an utterance is considered “more relevant” when it yields more “positive cognitive effects”, i.e. informative connections with other available contextual cues or knowledge; the stronger the sum of informative connections made, the more relevant the interpretation is, until some satisfying threshold of relevance is reached and an intended meaning is arrived upon. All implicatures are subject to contextual interpretation and are therefore intrinsically effortful, tending to have a higher cognitive cost than what are termed “explicatures”, which only convey a literal or logical meaning. But unlike the “literal first” hypothesis, this process is not staged or modular in the “code model” sense: it is essentially a network model in which interlocutors make parallel and simultaneous connections across web-like information structures.

Relevance Theory has occasionally been criticised as being difficult to verify experimentally, and even its authors expressed concerns in that regard in the early years of its development (Sperber & Wilson, 1987). Nonetheless, a number of experimental studies have cited either Relevance Theory specifically or more broadly “contextual” accounts generally as providing a better, if not complete, explanation for results found than either the “literal first” or “default” hypotheses are capable of (Bott & Noveck, 2004; Chevallier et al., 2008; De Neys & Schaeken, 2007; Fekete et al., 2015; Kurumada & Clark, 2017; Noveck & Posada, 2003).

4.1.3 Probabilistic accounts

In the last decade or so, a number of newer theories and models have been developed which have been collectively referred to as “probabilistic” accounts of pragmatics, variously drawing on parallel research in the fields of semantics, pragmatics, psycholinguistics, cognitive psychology, and computational linguistics. Such accounts tend to share the “contextual” view above that the process of deriving an implicature can take a range of cues as input, including any form of linguistic or epistemic context. They also share the view that it is difficult to fully formalise a model of internal reasoning from a maxim-based account alone. However, they may take points of issue with the assertion that implicature is *intrinsically* more effortful than explicature. Furthermore, the measurement of implicature derivation as a binary outcome (either made or not made) is comprehensively challenged:

probabilistic accounts characterise beliefs in intended and alternate meanings as a *likelihood*, essentially a degree or strength of belief, and the informativeness of contextual influences also have a gradient value. Some accounts also directly confront the "traditional" view that lexicalisation has a distinct and special role in scalar implicature reasoning in comparison to other kinds of pragmatic inference.

4.1.3.1 The "constraint-based" account

On that last point, the "constraint-based" account offered by Degen (2015) and Degen & Tanenhaus (2015, 2019) succinctly lays out its objections to what it describes as a widespread "homogeneity assumption" in some earlier accounts of implicature, comparable on some points to a "uniformity assumption" discussed by van Tiel, Miltenburg, Zevakhina & Geurts (2014). The assumption is said to originate in the Gricean and Levinsonian classification of lexicalised scalar implicatures as GCIs (§4.1.1): these were originally defined as "strength invariant" and "context independent", and each scalar term is required to be lexicalised to the same degree (as opposed to PCIs, which are not required to be lexicalised and are more directly reliant on context). The authors argue that these assumptions persist throughout many previous and contemporary debates on the nature of scalar implicature, even within arguments which otherwise reject key planks of "traditional" models (such as the processing order of literal and pragmatic meaning), since they continue to assume that scalar implicatures are strength invariant and context independent, and are therefore directly comparable across utterances and producers.

According to Degen's argument, given the assumption that beliefs are better characterised in terms of probability or strength, it follows that if the reported strength of a "generalised" scalar implicature can be shown to vary significantly with the receiver within a fixed discourse context, then, all else being equal, it is not strength invariant; similarly, if particular linguistic contexts can be shown to consistently affect the strength of scalar implicatures across receivers, then they are not context independent. If they are not strength invariant and not context independent, then by definition they cannot be GCIs, and are starting to resemble PCIs; in which case, the GCI/PCI distinction may be an arbitrary one without actual grounding in cognition. In Degen (2015), this issue was experimentally tested in a relatively large corpus-based judgement task study with a non-binary rating paradigm (details of the methodology are discussed in §4.2.1, §4.2.2). Analysis of the results suggested statistically significant associations between factors, such that variations in implicature strength correlated with specific linguistic and discourse contexts across all items. Furthermore, the strength of specific "some but not all"

implicature items significantly varied between participants despite the fixed context. The conclusion frames this as evidence against both the strength invariance and context independence definitions of GCIs, supporting the argument against the GCI/PCI distinction, the argument for an account of belief as degree, and the argument for consideration of a broad range of contextual cues that can vary in informative strength.

4.1.3.2 The Rational Speech Act framework

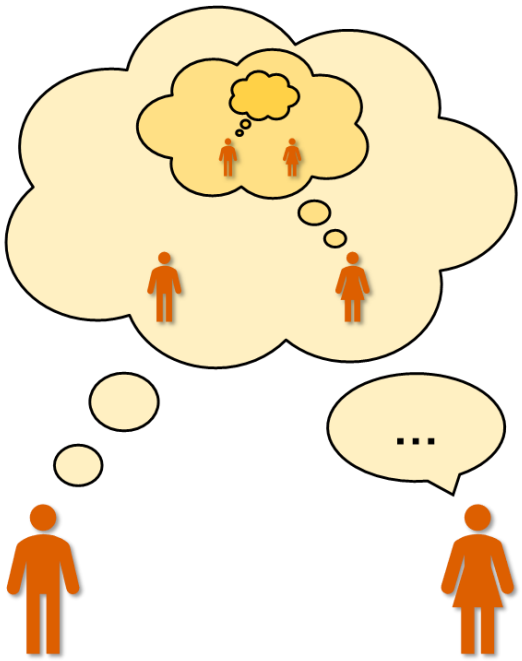


Figure 15: Simple illustration of recursive pragmatic reasoning, adapted from Goodman & Frank (2016)

Another common theme across "probabilistic" models is that there are no linear "stages" to the underlying cognitive architecture of pragmatic inference, and therefore no "interfaces" between semantic and pragmatic "modules": the question of whether it is the literal or pragmatic meaning that is derived first does not arise. Instead, these models tend to characterise the interlocutors as developing an internalised model of each other from which, given a variety of linguistic and contextual information as parallel inputs, inferences across the relative probabilities of every alternative reading of the producer's intended meaning being true can be drawn as output. This process of probabilistic inferential reasoning has been formally modelled in various ways which can be applied not only to implicatures but a range of linguistic phenomena. Notable areas of endeavour are models inspired by Game Theory (the study of strategic decision-making by rational individuals: for an overview of the application to pragmatic reasoning, see Benz & Stevens, 2018); non-Bayesian probabilistic models (e.g. Russell, 2012); and the fully Bayesian approach of

the Rational Speech Act (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016). The latter has arguably gained the most traction within the field, and is the formal school we will briefly visit here. For a comprehensive explanation and review of RSA's mechanisms, development, and future directions, with reference to a range of theoretical and experimental works informed by the framework, see Degen (2023), from which the following summary is partly drawn.

Under RSA, several Gricean assumptions such as interlocutor rationality, reasoning about alternatives, and expectations of a productive balance between informativeness and economy are more or less preserved, but in a substantially different paradigm. In common with other probabilistic accounts, a range of potentially relevant contexts (beyond just lexical choice) are assumed, as well as the characterisation of both contextual informativeness and strength of belief in intended meanings as gradients rather than categorical outcomes. Crucially however, what is elsewhere termed "world knowledge" is handled here not as a shared bank of objectively true information, but as subjective and mutable "prior beliefs" about the true state of things. These "priors" are iteratively updated by the interlocutors, who recursively model each other's reasoning during the course of interaction (see Figure 15) towards deriving "posterior beliefs" in intended meanings. This process of iteration is one of the defining features of Bayesian inference, in which prior probabilities are incrementally updated by new information until they converge on a stable posterior outcome. The RSA application of Bayesian inference runs as follows (paraphrased to use relatively accessible terms and avoiding the use of equations):

- The process of modelling begins with a hypothetical "literal listener" who infers a logical semantic meaning as the initial prior belief. Note that this is not the same concept as the "literal first" hypothesis described previously, since in this case the literal listener does not exist beyond the model: the literal meaning is never settled upon independently of any alternative meaning or cue that might contribute.
- A "pragmatic speaker" is then introduced as considering a range of alternative utterances as they might be understood by the listener; alternatives are preferred on the basis of a balance between informativeness and cost. An utterance is more informative if it increases the probability that the *intended* meaning will be inferred; in comparison, considerations of cost are more difficult to define but typically include retrieval difficulty and the phonetic and grammatical complexity of the utterance (as a prototypical example, "some but not all" is more complex than "some").

- To complete the model, a “pragmatic listener” derives the speaker’s most likely intended meaning by “reverse-engineering” the reasoning process above, based on their own prior beliefs about the most likely intended meaning and their expectations about the pragmatic speaker’s likely production choices for any different possible meaning they might want to communicate.

Due to its computational tractability, RSA is able to make very specific and experimentally verifiable predictions about derived pragmatic meanings from specific utterances in specific contexts, as long as prior beliefs can be justifiably estimated, the literal semantics of all possible alternative utterances can be specified, and the contributing informativeness of various contextual influences can be identified and quantified against specified cognitive costs. These are not trivial requirements, and RSA has therefore been criticised for being more easily applicable to contrived experimental paradigms that test limited responses to “one-shot” utterances than it is to natural conversational exchanges (although to be fair, this is likely to be at least equally true of every previous and competing framework). However, it has made rapid progress over a relatively short period, and continues to be developed and applied.

In terms of direct relevance to the current study, questions that are still open for probabilistic accounts of implicature (including RSA) include the nature and range of linguistic, sensory, and epistemic phenomena which have the potential to contribute as relevant inputs to the derivation of pragmatic meaning (Bergen & Goodman, 2015; Goodman & Stuhlmüller, 2013; Russell, 2012), although it does not seem to be contested that lexical, grammatical and prosodic cues are all valid examples. See Degen (2023) for an overview of other current challenges and future directions.

4.1.3.3 Recursive reasoning and theory of mind

Another interesting point of comparison between various probabilistic accounts concerns the recursive depth of reasoning required: this in turn relates to the acquisition of pragmatic competence in the psychological context of “theory of mind” (the ability to internally represent and predict the mental state of others). This issue is particularly relevant to the current study given that language acquisition delays in deaf children were associated with either delays or lasting deficits in theory of mind ability by several twentieth century studies, and investigations have continued in the current century (Blose & Schenkel, 2022; Corina & Singleton, 2009; Courtin & Melot, 2005; Durrleman, Dumont, & Delage, 2021; Falkman, Roos, & Hjelmquist, 2007; Morgan & Kegl, 2006; Peterson,

Wellman, & Slaughter, 2012; Richardson et al., 2020; Schick, de Villiers, de Villiers, & Hoffmeister, 2007; Smogorzewska, Szumski, Bosacki, & Grygiel, 2022; Wellman & Peterson, 2013; Woolfe, Want, & Siegal, 2002). The essence of the argument has been that the development of both theory of mind and pragmatic competence requires passive exposure to complex adult-level linguistic interactions from infancy onwards, and that deaf children who do not have signing parents are less likely to have early access to this. However, the argument as stated here may be overly reductive (Marschark, Edwards, Peterson, Crowe, & Walton, 2019), and reliance on child performance in "false belief" tasks (an experimental paradigm for assessing theory of mind ability) as a predictor for lasting social and cognitive difficulties into adulthood has been criticised (Begby, 2023). Furthermore, while several of the psychological studies cited above make passing references to assumed links between theory of mind and general pragmatic ability, such links are rarely explored in terms of a linguistic framework for pragmatic communication.

However, such links have been explicitly made in studies of scalar implicature and autism, which has also been associated with theory of mind deficits. For example, Pijnacker et al. (2009) experimentally compared rates of scalar implicature derivation between two groups of adults with contrasting Autistic Spectrum Disorder (ASD) diagnoses (but no diagnosis of intellectual disability or global language delay) and a neurotypical control group; some points of the methodology used will be covered later. Statistically significant differences were found for one of the ASD groups in comparison to controls (fewer pragmatic readings of implicatures, and longer response times to make them), although their performance was characterised as "actually quite good" on the whole, reflecting a low-to-moderate effect size; the other group, with a less severe ASD diagnosis, did not show any significant difference. Towards accounting for this, the authors describe an argument put forward by Happé (1993) in the context of Relevance Theory's possible predictions for ASD and implicatures. The proposal is that scalar implicatures may only require "first order" theory of mind abilities ("I know that you know"), associated with recognising the speaker's *informative* intent, rather than the "second order" reasoning ("I know that you know that I know") associated with recognising the speaker's *communicative* intent[5]. If this argument actually holds, and if adults with specific ASD diagnoses could be more directly shown to have relative facility with first order reasoning

[5] The distinction between informative and communicative intent was made by Grice (1957) and was further explored by Relevance Theory (Sperber & Wilson, 1986, 1995). The former is the intent to make information known, while the latter is the intent to make it known that you intended to make the information known.

compared to second order, it might help to explain the partial or total preservation of scalar implicature derivation (and other inferential skills) in some individuals with ASD. Another ASD study of the period (Chevallier, Wilson, Happé, & Noveck, 2010) found similar evidence and also suggested that different pragmatic skills require different orders of reasoning, but the question of which order scalar implicatures require was not settled.

These studies (Chevallier et al., 2010; Happé, 1993; Pijnacker et al., 2009) were published before the maturity of "probabilistic pragmatics" accounts. In the light of more recent proposals, it is interesting to note that some early work in the RSA paradigm (Goodman & Stuhlmüller, 2013) similarly suggested that RSA might only require a single level of recursive reasoning, whereas Game Theory approaches of the time optimised probabilities to an infinitely recursive depth, requiring higher order reasoning (Franke, 2009). However, more recent implementations of RSA also seem to allow for recursion to proceed to any depth (Franke & Jäger, 2016; Goodman & Frank, 2016). These accounts do not yet seem to have been explored in ASD studies. In a more recent experimental study which directly explored the first-order/second-order contrast (Hochstein, Bale, & Barner, 2018), individuals with ASD performed close to the level of controls in scalar implicature tasks once again (in contrast to their performance in other aspects of epistemic reasoning), but the pragmatics argument is still framed in terms of a Gricean account.

Extending the arguments above away from autism and back to the theoretical link between delayed language acquisition and theory of mind deficits, if only first order reasoning is required for scalar inferences, then one might expect to see no significant effect of delayed acquisition on the strength of implicatures made, whereas if it requires deeper recursive reasoning, effects might be expected and a contemporary probabilistic account of implicature might be further supported. The first-order/second-order contrast has been directly explored in deaf children, finding that they were more successful with the former than the latter (Jones, Gutierrez, & Ludlow, 2015), but again, this was done via "false belief" paradigms and not directly in terms of any linguistic-theoretical account of pragmatic reasoning. Further interdisciplinary study is likely required.

4.2 Experimental approaches

As alluded to above, experimental studies have attempted to gather evidence for or against a number of theoretical positions on scalar implicatures. In this section we turn to methodological issues and consider elements of judgement task design, preferring examples of studies investigating inferences from the contrast between conjunction and disjunction where possible (seemingly less common than those which examine the "some

but not all" implicature). We will also examine experimental design issues relating to language acquisition and the development of competence in handling potential implicatures. With regard to the question of how scalar implicatures in a sign language might be studied, we will look at a rare example of a study in ASL and English in some depth. Finally, given that rarity, we will briefly summarise some issues around language background in studies of BSL that are not specific to pragmatics, finishing with learning points from a study in which effects of age of BSL acquisition and the first or second language status of English on grammatical sensitivity were found.

4.2.1 Judgement task paradigms

Experimental judgement task experiments in linguistics typically expose participants to a series of linguistic stimuli, with or without some other overt non-linguistic context, and ask them to provide some kind of response or rating for each one individually or in comparison to each other. Contrasts within the stimuli are designed to illuminate the specific research questions involved; this often concerns grammatical acceptability or "naturalness", but the method is applicable to any linguistic area, including pragmatics. The rating mechanism can be a closed-ended response, in which the participant selects from a range of alternative ratings provided for them (such that the responses are amenable to quantitative analysis), or they might be free to provide open-ended responses in their own words (requiring at least some level of qualitative interpretation by the researcher, whether or not that takes the form of principled coding towards a quantitative analysis). Quantitative rating paradigms include:

- Binary choice paradigms, in which the participant selects between "true" and "false", or "right" and "wrong", or "accept" and "reject", or any pair of opposing absolutes.
- Multiple choice paradigms, such as a Likert scale, representing an ordinal ranking or discrete degree, either numerically or descriptively labelled: for example, a five item scale often used in opinion polls is "strongly disagree", "disagree", "neither agree nor disagree", "agree", "strongly agree"; alternatively, only the extreme end-points might be explicitly labelled.
- Continuous scale paradigms, optionally labelled at the end-points or at gradations along the scale; for paper-based studies, participants might make a mark at some point along a line of fixed width (requiring the researcher to measure and record the distance from an endpoint for every response), but for screen-based studies this usually takes the form of a draggable slider.

- Forced choice paradigms; there seems to be some confusion around this terminology, as some have used it to mean a closed response paradigm (i.e. the provision of a fixed and pre-populated rating scale as above); others use the term to mean a paradigm in which items are not rated individually but are presented in pairs, with participants indicating which of the pair meets a criterion better than the other (e.g. “Which is more natural?”); a ranking of items is then inferred from systematically working through the space of possible comparisons, ideally in a randomised order.

For a quantitative study, the selected rating system places direct limitations on the choice of statistical method employed to analyse the results, but may also reflect theoretical motivations or assumptions: we will examine such issues in more detail shortly. But whichever paradigm is used, the purpose of the rating in studies of scalar implicature is to capture some indication of whether or not, or how strongly, a participant has inferred an implicature from a particular stimulus or combination of stimuli. An appropriate form of statistical analysis can then determine whether or not there is a dependable correlation between the rates or strengths of implicatures made and potentially predictive factors (per stimulus item, or per participant, or per interaction between item and participant), providing evidence for arguments around the causes that might be driving such correlations.

Historically, some scalar implicature studies have asked participants to make judgements on bare linguistic stimuli without any further explicit context, relying (at least in part) on the participant’s “world knowledge”, which can be characterised as a form of implicit context. As an example, one of the experiments in Noveck (2001) examines the difference in responses between adults and two different age groups of children (8 and 10) to short statements in French with the potential to trigger a quantity implicature, in which the rating mechanism was a logical true/false binary choice. The statements were absurd distractors (e.g. “All chairs tell time”), appropriately true (“All elephants have trunks”), or inappropriately false (“All dogs have spots”); the quantity term in each stimulus was either “some” or “all”; this led to six possible sentence conditions. One of these, the “inappropriate some” stimuli, were taken to represent pragmatically infelicitous but logically true statements (e.g. “Some giraffes have long necks”), the response to which would be taken as directly indicating whether a scalar implicature was made or not. For example, if “Some giraffes have long necks” were rejected as false, it can be assumed that the participant had determined that the upper-bounded “Some but not all giraffes have long necks” reading was intended: since knowledge of giraffes is likely to suggest that absolutely all giraffes have long necks, that upper-bounded meaning is false, and so the

rejection of “some” can be taken as evidence of a “some but not all” implicature made. This kind of design relies on the participants knowing (for example) the features that characterise an archetypical giraffe. It may be possible to exclude those without the required knowledge via pre-screening or by principled exclusions following examination of their responses to the control conditions (e.g. if a participant consistently rejects sentences like “All giraffes have long necks”, there may be a problem). Noveck (2001) did not apparently set any exclusion criteria, but did assess the children’s performance with the control statements, finding them to be generally very competent (although this does raise the question of how the test condition results would have been interpreted if the children had in fact been found to fail too many controls). Pijnacker et al. (2009) used a comparable set of bare linguistic stimuli (featuring both quantifier and coordinator stimuli), previously discussed above in the context of autism and theory of mind (§4.1.3.3). The primary research questions were different and the statistical methods employed were a little more nuanced (§4.2.2), and a principled approach to data exclusion was taken, but the presentation of stimuli and the world knowledge requirement were along similar lines (e.g. “Some sparrows are birds”, “Snakes have paws or wings”) and the rating paradigm was the same binary truth value choice.

In contrast, other implicature studies have provided some form of explicit extra context, allowing for a more direct manipulation of non-linguistic variables. The majority take the form of a “sentence/picture” judgement task, where each linguistic stimulus describes an accompanying visual stimulus, and the participant is asked to rate the accuracy or appropriateness or naturalness of the description. A range of illustrative studies follows, some of which will be further explored (§4.2.2, §4.2.3, §4.2.4):

- Chevallier et al. (2008), in which the participants gave binary true/false ratings to either written or spoken descriptions of printed words and non-words (e.g. the visual stimulus “RSOUD” and the descriptive linguistic stimulus “There is a T or an R”). Contrastive stress on the scalar terms (intonational emphasis for the spoken stimuli, orthographic emphasis for the written stimuli) was also manipulated per item, and correlations between the presence of stress and the rejection rates were analysed.
- Davidson (2013), a between-groups study of quantifier and coordinator scale implicatures derived by ASL and English users. Photographs of various arrangements of household objects on a table were combined with a spoken or signed sentence stimulus (e.g. “A spoon is in the mug or a spoon is in the bowl”); the rating paradigm was a logical true/false binary choice.

- Degen & Tanenhaus (2015), a set of experiments employing a “gumball paradigm”, in which the participant rated the appropriateness of spoken descriptions of an animated machine dropping gumballs (e.g. “You got some of the gumballs” when either some, all or none of the gumballs actually dropped) via either a seven point Likert scale rating or a binary rating coupled with a recording of response time.
- Kurumada & Clark (2017), in another study of contrastive stress, spoken only in this case, as interpreted by young children versus adults; the linguistic stimuli were combined with pictures of various animals and objects (and additionally some puppetry and other props with the children); the child responses were open but manually coded by the research team, while the adults gave a closed binary response; the chosen animals were previously rated for familiarity (i.e. for world knowledge accessibility) with a separate group of children in a similar school setting who did not go on to participate in the main study.
- Van Tiel & Schaeken (2017) and van Tiel & Kissine (2018), two methodologically similar studies but with different target participants and research questions; participants gave a binary true/false rating to descriptions of a collection of different geometric shapes in a range of colours (e.g. “Some of the shapes are red”); the motivation for this choice of visual element is explicitly described as an attempt to avoid participants having to draw upon world knowledge (by which we might assume they mean that familiarity with colours and shapes is a more fundamental experience than familiarity with, for example, exotic animals).
- Jasbi, Waldon & Degen (2019), a meta-methodological study with a “guessing game” paradigm, which set out to test the interaction between differently composed rating scales and other theoretical assumptions on participant judgements of potential exclusivity and exhaustivity inferences; the picture stimuli were illustrations of a “playing card” featuring various combinations of cartoon cats, dogs, and elephants; participants were introduced to Bob, fictitious character who made guesses (in the form of scalar statements) about what might be on the card without having seen it, such as “There is a cat or a dog on the card”; we will return to this particular study more than once in both this chapter and the next.

There are also non-pictorial ways to provide additional context or contrast. As mentioned in our discussion of “probabilistic” accounts above, in a corpus-based experimental study, Degen (2015) exposed participants to a large total number of potential “some but not all” implicatures taken from the Switchboard Corpus, an annotated database of written

transcripts of spontaneous telephone conversations between strangers. Each extract inherited corpus annotation metadata which contributed to factor coding for some of the experimental conditions (syntactic partitivity, determiner strength, and discourse accessibility). The extracts were presented to participants as a written transcript of bilateral dialogue, including several preceding conversational turns as discourse context. The final utterance in each stimulus contained the target scalar phrase: this was highlighted alongside a modified version which explicitly encoded the upper-bounded interpretation of the original (for example, the original “I like to read some of the philosophy stuff” in red text with the modification “I like to read some, but not all, of the philosophy stuff” in green text). Participants were then required to rate the similarity in meaning of the two versions (within that particular discourse context) using a seven point Likert scale. The inclusion of corpus-attested spontaneous conversation data was particularly novel (even in transcript form), and the non-binary response capture method was appropriately designed around the theoretical account of implicature presented, although the statistical method might be mildly criticised (to be discussed in the next section).

Other studies, perhaps especially those focused on the cognitive architecture involved in deriving implicatures, do not capture a judgement rating at all, but are concerned instead with the costs of processing the linguistic input (as previously described in §4.1.1). For example, in Breheny, Katsos & Williams (2006), participants were required to read a series of short passages (balanced for length) and press a button when they had finished reading each one. In this case, the length of time taken per item was analysed for correlation with experimental variables, rather than any examining any direct judgements of the passage’s intended meaning. Degen & Tanenhaus (2015), in the “gumball” series of related experiments mentioned above, also recorded response times in one experiment, in addition to a binary judgement rating; other experiments in the same series used a seven point scale, but the binary rating was presumably used in this case so as to simplify the choice between rating options and thereby reduce inflation of the time taken to choose.

4.2.2 The consequences of rating scale granularity

As we have seen, in quantitative judgement tasks, the type of response data gathered (binary, ordinal ranking, or continuous measure) is a direct consequence of the rating paradigm. This places limitations on the choice of statistical analysis method. However, the chosen paradigm may also be entangled with assumptions from within a particular theoretical account of implicature (examples of which were summarised in §4.1). In Gricean and “traditional” accounts, the chain of reasoning involved in deriving implicatures

is dependent on the underlying binary logic of quantifiers (set theory) or coordinators (Boolean algebra), not to mention the broader assumption that every proposition carries a binary truth value (true or false). It is therefore not surprising that studies aligned with such accounts have deemed the binary rating paradigm to be an appropriate reflection of participant judgements, especially given that statistical methods for analysing such data have been available for much longer. Some of the earlier experimental pragmatics studies designed with binary choice ratings employed relatively simple analyses of variance between proportions of response (ANOVAs and related methods) with additional post-hoc tests (e.g. Chevallier et al., 2008; Noveck, 2001), although it has since been argued that ANOVAs are not an appropriate analysis method for closed choice judgement tasks (especially on arbitrarily transformed data) and should be avoided (Jaeger, 2008). Other experimental pragmatics studies of the time preferred non-parametric methods such as the Kruskal-Wallis or Mann-Whitney tests (e.g. Pijnacker et al., 2009; Slabakova, 2010): these have the benefit of accounting for the non-parametric nature of the collected responses (discussed below), but are still relatively unsophisticated tools by today's standards. In the last decade or so, following the gradual refinement of statistical tools and non-commercial public domain software that implements them, mixed effects analyses that more accurately and thoroughly account for the underlying distributions and variations in judgement task response data have become feasible even for researchers without deep statistical knowledge. For binary response data, more recent implicature studies have appropriately employed various forms of binomial or logistic mixed effects regression, which have the advantage of accounting for repeated measures and the so-called "random" effects of participant variation (Davidson, 2013; Kurumada & Clark, 2017; Skordos & Papafragou, 2016; van Tiel & Kissine, 2018).

However, as explored above (§4.1.3), there are accounts of scalar implicature that take issue with binary judgement paradigms on a theoretical level. "Probabilistic" accounts characterise the belief that the stronger alternative warrants rejection as a matter of degree, rather than a binary absolute. Experimental pragmatics studies on such lines have therefore been more likely to favour an ordinal ranking paradigm, allowing for a granular range of responses; Degen (2015) and Degen & Tanenhaus (2015) both employ seven point Likert scales, for example. However, a common pitfall when using such scales as the basis for an effects analysis is to assume that they are metric (evenly-spaced linear quantities) when they are in fact a ranking. Regardless of whether each point on the scale is explicitly labelled with a numeric value (e.g. 1 to 7 or -3 to +3), or represented to the participants with descriptive or symbolic labels but treated as numeric values internally,

parametric analysis methods that assume an unbounded measurement and a normal distribution of residuals (such as linear mixed effects modelling) are arguably inappropriate. On a conceptual basis alone, it is hard to argue that Likert scales represent and capture evenly-spaced degrees of judgement for even one participant, let alone that the spacing is objectively the same for everyone (e.g. the distance between “agree” and “neither agree nor disagree” is not necessarily the same as the distance between “agree” and “strongly agree”, whether these options are explicitly presented as 1, 2 and 3 or as labels). Such responses should not be subjected to mathematical operations that treat them as if they had a metric value (e.g. trying to calculate the mean of “strongly disagree” and “slightly agree”, or the standard deviation of a set of emoji; the median and mode can be used less inappropriately). More importantly however, with regard to aggregating and analysing ordinal Likert scale response data for correlations with predictive factors, it has been demonstrated that metric methods make invalidating assumptions about the distribution and boundedness of responses which can potentially lead to effect detection errors (Type I and II), distortions in effect size, even *inversions* of effects, and (theoretically, if rarely in practice) impossible predictions that are off the scale entirely (Liddell & Kruschke, 2018; Vuorre, 2019). Likert scales are a form of discrete ordinal ranking, and as such the current methods arguably best suited to their analysis are some form of ordinal regression, such as cumulative link or adjacent category models (Bürkner & Vuorre, 2019). In a footnote, Degen (2015, p. 21) does note that an ordinal regression performed on the same data (as opposed to the linear mixed effects modelling presented in the main body of the paper) yielded the same qualitative results “in terms of significance of effects” (which might be interpreted as implying that the quantitative effect *sizes* may still have differed); but this was to some extent lucky.

With further regard to the impact of theoretical background on selecting a rating scale paradigm for scalar implicature judgement tasks, the Jasbi et al. (2019) experiment mentioned briefly above is illuminating. The authors discuss the widespread references in experimental pragmatics to “implicature rates”, a consequence of an assumption that implicature derivation is a binary state (made or not). They further hypothesise that in the context of a study of pragmatic reasoning, the chosen number of ranks for a multiple choice scale has the potential to radically change the interpretation of results, depending on the definition of what constitutes a “pragmatic response”, which can be related in turn to the underlying theoretical account of implicature. Binary, trinary, quaternary and quinary rating scales are experimentally compared in the study. In each trial, groups of participants played a “playing card guessing game” in which they rated the appropriateness of

descriptive linguistic stimuli via Likert scales of varying length per group. The test condition stimuli pairs had the potential to elicit either a level of belief in an exclusivity inference from the scalar contrast (“either or both” versus “either but not both”), or an exhaustivity inference (“There is an X here” versus “There is only an X here”). The first and last ranks were always labelled “Right” and “Wrong”; if there were an odd number of points in the scale, then the central point was labelled “Neither”; if there were more than three points in the scale, then the intermediate off-centre points were labelled “Kinda wrong” and “Kinda right” (an American English contraction meaning “kind of” or “approximately”). If you were to subscribe to a “traditional” account in which an implicature is either made or not, you might interpret anything less than total acceptance (“Right”) as some level of “enriched” pragmatic response, i.e. that a scalar implicature was made: this is described by the authors as a “weak” linking hypothesis. Alternatively, your theoretical assumptions might lead you to determine that anything other than total rejection (“Wrong”) is indicative that a scalar implicature was *not* made, labelled as a “strong” linking hypothesis. Other linking hypothesis options are available (such as setting the boundary between rejection and non-rejection at an intermediate point in the scale, as well as the fully gradient “degrees of belief” consideration favoured in probabilistic accounts of implicature), but the experiment confines itself to these two extremes for the sake of illustration. Accordingly, trial responses that were not already binary were collapsed into binary values (“SI” and “not SI”) according to the specification of the respective linking hypothesis, and then collectively analysed using Bayesian binomial mixed effects modelling with the rating scale length, trial type (exclusivity versus exhaustivity) and linking hypothesis (weak versus strong) as predictive factors. The results indicated a credible predictive effect on so-called “implicature rates” arising from interactions between the number of response options, the linking hypothesis, and the trial type, all within a study otherwise carried out under the same conditions. This evidence deftly demonstrates that given an assumption that implicature derivation is a binary state (made or not made), the combination of the number of discrete response options you give to participants and your definition of “pragmatic response” have the potential to radically change conclusions drawn from the results. The authors recommend a linking hypothesis which supports the treatment of implicatures as a gradient of belief instead.

The ultimate extension of the argument for a move away from a binary response option towards a more granular rating scale, particularly if your theoretical motivation is to try to reflect a probabilistic degree of belief in a pragmatic reading, is to use a continuous rating scale: infinitely granular, in theory if not in practice. An ungraduated continuous

scale (one without any intermediate labels or tick marks at all, but allowing for labels at each extreme) has the further advantage of side-stepping the non-parametric statistical analysis issue discussed above, since such scales can be appropriately treated as metric (but bounded). However, during the literature search for this review, no published experimental pragmatics judgement task studies which used a continuous rating paradigm were located. This may be partly explained by the historical difficulty of finding an appropriate statistical method to handle such data. In the past, continuous rating variables have sometimes been modelled with methods which assume an unbounded measurement and a normal distribution of residuals, but aggregated judgement task responses on a bounded continuous scale do not meet this requirement (Vuorre, 2019). In more recent years, however, other families of distribution have become available for driving mixed effects analyses: arguably, derivations from the beta distribution are particularly appropriate for continuous judgement task data (more detail and justification follows on this point in §5.2.3). As further incentive, there have also been suggestions that continuous scales may have a statistical edge over binary ratings and shorter Likert scales (less than seven options) in situations where sample size is necessarily limited, as is likely to be the case in minority language studies with relatively limited recruitment potential (such as the current study). For example, Marty, Chemla & Sprouse (2020), building on judgement task design advice in Sprouse & Almeida (2017), demonstrate in a simulation study that if that there is no theoretical or practical requirement for a binary choice paradigm, the minimum sample size required for adequate statistical power (or alternatively the detectable effect size) can be reduced, substantially in some cases, by using more granular judgement scale paradigms (such as a seven point Likert scale or a continuous scale, with careful attention paid to the details of labelling the scale).

4.2.3 The acquisition of pragmatic competence

Rating paradigms are not the only methodological issue with a direct connection to the supporting theoretical account: the details of how the linguistic stimuli are presented can also influence the results, and studies which have explored this issue have led to theoretical advances. For example, young children are often assumed to have less pragmatic competence than adults, not least because they are likely to have less linguistic competence in general. Formal language proficiency (vocabulary and grammar), executive function (attention, inhibition, and working memory), and theory of mind (see §4.1.3.3) have all been associated with pragmatic development, although the need for further research has been highlighted (Matthews, Biney, & Abbot-Smith, 2018). While pragmatic

competence covers a range of abilities, it includes handling scalar implicatures: in that regard, it has been said that children can be more logical than adults (Noveck, 2001). A number of studies over the last few decades have investigated this intuition, comparing early age performance to that of adults and attempting to account for differences observed. The utility of examining how children acquire competence with implicatures is that, as well as improving our understanding of the developmental process, it may reveal insights into how implicatures actually work in terms of the contrast between weakly and strongly informative terms (§4.1.1). Recall, however, the relative prevalence of language deprivation or acquisition delays in deaf children, particularly for the large majority who do not have signing parents or guardians as language models (§1.4.1). Deprivation can have consequences for general language skills, neurocognitive development, and quality of life (e.g. Cheng et al., 2019; Hall et al., 2017, 2019; Twomey et al., 2020), as well as for pragmatic competence (Matthews & Kelly, 2022; Szarkowski et al., 2020). A study of scalar implicatures in BSL therefore needs to account for a diversity of language acquisition experiences in addition to any linguistic factors of interest. In this section, we examine the findings and arguments of three studies that provide key notes for theory and experimental design.

Noveck (2001) documents a relatively early but influential (if contested) set of experiments, following up on works which previously claimed that young children tend to derive the logical (non-pragmatic) meaning of scalar terms more often than adults. Three experiments were designed to compare rates of scalar implicature derivation between two groups of children (aged 8 and 10) and a group of adults, based on French equivalents of the quantifier <*some, all*> scale and a modal certainty <*might, must*> scale; we have referred to the experimental and statistical methodologies of the third experiment in the series previously (§4.2.1, §4.2.2), for which an analysis of variance indicated one main effect of age. Both groups of children were said to be significantly more likely to accept pragmatically infelicitous statements as true (i.e. they often appeared to interpret “some” as being fully compatible with “all”, and “might” with “must”) in comparison to the adults, who were very likely to reject such statements. Post-hoc tests suggested that the difference between the 10 year olds and the adults was the main contributor to the effect.

Discussing the results and suggesting possible avenues of future research, Noveck (2001) compares a neo-Gricean theoretical account to that of Relevance Theory as applied by Carston (1998), suggesting that either one might be capable of predicting the findings, but only if certain assumptions within each account could be validated. In the neo-Gricean “default” story (§4.1.1), “generalised” (i.e. lexicalised or conventionalised)

implicatures are automatically derived as the pragmatic upper-bounded meaning first of all; a second stage of deriving the literal interpretation only later cancels the pragmatic meaning if context warrants it (entailing additional processing costs). Noveck suggests that the neo-Griceans would "probably" assume that the children only *appear* to be proceeding to the literal stage with significantly increased regularity because they have not yet "generalised" the implicature, due to lacking sufficient experience of deriving it, and they are therefore going straight to the logical semantic meaning. The Relevance Theory version of events, on the other hand, would hold that the pragmatic interpretation is only ever derived if it is contextually required to achieve the expected level of "relevance" (§4.1.2), otherwise the purely semantic interpretation satisfies; the findings would therefore be directly explained if it could somehow be demonstrated that the children had their expectations of relevance immediately satisfied by the literal meanings of "some" and "might", or that the cost of deriving the implicature is for some reason greater for children (reducing the net cognitive benefit of relevance gained from making the implicature, making it more likely that the semantic meaning satisfies). It is not explicitly suggested how either of the internal processes offered here could be demonstrated directly, particularly for the Relevance Theory issue, for which retrieval of the stronger alternative from a hypothetical lexicon would need to be demonstrated as a measurable cost (issues around retrieval and accessibility are to be discussed shortly). With regard to the neo-Gricean "default" account explanation, we have previously covered other experimental findings that have thrown reasonable doubts on the model in general (§4.1.1, §4.1.3.1), but all were published in the years following Noveck's study.

Skordos & Papafragou (2016) follow up on similar questions around the difference between child and adult performance, citing a range of experimental implicature studies from Noveck (2001) onwards. In particular, they highlight arguments around the role of the "accessibility" or "retrievability" of scalar alternatives, in the sense of the ability to retrieve semantic meanings from a hypothetical mental lexicon. Some of these accounts are more naturally aligned than others with the characterisation of a certain class of implicatures as "generalised", roughly compatible with (but further qualifying) the neo-Gricean hypothesis offered in Noveck (2001) above. For example, the authors cite a "restricted alternatives" hypothesis put forward by Tieu, Romoli, Zhou, & Crain (2016) which suggests that if in any specific case a child does not derive an implicature, it could be for one of two general reasons. One possibility is that they have not internalised the "co-scalar status" of the alternative terms (i.e. they have not yet fully learned that the semantic meanings of all of the alternatives in the scale are contrastively salient, and therefore cannot "generalise" that

contrast). Alternatively, they may in fact have internalised the contrast to at least some extent, but on being presented with the under-informative term, they still find it more difficult to retrieve the stronger alternative from their hypothetical lexicon than adults do. To test this latter hypothesis (that to make an implicature, the semantics of the stronger alternative must be retrievable), the authors predict that for those children who can be demonstrated to understand the semantic meaning of the individual terms via responses to control conditions, increasing the accessibility of the stronger alternative (by making it explicitly available in some way) might increase the rate of implicatures made.

The study consists of three experiments, with the second and third making relatively small adjustments to the paradigm to follow up on findings from the previous: we will focus on the first experiment here. The participants were a group of five-year-old children and a group of adult controls (college students), all monolingual English users. The experimental paradigm employed has several noteworthy features not always seen in earlier pragmatic judgement task studies, attempting to rigorously control for effects specific to the driving research questions around lexical retrieval and the scope of alternatives. Principled exclusion criteria were set, based on responses to control statements. Quantifier *some* and *all* stimuli in the main trials were restricted to the scope of four novel and unique creatures (“blickets”), allowing for a precise establishment of visual and epistemic context and controlling the possible range of alternative situations. For the children, the experiment featured a short pre-trial screening section, designed to accustom the children to the tasks and encourage responses; unlike the main task, for the pre-trials the children received some feedback if they didn’t reject pragmatically infelicitous descriptions. Both the pre-trial section and the main task featured puppets as the providers of the linguistic stimuli (respectively, “Max the silly gorilla” who “says silly things sometimes”, and “Ben the Wizard” who uses magic to create “the only blickets in the world”). After a statement which described an accompanying visual stimulus, the children were asked if the puppet had “said it well”, with their open responses recorded (later coded as a binary “yes/no” response). For the adults, the puppet was replaced with a cartoon animation providing the same combinations of stimuli, and they recorded their own answers (a binary “yes/no” response with an optional space for giving more information about any “no” decision).

The linguistic/visual stimuli combinations produced four truth conditions: three semantic controls and a pragmatically infelicitous test condition. Trial sets were further divided into three distinct trial orders, balanced across participants: “mixed”, where *some* and *all* stimuli were presented in a pseudo-random order (but ensuring that the main truth condition did not repeat itself in any adjacent pair of trials); “some-first”, in which the *some*

and *all* stimuli were presented in separate blocks, with the *some* block always coming first; and "infelicitous-some-first" in which the *some* block also always came first, but was furthermore ordered such that the pragmatic test conditions were always presented before the semantic control conditions. These combinations of condition and order were designed to allow for accessibility to be treated as a predictive factor: in the "mixed" order where *all* (as the stronger scalar alternative) was permitted to feature early in the trial sequence, it was deemed to be "more accessible" in the sense of being more easily retrievable, since the participant would have recent activation of it; in the orders where *some* was front-loaded, accessibility to *all* might be relatively reduced.

For the analysis, "yes/no" responses were re-coded as "correct/incorrect" depending on the truth condition. The descriptive results demonstrated that adults almost always performed at ceiling; the children performed below ceiling but not poorly in most of the semantic control conditions, and for the semantically true *all* condition specifically, they also performed close to ceiling. However, they did noticeably less well for most of the different trial orders in the pragmatic test condition. Statistical analysis (Fisher's exact test and a confirmatory logistic regression) confirmed that while the children did not significantly differ in their responses between trial orders for any of the semantic control conditions, there was a significant difference for the pragmatic test condition: where the order was "mixed" (i.e. the stronger alternative was more accessible), the children performed relatively well, rejecting the description most of the time; for "some-first" (less accessible), they performed much less well; for "infelicitous-some-first" (least accessible), they performed poorly. A follow-up analysis applied the participant exclusion criteria, removing those children (roughly a third) who could not be guaranteed to have understood the semantics of *some* and *all* (as interpreted from their responses to the semantic controls). This produced qualitatively similar results, not essentially changing the main conclusions drawn. Finally, a between-groups analysis of the difference between adult and child performance in the pragmatic test condition suggested an effect of accessibility, such that the difference for the "mixed" order (in which the stronger alternative was most accessible) was not significant: only in the orders where the stronger alternative was less accessible did the children perform significantly worse than adults.

The authors conclude that their findings are not incompatible with previous studies suggesting some childhood difficulties with scalar implicatures, but with important qualifications gleaned from the methodological approach: young children can in fact handle at least some forms of prototypical scalar implicature with a nearly adult level of performance, as long as the scalar alternative is made accessible or retrievable, and as

long as steps are taken to ensure they are familiar with the goals of the interaction, and as long as the scope of alternatives is both finite and established. While some children may take longer to learn the semantic meanings of scalar alternatives than others, contrasting scalar term accessibility is also a requirement for developing facility with pragmatic inferences in the early stages of language acquisition. A broader ramification of this finding is that judgement task design clearly has the potential to significantly affect the results.

Finally, Kurumada & Clark (2017) describe a study of intonational contributions to scalar implicature, making a developmental argument that can be taken as broadly supporting a "contextual" account of implicature. The responses of four-year-old children to the presence and absence of contrastive stress are explored and compared to those of adults, following research which suggested that children do not reliably understand contrastive contours of intonation until age 6-10. The specific implicatures examined, based on the ad hoc[6] scale *<X looks like a Y, X is a Y>,* are perhaps less relevant to the current project; however, the general question of the contributions of contrastive stress to implicature derivation, and how this might interact with the development of pragmatic ability, is highly relevant. We have already briefly encountered the experimental paradigm above (§4.2.1) and will not examine it much more closely here; as in the previous study, it was designed around the practical needs of working with very young children.

The study compares the results of a series of experiments, variously assessing the effects of contrastive stress on the verb ("It ***looks*** like a zebra" can imply "I don't think it is a zebra") or the noun ("It looks like a ***zebra***" can imply "I think it is a zebra") or without any particularly marked stresses, as well as effects from providing additional discourse context ("It ***looks*** like a rope. It's a cord" can imply "I don't think it is a rope"). The experiments are not large in scale but are principled in terms of the strength of claims made. A great deal of care is paid to the detail and consistency of the pitch contours of the intonations used in the stimuli, with extensive reference to the literature in that area. The overall conclusion drawn is that four-year-olds can be shown to reason pragmatically, significantly better than chance and in some cases approaching an adult level of competence, if they have access to *both* contrastive stress and additional context, but less so if they only have one or the other (i.e. aligned with but qualifying previous findings which suggested that young children demonstrate less competence than adults with contrastive stress alone). It is

[6] An "ad hoc" scale usually means that it is relatively unconventional, in the sense that the contrast between alternatives is based somewhat less on literal semantics and more on situational context; for a study comparing child responses to conventional versus ad hoc scales, see Stiller, Goodman, & Frank (2011).

argued that via learning from repeated exposure, specific intonational contours (initially novel to children) are increasingly recognised as conventional for specific forms of meaningful contrast, and that this learning process is scaffolded or reinforced in the early stages of language development by access to additional discourse context, a scaffold which becomes increasingly redundant as pragmatic competence develops.

4.2.4 Scalar implicatures in American Sign Language (ASL)

We now turn to the exceptionally rare cases of experimental scalar implicature studies in a sign language. The first such published studies were most likely those originally carried out within a doctoral thesis on the semantics of scalar terms in ASL (Davidson, 2011), and no others could be located during the current project's literature search. Three individual experiments within the thesis were later adapted for separate peer-reviewed publication (Davidson, 2013, 2014; Davidson & Mayberry, 2015). One of the experiments is directly concerned with the contrast between conjunction and disjunction forms in a sign language, as is the current project, and it therefore deserves a detailed review.

Figure 16: ASL "body shift" (Vicars, 1997)

Davidson (2011, Chapter 3) presents a felicity judgement task experiment with a mixed factorial between-groups design, comparing scalar inferences made by deaf ASL signers (n = 10) and hearing English speakers (n = 12) from stimuli based on quantifier *<some, all>* scales and coordinator *<or, and>* scales; we will tend to focus on the latter here. To establish a linguistic background, the author describes several coordination strategies in ASL, supported primarily by the testimony of expert consultants (two deaf ASL users, one with a linguistics background). The larger part of this documentation is on two different constructions that Davidson classes as "general use coordinators": list buoys (described and illustrated previously in §2.2.3, glossed as COORD-L), and a form described as a "body shift" or lean/turn in signing space (COORD-SHIFT, illustrated in Figure 16). "General use" in this context means that these coordinators are said to have the potential for both conjunctive and disjunctive readings, depending on context; an analogy is made

with similarly generalised coordinators in Maricopa (in the Yuman language family, Arizona, USA) and Japanese. The resolution of these alternative readings is said to be afforded either by lexical context, such as the addition of "discourse particles" glossed as EITHER (for disjunction) and BOTH (for conjunction), or by non-manual context via specific additional marking over the coordinates (head nods for conjunction, a "squint and bit lip" for disjunction). This particular example of non-manual marking is asserted to be "the equivalent of intonation in spoken languages", in the functional sense that they may serve to emphasise affective or grammatical distinction; the current study has also made this comparison (§3.5).

The given description of ASL list buoys does overlap with the current study's documentation of list buoys drawn from a BSL Corpus sample (§3.3.1.7), which matches Davidson's documentation to the extent that BSL list buoys are corpus-attested in both conjunctive and (much less frequently) disjunctive constructions, and that it is additional context that affords a disjunctive reading to a list buoy episode (§3.4.3). However, to recap, the current project found no reliable pattern of use for any *specific* co-occurring non-manual feature in either conjunctive or disjunctive list buoy coordination phrases in the BSL Corpus, as is claimed by Davidson to be a feature of ASL list buoys (see paragraph above). It is worth noting here some supplementary findings from a later qualitative study (Podlesny, 2015) on the role of non-manuals and the influence of English contact on ASL disjunction forms, informed by the judgements of five deaf Canadian ASL users. There, it is suggested that both ASL list buoys and the COORD-SHIFT body lean tend towards an additive reading in the absence of additional lexical or non-manual context, i.e. it is only the disjunctive readings of both forms that seem to *require* disambiguating context, which is a slightly different account to Davidson's and more strongly reminiscent of the current study's general conclusions from BSL Corpus data (§3.5).

Further minor differences between Davidson's account of ASL list buoys and the findings from the current project include an assertion that "the first COORD-L is never optional, and also must immediately precede the first coordinate". This was strongly indicated not to be the case for the current project's BSL Corpus sample: over a quarter of the list buoy episodes examined lacked an initial buoy before the first coordinate (§3.3.1.7). The study also suggests that for some (but not all) of the informants consulted, ASL list buoys were felt to be too "prosodically heavy" to coordinate a pair of light nouns; however, in the current study's BSL Corpus sample, two-thirds of the identified BSL list buoy episodes had only two coordinates (e.g. a pair of noun phrases), and the mean number of manual tokens per coordinate was less than two (both in cases where there

were only two coordinates and more than two). This can be taken as evidence that relatively short and light list buoy episodes can and do happen in BSL. These inconsistencies between the respective accounts of list buoys in ASL and BSL may reflect actual cross-linguistic differences between two sign languages that lack a familial relationship, or may be due to the nature of the respective evidence bases (expert individuals' intuitions versus analysis of natural language usage via corpus attestation).

ASL is further described in the study's documentation as having a small inventory of lexicalised coordinators available which are not "general use", i.e. they each have a manual form which conveys either conjunction or disjunction but not both. However, none of them were selected as a basis for sentence stimuli in the study's judgement tasks, and nor was the list buoy form; the reason for not using the latter as the basis for exclusivity inferences was not made fully explicit, unless it was the previously mentioned intuition that list buoys were deemed to be unconventional for coordinating pairs of light phrases in ASL. The only coordination contrast in the ASL stimuli was between differently marked versions of the non-manual body lean COORD-SHIFT construction. Relatively few examples of body lean (in any direction) were found within the current project's BSL Corpus data sample in the context of within-sentence coordination (§3.4.4), and I argued that non-manual accompaniments to coordination in BSL such as body leans and brow movements are essentially prosodic in nature, having a relationship with grammar rather than being a formal part of it. Bearing the idea of body leans as prosodic in mind, Davidson's description of the COORD-SHIFT lean is somewhat reminiscent of accounts in the cross-linguistic sign language literature concerning non-manual anaphoric reference control mechanisms and other forms of contrast that exploit spatial loci, perhaps more strongly associated with discourse-level coordination (in the sense of reference tracking) than within-sentence coordination of alternatives (Bel, Ortells, & Morgan, 2015; Crasborn & van der Kooij, 2013; Hodge, Ferrara, & Anible, 2019; Morgan, 2000; van der Kooij et al., 2006). There is also some surface resemblance to the iconic and spatial use of demonstrative body turns to represent contrasting viewpoints in what has been termed "constructed action" or formerly "role shift" (Cormier, Smith, & Zwets, 2013; Engberg-Pedersen, 2015; Perniss, 2007). Even more to the point, in the particular context of coordination, we previously noted (§2.4) that left-right body leans have been identified as co-occurring with both list buoy constructions and asyndetic coordination as a form of focal stress in ASL (Wilbur, 2012; Wilbur & Patschke, 1998). There is some speculative support for these general lines of thinking within Davidson's study itself: in the paper's post-experimental reflections on future directions, some similarities between COORD-SHIFT and list buoys

(COORD-L) are explored in terms of their iconicity (the extent of apparent similarity between form and meaning). Both forms exploit the 3D nature of signing space to facilitate a form of referential contrast: the loci of control are the numeric buoys on the non-dominant hand in the case of list buoys, and more generalised spatial loci on relative leftward and rightward sides of signing space in the case of COORD-SHIFT. Crucially however, while list buoys are reasonably well conventionalised in manual form, COORD-SHIFT appears to lack *any* phonological form as the term is usually understood in sign linguistics, given that it is entirely non-manual; we will return to this issue shortly.

Turning to the experimental paradigm, the judgement task trials consisted of a binary choice (accept/reject) sentence-picture verification task, untimed. Participants were shown a series of photographs displaying recognisable household objects on a tabletop in varying configurations. Participants were asked to first examine the picture and to press a key when ready, which triggered a video recording of a person describing the picture (in ASL or spoken English depending on group); for example, “Some of the cans are red” or the ASL equivalent as a quantifier scale stimulus, “A spoon is in the mug or a spoon is in the bowl” or ASL equivalent as a coordinator scale stimulus. Along similar lines to other studies described in this chapter, the description could be a match condition (it clearly described the image and should be accepted as true), a mismatch condition (it clearly did not describe the image and should be rejected as false) or a test condition (it was expected to be rejected in cases where the participant inferred an upper-bounded “pragmatic” reading, and accepted in cases where the literal meaning was inferred). It was not possible to replay the videos: if this was a deliberate design feature rather than a platform limitation, the motivation for it is not given. There were a series of practice items before the main tasks, followed by a reiteration of the instructions and a request for confirmation that the instructions were understood. The main body of tasks consisted of 48 trials, of which 24 were “fillers” (conditions for a different study); the remainder were divided into 12 quantifier trials and 12 coordinator trials; these were each further divided into 4 matches, 4 mismatches, and 4 test conditions.

Statistical analysis of the results consisted of a comparison between separately constructed mixed logit models for the ASL and English group results per sentence type (quantity-based “some” and exclusivity-based “or”). The pattern of responses across the three sentence conditions for each scale type were taken to be broadly similar across groups, with the exception of the ASL group’s test condition rejection rate (representing the assumed rate of scalar implicatures derived) for the coordinator scale, which was significantly lower than in the other three combinations of group and sentence type. To

spell that out, test condition stimuli in ASL that contained a COORD-SHIFT with squint and bit lip were less likely to be taken as “either but not both” than the lexical SOME stimuli were taken as “some but not all”, and less likely than for the English-speaking group with either scale type. The study concludes from this analysis that COORD-SHIFT may not provide a sufficient lexical contrast to reliably “trigger” scalar implicatures, and relates this finding to the theoretical definition of GCIs proposed by Levinson (2000), such that scale members must be in “salient opposition” but also “lexicalised to the same degree” (discussed above throughout §4.1); this position also entails that scalar implicatures are either “triggered” or not (i.e. a binary outcome). The argument here can be interpreted as suggesting that conveying disjunction with a “general use coordinator” form may not provide enough of a salient contrast (i.e. the alternatives are not sufficiently lexicalised and/or contrastive) to satisfy the Levinsonian definition of a GCI, and such a form is therefore intrinsically less likely to “trigger” a scalar inference than a fully lexicalised disjunction might (although this is not tested directly within the ASL stimuli, e.g. via contrast with a lexicalised ASL coordinator, but by comparison to lexicalised English coordinators as interpreted by a separate group of hearing people).

Returning to linguistic issues, the selection of COORD-SHIFT as the “general use coordinator” of interest, rather than (or in addition to) the list buoy COORD-L**,** may be problematic, even by the study’s own account of non-manuals. To recap, the study emphasises the role of the specific non-manual contexts (head nods versus squint plus bit lip) that are said to distinguish a disjunctive COORD-SHIFT reading from a conjunctive one, and suggests that this kind of non-manual distinction is analogous to intonational modulation in spoken languages. But the COORD-SHIFT construction is itself an entirely non-manual body lean: it is not clear why that particular non-manual form justifies being characterised as a coordinator (lexicalised or otherwise) rather than another form of modulation. As mentioned earlier, experiments manipulating contrastive stress in spoken languages (Chevallier et al., 2008; Kurumada & Clark, 2017) have demonstrated that implicatures can be affected by intonational contours, but in such experiments it was always clearly lexical items undergoing the stress. We might hesitate to label an intonational modulation pattern in a spoken language as “a coordinator”; if so, it might be argued that we should not class a purely non-manual form in a sign language as such.

This general question of whether purely non-manual forms can be considered as “lexical” (i.e. effectively treated as “non-handed signs” or “non-manual signs”) is not easy to answer. However, it is fair to say that even highly inclusive usage-based approaches to sign language linguistics (e.g. those that define the properties of a lexical item as

consisting of conventionalised form, function and construed meaning, subject to change over time) are still likely to define the form of “a sign” in terms of a manual phonology first and foremost. In the context of syntax, non-manuals have had their status debated (§2.4) as either prosodic contributions (as opposed to lexical) that may or may not have some kind of semi-conventionalised but non-obligatory relationship with a grammar, in contrast with stricter views that they are analogous to bound morphemes (still not lexical, at best morphosyntactic, but lacking a model of phonological specification). A very rare example of the explicit identification of a number of “non-handed signs” as *free* morphemes was given by Dively (2001), based for the most part on qualitative elicitation interviews with two deaf ASL signers and self-elicitation. However, most of the given candidates have been described elsewhere as emblematic gestures rather than signs, such as nodding the head to indicate affirmation. For example, while describing various functions of head nods in BSL, Sutton-Spence & Woll (1999) do not at any point explicitly identify them as a “sign”, but as a gesture conventionalised for various discourse and grammatical functions and an emblem that signals affirmation. Brennan (1992) does assert that “non-manual signs” are possible in BSL, but that they typically have a discourse interaction function (e.g. affirmation/negation, agreement), and they can be identical to non-manual features that also co-occur with manual signs in a grammatical context (e.g. raised or lowered brows for questioning). In either case, they are also sometimes directly comparable to conventional gestures in use by the British hearing community for the same functions (e.g. nodding for affirmation, brow movements for questioning); furthermore, they “rarely, if ever, occur as content words”. Johnston & Schembri (2007) cite this view directly, but also give two examples of “non-manual signs” in Auslan that (from the description) lack any apparent resemblance to cross-linguistic gestures. They are both potentially taboo (or at least require discretion), which may or may not be a coincidence; at any rate, the phenomenon is characterised as “rare”, which might be an understatement (pending frequency data).

To be clear, it is not being disputed here whether forms produced with the mouth, face, head, body or any non-manual articulator can become functionally and meaningfully conventionalised within a community of sign language users: that is well known (§1.4.2). But whether purely non-manual forms can be considered “lexical”, and whether they can substitute for manual alternatives in any grammatical role, are different and much more specific questions. Pending usage-based evidence and a supporting framework (e.g. a phonological model of “non-manual signs”), it will be assumed for the purposes of the current study that in BSL, a purely non-manual form is highly unlikely to meet the requirements of a lexicalised coordinator. With regard to the relatively infrequent body

leans observed in the current study's BSL Corpus sample (§3.3.2), the description we encountered earlier in the thesis (§2.4) of body leans over ASL coordination as marking focus or stress (Wilbur, 2012; Wilbur & Patschke, 1998) was felt to give a better explanation for the observations made.

We have also previously covered several theoretical and/or evidence-based objections to the Levinsonian "default" model argument cited by Davidson. In particular, the study essentially carries the "homogeneity" and "uniformity" assumptions described in §4.1.3 above (Degen, 2015; van Tiel et al., 2014), whereby lexicalised scales are assumed to be "strength invariant" and "context independent" and beliefs in intended meaning are a binary outcome. Such assumptions lead to an expectation of similar "rates" of implicature between different prototypical scales (<*some, all*> versus <*or, and*>) and also between different participants using different languages (ASL versus English). "Rates" of implicature is a description we have also identified as potentially problematic (§4.2.2). A probabilistic pragmatics account would not necessarily predict such outcomes, expecting or allowing for variation in degrees of belief in intended meaning that interacts with a range of other linguistic and epistemic contexts. Recalling that both ASL and BSL also have coordinator forms available that are more clearly lexicalised (or at least manual), perhaps an argument about the extent to which non-manual modulation and lexical context contribute to salience in scalar implicature derivation could be better made by directly comparing responses to differently composed manual disjunctions with varied non-manual prosodic stress within a single group of sign language users. This is the approach taken for the current project's experimental implicature study in the next chapter.

Beyond the three experiments originally undertaken as sections of Davidson (2011), no other examples of experimental scalar implicature studies of a sign language were located during the literature search for the current study. The only potential exception is Schlenker et al. (2016), a "squib" (a peer-reviewed discussion which is not required to propose any solutions to problems described) which is arguably not experimental (depending on definition). This study never uses the term "scalar implicature" directly, identifying as an exploration of focus marking in ASL and LSF using examples elicited from two of the co-authors, but it includes discussion of exclusivity and exhaustivity inferences from disjunction, phenomena which have been treated as implicature elsewhere. The authors do contrast their arguments with those of a "grammatical" account of scalar implicatures proposed by Chierchia, Fox, & Spector (2011); the latter invokes a covert (i.e. silent/invisible) operator in a generative linguistics framework. Theoretical debates of this kind were not deemed directly relevant to the current project.

4.2.5 Language background issues and grammaticality judgements in BSL

As previously discussed in the sociolinguistic context of BSL (§1.4.1), the transmission of sign languages tends to be more horizontal (peer to peer) than the typically vertical transmission of spoken languages between generations (parent to child). The exact proportion of deaf BSL users who had at least one deaf BSL-using parent or guardian as a language model during infancy is not known, but it is not likely to be much more than 5%, with the large majority being born into hearing families. It is perhaps not surprising that there have been perceptions of prestige and heritage around the deaf family language background minority, who have often been described as “native signers”. However, for some, the term “native” is alienating: it may be uncomfortably associated with elitist concepts of language purity, or colonial attitudes (as in “the natives”), even if only the literal meaning is innocently intended (“originating from birth”). For example, Hou (2024) has argued that binary “native” versus “non-native” labels are inherently Othering and erase the heterogeneity of deaf signing communities. Birkeland et al. (2024), following many others, have argued at some length that linguists have a duty not to avoid or replace such concepts but to openly challenge and interrogate them.

In terms of the problem as it manifests in linguistics, Hou (2024) and Schembri (2024) have discussed “native signer ideologies”, the assumption that experimental or usage-based studies (and therefore the community) are best served by excluding “non-native” deaf sign language users. This may have been done with the well-meaning intention of controlling for effects of later language acquisition or cross-linguistic interference. However, I would argue that the benefits of such a severe limitation (essentially excluding 95% of the community) may be imaginary under the scope of at least some research questions, and perhaps all; in any case, exclusion is not the only means of controlling analysis factors. Schembri (2024) also points out that for the great majority of usage-based analyses of sign language corpora in BSL and Auslan, no significant effects of family language background were found. For BSL, this includes studies of phonological variation (Fenlon, Schembri, Rentelis, & Cormier, 2013); fingerspelling nativisation (Brown & Cormier, 2017); the use of indicating verbs (Fenlon et al., 2018); mouthing (Proctor & Cormier, 2022); sign duration and rates (Börstell et al., 2024); and question forms (Hodge et al., 2025). Only one study of BSL Corpus data has found a significant effect of family language background, in which BSL users from deaf families were found to be more likely to use traditional regional variants of certain signs, explainable as being due to the vertical transmission of such signs from previous generations (Stamp et al., 2014). However, while

the BSL Corpus included participants with both deaf and hearing family language backgrounds and therefore cannot be technically included under the “native signer ideology” umbrella, it did restrict age of BSL acquisition to ages 0-7.

Research which is broadly representative of language acquisition experiences can reveal rich interactions of effects. Cormier, Schembri, Vinson, & Orfanidou (2012) is one such study which explored the effects of age of BSL acquisition, language background, English reading proficiency and non-verbal IQ on grammatical acceptability judgements of BSL sentence stimuli. The theoretical context was the “critical period” hypothesis of language development (Lenneberg, 1967) which holds that there is a critical or sensitive period in childhood for acquisition of a first language (L1), linked to neural plasticity, which was assumed to decrease as an individual grows older. This hypothesis is essentially untestable in the general population since practically all neurotypical hearing children have access to a linguistic environment from birth. Studies of the critical period hypothesis have therefore relied either on extremely rare and isolated cases of language deprivation, or the language acquisition of deaf children with non-signing parents. Questions around grammatical sensitivity have non-trivial relevance to the current project, given that conjunction and disjunction are grammatical phenomena, even though Cormier et al. (2012) did not examine coordination directly. We will therefore examine the methodology and findings briefly here.

Acceptability judgements of manipulated BSL stimuli were elicited from deaf adult participants with a range of family language backgrounds and self-reported ages of BSL acquisition: 10 with deaf parents who reported BSL as L1, and 20 from hearing families who reported exposure to BSL from between 2 and 18 years of age (variously reporting BSL as L1 or L2). All participants reported that they currently preferred to use BSL for everyday communication and had been using BSL for a minimum of 10 years. English reading performance scores and non-verbal IQ scores (based on visual-spatial skills) were assessed in supplementary tasks; the authors justify the benefit of collecting this data as being able to more directly explore L1 BSL effects by factoring out the effects of non-verbal reasoning proficiency, and also by treating high English reading proficiency as a possible alternative L1 (rather than relying entirely on self-reported L1). In the main task, the BSL video stimuli were divided into six grammatical types; each type was presented in both grammatical and ungrammatical constituent orders, with the latter having one constituent moved to an incorrect position (based on existing knowledge and findings of the time around typical BSL constituent order, but as a precaution, stimuli items were further subjected to principled exclusions after data collection was completed). The

response paradigm was a binary accept/reject rating (a red keyboard button for reject, green for accept). Response times were also recorded. Following data collection, principled exclusions were applied to both participants and items. Statistical analyses took the form of various mixed effects regressions with crossed random effects for participants and items, performed separately for response accuracy and response times. Given the large range of possible participant and sentence factor combinations involved, we will not go into the details of the analysis of specific grammatical constructs here (see the paper) but confine ourselves to the main conclusions drawn around effects of language background, age of acquisition, and bilingualism.

For descriptive purposes, participants with a hearing family background were divided into early BSL learners (age of acquisition between 2 and 8 years) and later BSL learners (9 to 18 years). The analysis indicated that those with a deaf family background (age of acquisition zero) were the most accurate overall in identifying both grammatical and ungrammatical BSL sentences. For hearing family early learners, accuracy decreased as age of acquisition increased; but interestingly, for hearing family late learners, accuracy *increased* as age of acquisition increased. An illustrative graph of age of acquisition against accuracy reveals a roughly U-shaped trend, where participants with an intermediate age of acquisition gave the least accurate responses. Towards explaining this, analysis further found that increasing English reading scores correlated with increasing age of BSL acquisition; the authors choose to interpret relatively high English proficiency as evidence towards regarding English as L1 (regardless of self-reported L1).

Given both the U-shaped accuracy curve and the finding of higher English proficiency with later BSL acquisition in this particular sample, two conclusions can be drawn from the analysis: that an early age of BSL acquisition increases BSL grammatical sensitivity the most, but also that for relatively late acquirers of BSL, L1-like English reading proficiency can potentially act as a “scaffold” for increased grammatical sensitivity in BSL, constituting a cross-modal transferrable benefit. The question that remains for the current project is whether facility with pragmatic skills such as scalar implicatures might also be predicted by a pattern of interaction between BSL and English acquisition experiences.

4.3 Conclusions

We have discussed some of the key differences between alternative accounts of scalar implicature, including Gricean and “traditional” accounts (§4.1.1), Relevance Theory and “contextual” accounts (§4.1.2), and “probabilistic pragmatics” accounts (§4.1.3). Experimental evidence has thrown some reasonable doubts on specific “traditional”

accounts (such as the “default” model), while Relevance Theory appears intrinsically quite difficult to test experimentally. In comparison, probabilistic accounts such as the Rational Speech Act framework are relatively new and have considerable explanatory potential, but are still maturing. With particular regard to the current project’s interest in the contextual and cognitive factors that might contribute to scalar implicature derivation in BSL, the key difference in probabilistic accounts is the characterisation of beliefs in intended meaning as a gradient likelihood, rather than the binary outcome assumed by “traditional” accounts. A broad range of contextual contributions (lexical, grammatical, prosodic, epistemic) to the reasoning process are also characterised as gradient strengths, as are any counterbalancing cognitive costs (such as utterance complexity and semantic retrieval).

Regarding experimental design, we audited a number of judgement task paradigms, focusing where possible on experiments looking at the scalar contrast between conjunction and disjunction. In particular, we have seen that the rating scale given to the participants (binary, multiple choice, or continuous) has theoretical implications for the interpretation of results obtained, as well as for statistical sensitivity (§4.2.2). Bearing these findings in mind, the experimental paradigm documented by Jasbi et al. (2019) presented several advantages for a web-based study of exclusivity inferences in BSL. It was built specifically for the conjunction/disjunction contrast, does not have any modality-specific linguistic stimuli restrictions (e.g. no use of orthography), and utilises relatively simple and flexible visual stimuli. It was decided to adapt this paradigm for the BSL experiment that follows in chapter 5, with some necessary modifications (§5.2.2.1). It was further decided that the argument put forward for more granular rating paradigms (rather than a binary treatment or a small number of multiple choice options) and a linking hypothesis which supports the characterisation of belief as a gradient could be taken to its logical conclusion: a continuous scale rating paradigm supported by an appropriate statistical method (§5.1.3, §5.2.3).

We also briefly explored issues around the development of pragmatic competence, given the context of the atypical language acquisition experiences of many deaf children in terms of delays or deprivation. We noted that a number of psychological studies have associated acquisition delays with delays in theory of mind development, and examined related pragmatics discussions around the depth of recursive reasoning required to handle implicatures (§4.1.3.3). In this regard, the main conclusion drawn was that the need for further multidisciplinary study is indicated. We also examined some theoretical and experimental approaches concerned with the difference between adult and child performance with scalar implicatures (§4.2.3). In particular we encountered a “restricted

alternatives" hypothesis (Tieu et al., 2016), which proposes that young children who have learned the semantic meaning of individual scalar terms, and may have begun to internalise the contrast between them, may still derive fewer implicatures than adults because the cognitive cost of retrieving the semantics of the stronger term when presented with the weaker is still relatively high. The implication to be drawn here is that any lasting effects of language acquisition delays might possibly be revealed in an experimental scalar implicature study as patterns of adult response which are reminiscent of early years performance; that said, there are many other factors to be taken into account. In terms of consequences for experimental design, the requirement of controlling the scope of alternatives was highlighted (§4.2.3): the range of visual stimuli should be pre-established as fixed and finite. Furthermore, the trial order of the main task (including a "practice" section) should also be designed to ensure that the stronger scalar alternative (conjunction) is relatively accessible.

We also reviewed evidence (§4.2.5) which suggests that for deaf adult users of BSL (who are typically bilingual in BSL and English to at least some extent), their age of BSL acquisition and English reading age are potentially interacting factors that could partially explain some of the variation in results obtained from linguistic judgement tasks. Given a desire to avoid a problematic "native signer ideology" approach (§4.2.5), it was decided that family language background (deaf or hearing parents) would not be a participant variable. Instead, a decision was made to design the study directly around age of BSL acquisition and English reading accuracy as the participant-level language background factors of interest, in addition to item-level linguistic factors (§5.1.2). Regarding those potential contributions from linguistic context, we also reviewed a proposal that early childhood performance with prosodic contributions to intended meanings may be "scaffolded" by access to increased lexical context (§4.2.3). This has direct relevance to the current project's previous findings (§3.4) that conjunction and disjunction can be expressed in BSL with different levels of lexical context (asyndesis versus gesturally or lexically marked coordination). It was also previously conjectured that certain non-manual contributions to coordination in BSL might be better explained as prosodic rather than morphosyntactic (§3.5). To test these conjectures, it was therefore decided that the linguistic stimuli variables should manipulate levels of both manual coordinator context and non-manual emphasis, and examine their interactions with each other and with the participant-level variables of interest (§5.1.1). It was also decided that, based on the assumptions of a probabilistic pragmatics account, there would be little value in directly comparing the ratings of deaf BSL users and hearing English users, or comparing two

semantically different scales within BSL, given the “homogeneity” and “uniformity” assumptions that have been argued to be present in “traditional” accounts of scalar implicature (§4.1.3.1, §4.2.4). We have also noted the value of setting principled exclusion criteria before data collection begins (§4.2.3, §4.2.5).

Informed by these conclusions, an experimental study of the contributions from linguistic context and language acquisition experience to exclusivity inferences from disjunction in BSL was undertaken, presented in the following chapter.

5 Experimental study: the effects of linguistic context and language acquisition on the strength of exclusivity inferences in BSL

5.1 Introduction

In previous chapters, we have explored conjunction, disjunction, and the realisation of contrastive stress in BSL and other sign languages (§2); conducted a frequency and correspondence analysis of conjunction, disjunction, and non-manual features in BSL Corpus conversational data (§3); and taken a tour of theory and practice in experimental scalar implicature studies, with a focus on those eliciting exclusivity inferences from disjunction (§4). This has all been preparing the ground for the following experimental study of contextual and language background influences on scalar implicatures in BSL. Before getting into the method and results of the study itself, we should briefly recap the key issues and positions that influenced the experimental design, and establish research questions and predictions.

5.1.1 Conjunction, disjunction, and stress in BSL and other sign languages

Earlier in the thesis (§2.4), we encountered a handful of assertions, not always strongly backed by usage-based evidence, along the lines that coordination structures in sign languages may have specific and compulsory non-manual marking, either in identified languages such as ASL and HKSL (Padden, 1988; Tang & Lau, 2012), or perhaps as a briefer and more speculative statement about sign languages in general (Pfau, 2016). By way of contrast, we also covered empirical findings from NGT and RSL studies (Crasborn & van der Kooij, 2013; Kimmelman, 2014, 2019) which leant towards characterising similar features as prosodic and pragmatic markers of focus or contrast, and as optional (or at least highly subject to variation); the conclusions tentatively drawn were that perhaps the precise form of contrastive non-manual marking is less important than how strongly or overtly it is marked. We also discussed Davidson (2011) in some detail (§4.2.4), an experimental study of scalar implicatures in ASL which investigated the potential for a non-manual body lean form with varying facial gestures to act like a lexicalised scalar contrast between conjunction and disjunction without any manual coordination (it apparently did not); however, we further noted descriptions of focal stress in ASL (Wilbur, 2012; Wilbur & Patschke, 1998) which documented the use of a very similar body lean in the focal

marking of disjunctive and adversative coordination (for asyndetic paired items and also list buoy constructions). However, for BSL, no study was found which explicitly makes the case either for or against the morphosyntactic or prosodic nature of non-manual marking of coordination, evidence-based or otherwise. We also discussed (§2.2, §2.3) the relatively small amount of existing documentation of manual conjunction and disjunction forms in the BSL language family (including Auslan and NZSL), covering lexicalised coordinators, asyndetic coordination, list buoys, and the *palm-up* semi-lexicalised/semi-gestural form that has been associated with disjunction (among several other things).

This literature (and gaps in it) informed an exploratory study of conjunction and disjunction in a sample of BSL Corpus conversational data (§3), in which a hierarchical clustering analysis was used to explore the associations between various within-sentence coordination structures and non-manual features. Key findings were that lexicalised BSL conjunctions (e.g. AND, NEXT, PLUS) appear to be much less frequently employed than asyndetic conjunction and list buoy forms; that the lexicalised disjunction OR (a borrowing from English via fingerspelling) is quite rare, and the *palm-up* form is much more likely to be used as a disjunctive coordinator; that it is possible (but relatively infrequent) for list buoy forms to have a disjunctive reading, given sufficient context; and that there was no robust association between either type of coordination and any specific non-manual feature. It was concluded with some confidence from these findings that within the conversational data sample at least, body leans and brow movements co-occurring with coordination in BSL are certainly not compulsory and are therefore unlikely to be morphosyntactic, given the relatively low frequencies and high variation of their occurrences across users. It was conjectured that instead they are more likely to be prosodic and pragmatic markers, most likely of contrast, emphasis or stress, and that (as argued above for NGT and RSL) their precise form is less important than their markedness. If they are in fact a prosodic stress-like contrast, they could contribute to exclusivity inferences in a way that is analogous to the effect of intonational or emphatic stress marking of disjunction in spoken languages (§4.2.3). The following experiment attempts to test that conjecture by eliciting ratings of disjunctive exclusivity, using BSL statements with varying levels of lexical and prosodic context.

5.1.2 Age of BSL acquisition and English reading ability

In the previous chapter (§4.2.5) we briefly described Cormier et al. (2012) as an example of a BSL grammaticality judgement task study which included both BSL and English acquisition factors as predictive variables, uncovering a rich interaction. While those who

acquired BSL from birth made the most accurate BSL grammaticality judgements, it was demonstrated that for later learners of BSL, increased English skills also correlated with increased BSL grammatical sensitivity, approaching the competence of the early acquirers in some cases. This suggested a cross-modal transfer of metalinguistic awareness despite the grammars of BSL and English being different in nature.

It was speculated that a similar general effect might be evident in an experimental pragmatics study such as the one presented here. For that reason, it was decided to explore both age of BSL acquisition (AoA) and English reading accuracy skills (RA) as separate participant variables. The latter requires some form of assessment to be taken as a supplementary task; after researching the limited range of available assessments that are short enough to be practical but long enough to provide a reasonable estimate, the Revised Vernon-Warden Reading Test (Hedderly, 1996) was selected and adjusted (details in §5.2.2.2). In practical terms, BSL AoA could only be self-reported: there was no equivalent BSL assessment method available for healthy BSL-using adults.

5.1.3 Judgement task design

We also examined a number of experimental pragmatics methodologies (§4.2), paying the most attention to those examining exclusivity inferences from disjunction, with the aim of exploring methods adaptable for BSL sentence stimuli. Designs which were reliant on orthography were not an option for a language that has no written form. Sentence/picture judgement tasks seemed to be an appropriate and accessible visual paradigm, but while designs employing symbolic stimuli constructed from geometric shapes and hues were initially attractive, these turned out to be problematic to adapt for BSL (see the picture symbol selection criteria in §5.2.2.1 for more detail). The paradigm judged to be the most immediately useful and accessible was the “playing card guessing game” design demonstrated by Jasbi, Waldon & Degen (2019). This documents a judgement task paradigm specifically designed for eliciting inferences from disjunction which does not rely on orthography, prescribed categories of pictorial symbol, or a specific modality of language stimuli: it was adapted for use with BSL with some modifications. The decision was taken to use a continuous scale for the judgement mechanism, labelled only at the end points with no intermediate markings, rather than any length of discrete Likert scale (i.e. infinite levels, although in practical terms, the granularity is limited by the pixel density of the participant’s screen; for consistency across users, the experimental platform was set up to capture a score of 0-100). This was felt to be the most appropriate rating method, following recommendations that the use of continuous scales may increase sensitivity in

mixed effects modelling analyses of judgement tasks for relatively small sample or effect sizes (Marty et al., 2020); that a low granularity paradigm intrinsically fails to reflect a listener's *degree* of belief in an utterance, given the "probabilistic pragmatics" view (§4.1.3) that any kind of contextual influence (e.g. linguistic, prosodic, epistemic) can potentially contribute differing degrees of effect per utterance and per interlocutor (Degen, 2015; Jasbi et al., 2019); and a personal disinclination to treat ordinal Likert scale data as a linear degree of judgement in any case, on the basis that such scales (particularly those labelled at every interval with adjectives or sentiments) are open to subjective interpretation and are intrinsically unlikely to represent evenly-spaced degrees of judgement (§4.2.2). An appropriate statistical method for handling continuous scale judgement task ratings was then researched and implemented (Bayesian mixed effects modelling with an ordered beta distribution, described in §5.2.3).

There was then the key matter of how the conjunctive and disjunctive BSL sentence stimuli should be constructed. A sentence/picture judgement task paradigm requires the items to have some principled level of balance across all levels of all variables in combination. If too many factor combinations are required then the number of stimuli required for balance and repetition will explode, creating a lengthy and possibly disorienting or tedious task that the participants will struggle to complete with full attention. Therefore, a minimal per-item structure of two factors with two levels each were chosen as dependent test condition variables: sentence structure (less versus more lexical context) and contrastive stress (neutral versus marked non-manuals).

For the sentence structure variable representing two levels of lexical context for coordinated items, asyndesis (less context) and list buoys (more context) were chosen to represent conjunctions, due to being the two most frequently attested forms of conjunction in the BSL Corpus conversational data analysis (§3.3). However, the basis for the decision on disjunctions was not as clearly cut. Asyndetic disjunction appears to be possible in BSL but the exploratory analysis suggested that it seems to be associated with exclusive "either but not both" readings from epistemic context (i.e. it did not seem to be at all open to degrees of exclusivity inference). The *palm-up* form was the most frequently attested manual articulation of disjunction in the sample, and a few examples of *palm-up* within list buoy episodes were also attested, while use of the lexicalised loan sign OR was rare. Therefore, and partly in the interests of symmetry, the decision made was to effectively "add" the *palm-up* form to the asyndetic and list buoy forms of the conjunctive items, producing lexically contrasting disjunctions composed of *palm-up* alone (less context) and list buoys combined with *palm-up* (more context).

For the contrastive stress variable, as previously indicated, the BSL Corpus data did not indicate robust associations between disjunctions and any specific form of non-manual marking. The argument outlined above (§5.1.1) for the prosodic and pragmatic nature of this kind of marking (that the specific form is less important than the extent of its markedness) led to a decision to test that conjecture with as overt a contrast as could still be deemed natural usage (i.e. with due reference to the corpus attestation). The two levels were therefore composed of an “unstressed” absence of non-manual marking (neutral brow position, no body lean, no use of spatial contrast, relatively unmarked amplitude) and an emphatically “stressed” contrast (consistently raised brow movement, clear left-right body lean, spatial left-right contrast of the two coordinated nouns, increased amplitude).

In terms of the balance between test and control conditions in a scalar implicature study, it is necessary to expose the participants to both the least and most informative members of the scale (in this case, disjunctions and conjunctions respectively), so that for test conditions where the strength of implicature from the less informative alternative is being probed, the participant has the more informative alternative in recent memory as a salient contrast (§4.2.3). It was also necessary to include a number of uncoordinated single item sentences: the chosen paradigm unavoidably requires some picture stimuli to display only a single item due to this being the only way of creating certain truth conditions in control items, and so simple (uncoordinated) control condition sentences were also included as further epistemic context, so that there were overt linguistic references to all possible configurations of picture items within the “playing card” pack.

5.1.4 Research questions and predictions

In essence, this experiment is asking whether linguistic factors (levels of lexical and prosodic context) can influence the strength of an exclusivity inference drawn from BSL disjunctions, or whether language acquisition factors are more influential, or whether some kind of interaction between linguistic and acquisition factors will dominate. Given the probabilistic pragmatics view discussed above (whereby a range of different contextual cues can all contribute to the strength of inference made), if the non-manual marking of BSL coordination is in fact a prosodic realisation of contrastive stress, we might expect to see it increasing the strength of scalar implicature derivation from disjunctions that are open to an exclusivity inference (analogous to the way that vocal modulation has been shown to affect strength of similar implicatures in spoken languages). However, it is also possible that the level of lexical or structural context in the marking of disjunction might be

more influential than prosodic context for those with relatively late ages of acquisition in either or both modalities of language (BSL and English).

The statistical hypotheses for the analysis of results are, therefore, that when deaf adults who use BSL as their main or preferred language are given differently constituted disjunctive statements which are open to an exclusivity inference, the strength of scalar implicature derivation (i.e. the degree of receiver belief that the producer does not believe the alternative, a conjunction, to be true) will exhibit a credibly predictable variation with any or all of the following:

- Hypothesis 1: Interactions between the per-item factors of:
    a. the presence/absence of contrastive stress on the coordinated items;
    b. more/less lexical context in the structural form of the disjunction.
- Hypothesis 2: Interactions between the per-participant factors of:
    a. age of first BSL acquisition;
    b. English reading accuracy.
- Hypothesis 3: Interactions between the per-item and per-participant factors above.

5.2 Method

5.2.1 Participants and exclusion criteria

The eligibility criteria for participation required each volunteer to be a deaf adult (aged 18+) user of BSL as their main or preferred language. They were also required to be born and currently resident in the UK, to reduce the chance that their responses might be affected by early acquisition of, or recent immersion in, a different sign language. At the start of the web task, participants were asked to confirm that they met the eligibility criteria and that they consented to their anonymised responses being used and stored for the purposes of the study (see Appendix A for the study information sheet and consent form, including data protection legislation compliance). Demographic information was collected per participant at the end of the web task: age in months, gender, ethnicity, current UK region of residence, and a banded category for age of BSL acquisition (AoA); with the exception of BSL AoA, this information was only used to monitor participant balance and not as predictive variables in the study itself. Participants also had the option to give feedback on their experience of undertaking the study. Response times per trial were recorded, but only for the purpose of verifying attention to the task.

In terms of recruitment, a number of deaf adults who had volunteered as study participants via the DCAL Participant Database were invited directly. The task was also promoted within the UK deaf community via personal and departmental contacts and through social media. A small number of deaf organisations and individuals were later contacted with a polite request to promote the study, mainly targeting regional and ethnic balance. The study was also promoted at a large community outreach event (Deaf Day 2024 at City Lit adult education college in central London). In an attempt to compensate for gender imbalance (a large majority of the early volunteers were female), only male volunteers were enrolled during the latter weeks of the recruitment period. These drives resulted in a total of 67 individuals attempting the online task.

Principled exclusion criteria for participant task data were decided in advance of the study. The majority of the trials in the main task represented a control condition with an expected response relating to a logical truth value (described below in §5.2.2.1). Control items served primarily as salient contrasts to test items (in line with the theoretical motivations of the study), but also functioned as a measure of attention to the task and as a proxy for BSL fluency. A participant entering uninformed guesses, due to a lack of BSL proficiency or any other reason, would tend towards answering the randomly distributed control items at chance levels of accuracy (for example, if the intended response to a control item was at the very high end of the continuous rating scale, an unexpected response would be a rating at the low end, and vice versa). A threshold was set such that achieving less than 75% of expected control condition responses would be grounds for excluding a participant from the study. In addition, any participant whose rushed through the main trials without considering their responses (a mean time spent per trial item below 6 seconds, with the BSL sentence stimuli videos being between 3.5 and 7.0 seconds long, mean 5.6 seconds) was to be deemed completely inattentive and excluded. After these criteria were checked against, 18 individuals were excluded (for near total inattention in 17 cases, for not achieving the control condition accuracy threshold in the remaining case). This left data from 49 participants for analysis. Those who gave the task the required attention (including the one case where the control condition threshold was not met) were recompensed (as advertised) with an electronic gift voucher worth £15 at a range of UK high street retailers for their time (roughly 30-45 minutes for the entire sitting).

5.2.2 Materials and procedure

Following the information, consent, and instruction sections, each sitting had two main tasks: a sentence/picture judgement task (including a brief practice session) designed to

probe the strength of scalar implicature derivation for differently constituted disjunctive BSL sentences, followed by a written English accuracy task. The tasks were constructed with the Gorilla Experiment Builder platform (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). Supporting information was provided as both BSL video (produced by a professional deaf translator) and written English: this included the pre-study information and data protection rights notice, the consent form, and all pre-task instructions (see Appendix A for the English version of the information sheet and consent items, Appendix B for the English version of the task instructions; the BSL versions are viewable in the open materials linked to below). Care was taken not to reveal anything to participants about the study's main variables of interest. The tasks were piloted twice (a technical pilot undertaken with colleagues to assess the platform and design, followed by a trial run with a small group of deaf and hearing participants with BSL skills who were relatively naive to the goals of the study) with resulting minor revisions to the final public version.

The experiment can be viewed online at https://app.gorilla.sc/openmaterials/944777 by clicking the button marked "Preview" next to "BSL Guessing Game Experiment 2024". No responses are recorded in this "dummy" open materials version. To advance to the instructions and the tasks, the consent and eligibility confirmation items on the initial briefing page need to be ticked (even if they do not apply to the previewer).

5.2.2.1 BSL sentence/picture task

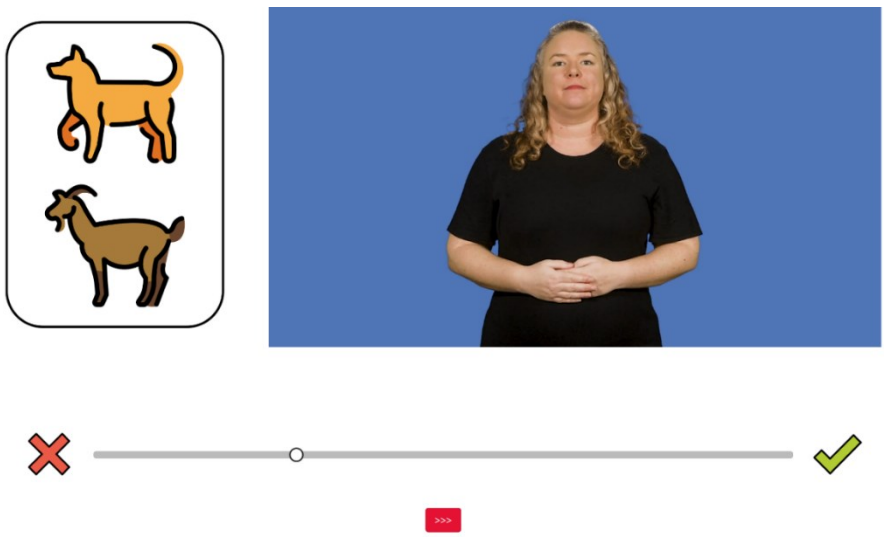


Figure 17: Screenshot of a sentence/picture trial

The main section of the experiment was a web-based sentence/picture judgement task (Figure 17). While the pre-task information was provided in both BSL and English, the main trials deliberately excluded all use of English: the sentence stimuli were only ever displayed in BSL video form and no written words or letters were associated with task

items or screen controls. Participants were instructed that they should imagine themselves playing a guessing game with another person (the "partner"). In each trial, they would be shown an illustration of a "playing card" displaying either one or two simple symbolic pictures representing recognisable items. Next to this picture, they would see a video of the partner signing a "guess": the task instructions reinforced (more than once) that this person had not seen and could not see the card[7]. The partner was described as trying to guess what each card was showing, giving their guess as one short descriptive BSL sentence. The task instructions displayed the full "pack" of cards as a looping animation in a randomised order, in order to give the participant an impression of the entire set of possible combinations (i.e. to establish the scope of possible alternatives). Participants were told that their assignment was to watch each BSL sentence as many times as they wished and then to give a rating of the signer's guess, when compared to the shown card, on a continuous sliding scale between "right" (a green tick) and "wrong" (a red cross) that had no other intermediate markings or labels. The instructions also reinforced to participants (more than once) that they might decide a sentence was neither fully right nor fully wrong, in which case an intermediate position on the slider should be used.

Participants first undertook a short practice session consisting of three trials in the following fixed order: a control item with no coordination that was logically true, a control item featuring a conjunction that was logically false, and a test condition item featuring a disjunction whose interpretation would depend on whether an exclusivity inference was derived (examples to follow). No suggested rating for the examples was given (due to priming concerns) and the responses to these practice items were discarded. The practice session was mainly intended to allow participants to become familiar with the task, but also served the purpose of increasing the retrievability of the contrasting scalar terms by exposing them to all three types of coordination and condition at the very start, to address the rare case where the randomised order of trials in the main task might end up "front-loaded" with any particular condition.

---

[7] This was done in order to avoid violations of "ignorance expectations" associated with the use of disjunction (Jasbi et al., 2019). In a co-operative conversational interaction, use of a disjunction in a description of a world-state typically implies some degree of epistemic uncertainty, i.e. that the actual state is not (fully) known by the speaker; if it were fully known, it would have been described with more certainty, e.g. by using a conjunction. In the Jasbi et al. study, the game's guessing "partner" was depicted as blindfolded, making it impossible for them to see the card. Blindfolding cannot be done for any sign language study, given that the modality of the language is visual-gestural. In the current experiment it would also have obscured a key element of one of the main variables (brow movement). Instead, the study relied on repeated reinforcement of the partner's ignorance in the task instructions.

Participants then began the task proper, consisting of 32 sentence/picture trial items, of which 8 were test conditions and the remaining 24 were salient contrasts and controls. The order of the trials was randomised per participant. Each trial consisted of a short BSL video stimulus (the "sentence") paired with an image of a playing card (the "picture") displaying either one or two pictorial symbols from within one of four categories: smaller animals (DOG/GOAT/FISH), larger animals (HORSE/TIGER/CAMEL), food items (BREAD/CAKE/BANANA), or stationery items (PENCIL/RULER/SCISSORS). The third item in each category was only used in the picture stimuli and was not produced in the BSL sentences (in order to limit the required duration and cost of filming all of the possible noun, stress, and structure combinations). The picture elements were sourced from OpenMoji (license CC BY-SA 4.0, https://openmoji.org/) and were selected with the following criteria:

- Each picture within a category was visually distinctive.
- The BSL noun signs chosen to represent the pictures within a category had a minimum level of phonological contrast, such that at least two phonological parameters differed (handshape, location, orientation, movement).
- The noun signs were sufficiently lexicalised (i.e. core vocabulary, listed in BSL SignBank as corpus-attested and not derived from cross-cultural emblematic gestures, fingerspelling loans from English, or graphical symbols; for example, this ruled out most geometric shapes).
- The noun signs were not phonologically specified for a final body-anchored location, in order to remove body anchoring as a confounding effect[8] (for example, this ruled out CAT and MOUSE, which are anchored on or close to the face).
- There were relatively few or no lexical variations listed in BSL SignBank for the chosen concepts; for example, this ruled out references to colour, which is known to exhibit considerable lexical variation (Stamp et al., 2014).

[8] Body-anchored signs are phonologically specified such that the dominant hand makes physical contact with, or is produced in close proximity to, another part of the body (head, face, torso, arms or the non-dominant hand). If the sign also specifies movement then this contact location may be "initial", i.e. formed at the beginning of the sign and then broken by a movement away, or "final", i.e. beginning at some other location and then moving to make contact. A sign can have both initial and final anchors, either, or neither. If the body is also turning or moving, signs produced in neutral space tend to describe a much larger and obvious spatial arc compared with those that make contact close to the body. It was important to control for this, as body leans/turns were an element of one of the dependent variables (contrastive stress). An exception was made in the case where the non-dominant hand was the final anchored location if (and only if) the non-dominant hand was itself located in neutral space, e.g. BREAD.

The BSL sentence stimuli videos were produced by a deaf-led media company. The professional who modelled the sentences was a deaf user of BSL and a qualified BSL translator; they were consulted for their views on adjusting the production of the sentences with the aim of maximising naturalness of expression, with one important and one lesser change made as a result (the depictive aspect of SHOW across all sentence items, and the specific lexical choice for RULER).

	Coordination	**Stress**	**Coordinator(s)**	**Example BSL gloss** (underline = contrastive stress)
1	None	No	n/a	CARD SHOW DOG
2	Conjunction	No	asyndetic	CARD SHOW DOG GOAT
3	Conjunction	No	list	CARD SHOW LBUOY-1 DOG LBUOY-2 GOAT
4	Disjunction	No	*palm-up*	CARD SHOW DOG *palm-up* GOAT
5	Disjunction	No	list + *palm-up*	CARD SHOW LBUOY-1 DOG *palm-up* LBUOY-2 GOAT
6	Disjunction	Yes	*palm-up*	CARD SHOW <u>DOG *palm-up* GOAT</u>
7	Disjunction	Yes	list + *palm-up*	CARD SHOW <u>LBUOY-1 DOG *palm-up* LBUOY-2 GOAT</u>

Table 9: Sentence stimuli structures

The 7 different sentence structures used across the 32 sentence stimuli are listed in Table 9, using the DOG and GOAT pair as examples. Each sentence offered either one or two items. A quarter of the sentences were not coordinated, i.e. the sentences only featured one item and could only represent control items (sentence type 1). Another quarter represented conjunctions: these were split equally between asyndetic conjunction (sentence type 2) and list buoy constructions (sentence type 3). Conjunctive items could also only represent control items, further serving as salient contrasts (i.e. to ensure that the participant continued to have conjunction, the most informative member of the *<or, and>* scale, in mind as an alternative to disjunction after initially being exposed to it in the practice section). The remaining half of the sentence stimuli represented disjunctions, split equally between use of the *palm-up* form alone (§2.2.2) and a list buoy construction (§2.2.3) combined with *palm-up*. Each of the two forms of disjunction stimuli were further divided into a relatively "neutral" version with no marked non-manual stress (sentence types 4 and 5) and a second version with explicit stress (sentence types 6 and 7) in which the coordinator and coordinates were consistently marked with a spread of upward brow movement, an increased amplitude of signing (larger movements and a more emphatic use of space), an emphatic left-to-right body lean/turn, and anchoring of the two noun signs in contrasting left-right spatial locations. By way of example, the unstressed and stressed versions of the DOG *palm-up* GOAT elements in sentence types 4 and 6 are illustrated in Figure 18 and Figure 19 respectively. In all versions of the sentences, English

mouthings were co-produced with nouns but not with any other signs, including the verb and coordinators (e.g. the disjunctive BSL utterance CARD SHOW DOG *palm-up* GOAT would have mouthings for “card”, “dog” and “goat” only)[9].

Figure 18: Unstressed version of DOG *palm-up* GOAT (neutral brows and no spatial contrast or body lean)

Figure 19: Stressed version of DOG *palm-up* GOAT (brow raise, spatial contrast, and body lean)

			Conjunction		Disjunction		
		Single item	asyndetic	list	palm-up	list + palm-up	Total
Non-stressed	Control (false)	4	2	2	1	1	**10**
	Control (true)	4	2	2	1	1	**10**
	Test condition	-	-	-	2	2	**4**
Stressed	Control (false)	-	-	-	1	1	**2**
	Control (true)	-	-	-	1	1	**2**
	Test condition	-	-	-	2	2	**4**
	Total	**8**	**4**	**4**	**8**	**8**	**32**

Table 10: Weighting of conditions per coordinator and stress variables

[9] The decision to avoid the use of English mouth patterns with the chosen BSL coordinators was partly based on the balance of evidence from the exploration of BSL Corpus data (§3.3.2), but mainly a desire to control for the effects of English influence on readings of the coordinator (given the extent of bilingualism in the BSL community). With regards to the nouns and verb, there is broad evidence in the literature to suggest that mouthings often co-occur with nouns, but the situation with verbs is more complex; for an overview and recent BSL Corpus findings, see Proctor & Cormier (2022).

Table 10 shows how the seven sentence types were balanced across the batch of 32 trials to produce control and test conditions across the key variables of interest. 24 of the trials were control and contrast items, split between no coordination (single item utterances), conjunctions (as salient contrasts) and disjunctions (as relatively unambiguous controls); the remaining 8 trials represented test condition disjunctions and were equally balanced across the dependent variables of sentence structure and stress. As mentioned above, conjunction items could not ever represent test conditions; there was therefore no theoretical motivation to dilute the trial pool with any stressed conjunctions. The four categories of picture elements were also balanced as far as was possible across items and conditions (not tabulated here; for a full list of the 32 trials and their variable combinations, see Appendix C). The overall goal of this spread was to maximise the number of test condition data points per participant in order to gain as much statistical power as possible (given that only a moderately-sized pool of recruits was expected) and to keep the overall task length reasonable. Some previous scalar implicature studies have opted for designs where all conditions and sentence types are perfectly balanced, resulting in situations where the proportion of test conditions within the trials is a sixth (Noveck, 2001) or even an eighth (Chevallier et al., 2008) and the participants spend the large majority of their time responding to control conditions. In such cases, either the task needs to be relatively long (risking participant fatigue and inattentive answers), or relatively few test condition observations are collected per participant (risking statistical insensitivity, especially if the expected pool of recruits is not large). For the current study it was decided that as long as the salient contrasts were sufficiently accessible and every sentence type within the full range of combinations was fully explored above a minimum quota, there was no theoretical requirement for a precisely even balance. The balance chosen therefore allowed for a quarter of the stimuli to represent test conditions while still limiting the entire task to a reasonable length.

Table 11 illustrates some examples of control and test condition picture/sentence combinations. *Control (false)* conditions combined stimuli in such a way that the most reasonable judgement that could be made is that the guess was “wrong”, i.e. it described the card very poorly. Similarly, *control (true)* conditions were intended to be interpreted as appropriately “right”, i.e. the guess could only be taken as describing the card relatively well. Differing responses to *test* conditions, however, were designed to detect whether, and how strongly, a participant had derived an exclusivity inference. Test condition items all took the form of a picture showing items A and B combined with a disjunctive BSL utterance equivalent to “The card shows A or B”, balanced across all four combinations of

lexical context and contrastive stress (sentence types 4-7 in Table 9). If the participant rated a test condition item nearer to “right”, this indicated that they were more likely to interpret the sentence as an inclusive disjunction (“either or both”) and had not derived a scalar implicature; if they rated it nearer to “wrong”, this indicated that they were more likely to read the sentence as representing an exclusive disjunction (“either but not both”) and they had derived a scalar implicature (i.e. that to at least some extent, they believed that that the partner did not intend to mean that the more informative alternative, a conjunction, was true). Analysis of the test condition responses then tested for interactions between responses, the dependent item variables (lexical context and stress), and the independent participant variables (age of BSL acquisition and English reading accuracy).

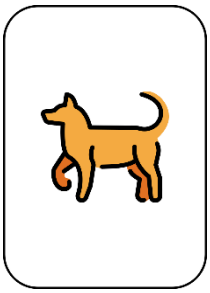
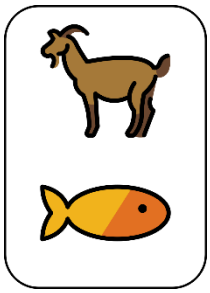
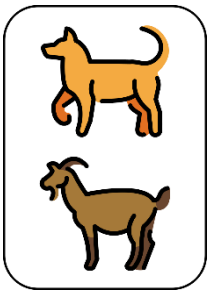
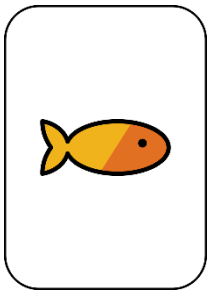
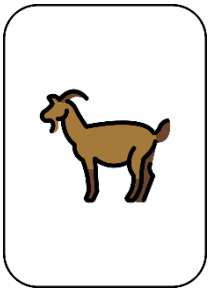
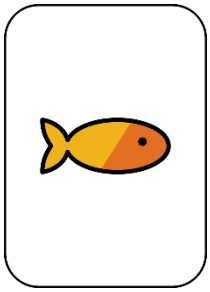
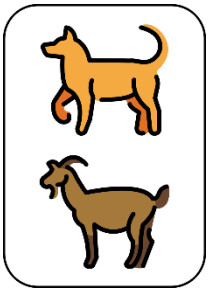
	Control (true)	Control (false)	Test condition
No coordination	 The card shows a dog.	 The card shows a dog.	n/a
Conjunction	 The card shows a dog and a goat.	 The card shows a dog and a goat.	n/a
Disjunction	 The card shows a dog or a goat.	 The card shows a dog or a goat.	 The card shows a dog or a goat.

Table 11: Examples of picture/sentence combinations for control and test conditions

5.2.2.2 English reading accuracy task

As described in §5.1.2, while the only practical way of gathering data about participant age of BSL acquisition was via self-reporting, standardised English reading assessments are more readily available. After completing the main sentence/picture implicature task, participants were tested on their English reading accuracy. The questionnaire selected was an adapted version of the Revised Vernon-Warden Reading Test (Hedderly, 1996), slightly shortened (40 items instead of 42) and with a number of replacements for anachronistic vocabulary (see Appendix D for the English version of the reading task instructions and Appendix E for a list of the adjusted questions and answers). Each item was a "missing word" question with a closed multiple choice response. The subject was asked to fill a blank in each sentence with one of five alternatives, of which only one was deemed correct. If they felt that more than one option was a possible answer, they were asked to select the "best one". The difficulty and complexity of the sentences gradually increased from first to last. One point was scored for every correct answer, and the total then had a normed penalty deducted in proportion to the total number of incorrect answers. The test also allows for answers to be left blank, which are counted as neither correct nor incorrect. The final score can be converted to an approximation of English reading age if desired; this was not done in this case as the group results were instead converted to a discrete two-level variable based on the median score (§5.3.1.3). This banded raw score is referred to here as "reading accuracy" due to the decision not to use the reading age mapping; furthermore, the assessment documentation notes that "the test is not strictly speaking of comprehension but of accuracy" (Hedderly, 1996).

5.2.3 Analysis

Anonymised experimental data was downloaded and stored securely in compliance with the data protection and processing notice given in the pre-study information sheet according to University College London regulations and UK data protection rights (see "Data protection privacy notice" in Appendix A). Following principled exclusions (§5.2.1), descriptive statistics were produced (§5.3.1). The demographic balance of the participants was noted and the participant variables of interest were converted to factors as planned. The judgement task test condition responses were then plotted against variables of interest and inspected for obvious patterns. Any feedback from individual participants that might inform discussion of the results was also noted.

Test condition responses were analysed using Bayesian mixed effects modelling with a weakly informative prior (§5.3.2). An ordered beta regression (Kubinec, 2023) was selected to model the response data, following the reasoning that derivatives of a beta distribution are a better representation of bounded continuous scale responses than methods such as linear mixed effects modelling which assume an unbounded measurement and a normal distribution of residuals. While the “standard” beta distribution is a flexible and useful model for non-normally distributed data mapped to a closed [0,1] continuous scale, it cannot accommodate values at the limits of an open (0,1) scale (values of exactly 0 or 1, both likely to be obtained from a judgement task slider). One approach to this limitation has been to arbitrarily transform the data so that it does not contain boundary values (Smithson & Verkuilen, 2006). An alternative is the zero-one-inflated beta (ZOIB) regression model (Liu & Kong, 2015; Ospina & Ferrari, 2010; Vuorre, 2019), which combines separate analyses of a beta distribution for the continuous [0,1] portion of the data and a Bernoulli distribution for the {0,1} absolute portion. However, Kubinec (2023) argues, with supporting simulation evidence, that the arbitrary transformation approach can affect conclusions unpredictably, while ZOIB models are relatively prone to overfitting due to the requirement of specifying multiple parameter sets. Building on the ZOIB approach, ordered beta regression employs a “cut point” technique to estimate the joint probabilities of lower-bounded values, continuous values, and upper-bounded values, allowing for all outcomes to be included in a single predictive model that is less computationally intensive and less prone to fitting issues than its predecessor.

Inferential analysis was carried out using R Statistical Software 4.3.3 (R Core Team, 2024) and the Bayesian modelling packages *brms* 2.21.0 (Bürkner, 2017), *cmdstanr* 0.7.1 (Gabry, Češnovar, & Johnson, 2023), and *ordbetareg* 0.7.2 (Kubinec, 2023), underpinned by Stan 2.34 (Stan Development Team, 2024). For supplementary statistics, *performance* 0.11.0 (Lüdecke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021) was used to assess model fit and variation explained, *marginaleffects* 0.18.0 (Arel-Bundock, 2024) was used to extract predictions and average slopes, and *coin* 1.4.3 (Hothorn, Van De Wiel, Hornik, & Zeileis, 2008) provided independence tests. Data visualisation was created with *ggplot2* 3.5.1 (Wickham, 2016).

5.3 Results

5.3.1 Descriptive statistics

5.3.1.1 Participants

Table 12 describes the 49 participants by age, gender and ethnic background per region. Balance was not ideal for some factors despite targeted recruitment efforts: female participants outnumbered male by approximately three to two, and non-White ethnic identities were not well represented outside of London. Regional distribution mostly compares adequately to the general UK population spread, with the exceptions of the Midlands (only one participant from West Midlands and none from East Midlands) and Northern Ireland (none). While the full range of ages was 18-68 years, the majority were in the range 30-50, and only one participant was below the age of 28.

Region	n	Age range	Gender		Ethnicity				
			Female	Male	Asian	Black	Mixed	White	Other
East of England	6	18.2 - 68.3	3	3	0	0	0	6	0
London	15	28.4 - 45.5	11	4	2	2	1	9	1
North East	1	35.7 - 35.7	0	1	1	0	0	0	0
North West	5	43.7 - 62.2	5	0	0	0	0	5	0
Scotland	7	29.0 - 54.6	2	5	0	0	0	7	0
South East	6	31.1 - 68.3	5	1	0	0	0	6	0
South West	4	33.4 - 49.2	1	3	0	0	0	4	0
Wales	3	32.4 - 66.6	2	1	0	0	0	3	0
West Midlands	1	36.2 - 36.2	0	1	0	0	0	1	0
Yorkshire & the Humber	1	34.5 - 34.5	1	0	0	0	0	1	0
ALL REGIONS	**49**	**18.2 - 68.3**	**30**	**19**	**3**	**2**	**1**	**42**	**1**

Table 12: Participant demographics

5.3.1.2 Age of BSL acquisition

For the self-reported age of BSL acquisition (AoA) variable, participants originally selected from four bands (“From birth”, “2-8 years old”, “9-18 years old”, “19+ years old”). However, only two participants selected the 19+ category (about 4% of the group), and the “From birth” group was relatively large (about 45%), with the remainder split fairly evenly between 2-8 years and 9-18 years. For the analysis, given this imbalance and that mixed effects models with smaller sample sizes benefit from having fewer levels for discrete categorical factors in any case, the AoA variable was collapsed into two levels by merging the three

smallest groups: “Early AoA” (22 participants, acquiring BSL before age 2) and “Later AoA” (27 participants, acquiring BSL after age 2).

5.3.1.3 Reading accuracy scores

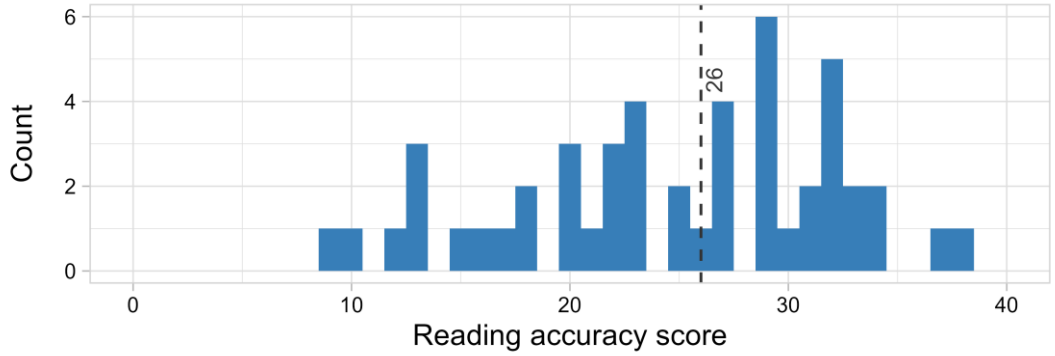


Figure 20: English reading accuracy (RA) scores

Figure 20 shows a histogram of participant English reading accuracy (RA) scores from the secondary task. The median score was 26, described by the assessment as equivalent to a reading age of 16 years (Hedderly, 1996). An approximative Fisher-Pitman permutation test was conducted to assess the difference in reading accuracy scores between the Early AoA (mean score 25.8) and Later AoA (mean score 23.8) groups: the result suggested no significant difference ($Z \approx -0.97$, $p \approx 0.34$, based on 10,000 permutations). As no dependent relationship was indicated between raw reading score and AoA band, the scores were converted into a discrete categorical factor of two levels, using the median as the dividing line. This produced a “Higher RA” group (24 participants) and a “Lower RA” group (25 participants). The two-by-two split of participants across the AoA and English RA groups is shown in Table 13.

	Higher RA	**Lower RA**	**TOTAL**
Early AoA	12	10	**22**
Later AoA	12	15	**27**
TOTAL	**24**	**25**	**49**

Table 13: Participant counts, split by age of BSL acquisition (AoA) and English reading accuracy (RA)

5.3.1.4 Test condition responses

All of the descriptive plots in this section are laid out horizontally to reflect the layout of the rating slider used by participants (Figure 17), with values to the right of the scale representing degrees of “right” judgement (suggesting an inclusive reading), and values to

the left of the scale representing degrees of “wrong” judgement (suggesting an exclusive reading). A violin plot is used to approximate the relative density of the distribution of responses. An overlaid box plot indicates the median response (the central thicker black line) and the first and third quartiles (the box). The mean is indicated with a cross.

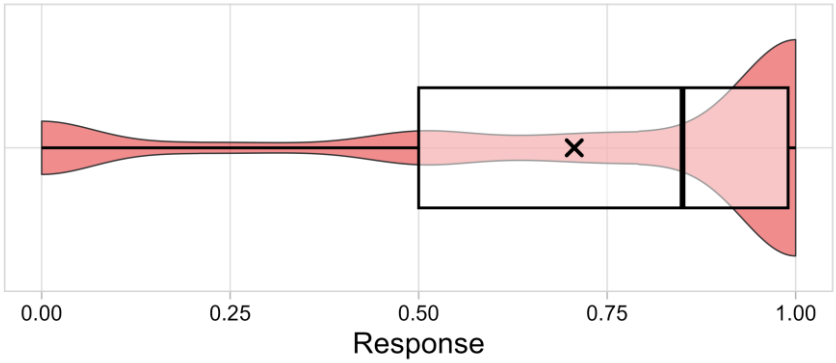


Figure 21: Test condition responses (all)

Figure 21 shows the general distribution of all test responses from all participants. The median value is towards the high end of the scale, and the first quartile is very close to the mid-point, indicating that three quarters of all responses were read as more “right” than not (representing some level of belief in an inclusive reading), and half quite strongly so. However, a quarter of the responses are below the mid-point, and there is a cluster of “hard wrong” responses at the low end of the scale, suggesting that strongly exclusive readings are being derived by some participants some of the time.

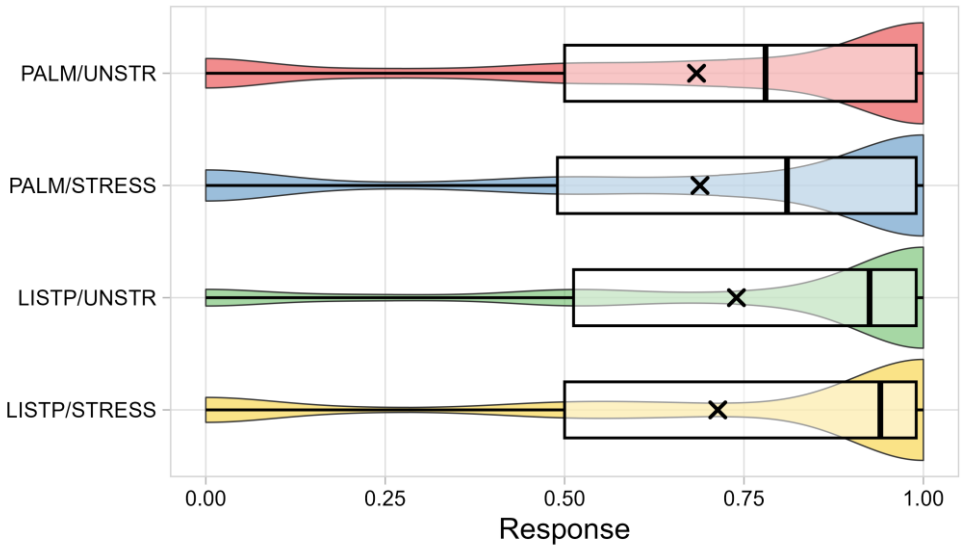


Figure 22: Test condition responses (lexical context × stress)

Figure 22 breaks the test condition responses down by the item factors of lexical context and contrastive stress, ignoring participant factors. The distributions all have a

reasonably similar pattern to the overall distribution. For both PALM-UP and LIST + PALM-UP, there are only small visual differences between the unstressed and stressed versions, suggesting that contrastive stress does not have much of an overall effect if participant factors are not taken into account. Comparing the two structures to each other, the LIST + PALM-UP versions are slightly more skewed towards fully "right" responses as indicated by a shift in the median response, indicating that this construction may be more likely than PALM-UP alone to be read as inclusive. However, the first and third quartile ranges are approximately the same across the combinations, suggesting that the proportions of inclusive readings are similar when averaged across the entire cohort.

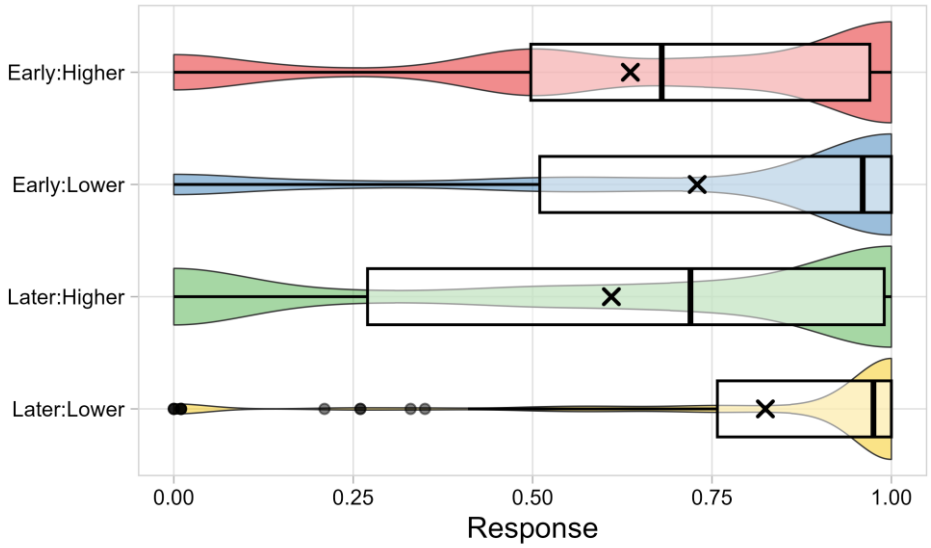


Figure 23: Test condition responses (BSL AoA × English RA)

Figure 23 breaks down the responses by participant age of BSL acquisition (AoA) and English reading accuracy (RA), ignoring item factors. There are much clearer differences between the participant groups than for the item breakdown. For the two Early AoA groups, the Higher RA group (top) makes a pronounced use of the mid-point in the rating scale, suggesting that this group quite often tended to rate items as fairly ambiguous ("neither right nor wrong"); this group might be characterised as having the most "language experience", and their apparent preference for mid-range responses suggests they are the group most likely to recognise that the sentences are inherently ambiguous. In contrast, the Later AoA and Higher RA group (second from bottom) has the broadest interquartile range of responses and has the highest relative density of "hard wrong" ratings, suggesting that they are the group most polarised in their responses. The Lower RA groups appear to favour "right" (i.e. inclusive) readings, especially the Later AoA and Lower RA group (bottom), for whom three quarters of responses fall within the top quarter

of the rating scale, and half of the responses are at the hard upper limit; there are only a few outlying observations of “hard wrong” responses at the lowest end of the scale, indicating that this group only weakly made exclusivity inferences on the whole.

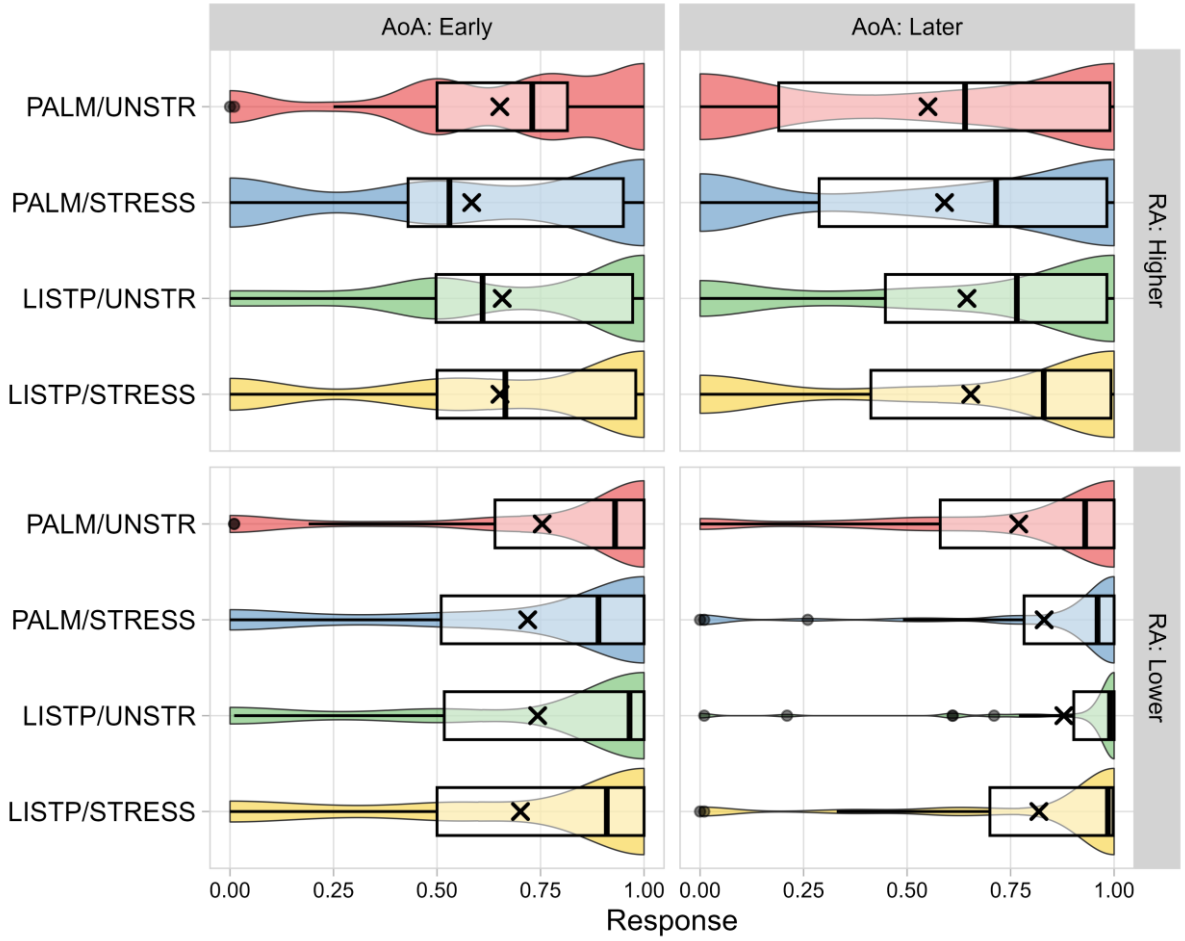


Figure 24: Test condition responses (AoA × reading accuracy × structure × stress)

Finally, Figure 24 breaks down the test condition responses by both item and participant variables. Visual observations are as follows:

- The early AoA and higher RA group (top left) has a lower mean rating (indicating a shift towards a less inclusive reading) in stressed PALM-UP sentences compared to unstressed, and in visual terms the central “lump” around the mid-point appears to shift to the bottom of the rating scale, possibly indicating an effect of stress there; but the effect of stress is less clear for the LIST + PALM-UP sentences.
- The early AoA and lower RA group (bottom left) may also have a tendency towards slightly lower ratings for the stressed version of the stimuli, this time for both sentence structures, but the difference is subtle.
- The later AoA and higher RA group (top right) again displays the most polarised distributions of the four participant groups and appears to be the most likely to give

the lowest ratings across all four sentence conditions. There is only a subtle difference between contrastive stress levels (but in the opposite direction, towards a *more* inclusive reading). The LIST + PALM-UP sentences appear to have elicited slightly more inclusive readings generally than PALM-UP**.**

- In contrast, as expected from the previous plot, the later AoA and lower RA group (bottom right) has the most strongly inclusive ratings overall. Some visual differences in mean and range of responses are suggested between the sentence variants, but the overall pattern is not particularly clear from visual inspection alone.

A mixed effects modelling analysis was then performed to determine whether or not any of these visual differences are borne out statistically when individual responses to the task and repeated measures are accounted for, and whether other less obvious trends were detectable and credible.

5.3.2 Quantitative analysis

5.3.2.1 Summary

A Bayesian mixed effects model based on an ordered beta distribution (see §5.2.3 for rationale) with a weakly informative prior (package default) was constructed to predict the effect of interactions between the per-item categorical variables (lexical context and contrastive stress) and the per-participant categorical variables (age of BSL acquisition and English reading accuracy) on the test condition responses. The following random effects justified by the design were included: random intercepts per participant, and random by-participant slopes for contrastive stress, lexical context, age of acquisition, English reading accuracy, and all interactions. Four chains converged after 2000 iterations each (1000 warm-up).

An initial summary of the model's fixed effects was examined (Table 14), mostly for diagnostic purposes: the estimates in this summary are in terms of the log odds of the responses, which are not intuitive to read, and they are expressed in terms of contrast with the intercept only (the default value of all four variables in the model). To explore interactions in more detail, pairwise marginal effects were calculated (sections to follow). Rhat values were very close to 1 in all cases, indicating that the model converged successfully. The effective sample size (ESS) values of well over 1000 in all cases suggested a robust estimation of the posterior distribution.

A Bayesian approximation of conditional R^2 for the fitted model was estimated at 0.79, suggesting that the model explains approximately 79% of the variation in the data

when both fixed and random effects are taken into account (a good amount for a social sciences study); marginal R^2 was estimated at 0.22, suggesting that the model accounts for approximately 22% of the variation when only fixed effects are accounted for. The relatively large difference between these two estimates suggests that individual participant reactions to the task are an important contribution to the model.

Term	Estimate	Std. error	CI low	CI high	Rhat	ESS
(Intercept)	1.39	0.37	0.68	2.14	1.00	3069
stress	-0.08	0.14	-0.35	0.19	1.00	3768
structure	0.13	0.15	-0.15	0.42	1.00	3796
AoA	0.06	0.52	-0.96	1.09	1.00	2842
RA	1.21	0.53	0.16	2.29	1.00	3335
stress:structure	-0.13	0.15	-0.44	0.16	1.00	4051
stress:AoA	0.32	0.20	-0.06	0.72	1.00	4063
structure:AoA	0.22	0.20	-0.16	0.63	1.00	3538
stress:RA	0.00	0.20	-0.41	0.39	1.00	3629
structure:RA	-0.31	0.21	-0.73	0.08	1.00	3736
AoA:RA	0.47	0.74	-1.03	1.97	1.00	3173
stress:structure:AoA	-0.26	0.21	-0.67	0.15	1.00	4368
stress:structure:RA	-0.27	0.21	-0.70	0.13	1.00	4576
stress:AoA:RA	0.09	0.27	-0.44	0.63	1.00	3428
structure:AoA:RA	0.08	0.29	-0.49	0.64	1.00	3334
stress:structure:AoA:RA	-0.05	0.30	-0.63	0.55	1.00	4182

Table 14: Summary of fixed effects (contrasts with intercept) and model fit diagnostics

Addressing each of the statistical hypotheses given in §5.1.4, the following sections describe the direction and magnitude of group average marginal effects in the data, in terms of slopes (contrasts between values of a single variable) and the strength and reliability of evidence they suggest. The reported slopes below are in terms of the original response scale, where a value of absolute magnitude 1.0 is equivalent to the difference between a fully “wrong” and a fully “right” rating at the extreme ends of the rating scale. A slope is usually considered to provide evidence of a reliable effect (marked with an asterisk) if its credible interval (indicated by the line-and-whiskers) does not include zero (marked by the dotted vertical line). The credible intervals used in this case are 90% highest density intervals. To avoid troubling the reader with several pages of numbers, the effects are shown using visual plots, and only the statistically credible estimates are quantified in the accompanying descriptions. For the full tables of the group average marginal effects including all estimates and credible intervals, see Appendix F.

5.3.2.2 Marginal effects of item factors

Relating to hypothesis 1, Figure 25 describes the group average marginal effects of the item variables in terms of the slope between levels of contrastive stress for each level of lexical context, and vice versa, for the cohort as a whole (i.e. averaged across all participant variables). The left hand plot can be read as the average effects of "adding" stress to each of the sentence structures. There is no credible effect for PALM-UP sentences, but there is for LIST + PALM-UP: the average slope is -0.03, suggesting that the stressed version is likely to be read as fractionally more exclusive than the unstressed version by the entire participant group on average. The right hand plot can be read as the average effects of "adding" a list buoy to the coordination structure, for each level of contrastive stress. There is no credible effect for stressed sentences, but there is for unstressed sentences: the average slope is +0.06, suggesting that unstressed disjunctions with a LIST + PALM-UP construction are likely to be understood as slightly more inclusive than unstressed disjunctions with PALM-UP alone by the entire participant group on average. However, in both cases, the effect sizes are clearly very small when participant variables are not taken into account.

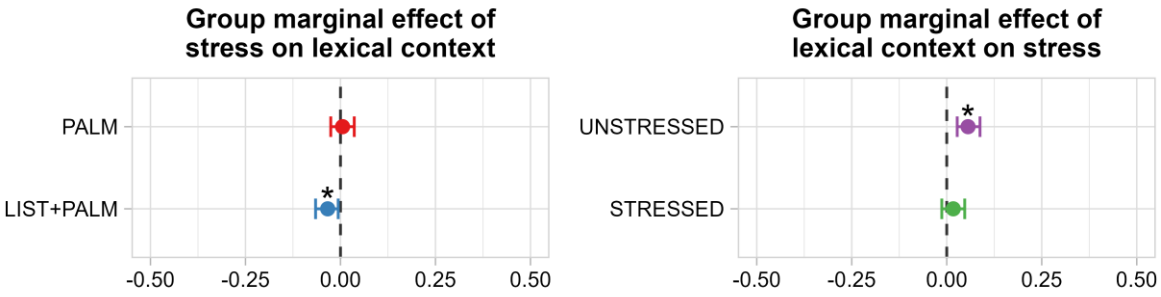


Figure 25: Group average marginal effects of item variables

5.3.2.3 Marginal effects of participant factors

Relating to hypothesis 2, Figure 26 describes the group average marginal effects of the participant variables in terms of the slope between levels of BSL AoA for each level of English RA, and vice versa, for the test condition stimuli as a whole (i.e. averaged across all sentence variables). The left hand plot can be read as the average effect of acquiring BSL later in comparison to earlier, for each level of English RA; no credible effect is indicated. The right hand plot can be read as the effects of having higher English RA in comparison to lower, for each level of BSL AoA. There is no credible effect for the Early AoA group, but there is for Later AoA: the average slope is -0.24 and the credible interval is fairly broad, suggesting that later BSL acquirers with higher English reading scores are

likely to rate any of the given forms of disjunction as substantially less inclusive than those with lower English reading scores.

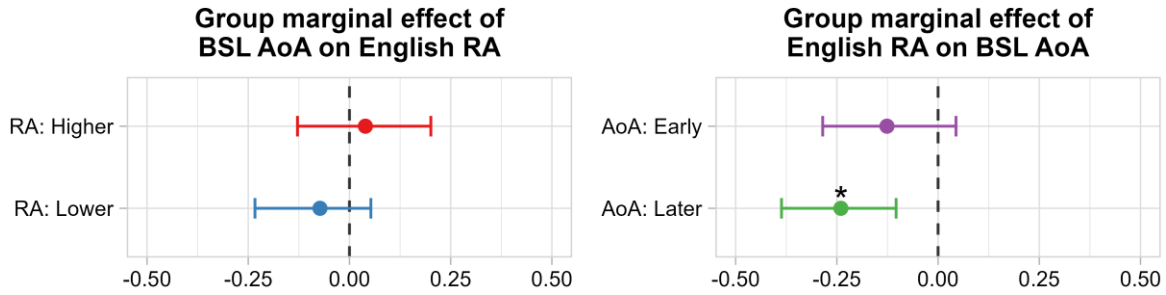


Figure 26: Group average marginal effects of participant variables

5.3.2.4 Marginal effects of factor interactions

Relating to hypothesis 3, we will now examine the group average marginal effects of the interactions between item and participant factors, in terms of the contrasts between levels of contrastive stress (Figure 27), lexical context (Figure 28), and English reading accuracy (Figure 29); the corresponding BSL age of acquisition contrasts are not shown as no credible effects were found (see Appendix F for full tables). While some of these interactions echo the averaged item and participant group contrasts analysed above, there are also specific interactions that differ.

Figure 27 can be read as the effects of "adding" contrastive stress to each level of lexical context, broken down by the interactions between BSL AoA and English RA. There is a credible effect for participants in the Early AoA and Higher RA group (top left) for sentences featuring PALM-UP alone: the average slope is -0.08, suggesting that this group is likely to rate stressed PALM-UP disjunctions as slightly less inclusive on average than unstressed. For the later AoA and lower RA group (bottom right), there are credible effects for both levels of lexical context, but in opposite directions. For PALM-UP disjunctions, the average slope is +0.08, suggesting at face value that the effect of added stress appears to make disjunctions slightly more inclusive for this group, but for LIST + PALM-UP, the average slope is -0.06, suggesting that the effect is to make them slightly less inclusive. For the other two group combinations (both with mixed levels of language experience), there were no credible effects: it is important to bear in mind that this does not mean that they were not affected by contrastive stress, just that they were not affected to a statistically credible *different* extent when compared to the entire participant group on average.

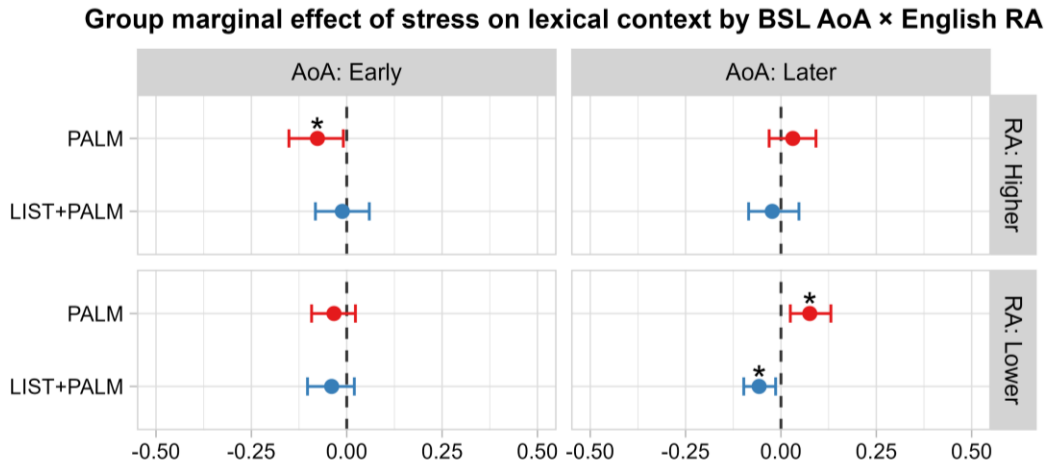


Figure 27: Group marginal effect of stress on lexical context by BSL AoA × English RA

Figure 28 can be read as the effects of "adding" a list buoy to the sentence structure for each level of contrastive stress, broken down by the interactions between BSL AoA and English RA. For those with Early AoA and Higher RA (top left), there is no credible difference for unstressed sentences, but stressed LIST + PALM-UP items were read as more inclusive than stressed PALM-UP (average slope +0.08). For those with Later AoA and Higher RA (top right), both levels of stress show a contrast towards more inclusive readings (average slope +0.12 for unstressed, +0.07 for stressed). Those with Early AoA and Lower RA show no credible effects, but for those with Later AoA and Lower RA (bottom right), both levels of stress show a credible difference from lexical context, again in opposite directions (average slope +0.08 for unstressed, -0.05 for stressed).

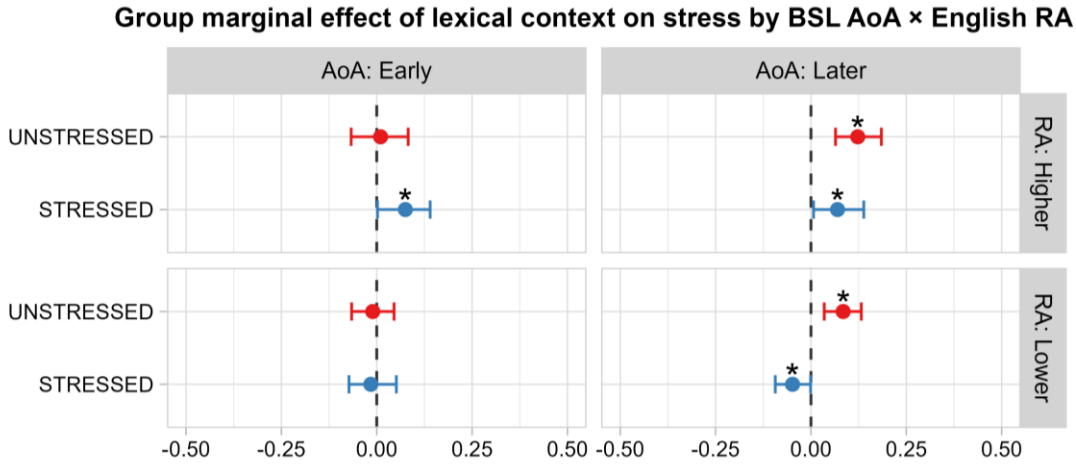


Figure 28: Group marginal effect of lexical context on stress by BSL AoA × English RA

Finally, Figure 29 can be read as the effect of having a higher English RA score in comparison to lower, for each level of BSL AoA, broken down by the interaction between lexical context and contrastive stress. These are generally the largest pairwise effects but they also have the broadest credible intervals, suggesting greater variation. For

unstressed PALM-UP items (top left), there is no credible effect for those with Early AoA, but for those with Later AoA, those with higher English RA gave much less inclusive ratings in comparison with lower RA (average slope -0.26). For stressed PALM-UP items (top right), both AoA groups showed a similar effect of having a higher RA (average slopes -0.18 for Early AoA, -0.30 for later AoA). However, for both unstressed and stressed LIST + PALM-UP items (bottom left and right), those with Early AoA showed no effect of English RA, while those with Later AoA showed the same effect of higher RA towards less inclusive readings (average slopes -0.22 for unstressed LIST + PALM-UP, -0.18 for stressed LIST + PALM-UP).

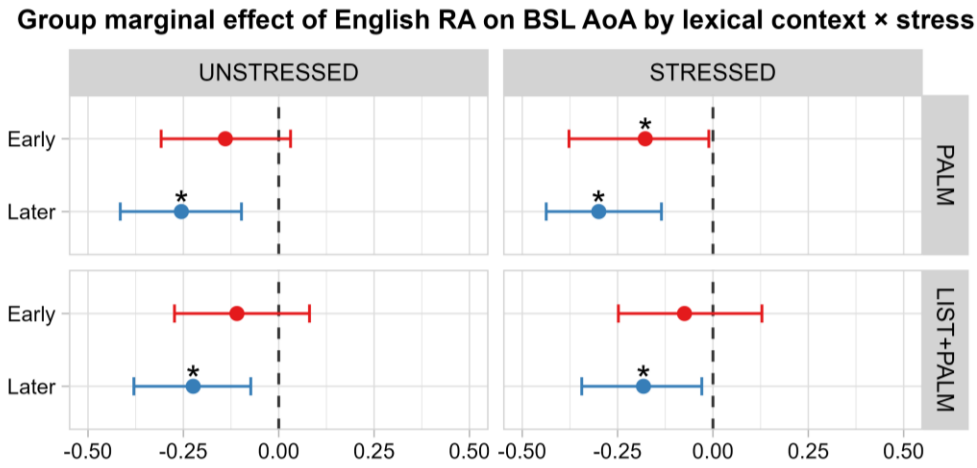


Figure 29: Group marginal effect of English RA on BSL AoA by lexical context × stress

We now turn to a discussion of implications drawn from these results and try to account for the patterns observed.

5.4 Discussion

5.4.1 Research questions addressed

The broadest possible headline from this experiment is that some users of BSL strongly derive a scalar implicature from a pragmatically infelicitous disjunctive construction some of the time. Although this may not be especially surprising to anyone who regularly converses in BSL, experimental evidence is important as no previous scalar implicature studies of BSL are known. However, more importantly, analysis of the results suggests that variations in the strength of exclusivity inferences can be statistically associated with particular combinations of linguistic context and language background.

The stated statistical hypotheses (§5.1.4) were that the strength of scalar implicature derivation suggested by exclusivity ratings of ambiguous disjunctive statements would vary for BSL users in predictable ways: by sentence factors such as the lexical context of the

disjunction's expression (use of PALM-UP with and without an accompanying list buoy) and/or the presence or absence of non-manual marking conjectured to function as contrastive stress (hypothesis 1); and/or by participant factors such as age of first BSL acquisition and/or English reading accuracy (hypothesis 2); and/or by interactions between the sentence and participant factors (hypothesis 3). The quantitative analysis above provides some evidence to support the first two hypotheses, in a technical sense: there were indeed some credible variations per sentence factor interaction regardless of participant factors, and one credible source of variation for a specific participant factor combination regardless of sentence, although the size of the sentence effects is small when participant factor interactions are not accounted for (giving them the opportunity to cancel each other out). However, the pattern of interactions between sentence and participant (hypothesis 3) tells a different and much more interesting story.

In the previous chapter, we cited experimental studies (Chevallier et al., 2008; Kurumada & Clark, 2017) which collected evidence towards demonstrating something that many non-linguists will intuitively grasp right away, if you give them concrete examples: that contrastive stress (intonational emphasis for specific spoken languages, and typographical emphasis for their written form) can have a detectable and reliable influence on implicature strength (e.g. for exclusivity inferences, that "There is a dog or a goat" might be less likely to be read as exclusive than "There is a dog *or* a goat"). In the studies cited, this was experimentally demonstrated to be the case for exclusivity inferences by adults and a type of ad hoc implicature by children respectively. In both studies, the presence of stress on key sentence constituents tended to produce a higher rate of "pragmatic" responses (i.e. rejections of the unstated alternative) in comparison to the absence of such stress, although in the children's case, they made the most rejections when they also had access to additional discourse context as well as the intonational emphasis. The number one concern of this thesis was how such a phenomenon might play out in BSL.

The model fitted to the data collected in this study also suggests an effect of contrastive stress away from a fully inclusive reading in BSL (expressed in this case via non-manual marking, increased amplitude of signing, and left-right spatial contrast), but only for participants with specific language backgrounds and in specific lexical contexts (Figure 27). Those who both began to acquire BSL early in life and also scored higher in English reading accuracy (i.e. those who might be understood as having the most language experience within the sample) did show the expected effect of additional contrast away from an inclusive reading, but only for a disjunction composed of PALM-UP alone (i.e. with less lexical context). In contrast, the later AoA and lower RA group (i.e. those who

might be said to approximate the least developed language background in the sample) showed a different pattern: for PALM-UP alone, contrastive stress reliably tended towards the opposite effect, a *more* inclusive reading, while for LIST + PALM-UP (i.e. with more lexical context), the expected effect away from an inclusive reading was again demonstrated.

In accounting for this group difference, we might consider the possibility that those participants without lifelong BSL experience and below-cohort-average English reading skills were relatively insensitive to the presence of PALM-UP alone (i.e. without any additional coordinating lexical context) and in at least some cases may not have interpreted it as a disjunction at all: in the descriptive statistics, the group's test condition results are clearly skewed towards very inclusive ratings generally regardless of the utterance composition (Figure 23, bottom right quadrant). In and of itself, this descriptive statistic cannot distinguish between a group tendency towards very inclusive readings ("either or both") and the alternative possibility that they were interpreting the disjunctions as conjunction ("both"). There have been some thematically comparable spoken language studies which suggest that children still in the critical period of language acquisition have a tendency to interpret disjunction as conjunction in some linguistic contexts, but also lean towards deriving inclusive interpretations of disjunction in other specific contexts (Pagliarini, Bill, Romoli, Tieu, & Crain, 2018; Singh, Wexler, Astle-Rahim, Deepthi Kamawar, & Fox, 2016). As previously described (§2.2.2) and demonstrated (§3.4), PALM-UP is corpus-attested for a disjunctive function in BSL, but the extent to which it is *lexicalised* as a coordinator is difficult to define; the form is also associated with a fairly large number of more and less related discourse marking functions (e.g. uncertainty marking, turn yielding). If a hypothetical BSL-using individual were not to interpret PALM-UP without further lexical context as a disjunction ("dog or goat"), it might effectively be read as asyndetic conjunction ("dog, goat"); in which case the effect of additional stress-like modulation might be to highlight and strengthen the conjunctive interpretation instead ("dog ***and*** goat"). If this were the case, it would account for the high frequency of very strongly inclusive ratings generally observed for that combination of participant and sentence factors, and also for the effect towards *more* inclusive readings for stressed PALM-UP without a list buoy. However, the same group *does* show the expected contrast for added non-manual stress (tending away from a fully inclusive reading to at least some extent), with an effect of moderate size, when there is also additional lexical context (both PALM-UP and a list buoy present as coordinators; see Figure 27 and Figure 28, bottom right quadrants). This suggests that given additional lexical context, a statistically detectable proportion of the members of this less language-experienced group can and do distinguish

between conjunction and disjunction, and also that they are sensitive to exclusivity inferences to at least some extent, i.e. a level of belief in producer intent, but that in this regard they generally benefit from a combination of increased lexical and prosodic context. Furthermore, all of the participants in the inferential analysis successfully passed the control condition exclusion criterion (responding as expected to both conjunctive and disjunctive controls), which further suggests that difficulties with identifying the semantic meanings of conjunction and disjunction are not the main issue.

The proposed explanation for this interaction between marked contrastive stress and additional lexical context is that the additional visual-structural contrast afforded by a coordinating list buoy, in combination with the spatial contrast produced by the contrastive stress and a reading of PALM-UP in its non-lexicalised uncertainty-marking role, makes the stronger scalar alternative clearer, enabling the possibility of an exclusivity inference. This would further suggest that increased sensitivity to the disjunctive coordination potential of PALM-UP *without* additional coordinative structure, as well as the general role of contrastive prosody in distinguishing between inclusive and exclusive intention, are abilities that come with additional language experience, certainly from early exposure to adult-level BSL, but otherwise perhaps also from the successful acquisition of spoken or written English (since neither the Later AoA group with Higher RA nor the Early AoA group with Lower RA show any credible differences from the cohort-average effects from levels of lexical context; see Figure 28, upper right and lower left quadrants). Furthermore, it supports an argument that pending the required language experience in either modality, the development of competence in interpreting contrastive non-manual stress on disjunctions in BSL can be "scaffolded" by additional lexical context, perhaps leading over time to increased and more nuanced awareness of the communicative intent of contrastive stress such that the extra context is no longer required. This would echo the conclusions made in the context of hearing children and stressed ad hoc scales by Kurumada & Clark (2017), described earlier, that the communicative intent of contrastive stress is learned during language acquisition via supporting context (in our case, that an exclusive disjunction is intended).

Tugging further on that particular thread by seeking confirmation from the general contrasts between language backgrounds across all sentence types (Figure 29), we find several credible effects for the contrast between higher and lower English reading accuracy, and they are relatively large effects in comparison to any of the sentence factor contrasts. There is a clear tendency towards less inclusive readings (i.e. more strongly "pragmatic" judgements) across all four sentence types for those with later BSL age of acquisition but higher English reading accuracy in comparison to lower, while those with

early BSL acquisition only show a credible relative effect of higher English reading accuracy for stressed sentences with minimal lexical context. This finding aligns with the argument above that the development of this particular form of pragmatic reasoning is dependent on, or at least strongly supported by, early access to language in either or both modalities. That said, the reading accuracy score does not actually tell us whether or not English acquisition began at an early age: we can only infer increased English experience from the accuracy test taken as part of the experiment.

Given the above, the analysis supports hypothesis 3, such that these findings suggest credible evidence that the strength of exclusivity inferences from disjunction by deaf users of BSL is influenced by specific interactions between language experience and linguistic expression. The language background of the individual seems to have the stronger influence in general, but the strength and (importantly) direction of effects towards or away from an inclusive reading can also depend on the lexical and prosodic context of the utterance. In general terms, early acquirers of BSL showed stronger and more reliable effects towards exclusive interpretations from additional prosodic contrast in minimal lexical contexts, and later acquirers seemed more dependent on the presence of additional lexical context in drawing inferences. Furthermore, there is some evidence to suggest that given either sufficient language acquisition or additional lexical context, contrastive non-manual emphasis does appear to function just as contrastive intonational stress has been shown to do in spoken languages (shifting the interpretation at least some way towards an exclusive reading), providing further evidence to that gathered by this project's earlier BSL Corpus analysis towards an assertion that such non-manuals are not well characterised as obligatory or morphosyntactic. This is not to say that they are "not grammatical" as such – they have a relationship with syntax and the derivation of communicative intent – but that this derivation is a pragmatic process and that this kind of non-manual contrast in this specific context is better characterised as a form of prosodic modulation.

All of this said, however, a caveat or two: these trends are tendencies within a sample, not absolutes. Individual reactions to the task were fairly diverse: this is indicated by the fact that while all effects highlighted above are to some extent credible (in the sense that they have a level of statistical predictability), they are based on means and are not often of very great magnitude when averaged across groups (with the exception of some larger effects from having higher English reading accuracy in the absence of early BSL acquisition). The difference between the contributions of fixed and random effects to variance explained (§5.3.2.1) indicates that individual differences accounted for a notable amount of the variation in responses, suggesting that any individual's language experience

cannot necessarily be defined by the group labels and effects that this analysis applies, and that there may be unexamined variables that have an influence. This is in fact very likely, since age of acquisition bands and generalised reading accuracy scores give very little *qualitative* description of lifelong experiences of linguistic and social environments. However, it was felt that the judgement task design (particularly the use of a continuous rating slider with only end-point labels) and the statistical method it afforded (a Bayesian mixed effects modelling analysis of item and participant variable interactions, based on an ordered beta regression) did prove to be very effective in identifying credible trends within the variation observed; a less granular rating paradigm with the same sample size might have missed some of these details. Finally, it should also be made absolutely clear that no definitive claim is being made here about pragmatic competencies in any segment of the general population of BSL users: this was a contrived non-conversational experimental setting that elicited responses from a self-selected sample, and the focus was on the *relative* effects of variables, not absolute determinations of scalar implicature derivation frequency. Interpretations of these findings must bear that in mind.

5.4.2 Supported accounts of scalar implicature, and future directions

In the previous chapter (§4.1), we described some historical and contemporary accounts of scalar implicature, often focusing on differences between earlier “traditional” accounts, particularly the “default” model, and more recent “probabilistic pragmatics” theories such as the Rational Speech Act framework (RSA). To recap very briefly, the traditional accounts tend to place more emphasis on the sequential order of the interpretation of pragmatic and literal meanings (one happens before the other, depending on the specific theory); the “generalisability” of highly lexicalised scales in terms of the retrievability of their semantic contrasts; and the characterisation of implicatures as a strength-invariant outcome related to binary logic (made or not). “Probabilistic” accounts, on the other hand, reject “code models” of literal and pragmatic meanings being derived in any particular order; describe the accessibility of alternative scalar terms as being just one of the balancing costs against the informativity of an utterance; characterise beliefs in intended meanings and the effects of contextual influences in terms of likelihoods or strengths (a gradient outcome from gradient influences offset by gradient costs, rather than binary); and model language producers and receivers as recursively reasoning about one another.

The findings as described above are at least broadly compatible with probabilistic pragmatics accounts in general, since different kinds of linguistic context (lexical and prosodic) have been shown to make differing contributions to the pattern of results, further

mediated by the individual's history of language acquisition, with at least some level of statistical credibility (limitations and future improvements notwithstanding). In contrast, "generalised/particularised conversational implicature" accounts (wherein the kind of scalar implicature judgements the current study's sentence stimuli were designed to elicit might be characterised as "strength invariant" and "context independent") are not well supported by this data, since the strength of judgements have been shown to vary in predictable ways with individual language backgrounds and with lexical and prosodic context.

Furthermore, at no point has the discussion of results above described any "rates" of implicatures made (see discussion of binary ratings and linking hypotheses in §4.2.2). It can be argued that the "game partner" paradigm of the current study is co-operative in nature and encourages a "charitable" interpretation of utterances, such that the ratings are skewed towards more inclusive interpretations. This tendency can also be seen (at face value) in the higher granularity (quaternary and quinary) scale results in the English language experiment on which the current study's paradigm was based (Jasbi et al., 2019), as opposed to the lower granularity (binary and trinary) scale results which produced a higher proportion of exclusive interpretations. The aim of that study was to demonstrate that rating scale granularity and alternative linking hypotheses have the potential to produce credibly different results and conclusions when all else is equal. However, the current study has very different research questions. What has been demonstrated is that in BSL, linguistic contexts and language acquisition experiences have the potential to affect the *likelihood* of belief in an intended meaning to varying degrees in varying directions, a point which is arguably one of the pillars of probabilistic pragmatics. The deliberate choice of a very high granularity rating paradigm to sensitively explore *relative* strengths of belief reflects that the findings were not intended to be compared directly to binary judgement paradigms, or to be compared with "rates" of implicature made by other users of other languages in other experimental settings (see also the discussion of issues arising from the "homogeneity assumption" in §4.1.3.1, §4.2.4).

One aspect of the "default" model argument that does bear further discussion concerns the lexicalisation extent of the *palm-up* form in BSL, and this has further relevance to the discussion of the effects of the interaction between context and language background in the previous section. As discussed in the previous chapter (§4.1.1), the "default" account put forward by Levinson (2000) suggested that for "generalised" scalar implicatures, the alternatives are required to be "in salient opposition: of the same form class, in the same dialect or register, and lexicalised to the same degree". The reader may have noticed by now that the current project has tended to hedge when referring to the

potential of *palm-up* as a lexicalised disjunctive coordinator form (e.g. “to some extent lexicalised” or “behaves as if lexicalised”), and the same issue applies to list buoys as well to an extent. As discussed more than once (§2.2.2, §3.4.1), the *palm-up* form is in widespread cross-modal, cross-linguistic and cross-cultural use as a gesture that can mark uncertainty, which does have relevance to disjunction, but also various forms of discourse regulation, which is less pertinent. In the BSL Corpus, it is the second most frequent token (after PT:PRO1SG, equivalent in meaning to “I” or “me”). That said, the current project did identify *palm-up* as the backbone of a distinctive cluster of disjunctions within BSL Corpus conversational data, and it was argued that it is (at the very least) semi-lexicalised for that function (§3.3.2). However, given the frequency and variability of the form and the overlapping functions and labels associated with it, it remains relatively difficult to establish the extent to which *palm-up* is lexicalised as a disjunction, and this is likely to vary between users in any case. The visual-gestural modality of BSL entails that, as a potential co-sign gesture, it is more difficult to draw a line between *palm-up* and lexicalised signs than it would be for a spoken language (where manual gestures and lexical items are in different modalities). The solution may be to not even try, instead regarding gesture and sign as a continuum, with *palm-up* being somewhere on a “grammaticalisation pathway” between those idealised categories (van Loon et al., 2014).

Regardless, the current study used *palm-up* as the single consistent coordinating feature of the various disjunctive sentence stimuli; furthermore, one of the stronger scalar alternatives made explicitly accessible was an asyndetic conjunction (the other was a conjunctive list buoy structure). Asyndesis is the exemplar of non-lexicalised constructions that are nonetheless conventional for a function, as supported by the current project’s corpus findings (§3.3.2). Some of the participants in the experimental study indicated levels of belief that the stronger scalar alternative warranted rejection either all or some of the time, despite half of the conjunctions being asyndetic, and the disjunctions (featuring *palm-up*) being of debatable lexicalisation extent. Furthermore, linguistic and language background factors in interaction were statistically associated with variations in response. How does this square with pragmatic-theoretical assumptions around lexicalisation?

Arguably, the most relevant aspect of the role of lexicalisation in inferential reasoning is the part of it that concerns accessibility (the extent to which, in the mind of the language user, the stronger scalar term is semantically retrievable as a salient contrast to the presented weaker term). On that line of thinking, we previously examined experimental findings (§4.2.3) that tested the effects of accessibility in the context of a “restricted alternatives” hypothesis (Skordos & Papafragou, 2016; Tieu et al., 2016). In those studies,

increased “rates” of implicatures made by young children still in the critical period of language acquisition were demonstrated to correlate with increased lexical accessibility (via recent exposure to the stronger alternative), bringing them much closer to adult performance than some previous studies had given children credit for. The restricted alternatives hypothesis was originally put forward to qualify broadly neo-Gricean accounts of the “generalisation” of lexicalised scalar implicatures: it proposed that children who do understand the semantic meanings and co-scalar status of the alternative scalar terms may still find the intended meaning of a potential scalar implicature more difficult to process than experienced adults because, when presented with the weaker scalar term, it is still relatively costly for them to access or retrieve the contrasting meaning of the stronger alternative from their hypothetical “lexicon”. Some of the differences observed in the language background contrasts identified by the current study could be explained in terms of that general variety of argument, given that the group assumed to have a later age of acquisition of both BSL and English were broadly accepting of the stimuli, while still having been demonstrated to understand the literal semantic meaning of the terms (since they met the threshold for correct judgements of the control conditions).

However, in the RSA framework (which was only just emerging at the time the “restricted alternatives” hypothesis was put forward), accessibility is just one possible cost among many that can make a gradient contribution against informativity towards deriving posterior beliefs (Goodman & Frank, 2016). The RSA literature (which is relatively young) has more often seemed to focus on production costs (such as utterance complexity) than reception costs such as accessibility, with exceptions (Goodwin & Degen, 2024). That said, word frequency has also been suggested as a proxy for cost, and given that RSA describes producer and receiver as recursively modelling the reasoning of one another, this implies that the producer may also reason about the accessibility of the chosen utterance for the receiver*,* and that the receiver can reason about the producer’s reasoning. Therefore, in the terms of a probabilistic pragmatics account, counting accessibility or retrievability as a cost dependent on language experience might explain the reinforcement of scalar contrast accessibility via both recent and long-term language experience just as well as the “restricted alternatives” hypothesis. The meanings of asyndetic coordination and “not fully lexicalised” manual forms in BSL (such as *palm-up* and list buoys) are still presumably capable of incurring retrieval costs, since regardless of whether or not they can be formally labelled as “lexicalised”, they are clearly *conventionalised* for conjunctive and disjunctive semantics by many (§3.4.1). In the case of asyndesis, a construction with no overt lexical marking, it seems unlikely that the

semantic relevance would reside in a hypothetical “active lexicon”. But regardless of whether the semantics are retrieved from a “lexicon” or a more generalised linguistic/epistemic information structure, how accessible the contrasting stronger alternatives are, and therefore the cost of retrieving them, will vary with individual language experience. That cost, like every other, would be balanced against the utterance’s potential informativeness in combination with other informative contextual cues, which (if informative enough) might still be capable of tipping the cost/benefit balance towards some degree of belief that the stronger alternative warrants rejection, even if the retrieval cost were relatively high for that individual. If only speculatively, this might explain the patterns seen for the Later AoA and Lower RA group, where the expected effect of contrastive stress towards a less inclusive reading was only statistically credible in the presence of additional coordinating context (with both the stress and the more overt coordinating structure providing further informative benefits against the group’s relatively high cost of retrieving the stronger alternative).

Further study might look at formally applying the RSA paradigm to scalar implicatures in a sign language towards fully predicting the outcomes of a controlled experiment. Such a study would require the estimation of the strength of prior beliefs given the experimental paradigm, the specification of likelihoods for a range of semantic alternatives being understood with the intended meaning, and quantifying the contributions from various contextual cues against various cognitive costs.

5.4.3 Limitations

In this penultimate section, we should take a moment to reflect on the limitations of this study, in the spirit of lessons learned and suggestions towards future improvements in similar works.

5.4.3.1 Sample size and participant balance

By the standards of majority spoken language web-based studies which are able to recruit hundreds or thousands of participants via commercial platforms, this sign language investigation was relatively modest, with 49 participants contributing 392 test condition data points out of 1,568 test and control condition points in total. However, this is a relatively large engagement for a sign language study, given the difficulties of recruiting willing volunteers from a minority language community. Furthermore, design choices aimed at obtaining robust results for smaller sample and effect sizes (such as the

continuous rating scale design and the specific statistical method used) did consciously attempt to address this issue.

The demographic balance of participants was only adequate. The gender ratio (favouring female) was an issue of some regret, although there were only purely speculative reasons to think that deaf British women might respond to BSL implicatures differently from men. It was also regrettable that more members of Black, Asian and minority ethnic communities were not recruited, despite outreach. It would also have been pleasing to have been able to include more participants from the Midlands (efforts were made), given that it is a major population centre with a strong deaf community. Younger participants could also have been better represented (the 18-27 range in particular). The main justification for hypothesising that region and age might be factors in the strength of derived pragmatic effects in BSL is that standards and practices in deaf education, support for parents of deaf children, and language access policies in general, are all known to have varied over time and by region (§1.4.1).

For the theoretical background of the study, it was also not ideal that the age of BSL acquisition variable had to be banded in the manner required by the number and balance of presenting participants. As described above (§5.3.1.1), the “Early AoA” group necessarily represented those who selected the “From birth” option, while the “Later AoA” group comprised those who selected the three other bands, setting the limits of the two age of acquisition groups as ages 0-2 and 2+. Additional recruitment might have allowed for the more theoretically relevant age boundaries of 0-8 and 8+, more closely aligning with the limits of the “critical period” hypothesis of language development (§4.2.5); an even larger pool might have allowed for additional levels or a continuous measure, enabling a more nuanced analysis.

5.4.3.2 Participant reactions to the task

10 of the 49 participants opted to give feedback as part of the task questionnaire; a small handful of additional responses were sent directly to the researcher by e-mail following completion of the trials. There were no especially common themes. One anonymous responder queried the intended meaning of *palm-up*, and another gave a response which suggested they were uncertain about the intended meaning of the list buoy form, but both successfully passed the control condition exclusion criterion. A small number indicated technical difficulties, usually with screen size and orientation (most of whom appeared to be using a tablet), and in one case that the BSL instruction videos were failing to load for them, but in all cases indicated that they had overcome the issues (they would have been

excluded otherwise). A couple of responses indicated that they found the English task challenging: this was expected, as the questions gradually increase in difficulty, and very few English users are expected to achieve a perfect score. Finally, one response clearly implied frustration at not being able to revise previous answers (a desire that was also reported by more than one pilot participant). It was felt that this would not be advisable and would likely lead to the erasure of differences between the item variables of interest; but the task instructions could perhaps have been more reassuring around not needing to be concerned about giving “correct” responses.

5.4.3.3 Remote participation

In bygone decades, before web-based experiments became feasible, and long before commonplace access to video calling, researchers wishing to collect experimental data from members of a sign language community had two alternatives. They could stay in one place and invite people to travel to them, which risked the exclusion of those more distant and the magnification of regional biases in the data. Alternatively, the researchers might tour the country and bring the study to the community, perhaps in partnership with local organisations. Both options might consume significant resources and time. Today, web-based experiments appear at first glance to be a quick fix for these issues: while they might involve more preparation, they require a lesser investment of physical resources and time to run, and geographical distance between researcher and participant is not an issue. However, we might also reflect on what is lost by remote participation.

As mentioned above, commercial recruitment platforms can provide hundreds of oral-auditory modality participants at a relatively low cost, and many can also vet participants on your behalf, but they are much less useful resources for minority language studies. All of the demographic data collected was self-reported, and confirmation of the key participation criteria was taken on trust; none of the participants were met face-to-face. The experimental design did allow for detection of those with no or minimal BSL ability (see §5.2.1), but not whether BSL was genuinely used as “main or preferred language”, which may mean different things to different people in any case. However, with the exception of those who failed the attention criterion and were excluded, there was no cause for suspicion that any other participant was misrepresenting themselves.

A minority of participants did not spend particularly long on the task instructions, which was of some concern. It could be argued that remote participation in research has an implicit disconnection: it does not encourage engagement with the study, and may even discourage it. In that regard, perhaps more could have been done to encourage a sense of

community ownership of the study. It was also impossible to verify that participants completed the tasks without external help, or that they undertook the tasks in a distraction-free environment as requested. It is legitimate to speculate as to whether the task data would have shown less variation and larger effect sizes if it had been conducted in a more controlled setting (in accordance with scientific principle).

It is therefore tempting to argue that online experiments that attempt to engage with minority linguistic communities suffer from all of the disadvantages of remote participation while enjoying fewer of the benefits. The less cynical counter-argument is that where resources are limited and the community is widely spread, web-based studies allow for qualified insights to be gleaned, as opposed to none at all. For the current study, the increased outreach was felt to be valuable and important, despite the drawbacks.

5.5 Conclusions

The analysis of the data gathered by this experiment broadly demonstrates that when eliciting exclusivity inferences from deaf users of BSL using a continuous rating paradigm, the strength of belief in a scalar implicature (the likelihood that the stronger alternative was intended to be rejected) can vary with interactions between individual language experience and the linguistic context in statistically credible ways. The strongest contributing factor was an interaction between banded age of BSL acquisition and a banded score in an English reading accuracy task, but the magnitude and direction of this group-level effect also interacted in some cases with the lexical context of the utterance (the presence or absence of an additional coordinating list buoy form) and the presence of contrastive stress (in the form of body lean, spatial placement contrast, raised brow movement, and an increased use of signing space). In the case where the participants had both an early age of BSL acquisition (before the age of 2) and above-cohort-average English reading accuracy, the presence of contrastive stress had a predictable tendency to produce more and stronger exclusive readings for disjunctive *palm-up* statements that did not also contain a list buoy, in line with expectations based on the effects of intonational or typographical stress on disjunctions in spoken and written forms of language. This group also seemed to be the most conscious of the inherent ambiguity of pragmatically infelicitous statements (inferred from their relatively pronounced use of the mid-points of the rating scale). For participants with a later BSL age of acquisition, the presence of the additional lexical context of a list buoy generally had a detectable effect towards a more inclusive reading, with the exception of the case where they had lower English reading accuracy and the sentence was also stressed, where the expected effect away from a fully

inclusive reading was again demonstrated. It is concluded from this that non-manual contrastive prosody, conjectured in earlier chapters to function as contrastive stress in BSL when co-occurring with a disjunction, can indeed have a similar effect to forms of contrastive emphasis in spoken and written forms of language, i.e. producing at least some level of “pragmatic” rejection of the unstated alternative. I have also argued (following up on indications from the clustering analysis of BSL coordination undertaken in chapter 3) that this finding further supports the argument that non-manual marking of coordination in BSL is best understood and described as voluntary, prosodic and pragmatic (as opposed to obligatory or morphosyntactic).

However, whether or not the contrastive stress actually produced the expected effect towards an exclusive reading was also clearly dependent on individual language experience. Those with earlier BSL acquisition and higher-than-cohort-average English reading accuracy were the most sensitive to manipulation of prosodic (non-manual) stress when there was minimal lexical context (manual *palm-up* only), with such stress being reliably shown to shift their reading towards an exclusive interpretation. For these early acquirers, added stress had no predictable effect when the lexical context was greater (*palm-up* and a list buoy construction), suggesting that different kinds of context are taken into account simultaneously when deriving communicative intent, in line with the general predictions of a probabilistic account of pragmatics. In contrast, those with both later BSL acquisition and below-cohort-average English reading accuracy were generally more overtly influenced or supported by extra coordinating structure and lexical context in addition to the stress. I have argued that this pattern of responses constitutes further evidence towards the argument that learning the communicative intent of contrastive prosody (in this case, to signal that an exclusive interpretation of disjunction is intended) is supported during early language acquisition by having access to additional lexical or discourse context.

6 General conclusions

6.1 Summary

This thesis has examined conjunction and disjunction in British Sign Language (BSL), and explored factors influencing the strength of disjunction-based scalar implicatures in BSL. We began (§1) by defining key terms, in particular the difference between conjunction and disjunction, the characterisation of coordination as syndetic or asyndetic, distinctions in sign language linguistics between manual and non-manual contributions to sign structure, and a working definition of scalar implicatures.

In an initial literature review (§2), we first explored suggestions that conjunction and disjunction can be expressed in BSL with lexicalised coordinators. These included several items from the core BSL lexicon and a couple of contact products from English. One empirical study (Waters & Sutton-Spence, 2005) had previously documented the use of asyndesis in BSL, finding that it could represent both conjunctions and disjunctions, but there was no explicit account of how the two might be distinguished. The lexical status and function of the *palm-up* form (glossed in the BSL Corpus as G:WELL) was also of great interest. Although it has been associated with several gestural discourse functions in BSL, as it has been in many other signed and spoken languages (§2.2.2), it has also been identified as a coordinator. BSL Corpus findings (Arnold, 2019) and BSL Signbank documentation (Fenlon et al., 2014) associate it with both disjunctive and adversative meanings, but only the disjunctive function was observed by Waters & Sutton-Spence (2005), while only the adversative senses are listed in older BSL dictionaries (Brien, 1992). As for non-manual features, while some documentation of other sign languages has claimed that specific articulations can be associated with coordination or even be obligatory (although not always with a clear evidence base), no existing documentation of non-manual contributions to BSL coordination could be located.

These findings shaped the research questions for a usage-based study (§3) which explored conjunction and disjunction in a sample of spontaneous conversation data from the BSL Corpus, concentrating on the manual and non-manual means of expression. Almost all of the previously documented lexical items were found in the data, but they were relatively infrequent in comparison to other strategies. Asyndesis was quite common for conjunction. The list buoy reference strategy was the most common syndetic (i.e. manual) articulation of conjunction, and a few examples of other pointing-based reference strategies were also observed. For disjunction, the *palm-up* form was very clearly the

preferred manual coordinator, lending weight to an argument that it is lexicalised for this function to some extent. There was also an unexpected number of asyndetic disjunctions: on inspection, all of these constructions had a clear contextual sense of exclusivity, and most were associated with numeric hedging. A small number of disjunctive list buoy episodes were also attested, using *palm-up* within the list buoy structure to create the sense of alternatives. Regarding non-manual contributions, co-occurring brow movements were found fairly often across the sample (both raising and furrowing), and there was the occasional use of body leans. However, a hierarchical clustering analysis found no robust correlation for either brow movements or body leans with any particular conjunction or disjunction strategy, with one exception. Neutral brow movements were slightly disassociated from asyndetic disjunction, but neither upward nor downward movements were significantly preferred and the size of the effect was relatively mild in comparison to the sample as a whole: a third of asyndetic disjunctions still retained neutral brows. Mouthings were found to be associated with lexical coordinators and *palm-up*, and strongly disassociated from asyndetic conjunction. Body leans had no credible associations with any form or structure.

Taken all together, these findings suggested that in BSL, disjunction tends to be more overtly marked than conjunction (manually with *palm-up*, and perhaps non-manually) unless there is a strong pragmatic contribution from the semantic or discourse context. Lexical coordinators are sometimes used, but asyndesis is the most common conjunction strategy. Pointing-based reference strategies, mostly list buoys, also outnumber lexical conjunctions. However, brow movements and body leans are certainly not obligatory and not even conventional for marking conjunction or disjunction in BSL. The preferred account for this kind of non-manual contribution, concurring with a general argument made by Crasborn & van der Kooij (2013) for NGT (Sign Language of the Netherlands) among others, was that it is prosodic and pragmatic: it can contribute to readings of focus or contrastive stress, and the amplitude or “markedness” of the non-manual contribution is generally more salient than the precise form. It was conjectured that if this is the case, the effect of a generalised non-manual emphasis should be detectable in an experimental BSL scalar implicature study, along similar lines to the established effect of spoken language prosody towards more exclusive readings of disjunction.

Towards designing such an experiment, we reviewed contrasting theoretical accounts of scalar implicature and experimental pragmatics paradigms suitable for adaptation to a sign language study (§4). Gricean and “traditional” accounts of scalar implicature were compared to more recent “probabilistic pragmatics” accounts. The latter

retain some assumptions of the former around the co-operative and purposeful nature of conversational interaction, but characterise beliefs in intended meanings as a gradient likelihood or a variable strength rather than a binary outcome. Contextual informativeness and counterbalancing cognitive costs are also both characterised as making gradient contributions. Issues around the reflection of these differing positions in experimental judgement task designs were discussed, in particular the theoretical consequences and statistical benefits of more granular rating mechanisms (Jasbi et al., 2019; Marty et al., 2020). The role of the accessibility of contrasting scalar terms was also highlighted in the context of the childhood acquisition of pragmatic ability and a “restricted alternatives hypothesis” (Skordos & Papafragou, 2016; Tieu et al., 2016). This area of research was felt to be particularly relevant to a study of scalar implicatures in a sign language, since delayed language acquisition is more likely for the ~95% of the deaf signing community who did not have signing parents as language models during infancy. Previous studies of the effects of contrastive prosody towards increased rates of “pragmatic” interpretation were also highlighted (Chevallier et al., 2008; Kurumada & Clark, 2017). Finally, a study of the effects of language background and cognitive factors on grammatical sensitivity in deaf BSL users was examined (Cormier, Schembri, et al., 2012), which suggested a transferrable effect from English reading ability to BSL grammatical accuracy for those with later ages of BSL acquisition, despite the respective modalities and grammars being very different. Taking all of the above into account, it was decided that the rating mechanism for the following scalar implicature study should be a bounded continuous measure (a gradient rating); that the stimuli design should control the scope of alternatives and ensure that the stronger scalar term is accessible; that a “native signer ideology” approach should be avoided, recruiting participants with any age of BSL acquisition; and that both linguistic and language background factors should be controlled for the statistical analysis.

That experiment was then carried out and documented (§5). A “guessing game” paradigm (Jasbi et al., 2019) was adapted for use with conjunctive and disjunctive BSL stimuli, informed by the earlier BSL Corpus findings (§3.5). In the main task, the participants were exposed to relatively simple BSL video descriptions of “playing cards” presented in a random order, of which three quarters were control conditions (a mixture of uncoordinated, conjunctive, and disjunctive utterances) and the remainder were test conditions (disjunctive only). Each card featured either one or two illustrations of animals or objects, and participants were asked to rate the appropriateness of the accompanying description using a sliding scale between “wrong” (a red cross) and “right” (a green tick). Control condition items were designed to make the contrasting scalar terms accessible,

and also to verify a minimal level of BSL fluency as a principled exclusion criterion (in addition to an attention measure). Test condition items were designed to have the potential to elicit a degree of exclusivity inference, such that a fully “right” rating would indicate that a fully inclusive interpretation had been taken from the disjunction, and a fully “wrong” rating would indicate a fully exclusive interpretation (a scalar implicature), with intermediate ratings indicating intermediate levels of belief that the stronger scalar contrast was intended to be rejected. Trial items were balanced between two contrasting levels of lexical context (the presence or absence of a list buoy structure) and two levels of non-manual emphasis (conjectured above as having the potential to function as contrastive prosodic stress). Participant variables were derived from self-reported ages of BSL acquisition (“Early” being ages 0-2, “Later” being ages 2+) and English reading scores as determined by a supplementary assessment task (“Lower” and “Higher” than the cohort median). Lexical context and contrastive stress were treated as item-level variables in the statistical analysis, along with the participant-level variables of age of BSL acquisition (AoA) and English reading accuracy (RA).

The test condition results were first plotted and described (§5.3.1), and then subjected to inferential analysis via a Bayesian mixed effects model based on an ordered beta distribution (§5.2.3, §5.3.2). For the item variables alone, a credible effect of contrastive stress was found for sentences with increased lexical context towards a marginally less inclusive reading, as well as an effect of increased lexical context towards a slightly more inclusive reading in the absence of contrastive stress, although the effect sizes were small when participant variables were not taken into account. For the participant variables alone (ignoring linguistic factors), a general effect of Higher RA was found for those with Later AoA towards less inclusive ratings. However, the findings also suggested specific interactions between the language background and linguistic variables which differed from the group average marginal effects in terms of effect size and (in some cases) direction. Those with the most language experience (Early AoA and Higher RA) were observed to make the most use of the mid-points of the rating scale, which was taken as indicating a relatively increased awareness that the test condition stimuli were ambiguous between inclusive and exclusive readings. That group also showed a statistically credible effect away from an inclusive interpretation of disjunction where there was additional contrastive stress and less lexical context, in line with expectations based on the effects of intonational or orthographic emphasis in previous studies of spoken and written language forms. In contrast, those with the least language experience (Later AoA and Lower RA) gave generally very inclusive responses, and seemed more dependent on

the presence of additional lexical context in making any level of rejection of the literal meaning. Those with mixed levels of BSL and English language experience showed fewer differences from the cohort average responses.

It was argued (§5.4.1) that these findings further support the earlier conjecture (§3.5) that the non-manual marking of conjunction and disjunction in BSL is best understood and described as voluntary, prosodic and pragmatic (as opposed to obligatory or morphosyntactic), and that it functions as contrastive stress. It was further argued that the analysis constitutes evidence towards supporting the general argument offered by Kurumada & Clark (2017), that learning the communicative intent of contrastive prosody (in this specific case, to signal that an exclusive interpretation of disjunction was intended) is likely to be "scaffolded" during the early stages of language acquisition by access to additional lexical or discourse context. In general, the findings were argued (§5.4.2) to be broadly in line with a "probabilistic pragmatics" account of scalar implicature. More speculatively, it was suggested that contemporary accounts such as the Rational Speech Act framework can fully accommodate the earlier neo-Gricean "restricted alternatives" hypothesis by explicitly accounting for the differing costs of the retrievability of semantic contrasts by individuals with differing language acquisition experiences.

6.2 Future directions

This thesis has occasionally pointed out areas that might benefit from future research, and in this final section, they are briefly summarised.

In the usage-based exploration of within-sentence conjunction and disjunction presented in chapter 3, a decision was taken (§3.2.2) to exclude specific grammatical contexts known to change or cancel the derivation of scalar implicatures (such as modal qualifiers, negation, and question forms), given that the study was primarily undertaken towards informing the design of the broader project's scalar implicature study. The sample of natural conversation data from the BSL Corpus was also necessarily limited to the 40 out of 249 participants who at that time had clause-like unit annotations available, as this assisted (to an extent) with differentiating between sentence-level and discourse-level coordination. This in turn entailed that only participants from four cities in England were included, excluding data from Scotland, Wales, and Northern Ireland. A future study might productively analyse the manual, non-manual and prosodic features of BSL coordination (including adversative and illative coordination) in spontaneous conversation data, with a less restricted range of grammatical contexts, and from a broader range of informants.

The same usage-based study also collected information on the properties of list buoy episodes (§3.3.1.7), but this data subset was relatively small. While most of the identified features were generally in line with expectations based on existing documentation of list buoy forms in other sign languages, some BSL observations differed in minor ways from specific accounts (§4.2.4). Furthermore, given the focus on within-sentence coordination, no uses of discourse-level referential back-tracking to previously enumerated list buoy items were studied. It was recommended that a future usage-based study of list buoys would benefit from comparing the features of sentence and discourse level usage.

The findings of the experimental study documented in chapter 5 were argued to be broadly compatible with a "probabilistic pragmatics" account of implicature, particularly with regards to variable and gradient levels of belief that the stronger scalar term warrants rejection, and the gradient contributions of linguistic contexts and cognitive costs. Future studies of implicature in a sign language might consider fully applying a formal probabilistic model such as the Rational Speech Act framework towards predicting the outcomes of experiments. This would require the estimation of the strength of prior beliefs and the likelihoods of semantic alternatives being understood with the intended meaning, as well as quantifying contextual contributions and cost-based offsets (§5.4.2).

As a closing note, I would like to strongly underline the view that pragmatic competence is a fundamental factor in social and cognitive development, regardless of whether the childhood linguistic environment is signed or spoken, and that this has general implications for well-being (Matthews & Kelly, 2022). Recent calls to action have argued that pragmatic abilities are too often overlooked in discussions of the consequences of language deprivation in deaf children and adults (Szarkowski et al., 2020; Young et al., 2020). The current project noted (§4.1.3.3) that while there has been a substantial body of psychological research into the links between language acquisition and theory of mind ability in deaf children and adults, the theoretical links between pragmatic competence and theory of mind were often taken for granted. Furthermore, such links have rarely been explored in the terms of pragmatic or linguistic frameworks for inferential communication and reasoning, either from a contemporary approach (such as the Rational Speech Act framework) or more traditional accounts. Studies of scalar implicature are perhaps a narrow measure of pragmatic reasoning, but that does not make them uninformative. It is sincerely hoped that the discussion and new evidence presented in this first known exploration of scalar implicatures in BSL represents a step towards such goals, and that future multidisciplinary studies of sign language pragmatics are inspired to extend and improve upon the approaches I have taken.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Arel-Bundock, V. (2024). *marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. Retrieved from https://marginaleffects.com/

Arnold, B. (2019). *Discourse functions of palm-up in British Sign Language*. MSc dissertation, University College London.

Asada, Y. (2019). General use coordination in Japanese and Japanese Sign Language. *Sign Language & Linguistics*, *22*(1), 44–82. https://doi.org/10.1075/sll.18003.asa

Battison, R. (1978). *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstock Press.

Begby, E. (2023). Theory of mind deficits in language delayed deaf subjects? *Philosophy and the Mind Sciences*, *4*. https://doi.org/10.33735/phimisci.2023.10269

Bel, A., Ortells, M., & Morgan, G. (2015). Reference control in the narratives of adult sign language learners. *International Journal of Bilingualism*, *19*(5), 608–624. https://doi.org/10.1177/1367006914527186

Benz, A., & Stevens, J. (2018). Game-theoretic approaches to pragmatics. *Annual Review of Linguistics*, *4*, 173–191. https://doi.org/10.1146/annurev-linguistics-011817-045641

Bergen, L., & Goodman, N. D. (2015). The Strategic Use of Noise in Pragmatic Reasoning. *Topics in Cognitive Science*, *7*(2), 336–350. https://doi.org/10.1111/tops.12144

Birkeland, A., Block, A., Craft, J. T., Sedarous, Y., Wang, S., Wu, A., & Namboodiripad, S. (2024). Rejecting nativeness to produce a more accurate and just Linguistics. *Language*, *100*(3), e156–e194. https://doi.org/10.1353/LAN.2024.A937197

Blose, B. A., & Schenkel, L. S. (2022). Theory of Mind and Alexithymia in Deaf and Hard-of-Hearing Young Adults. *Journal of Deaf Studies and Deaf Education*, *27*(2), 179–192. https://doi.org/10.1093/deafed/enac001

Börstell, C. C., Crasborn, O., & Schembri, A. (2024). Sign duration and signing rate in British Sign Language, Dutch Sign Language, and Swedish Sign Language. *Glossa Psycholinguistics*, *3*(1). https://doi.org/10.5070/G60111915

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. https://doi.org/10.1016/j.jml.2004.05.006

Breheny, R. (2008). A New Look at the Semantics and Pragmatics of Numerically Quantified Noun Phrases. *Journal of Semantics*, *25*(2), 93–139. https://doi.org/10.1093/JOS/FFM016

Breheny, R. (2019). Scalar Implicatures. In *The Oxford Handbook of Experimental Semantics and Pragmatics* (pp. 38–61). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198791768.013.4

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463. https://doi.org/10.1016/j.cognition.2005.07.003

Brennan, M. (1992). The visual world of British Sign Language: an introduction. In D. Brien (Ed.), *Dictionary of British Sign Language / English* (pp. 2–133). London: Faber & Faber.

Brennan, M. (2001). Making borrowings work in British Sign Language. In *Foreign vocabulary in sign languages: cross-linguistic investigation of word formation* (pp. 49–86). Mahwah, NJ: Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410601513-9

Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brien, D. (Ed.). (1992). *Dictionary of British Sign Language/English*. Boston: Faber & Faber.

British Sign Language Act. (2022). Retrieved from https://www.legislation.gov.uk/ukpga/2022/34/contents

Brown, M., & Cormier, K. (2017). Sociolinguistic variation in the nativisation of BSL fingerspelling. *Open Linguistics*, *3*(1), 115–144. https://doi.org/10.1515/opli-2017-0007

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. https://doi.org/10.1177/2515245918823199

Buyle, M., & Crollen, V. (2022). Deafness and early language deprivation influence arithmetic performances. *Frontiers in Human Neuroscience*, *16*, 1000598. https://doi.org/10.3389/fnhum.2022.1000598

Carneiro, B. G., El Khouri, J. I. B., Santos, T. C., & Ludwig, C. R. (2023). Parataxis, hypotaxis, and subordination in Brazilian Sign Language: a brief introduction. *Linguística: Revista de Estudos Linguísticos Da Universidade Do Porto*, *18*, 32–64. https://doi.org/10.21747/16466195/ling18a2

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: applications and implications* (pp. 179–238). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.37.11car

Cheng, Q., Roth, A., Halgren, E., & Mayberry, R. I. (2019). Effects of Early Language Deprivation on Brain Connectivity: Language Pathways in Deaf Native and Late First-Language Learners of American Sign Language. *Frontiers in Human Neuroscience*, *13*, 475975. https://doi.org/10.3389/fnhum.2019.00320

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology*, *61*(11), 1741–1760. https://doi.org/10.1080/17470210701712960

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar inferences in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*(9), 1104–1117. https://doi.org/10.1007/s10803-010-0960-8

Chierchia, G., Fox, D., & Spector, B. (2011). Scalar Implicature as a Grammatical Phenomenon. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (pp. 3–2297). De Gruyter Mouton.

Conlin, F., Neidle, C., & Hagstrom, P. (2003). A particle of indefiniteness in American Sign Language. *Linguistic Discovery*, *2*(1). https://doi.org/10.1349/PS1.1537-0852.A.142

Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The palm-up puzzle: meanings and origins of a widespread form in gesture and sign. *Frontiers in Communication*, *3*(June), 1–16. https://doi.org/10.3389/fcomm.2018.00023

Corina, D., & Singleton, J. (2009). Developmental social cognitive neuroscience: insights from deafness. *Child Development*, *80*(4), 952–967. https://doi.org/10.1111/j.1467-8624.2009.01310.x

Cormier, K., Fenlon, J., Gulamani, S., & Smith, S. (2017). *BSL Corpus Annotation Conventions version 3.0*. Deafness, Cognition & Language Centre, University College London. Retrieved from

http://bslcorpusproject.org/cava

Cormier, K., Quinto-Pozos, D., Sevcikova, Z., & Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: Comparing depicting constructions and viewpoint gestures. *Language and Communication*, *32*(4), 329–348. https://doi.org/10.1016/j.langcom.2012.09.004

Cormier, K., Schembri, A., & Tyrone, M. E. (2008). One hand or two?: Nativisation of fingerspelling in ASL and BANZSL. *Sign Language and Linguistics*, *11*(1), 3–44. https://doi.org/10.1075/sl

Cormier, K., Schembri, A., Vinson, D., & Orfanidou, E. (2012). First language acquisition differs from second language acquisition in prelingually deaf signers: Evidence from sensitivity to grammaticality judgement in British Sign Language. *Cognition*, *124*(1), 50–65. https://doi.org/10.1016/j.cognition.2012.04.003

Cormier, K., Smith, S., & Zwets, M. (2013). Framing constructed action in British Sign Language narratives. *Journal of Pragmatics*, *55*, 119–139. https://doi.org/10.1016/j.pragma.2013.06.002

Courtin, C., & Melot, A.-M. (2005). Metacognitive development of deaf children: lessons from the appearance–reality and false belief tasks. *Developmental Science*, *8*(1), 16–25. https://doi.org/10.1111/j.1467-7687.2005.00389.x

Crasborn, O., & van der Kooij, E. (2013). The phonology of focus in Sign Language of the Netherlands. *Journal of Linguistics*, *49*(3), 515–565. https://doi.org/10.1017/S0022226713000054

Davidson, K. (2011). *The nature of the semantic scale: Evidence from sign language research*. Doctoral thesis. University of California, San Diego.

Davidson, K. (2013). “And” or “or”: General use coordination in ASL. *Semantics and Pragmatics*, *6*(4), 1–44. https://doi.org/10.3765/sp.6.4

Davidson, K. (2014). Scalar implicatures in a signed language. *Sign Language & Linguistics*, *17*(1), 1–19. https://doi.org/10.1075/sll.17.1.01dav

Davidson, K., & Mayberry, R. (2015). Do Adults Show an Effect of Delayed First Language Acquisition When Calculating Scalar Implicatures? *Language Acquisition*, *22*(4), 329–354. https://doi.org/10.1080/10489223.2014.962140

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133. https://doi.org/10.1027/1618-3169.54.2.128

Degen, J. (2015). Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*(11), 1–55. https://doi.org/10.3765/sp.8.11

Degen, J. (2023). The Rational Speech Act Framework. *Annual Review of Linguistics*, Vol. 9, pp. 519–540. Annual Reviews Inc. https://doi.org/10.1146/annurev-linguistics-031220-010811

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: a constraint-based approach. *Cognitive Science*, *39*(4), 667–710. https://doi.org/10.1111/cogs.12171

Degen, J., & Tanenhaus, M. K. (2019). Constraint-Based Pragmatic Processing. *The Oxford Handbook of Experimental Semantics and Pragmatics*, 20–38. https://doi.org/10.1093/oxfordhb/9780198791768.013.8

Dik, S. C. (1989). *The Theory of Functional Grammar, Part 1: the structure of the clause*. Dordrecht: Foris Publication.

Dively, V. (2001). Signs without Hands: Nonhanded Signs in American Sign Language. *Signed Languages: Discoveries from International Research*, 62–73. Washington, D.C.: Gallaudet University Press.

Durrleman, S., Dumont, A., & Delage, H. (2021). Syntactic Strategy Training for Theory of Mind in Deaf

Children. *The Journal of Deaf Studies and Deaf Education*, *27*(1), 89–100. https://doi.org/10.1093/deafed/enab034

Engberg-Pedersen, E. (2002). Gestures in signing: the presentation gesture in Danish Sign Language. In R. Schulmeister & H. Reinitzer (Eds.), *Progress in sign language research / Fortschritte in der Gebärdenensprachforschung* (pp. 143–162). Signum.

Engberg-Pedersen, E. (2015). Perspective in signed discourse: the privileged status of the signer's locus and gaze. *Open Linguistics*, (1). https://doi.org/https://doi.org/10.1515/opli-2015-0010

Falkman, K. W., Roos, C., & Hjelmquist, E. (2007). Mentalizing skills of non-native, early signers: A longitudinal perspective. *European Journal of Developmental Psychology*, *4*(2), 178–197. https://doi.org/10.1080/17405620600574867

Fekete, I., Janebová, M., Emonds, J., Babarczy, A., Surányi, and B., & Gerőcs, M. (2015). Logical and Pragmatic Meaning in the Interpretation of Connectives: Scalar Implicatures and "Shallow" Processing. In *Language Use and Linguistic Structure* (pp. 171–180). https://doi.org/10.5507/ff.14.24440606.12

Fenlon, J., Cormier, K., & Brentari, D. (2019). The phonology of sign languages. In *The Routledge Handbook of Phonological Theory* (pp. 453–475). Routledge. https://doi.org/10.4324/9781315675428-16

Fenlon, J., Cormier, K., Rentelis, R., Schembri, A., Rowley, K., Adam, R., & Woll, B. (2014). *BSL Signbank: a lexical database of British Sign Language (First Edition)*. London: University College London. Retrieved from http://bslsignbank.ucl.ac.uk/

Fenlon, J., Cormier, K., & Schembri, A. (2015). Building BSL SignBank: The lemma dilemma revisited. *International Journal of Lexicography*, *28*(2), 169–206. https://doi.org/10.1093/ijl/ecv008

Fenlon, J., Schembri, A., & Cormier, K. (2018). Modification of indicating verbs in British Sign Language: a corpus-based study. *Language*, *94*(1), 84–118. https://doi.org/10.1353/lan.2018.0002

Fenlon, J., Schembri, A., Rentelis, R., & Cormier, K. (2013). Variation in handshape and orientation in British Sign Language: The case of the "1" hand configuration. *Language & Communication*, *33*(1), 69–91. https://doi.org/10.1016/J.LANGCOM.2012.09.001

Fenlon, J., Schembri, A., Rentelis, R., Vinson, D., & Cormier, K. (2014). Using conversational data to determine lexical frequency in British Sign Language: The influence of text type. *Lingua*, *143*, 187–202. https://doi.org/10.1016/j.lingua.2014.02.003

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998. https://doi.org/10.1126/science.1218633

Franke, M. (2009). *Signal to Act: Game Theory in Pragmatics*. PhD thesis, Universiteit van Amsterdam.

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift Fur Sprachwissenschaft*, *35*(1), 3–44. https://doi.org/10.1515/zfs-2016-0002

Gabry, J., Češnovar, R., & Johnson, A. (2023). *cmdstanr: R Interface to "CmdStan."*

Geurts, B. (2006). Take `Five': The meaning and use of a number word. *Non-Definiteness and Plurality*, (1), 311–329. https://doi.org/10.1075/LA.95.16GEU

Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press. https://doi.org/10.1017/CBO9780511975158

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science*, *5*(1), 173–184. https://doi.org/10.1111/tops.12007

Goodwin, E., & Degen, J. (2024). Production of Syntactic Alternations Displays Accessibility But Not Informativity Effects. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*.

GOV.UK. (2023). British Sign Language GCSE: Everything you need to know. Retrieved November 10, 2024, from https://educationhub.blog.gov.uk/2023/12/21/british-sign-language-gcse-everything-you-need-to-know/

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388. https://doi.org/10.2307/2182440

Grice, H. P. (1968). Utterer's Meaning, Sentence-Meaning, and Word-Meaning. *Foundations of Language*, *4*(3), 225–242. Retrieved from http://www.jstor.org/stable/25000329

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 41–58). New York: Academic Press.

Gu, Z. (2022). Complex heatmap visualization. *IMeta*, *1*(3), e43. https://doi.org/10.1002/imt2.43

Hall, M. L., Eigsti, I. M., Bortfeld, H., & Lillo-Martin, D. (2017). Auditory deprivation does not impair executive function, but language deprivation might: Evidence from a parent-report measure in deaf native signing children. *Journal of Deaf Studies and Deaf Education*, *22*(1), 9–21. https://doi.org/10.1093/deafed/enw054

Hall, M. L., Hall, W. C., & Caselli, N. K. (2019). Deaf children need language, not (just) speech. *First Language*, *39*(4), 367–395. https://doi.org/10.1177/0142723719834102

Hall, W. C. (2017). What You Don't Know Can Hurt You: The Risk of Language Deprivation by Impairing Sign Language Development in Deaf Children. *Maternal and Child Health Journal*, *21*(5), 961–965. https://doi.org/10.1007/s10995-017-2287-y

Halliday, M., & Matthiessen, C. (2004). An Introduction to Functional Grammar. In *An Introduction to Functional Grammar* (3rd editio). Routledge. https://doi.org/10.4324/9780203783771

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

Haspelmath, M. (2004). Coordinating constructions: an overview. In M. Haspelmath (Ed.), *Coordinating Constructions*. Amsterdam: John Benjamins. Retrieved from https://doi.org/10.1075/tsl.58

Haspelmath, M. (2007). Coordination. In T. Shopen (Ed.), *Language Typology and Syntactic Description, Second Edition: Volume II: Complex Constructions* (2nd ed., Vol. 2, pp. 1–51). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511619434.001

Hecht, J. L. (2020). Responsibility in the Current Epidemic of Language Deprivation (1990–Present). *Maternal and Child Health Journal*, *24*(11), 1319–1322. https://doi.org/10.1007/s10995-020-02989-1

Hedderly, R. (1996). Vernon-Warden Reading Test (Revised). *Dyslexia Review*, *7*, 11–16.

Heine, B., & Kuteva, T. (2007). *The Genesis of Grammar: A Reconstruction*. Oxford University Press.

Hochgesang, J. A., Crasborn, O., & Lillo-Martin, D. (2020). *ASL SignBank*. New Haven, CT: Haskins Lab, Yale University. Retrieved from https://aslsignbank.haskins.yale.edu/

Hochstein, L., Bale, A., & Barner, D. (2018). Scalar Implicature in Absence of Epistemic Reasoning? The Case of Autism Spectrum Disorder. *Language Learning and Development*, *14*(3), 224–240. https://doi.org/10.1080/15475441.2017.1343670

Hodge, G. (2014). *Patterns from a signed language corpus: Clause-like units in Auslan (Australian sign language)* (Macquarie University). Macquarie University. https://doi.org/10.25949/19442855.V1

Hodge, G., Ferrara, L. N., & Anible, B. D. (2019). The semiotic diversity of doing reference in a deaf signed language. *Journal of Pragmatics*, *143*, 33–53. https://doi.org/10.1016/j.pragma.2019.01.025

Hodge, G., Manrique, E., Winter, B., & Cormier, K. (2025). Manual wh-signs and English wh-mouthings differentiate BSL content and polar questions. *Linguistics*. https://doi.org/10.1515/ling-2023-0238

Hopkins, J. (2008). Choosing how to write sign language: a sociolinguistic perspective. *International Journal of the Sociology of Language*, (192), 75–89. https://doi.org/10.1515/IJSL.2008.036

Horn, L. R. (1972). On the semantic properties of logical operators in English. University of California, Los Angeles. Retrieved from https://www.proquest.com/dissertations-theses/on-semantic-properties-logical-operators-english/docview/302676533/se-2?accountid=14511

Horn, L. R. (2009). WJ-40: Implicature, truth, and meaning. *International Review of Pragmatics*, *1*(1), 3–34. https://doi.org/10.1163/187731009X455820

Hothorn, T., Van De Wiel, M. A., Hornik, K., & Zeileis, A. (2008). Implementing a Class of Permutation Tests: The coin Package. *Journal of Statistical Software*, *28*(8), 1–23. https://doi.org/10.18637/JSS.V028.I08

Hou, L. (2024). Interrogating “deaf” native signer ideologies in linguistics research. *The ROLE Collective: Rejecting Harmful Speaker/Signer Ideologies*. Retrieved from https://www.youtube.com/watch?v=BHHWOHfGvR8

Hoza, J. (2011). The discourse and politeness functions of HEY and WELL in American Sign Language. In C. Roy (Ed.), *Discourse in signed languages* (pp. 69–95). Washington, D.C.: Gallaudet University Press.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Husson, F., Josse, J., & Lê, S. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, *25*. https://doi.org/10.18637/jss.v025.i01

Husson, F., Josse, J., & Pagès, J. (2010). *Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data?* Agrocampus Ouest, Rennes, France. Retrieved from http://factominer.free.fr/more/reference.html

Hutton, G. (1855). *Specimens of a dictionary of natural signs for the deaf and dumb*. Retrieved from https://archive.org/details/gu_specimensdict00hutt/page/n1/mode/2up

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jasbi, M., Waldon, B., & Degen, J. (2019). Linking hypothesis and number of response options modulate inferred scalar implicature rate. *Frontiers in Psychology*, *10*(FEB), 189. https://doi.org/10.3389/fpsyg.2019.00189

Johnston, T. (2018). A corpus-based study of the role of headshaking in negation in Auslan (Australian Sign Language): Implications for signed language typology. *Linguistic Typology*, *22*(2), 185–231. https://doi.org/10.1515/lingty-2018-0008

Johnston, T., & Schembri, A. (2007). *Australian Sign Language: an introduction to sign language linguistics*. Cambridge: Cambridge University Press.

Johnston, T., van Roekel, J., & Schembri, A. (2016). On the Conventionalization of Mouth Actions in Australian Sign Language. *Language and Speech*, *59*(1), 3–42. https://doi.org/10.1177/0023830915569334

Jones, A. C., Gutierrez, R., & Ludlow, A. K. (2015). Confronting the language barrier: theory of mind in deaf

children. *Journal of Communication Disorders*, *56*, 47–58. https://doi.org/https://doi.org/10.1016/j.jcomdis.2015.06.005

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. Comprehensive R Archive Network (CRAN). https://doi.org/10.32614/CRAN.PACKAGE.FACTOEXTRA

Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.

Khorsheed, A., Price, J., & van Tiel, B. (2022). Sources of cognitive cost in scalar implicature processing: A review. *Frontiers in Communication*, *7*, 990044. https://doi.org/10.3389/fcomm.2022.990044

Kimmelman, V. (2014). *Information Structure in Russian Sign Language and Sign Language of the Netherlands* (University of Amsterdam). University of Amsterdam. Retrieved from https://hdl.handle.net/11245/1.432175

Kimmelman, V. (2019). *Information Structure in Sign Languages*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781501510045

Kubinec, R. (2023). Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds. *Political Analysis*, *31*, 519–536. https://doi.org/10.1017/pan.2022.20

Kurumada, C., & Clark, E. V. (2017). Pragmatic inferences in context: Learning to interpret contrastive prosody. *Journal of Child Language*, *44*(4), 850–880. https://doi.org/10.1017/S0305000916000246

Kyle, B., & Woll, B. (1985). *Sign language: the study of deaf people and their language*. Cambridge: Cambridge University Press.

Ladd, P. (2003). *In Search of Deafhood*. Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/doi:10.21832/9781853595479

Leahy, A., & Brown, P. R. (2020). Naming British Sign Language 1960–1975. *Sign Language Studies*, *20*(4), 691–698. Retrieved from https://www.jstor.org/stable/26984292

Leech, G., Raysen, P., & Wilson, A. (2001). *Word frequencies in written and spoken English*. Longman. https://doi.org/10.4324/9781315840161

Lenneberg, E. H. (1967). Biological foundations of language. In *Biological foundations of language.* Oxford, England: Wiley.

Levinson, S. C. (2000). *Presumptive meanings: the theory of generalized converstional implicature*. Cambridge, MA: MIT Press. https://doi.org/10.1017/S0022226703272364

Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Liddell, S. K., Vogt-Svendsen, M., & Bergman, B. (2007). A crosslinguistic comparison of buoys: evidence from American, Norwegian, and Swedish Sign Language. In M. Vermeerbergen, L. Leeson, & O. Crasborn (Eds.), *Simultaneity in signed languages: form and function* (pp. 187–215). John Benjamins. https://doi.org/10.1075/cilt.281.09lid

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348. https://doi.org/10.1016/J.JESP.2018.08.009

Liu, F., & Kong, Y. (2015). zoib: An R package for Bayesian inference for beta regression and Zero/one inflated beta regression. *R Journal*, *7*(2), 34–51. https://doi.org/10.32614/RJ-2015-019

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). {performance}: An {R} Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source*

Software, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Marschark, M., Edwards, L., Peterson, C., Crowe, K., & Walton, D. (2019). Understanding theory of mind in deaf and hearing college students. *Journal of Deaf Studies and Deaf Education*, *24*(2), 104–118. https://doi.org/10.1093/deafed/eny039

Marty, P., Chemla, E., & Sprouse, J. (2020). The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: A Journal of General Linguistics*, *5*(1), 72. https://doi.org/http://doi.org/10.5334/gjgl.980

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual Differences in Children's Pragmatic Ability: A Review of Associations with Formal Language, Social Cognition, and Executive Functions. *Language Learning and Development*, *14*(3), 186–223. https://doi.org/10.1080/15475441.2018.1455584

Matthews, D., & Kelly, C. (2022). Pragmatic development in deaf and hard of hearing children: A review. *Deafness and Education International*, *24*(4), 296–313. https://doi.org/10.1080/14643154.2022.2140251

Mazzarella, D., Trouche, E., Mercier, H., & Noveck, I. (2018). Believing what you're told: Politeness and scalar inferences. *Frontiers in Psychology*, *9*(JUN), 908. https://doi.org/10.3389/fpsyg.2018.00908

McKee, R., & Wallingford, S. (2011). 'So, well, whatever': Discourse functions of *palm-up* in New Zealand Sign Language. *Sign Language & Linguistics*, *14*(2), 213–247. https://doi.org/10.1075/sll.14.2.01mck

Mill, J. S. (1865). *An examination of Sir William Hamilton's philosophy and of the principal philosophical questions discussed in his writings*. London: Longmans, Green and Co.

Mitchell, R. E., & Karchmer, M. A. (2004). Chasing the mythical ten percent: parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies*, *4*(2), 138–163. https://doi.org/10.1353/sls.2004.0005

Mithun, M. (1988). The grammaticalization of coordination. In J. Haiman & S. Thompson (Eds.), *Clause Combining in Grammar and Discourse*. John Benjamins. https://doi.org/10.1075/tsl.18.13mit

Morgan, G. (2000). Discourse cohesion in sign and speech. *International Journal of Bilingualism*, *4*(3), 279–300. https://doi.org/10.1177/13670069000040030101

Morgan, G., & Kegl, J. (2006). Nicaraguan Sign Language and Theory of Mind: the issue of critical periods and abilities. *Journal of Child Psychology and Psychiatry*, *47*(8), 811–819. https://doi.org/10.1111/j.1469-7610.2006.01621.x

Müller, C. (2004). Forms and uses of the Palm Up Open Hand: a case of a gesture family? *Proceedings of the Berlin Conference April 1998*, *79*(1), 189–196.

Neidle, C., Kegl, J., Maclaughlin, D., Bahan, B., & Lee, R. G. (2000). *The syntax of American Sign Language: functional categories and hierarchical structure*. Cambridge, MA: MIT Press.

Neidle, C., & Lee, R. G. (2008). Well, "what" is it? Discovery of a new particle in ASL. In K. Lindgren, D. DeLuca, & D. Napoli (Eds.), *Signs and Voices* (pp. 100–116). Gallaudet University Press.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. https://doi.org/10.1016/S0010-0277(00)00114-1

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210. https://doi.org/10.1016/S0093-934X(03)00053-1

Ospina, R., & Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers*, *51*(1), 111–126. https://doi.org/10.1007/S00362-008-0125-4/METRICS

Padden, C. A. (1988). *Interaction of Morphology and Syntax in American Sign Language*. Routledge.

Pagliarini, E., Bill, C., Romoli, J., Tieu, L., & Crain, S. (2018). On children's variable success with scalar inferences: insights from disjunction in the scope of a universal quantifier. *Cognition*, *178*, 178–192. https://doi.org/10.1016/j.cognition.2018.04.020

Papafragou, A., & Skordos, D. (2016). Scalar Implicature. In *The Oxford Handbook of Developmental Linguistics*.

Perniss, P. (2007). *Space and Iconicity in German Sign Language (DGS)*. Doctoral thesis, Radboud Universiteit Nijmegen.

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The Mind Behind the Message: Advancing Theory-of-Mind Scales for Typically Developing Children, and Those With Deafness, Autism, or Asperger Syndrome. *Child Development*, *83*(2), 469–485. Retrieved from http://www.jstor.org.libproxy.ucl.ac.uk/stable/41416099

Pfau, R. (2016). Syntax: complex sentences. In A. Baker, B. van den Bogaerde, R. Pfau, & G. M. Schermer (Eds.), *The linguistics of sign languages: an introduction* (pp. 149–167). John Benjamins Publishing Company.

Pfau, R., & Quer, J. (2010). Nonmanuals: their grammatical and prosodic roles. In D. Brentari (Ed.), *Sign Languages* (pp. 381–402). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511712203.018

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J. P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-008-0661-8

Podlesny, O. (2015). *Investigating disjunction in American Sign Language: the importance of nonmanual signals and the influence of English*. MCogSc thesis. Carleton University, Canada.

Proctor, H., & Cormier, K. (2022). *Sociolinguistic Variation in Mouthings in British Sign Language (BSL): A Corpus-Based Study*. https://doi.org/10.1177/00238309221107002

Quinn, G. (2010). Schoolization: An account of the origins of regional variation in British sign language. *Sign Language Studies*, *10*(4), 476–501. https://doi.org/10.1353/sls.0.0056

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.r-project.org/

Richardson, H., Koster-Hale, J., Caselli, N., Magid, R., Benedict, R., Olson, H., … Saxe, R. (2020). Reduced neural selectivity for mental states in deaf children with delayed exposure to sign language. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-17004-y

Rowley, K., & Cormier, K. (2023). Accent or not? Language attitudes towards regional variation in British Sign Language. *Applied Linguistics Review*, *14*(4), 919–943. https://doi.org/doi:10.1515/applirev-2020-0144

Rudge, L. A. (2022). The Nominal Group in British Sign Language: A Preliminary Description. *Word*, *68*(2), 83–108. https://doi.org/10.1080/00437956.2021.2024351

Russell, B. (2012). *Probabilistic reasoning and the computation of scalar implicatures*. PhD thesis, San Francisco State University.

Sandler, W. (1989). *Phonological representation of the sign: linearity and nonlinearity in American Sign Language*. Dordrecht: Foris Publication.

Sandler, W. (1993). A sonority cycle in American Sign Language. *Phonology*, *10*(2), 243–279. https://doi.org/10.1017/S0952675700000051

Sandler, W. (2010). Prosody and syntax in sign languages. *Transactions of the Philological Society*, *108*(3), 298–328. https://doi.org/10.1111/j.1467-968X.2010.01242.x

Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139163910

Schembri, A. (2024). Native signer ideologies in Auslan and BSL linguistics. *The ROLE Collective: Rejecting Harmful Speaker/Signer Ideologies*. Retrieved from https://www.youtube.com/watch?v=kCqlkdL7_fk

Schembri, A., Cormier, K., Johnston, T., McKee, D., McKee, R., & Woll, B. (2010). Sociolinguistic variation in British, Australian and New Zealand Sign Languages. *Sign Languages*, 476–498. https://doi.org/10.1017/CBO9780511712203.022

Schembri, A., Fenlon, J., Rentelis, R., & Cormier, K. (2014). *British Sign Language Corpus Project: a corpus of digital video data and annotations of British Sign Language 2008-2014 (Second Edition)*. London: University College London. Retrieved from http://www.bslcorpusproject.org

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S., & Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation & Conservation*, *7*, 136–154. Retrieved from http://hdl.handle.net/10125/4592

Schembri, A., Rowley, K., & Leeson, L. (2024). Sign Languages in Britain and Ireland. In S. Fox (Ed.), *Language in Britain and Ireland* (pp. 439–452). Cambridge University Press.

Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: a study of deaf children. *Child Development*, *78*(2), 376–396. https://doi.org/10.1111/j.1467-8624.2007.01004.x

Schlenker, P., Aristodemo, V., Ducasse, L., Lamberton, J., & Santoro, M. (2016). The unity of focus: Evidence from sign language (ASL and LSF). *Linguistic Inquiry*, *47*(2), 363–381. https://doi.org/10.1162/ling_a_00215

Singh, R., Wexler, K., Astle-Rahim, A., Deepthi Kamawar, & Fox, D. (2016). Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics*, *24*, 305–352. https://doi.org/10.1007/s11050-016-9126-3

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*. https://doi.org/10.1016/j.cognition.2016.04.006

Slabakova, R. (2010). Scalar implicatures in second language acquisition. *Lingua*, *120*(10), 2444–2462. https://doi.org/10.1016/j.lingua.2009.06.005

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

Smogorzewska, J., Szumski, G., Bosacki, S., & Grygiel, P. (2022). Just listen to your mind: Consequences of theory of mind development for deaf or hard-of-hearing children. *Research in Developmental Disabilities*, *127*, 104261. https://doi.org/https://doi.org/10.1016/j.ridd.2022.104261

Sperber, D., & Wilson, D. (1986). *Relevance: communication and cognition*. Oxford, UK: Blackwell Publishing Ltd.

Sperber, D., & Wilson, D. (1987). Précis of Relevance: Communication and Cognition. *Behavioral and Brain Sciences*, *10*, 697–754.

Sperber, D., & Wilson, D. (1995). *Relevance: communication and cognition (second edition)*. Oxford: Blackwell Publishing Ltd.

Sprouse, J., & Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment

experiments. *Glossa: A Journal of General Linguistics*, *2*(1), 14. https://doi.org/10.5334/gjgl.236

Stamp, R., Schembri, A., Fenlon, J., & Rentelis, R. (2015). Sociolinguistic variation and change in British sign language number signs: Evidence of leveling? *Sign Language Studies*, *15*(2), 151–181. https://doi.org/10.1353/sls.2015.0001

Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., & Cormier, K. (2014). Lexical Variation and Change in British Sign Language. *PLOS ONE*, *9*(4), e94053. https://doi.org/10.1371/JOURNAL.PONE.0094053

Stan Development Team. (2024). *Stan Modeling Language Users Guide and Reference Manual, 2.34*. Retrieved from https://mc-stan.org

Stiller, A., Goodman, N. D., & Frank, M. C. (2011). Ad-hoc scalar implicature in adults and children. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 2134–2139.

Stokoe, W. (1960). Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. In *Studies in linguistics: Occasional papers* (Vol. 8). Buffalo, NY: Dept. of Anthropology and Linguistics, University of Buffalo.

Sutton-Spence, R. (1994). *The role of the manual alphabet and fingerspelling in British Sign Language*. Doctoral dissertation. University of Bristol.

Sutton-Spence, R., & Woll, B. (1999). *The linguistics of British Sign Language: an introduction*. Cambridge: Cambridge University Press.

Sutton-Spence, R., Woll, B., & Allsop, L. (1990). Variation and recent change in fingerspelling in British Sign Language. *Language Variation and Change*, *2*(3), 313–330. https://doi.org/10.1017/S0954394500000399

Szarkowski, A., Young, A., Matthews, D., & Meinzen-Derr, J. (2020). Pragmatics development in deaf and hard of hearing children: A call to action. *Pediatrics*, *146*. https://doi.org/10.1542/peds.2020-0242L

Tang, G., & Lau, P. (2012). Coordination and subordination. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: an international handbook* (p. 340). Berlin, Boston: De Gruyter Mouton.

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2016). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, *33*(2), 269–298. https://doi.org/10.1093/jos/ffv001

Twomey, T., Price, C. J., Waters, D., & MacSweeney, M. (2020). The impact of early language exposure on the neural system supporting language in deaf and hearing adults. *NeuroImage*, *209*. https://doi.org/10.1016/j.neuroimage.2019.116411

van der Kooij, E., & Crasborn, O. (2016). Phonology. In *The Linguistics of Sign Languages* (pp. 251–278). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/Z.199.11KOO

van der Kooij, E., Crasborn, O., & Emmerik, W. (2006). Explaining prosodic body leans in Sign Language of the Netherlands: Pragmatics required. *Journal of Pragmatics*, *38*(10), 1598–1614. https://doi.org/10.1016/j.pragma.2005.07.006

van Loon, E., Roland, P., & Markus, S. (2014). The grammaticalization of gestures in sign languages. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - Language - Communication* (Vol. 2). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110302028.2133

van Tiel, B., & Kissine, M. (2018). Quantity-based reasoning in the broader autism phenotype: A web-based study. *Applied Psycholinguistics*, *39*(6), 1373–1403. https://doi.org/10.1017/S014271641800036X

van Tiel, B., & Schaeken, W. (2017). Processing Conversational Implicatures: Alternatives and

Counterfactual Reasoning. *Cognitive Science*, *41*, 1119–1154. https://doi.org/10.1111/cogs.12362

van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2014). Scalar Diversity. *Journal of Semantics*, *33*(1), ffu017. https://doi.org/10.1093/jos/ffu017

Vicars, W. (1997). American Sign Language: “or.” Retrieved September 25, 2020, from https://www.lifeprint.com/asl101/pages-signs/o/or.htm

Vuorre, M. (2019). How to Analyze Visual Analog (Slider) Scale Data? Retrieved February 28, 2024, from https://vuorre.com/posts/2019-02-18-analyze-analog-scale-ratings-with-zero-one-inflated-beta-models

Waters, D., & Sutton-Spence, R. (2005). Connectives in British Sign Language. *Deaf Worlds*, *21*, 1–29.

Wellman, H. M., & Peterson, C. C. (2013). Deafness, thought bubbles, and theory-of-mind development. *Developmental Psychology*, *49*(12), 2357–2367. https://doi.org/10.1037/a0032419

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wilbur, R. (2012). Information structure. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: an international handbook* (pp. 462–489). Berlin, Boston: De Gruyter Mouton.

Wilbur, R., & Patschke, C. (1998). Body leans and the marking of contrast in American sign language. *Journal of Pragmatics*, *30*(3), 275–303. https://doi.org/10.1016/S0378-2166(98)00003-4

Wilson, D., & Sperber, D. (2002). Truthfulness and relevance. *Mind*, *111*(443), 583–632. https://doi.org/10.1017/CBO9781139028370.005

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006*, 1556–1559.

Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to Development: Theory of Mind in Deaf Children. *Child Development*, *73*(3), 768–778. https://doi.org/https://doi.org/10.1111/1467-8624.00437

Working Party on Signed English. (1984). *Signed English for Schools: Volume I*. Royal National Institute for the Deaf.

Yoel, J. (2009). *Canada’s Maritime Sign Language* (University of Manitoba, Winnipeg). University of Manitoba, Winnipeg. Retrieved from http://hdl.handle.net/1993/21581

Young, A., Szarkowski, A., Ferguson-Coleman, E., Freeman, D., Lindow-Davies, C., Davies, R., … Vesey, K. (2020). The lived experience and legacy of pragmatics for deaf and hard of hearing children. *Pediatrics*, *146*, 304–309. https://doi.org/10.1542/peds.2020-0242K

Zeshan, U. (2004). Interrogative constructions in signed languages: crosslinguistic perspectives. *Language*, *80*(1), 7–39. Retrieved from http://www.jstor.org/stable/4489610

Zimmer, J. (1989). Toward a description of register variation in American Sign Language. In C. Lucas (Ed.), *The Sociolinguistics of the Deaf Community*. Academic Press Inc.

Appendix A: Information sheet

[Note: in the web version sent to potential study participants, and within the study itself, BSL translations of the English version presented here were provided as embedded video clips. The BSL version in the study can be viewed online **here** by clicking the button marked “Preview”.]

Project title: The BSL Guessing Game Experiment 2024
Department: Deafness, Cognition and Language Centre / Linguistics
Institution: University College London
Researcher: Matt Brown (m.brown.13@ucl.ac.uk)
Supervisor: Prof Kearsy Cormier (k.cormier@ucl.ac.uk)
Ethics Committee Approval ID Number: LING-2019-10-07

About this study

This study is exploring a feature of communication which is common to all languages, both signed and spoken. So far, no-one has studied it in detail for British Sign Language (BSL). We think it is very important to have BSL data to compare with previous studies of American Sign Language and spoken languages.

Why was I invited? Do I have to take part?

We invited you because you have identified yourself as a deaf person who prefers to use BSL to communicate, who is at least 18 years old, who was born in the UK, and currently lives in the UK.

You do not have to take part. It is your choice if you want to participate or not. We will ask for your consent before starting.

You can talk about this project with other people if you want, or contact the researchers if you have any questions or if anything is unclear.

There is no penalty if you do not participate. There is no penalty if you begin the task and then decide not to continue. You can withdraw at any time by emailing the researcher (contact details above). However, we can only pay you if you complete the whole task.

Withdrawing from the project, if you choose to do so, will not affect you in any way. You are able to withdraw your data from the project up until your data has been combined with other data for research analyses. It will no longer be possible to withdraw participant data beyond this point because the data will have been anonymised.

What will I have to do if I decide to take part?

You will complete an online task, which has three parts. The first part is just a short practice session. In the second part, you will be shown 32 pictures and a video describing each picture in BSL. You will decide whether each BSL description is “right”, or “wrong”, or somewhere in between. The third part is a written English comprehension task. You will see 40 sentences. Each sentence has a gap and you must choose a word to fill each gap.

The time it takes to complete will be different for different people, but it should take less than 60 minutes. The task is not timed and you are welcome to take breaks.

You will need access to a computer with a web browser (a desktop computer, a laptop, or a tablet), an internet connection, and somewhere you can do the task in privacy without distractions. A smartphone will not be suitable as the screen would be too small. We will give you a link to a specific website address: please do not share this link with anyone else.

We will also ask for some personal information, for example, your approximate age, your ethnic background, which part of the country you live in, and how long you have been using BSL.

How will I be paid?

We will pay you after you complete the task. You will receive a £15 E-Gift Card by email. You can use this with participating retailers such as Amazon, Argos, Asda, Marks & Spencer and many others.

When you complete the task, you will receive a link via email. Click on it to go to the E-Gift Card reward page. You will then receive your E-Gift Card via email.

Data Protection Privacy Notice

The data controller for this project is University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk.

Your personal data will be processed for the purposes outlined in this notice. For more detail, see UCL's general privacy notice for research.

Data protection legislation including the General Data Protection Regulation (GDPR) requires us to have a valid legal reason to process and use personal data about you. This is often called a ‘legal basis’. GDPR requires us to clearly explain the legal basis that we rely on in order to process information about you.

The legal basis that will be used to process your personal data (your background information) is ‘Public task’. This means for general benefit of the public.

The legal basis that will be used to process your special category data (sensitive background information like your deafness and ethnicity) is ‘Research purposes’.

Your personal data will be processed so long as it is required for the research project. Your data will not be stored with your name. Your data will only be stored with an ID code to keep your personal data confidential.

Consent form

Before you start, please read the following statements and confirm that they are true by ticking the boxes. If you have any questions before you start, please contact the researcher by e-mail (Matt Brown, m.brown.13@ucl.ac.uk).

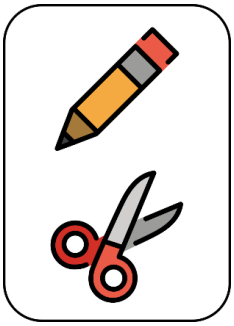
- ☐ I have seen or read the information provided about this study.
- ☐ I am 18 years of age or older.
- ☐ I was born in the UK, and I currently live in the UK.
- ☐ BSL is my main language, or I prefer to use it whenever possible.
- ☐ I understand that my personal information will be used only for the purposes explained to me. I understand that, according to data protection legislation, "public task" (for general benefit of the public) and "research purposes" will be the legal basis for processing.
- ☐ I understand that all personal information will remain strictly confidential; that any data gathered in this study will be stored securely with an ID code and not my name; that it will not be possible to identify me in any publications of the study's results.
- ☐ I understand that I am free to withdraw from this study at any time without having to give a reason, up until my data has been combined with other study data. If I want to withdraw, I can email the researcher (contact details above).

Appendix B: Main task instructions

[Note: in the task itself, this English version was accompanied by a BSL translation (an embedded video clip) provided by a qualified BSL/English translator. The BSL version in the study can be viewed online **here** by clicking the button marked “Preview” and proceeding past the initial briefing and consent section to the instructions.]

Main instructions

You are about to play a guessing game with a partner. Imagine that you are holding a pack of special playing cards. Each card shows one or two pictures of animals or objects. Your partner cannot see the cards, but you can.



[Note: in the live task, the image above was a slowly looping animation of the entire playing card set.]

We will pick out cards, one by one. You will see them appear on the screen in a random order. Your guessing partner will appear in a video clip next to the card. Remember: they cannot see the card and do not know what is on it, but they know what kind of pictures are in the pack. They will make a guess about what the card shows using a short description in BSL. They will use different ways of expressing what they think is on each card.

Below the card and video will be a slider. At one end is a green tick, which means “right”. At the other end is a red cross, which means “wrong”. Your task is to rate how right or wrong each guess is, using the slider.

For example, maybe you think a guess is accurate and the BSL matches the card very well. In that case, you would drag the slider towards the green tick, like this:

In another case, you might think a guess is badly wrong and the BSL sentence could not ever describe the card. In that case, you would drag the slider towards the red cross, like this:

But it is also possible that you think the guess is not quite right but not completely wrong either. Perhaps the sentence doesn't describe the card very well, but it could be worse. In that case, you should position the slider somewhere in between. For example:

We will first show you 3 practice videos, so that you can become familiar with the task. Your answers in the practice section will not be used. After that, the main guessing game task will begin. You will be shown a further 32 cards and videos, and your answers will be recorded.

You will then complete a short English reading comprehension test. You will receive instructions on that later.

Finally, we will ask you a few short questions about your background, and then the tasks will be complete. The time it takes to complete will be different for different people, but it should take between 30 and 60 minutes.

The task is not timed and you are welcome to take breaks. Your answers will be saved as you go along. If you need to take a short break, just leave the browser window open. If you need to stop the task and start again later, you can close the browser window or shut the computer down if you like. When you are ready to start again, click on the personalised link in your invitation e-mail, and you will be taken to where you stopped. While you are doing the tasks, please make sure that you are in a private space without many distractions.

When you are ready to start, click the Start button below.

Practice task preamble

This is a short practice session for the card guessing game. You will be shown 3 different cards and videos. Your answers to these practice questions will not be used. Use the slider to rate each video as right, wrong, or somewhere in between right and wrong. When you are ready to start, click Next.

Main task preamble

We will now show you 32 cards and videos. Your progress will be shown at the bottom. Your answers to these questions will be recorded. Use the slider to rate each video as right, wrong, or somewhere in between right and wrong. When you are ready to start, click Next.

Appendix C: Main task trial item combinations

Note: in the live study, the trials were presented in a randomised order.

#	Coordination	Stress	Condition	Coordinator	Utterance	Pictures	Group
1	None	No	FALSE	None	DOG	GOAT	A
2	None	No	FALSE	None	HORSE	TIGER CAMEL	B
3	None	No	FALSE	None	BREAD	CAKE	C
4	None	No	FALSE	None	PENCIL	RULER SCISSORS	D
5	None	No	TRUE	None	DOG	DOG	A
6	None	No	TRUE	None	HORSE	HORSE	B
7	None	No	TRUE	None	BREAD	BREAD	C
8	None	No	TRUE	None	PENCIL	PENCIL	D
9	Conjunction	No	FALSE	Asyndetic	DOG GOAT	FISH	A
10	Conjunction	No	FALSE	Asyndetic	HORSE TIGER	HORSE CAMEL	B
11	Conjunction	No	FALSE	List	BREAD CAKE	BANANA	C
12	Conjunction	No	FALSE	List	PENCIL RULER	RULER SCISSORS	D
13	Conjunction	No	TRUE	Asyndetic	BREAD CAKE	BREAD CAKE	C
14	Conjunction	No	TRUE	Asyndetic	PENCIL RULER	PENCIL RULER	D
15	Conjunction	No	TRUE	List	DOG GOAT	DOG GOAT	A
16	Conjunction	No	TRUE	List	HORSE TIGER	HORSE TIGER	B
17	Disjunction	No	FALSE	Palm-up	DOG GOAT	FISH	A
18	Disjunction	No	TRUE	Palm-up	HORSE TIGER	HORSE	B
19	Disjunction	No	**TEST**	Palm-up	BREAD CAKE	BREAD CAKE	C
20	Disjunction	No	**TEST**	Palm-up	PENCIL RULER	PENCIL RULER	D
21	Disjunction	No	FALSE	List + palm-up	BREAD CAKE	BANANA	C
22	Disjunction	No	TRUE	List + palm-up	PENCIL RULER	PENCIL	D
23	Disjunction	No	**TEST**	List + palm-up	DOG GOAT	DOG GOAT	A
24	Disjunction	No	**TEST**	List + palm-up	HORSE TIGER	HORSE TIGER	B
25	Disjunction	**Yes**	FALSE	Palm-up	DOG GOAT	FISH	A
26	Disjunction	**Yes**	TRUE	Palm-up	HORSE TIGER	TIGER	B
27	Disjunction	**Yes**	**TEST**	Palm-up	BREAD CAKE	BREAD CAKE	C
28	Disjunction	**Yes**	**TEST**	Palm-up	PENCIL RULER	PENCIL RULER	D
29	Disjunction	**Yes**	FALSE	List + palm-up	BREAD CAKE	BANANA	C
30	Disjunction	**Yes**	TRUE	List + palm-up	PENCIL RULER	RULER	D
31	Disjunction	**Yes**	**TEST**	List + palm-up	DOG GOAT	DOG GOAT	A
32	Disjunction	**Yes**	**TEST**	List + palm-up	HORSE TIGER	HORSE TIGER	B

Appendix D: English reading task instructions

[Note: in the task itself, this English version was accompanied by a BSL translation (embedded video clips) provided by a qualified BSL/English translator. The study can be viewed online **here**.]

This is the last task. You will be shown 40 English sentences. Each sentence has a missing word. Below each sentence is a choice of five words. Choose the word which you think makes the most sense if it was put in the gap. If it seems like there is more than one right answer, choose the best one. If you can't answer a question, don't worry: skip it and go to the next one.

When you are ready to start, click Next.

Appendix E: English reading task questionnaire

Note: this is the secondary task questionnaire described in §5.2.2.2. The changes mentioned to the original questions given in Hedderly (1996) are as follows:

2: parlour → cellar
3: charwoman → cleaner
3: housewife → army
6: Joan → Jane
9: removed reference to wife packing suitcase for husband
14: deleted
16: grammatical tweak
41: eyewash → nonsense
42: deleted

The answer deemed correct is indicated in **bold**.

1. Ducks swim in a [bucket, **pond**, yard, cage, garden].
2. Cooking is usually done in the [bedroom, cellar, fire, **kitchen**, street].
3. One of the water pipes burst, so the occupants of the flat sent for the [landlord, **plumber**, cleaner, army, bucket].
4. A burglary should be reported to the [friends, school, home, office, **police**].
5. A salesperson works in a [hospital, factory, **shop**, garden, house].
6. Tom and Jane are happy. In two weeks' time school will stop, and they will go away for their [home, teacher, **holidays**, fun, seaside].
7. She was the kindest woman in town: she never tired of [seeing, dressing, giving, **helping**, doing] others.
8. The head teacher was held responsible for the discipline and general behaviour of the children in the [class, **school**, town, street, family].
9. Mr Jones packed his suits, shirts, ties and other necessary articles of [**clothing**, toilet, travel, socks, food] into a suitcase.
10. Sam settled down to an evening of mystery and thrilling adventure with a [new, library, love, humorous, **detective**] story.
11. Farmers who till the fields or feed their sheep and cattle do not live in towns. They are [homely, dull, **country**, nice, fat] folk.
12. To trap a wild animal, some hunters make a big hole in the earth and cover it with branches and leaves. As the animal cannot see its [life, hunter, bottom, **danger**, path] it falls in.
13. The bare trees made a lacy pattern against the [summer, empty, wide, blue, **wintry**] sky.
14. When she learnt of her failure in the examination despite her hard work, the girl was completely [**disheartened**, exhausted, sick, relieved, rejected].

15. They scanned the “accommodation” column of the advertisement page, searching for a [**flat**, cook, motor car, dog, job].
16. The level of the water was at its highest; after some weeks it began to fall, at first slowly, but afterwards quickly. By the end of the month the river was quite [high, wet, wide, **low**, swift].
17. His parents suspended his monthly allowance in what they considered an effective step to curb his [thrifty, monthly, studious, light, **extravagant**] habits.
18. It was like a morning in May. The mellow November sun came streaming into the room. The sky was bright and there was a genial [dampness, **warmth**, breeze, cold, sun] in the air.
19. The art student worked hard and scrupulously at his task. Yet he failed to produce a master-piece for the obvious reason that he had no [paints, duty, need, support, **talent**].
20. It is not sufficient to know the means of preventing and curing disease; it is equally [bad, **necessary**, useful, sufficient, good] to provide these means and even to compel their use.
21. The student roamed from library to library in search of the [studies, librarian, professor, library, **book**] recommended for reading.
22. No film is open to public entertainment until it has been [seen, filmed, distributed, paid for, **censored**].
23. Considering the great capacity that a human being has for love, it is astounding that the history of mankind should be frequently blackened by deeds of [love, manliness, consideration, **hate**, history].
24. In the days when people lived in caves and rough shelters, there were no fields or farms. They got their food by hunting animals and by [finding, planting, **gathering**, growing, eating] berries and wild fruits.
25. The old man walked down into the wine cellar and looked for his cherished bottle of [medicine, lemonade, paraffin, **claret**, liquid].
26. After the long hours and strain they had gone through at work, their holiday made them all feel very much [sun-burned, **rested**, worse, overworked, energetic].
27. However brilliant students may be, strong emotional disturbances are bound to affect their [**results**, character, happiness, health, comfort].
28. In a totalitarian state, books and all other forms of printed matter are subjected to government [**censorship**, dictatorship, publication, printing, subsidies].
29. No scheme for a change of society can be made to appear immediately palatable, except by falsehood, unless society has become so desperate that it will accept any [lie, help, falsehood, **change**, politician].
30. Every stockman knows that he can breed sheep for fatness or for length of wool; he can breed horses for strength or for swiftness. He can thus [breed, have, create, **change**, fix] the size and shape of animals from generation to generation.

31. Few things are more irritating than to be accused of intellectual dishonesty when one is trying hard to fight against [self-improvement, **self-deception**, cheating, lying, irritation].
32. History has known several cases when a mode of thought has seized the imagination of a vast majority of the intelligentsia, as existentialism has today in France; hence the great [imagination, eagerness, **caution**, fear, popularity] with which a true philosopher approaches such highly popularised philosophies.
33. To write a successful book requires as much time and energy as to write an unsuccessful one, the difference lying in the fact that the energy in the first case reaps its [success, books, **rewards**, writing, energy].
34. She hardened herself against any emotional entanglements in an effort to spare herself pain, little realising that pain and hurt are an essential part of full [emotion, **living**, efforts, enjoyment, hardening].
35. A highly cultural environment often boosts up the academic career of an otherwise [practical, absent, **mediocre**, brilliant, educated] student.
36. Whilst self-restraint is a praiseworthy quality one should try to cultivate, it runs as much danger of being exaggerated as its [**counterpart**, equal, quality, virtue, denial] - a free expression of personal feelings and emotion.
37. The boss firmly denied any accusation of discrimination. It was claimed that the policies of the firm were not dictated by [**prejudice**, incrimination, accusations, profits, shareholders].
38. To allow the standards of a new play to be assessed by the box-office is as dangerous to art as to insist on the play's high calibre to the complete neglect of [prices, criticism, **popular appeal**, cinemas, art].
39. True friendship is a relationship so rare that most mortals go through life without realising that so precious and idealistic a [**communion**, truth, mortality, communication, friendship] between two persons can exist.
40. Despite strong claims to the contrary no nation colonises another with an altruistic motive, for the very basis of colonisation is exploitation, economic or political; to feign [colonial, **idealistic**, economic, strong, contrary] motives is sheer nonsense fit for the ignorant.

Appendix F: Tables of group average marginal effects

The tables below fully break down the findings described and illustrated in §5.3.2. The credible intervals indicated (conf.low to conf.high) are 90% highest density intervals. Intervals which do not include zero can be taken as suggesting a level of credible evidence, and are marked with an asterisk.

H1: Marginal effects of stress on structure: mean(STRESSED) - mean(UNSTRESSED)

structure	estimate	conf.low	conf.high	Evidence
PALM	0.01	-0.03	0.04	
LIST_PALM	-0.03	-0.07	-0.01	*

H1: Marginal effects of structure on stress: mean(LIST_PALM) - mean(PALM)

stress	estimate	conf.low	conf.high	Evidence
UNSTRESSED	0.06	0.03	0.09	*
STRESSED	0.02	-0.01	0.05	

H2: Marginal effects of AoA on RA: mean(Later) - mean(Early)

RA	estimate	conf.low	conf.high	Evidence
Higher	0.04	0.20	-0.13	
Lower	-0.07	0.05	-0.23	

H2: Marginal effects of RA on AoA: mean(Higher) - mean(Lower)

AoA	estimate	conf.low	conf.high	Evidence
Early	-0.13	0.04	-0.28	
Later	-0.24	-0.10	-0.39	*

H3: Marginal effects of stress on structure × AoA × RA: mean(STRESSED) - mean(UNSTRESSED)

structure	AoA	RA	estimate	conf.low	conf.high	Evidence
PALM	Early	Higher	-0.08	-0.15	-0.01	*
PALM	Early	Lower	-0.03	-0.09	0.02	
PALM	Later	Higher	0.03	-0.03	0.09	
PALM	Later	Lower	0.08	0.02	0.13	*
LIST+PALM	Early	Higher	-0.01	-0.08	0.06	
LIST+PALM	Early	Lower	-0.04	-0.10	0.02	
LIST+PALM	Later	Higher	-0.02	-0.08	0.05	
LIST+PALM	Later	Lower	-0.06	-0.10	-0.01	*

H3: Marginal effects of structure on stress × AoA × RA: mean(LIST_PALM) - mean(PALM)

stress	AoA	RA	estimate	conf.low	conf.high	Evidence
UNSTRESSED	Early	Higher	0.01	-0.07	0.08	
UNSTRESSED	Early	Lower	-0.01	-0.07	0.05	
UNSTRESSED	Later	Higher	0.12	0.06	0.18	*
UNSTRESSED	Later	Lower	0.08	0.03	0.13	*
STRESSED	Early	Higher	0.08	0.00	0.14	*
STRESSED	Early	Lower	-0.02	-0.07	0.05	
STRESSED	Later	Higher	0.07	0.01	0.14	*
STRESSED	Later	Lower	-0.05	-0.09	0.00	*

H3: Marginal effects of AoA on RA × structure × stress: mean(Later) - mean(Early)

RA	structure	stress	estimate	conf.low	conf.high	Evidence
Higher	PALM	UNSTRESSED	0.12	0.31	-0.06	
Higher	PALM	STRESSED	0.01	0.20	-0.18	
Higher	LIST+PALM	UNSTRESSED	0.02	0.19	-0.16	
Higher	LIST+PALM	STRESSED	0.01	0.18	-0.18	
Lower	PALM	UNSTRESSED	0.00	0.15	-0.16	
Lower	PALM	STRESSED	-0.11	0.04	-0.27	
Lower	LIST+PALM	UNSTRESSED	-0.09	0.06	-0.24	
Lower	LIST+PALM	STRESSED	-0.10	0.06	-0.27	

H3: Marginal effects of RA on AoA × structure × stress: mean(Higher) - mean(Lower)

AoA	structure	stress	estimate	conf.low	conf.high	Evidence
Early	PALM	UNSTRESSED	-0.14	0.03	-0.31	
Early	PALM	STRESSED	-0.18	-0.01	-0.38	*
Early	LIST+PALM	UNSTRESSED	-0.11	0.08	-0.27	
Early	LIST+PALM	STRESSED	-0.07	0.13	-0.25	
Later	PALM	UNSTRESSED	-0.26	-0.10	-0.42	*
Later	PALM	STRESSED	-0.30	-0.13	-0.44	*
Later	LIST+PALM	UNSTRESSED	-0.22	-0.07	-0.38	*
Later	LIST+PALM	STRESSED	-0.18	-0.03	-0.34	*